Technische Universität München

TUM School of Engineering and Design

Advanced methodologies for large-scale gravity field modelling

Philipp Zingerle

Vollständiger Abdruck der von der TUM School of Engineering and Design der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender:	Prof. Dr.rer.nat. Thomas H. Kolbe
Prüfer der Dissertation:	1. Prof. Dr.techn. Roland Pail
	2. Prof. Dr.techn. Wolf-Dieter Schuh
	3. Assoc. Prof. Dr. Mirko Reguzzoni

Die Dissertation wurde am 19.10.2021 bei der Technischen Universität München eingereicht und durch die TUM School of Engineering and Design am 17.01.2022 angenommen.

Zusammenfassung

Die Darstellung globaler Schwerefeldmodelle (GGMs) durch sphärische Harmonische (SHs) hat sich über die letzten Jahrzehnte als de facto Standard behauptet. Basierend auf deren Inhalt, können GGMs unterteilt werden in Modelle, welche lediglich Satelliteninformation enthalten (SGGMs) und Modelle, welche zusätzlich terrestrische Beobachtungen beinhalten (EGGMs). Die Prozessierung von EGGMs wird technisch zumeist durch ein Zweischrittverfahren realisiert: In einem ersten Schritt werden terrestrische Beobachtungen auf ein reguläres Gitter vorprozessiert (im Allgemeinen als regionale Schwerefeldmodellierung bezeichnet), welche anschließend, in einem zweiten Schritt effizient analysiert und mit einem SGGM kombiniert werden können (als globale Schwerefeldmodellierung bezeichnet). Bis jetzt werden üblicherweise beide Schritte, dh. regionale und globale Schwerefeldmodellierung, unabhängig voneinander behandelt, sodass globale Modelle oft aus einem Flickenteppich von regionalen Modellen bestehen, welche ihrerseits üblicherweise von verschiedenen Gruppen unter Verwendung verschiedener Methoden prozessiert werden (und daher unterschiedliche Eigenschaften aufweisen). Aus der Erkenntnis, dass dieses Vorgehen zu verschiedenen Inkonsistenzen und Nachteilen im resultierendem EGGM führen kann, ergibt sich der Gegenstand dieser Arbeit, nämlich die Verbesserung der Situation durch die Einführung eines integrierten Ansatzes, welcher beide Teilschritte umfasst. Das ganzheitliche Vorgehen versucht (1) verschiedene methodische Aspekte zu verbessern, (2) Synergien zwischen den Verarbeitungsschritten zu erzeugen und (3) auf verbleibende Einschränkungen hinzuweisen und mögliche Lösungen aufzuzeigen. Diese drei Aspekte definieren die Hauptthemen dieser Dissertation, welche (kumulativ) auf drei (Haupt-)Publikationen basiert. Zur Veranschaulichung wird die Vorstellung der Methodik von einem Testszenario begleitet, in dem Fluggravimetriedaten (von kontinentalem Ausmaß) der GRAV-D Schwerefeldkampagne nahtfrei in ein EGGM integriert werden (welches als SGDTv1 bezeichnet wird). Der ganzheitliche Ansatz selbst kann in drei Hauptprozessierungsschritte unterteilt werden (welche ihrerseits weiter unterteilt werden können), welche jeweils durch den Inhalt einer Publikation abgedeckt werden und auch gleichzeitig jeweils ein Hauptkapitel dieser Arbeit darstellen: Der erste Schritt besteht aus der Validierung und Reduktion der Eingangsbeobachtungen. Hauptaspekte hier sind die Herleitung des SATOP Modells, die Präsentation einer schnellen Synthesemethode und Strategien zur Identifizierung und Beseitigung von Problemen in den Daten. Im zweiten Schritt, der lokalisierten Schwerefeldmodellierung, wird ein gegitterter Datensatz unter Verwendung von Kollokation (LSC) abgeleitet. Unter der Einführung der PE-LSC Methode werden eine Reihe von Innovationen vorgestellt, welche dazu dienen, die Anwendbarkeit von LSC in mehrerer Hinsicht zu verbessern (automatisierte Datenausdünnung, Einführung einer 2D Kovarianzfunktion, Anpassung der ellipsoidischen Geometrie, Lokalisierung des Kernels und eine automatisierte Partitionierung). Zusätzlich werden Rückkopplungsmöglichkeiten aufgezeigt (zurück zur Validierungsphase), um Probleme mit den Beobachtungen (nach der Kollokation) noch besser identifizieren zu können und um das Ergebnis der Kollokation unter Umständen iterativ zu verbessern. Der dritte und letzte Prozessierungsschritt, die globale Schwerefeldmodellierung behandelt die SH Analyse und Möglichkeit zur Kombination mit SGGMs zur Herleitung des finalen EGGMs. Besondere Aufmerksamkeit wird dabei den Eigenschaften der SH Analyse auf geographischen Gittern zuteil, dem Blockdiagonalansatz und der sog. Kite-Methode als Kombinationsstrategie mit SGGMs. Abschließend werden Strategien zur Validierung des resultierenden EGGMs vorgestellt zusammen mit weiteren Möglichkeiten,

das finale Modell in die Validierungsphase zurückzukoppeln, um die Reduktionen (und damit auch das finale EGGM) iterativ zu verbessern. Ein Vergleich des SGDTv1 Modells (als Resultat des GRAV-D Testszenarios) zu anderen EGGMs durch GNSS/Nivellements zeigt, dass das experimentelle Modell (in Bereichen mit Datenüberdeckung) gleich oder besser abschneidet als die konkurrierenden Modelle. Dies unterstreicht die korrekte Funktionsweise des vorgestellten ganzheitlichen Ansatzes.

Abstract

The representation of global gravity field models (GGMs) through spherical harmonics (SHs) has de facto been established as standard over the last few decades. GGMs can be subdivided into models which solely contain satellite information (SGGMs) and models which extend SGGMs through additional (ground) gravity observations (EGGMs). The processing of EGGMs is due to its nature commonly a two-step approach: in a first step (ground) gravity observations are preprocessed onto a regular grid (denoted as regional gravity field modelling) which can then efficiently be analyzed and combined with a SGGM in a second step to obtain the final SH representation (classified as global gravity field modelling). Currently, both steps, i.e., regional and global gravity field modelling are usually treated independently of each other, so global models often consist of a patchwork of regional models with different properties processed by different groups using different methodologies. Acknowledging that this procedure may lead to various inconsistencies and disadvantages in the final EGGM, this thesis tries to tackle these deficiencies by incorporating the entire gravity field processing into a single integrated approach. The resulting holistic approach attempts to (1) improve various methodological aspects, to (2) create synergies among the various processing steps, and to (3) highlight remaining limitations and give hints towards possible solutions. These three aspects build the main objectives of this work which is of cumulative nature and based on three main publications. For illustration, the presentation of the whole procedure is accompanied by a test case where the continent-scale GRAV-D airborne gravity mission is integrated into a seamless EGGM (denoted as SGDTv1). The holistic approach itself can be subdivided into three main processing steps (which can be further divided into substeps), each covered by the content of one of the three publication and each representing one main chapter of this thesis: the first step consists of the validation and reduction of the input observations. Key aspects here are the derivation of the SATOP model, the presentation of a fast synthesis method and strategies to identify and remove problems in the data. In the second processing step, the localized gravity field modeling, a gridded dataset is derived by means of least squares collocation (LSC). Introducing the PE-LSC method, a set of innovations is presented which improves the applicability of LSC in several aspects (automated data thinning, introduction of 2D covariance functions, adjustment of the spheroidal geometry, kernel localization and automated partitioning). Additionally, feedback possibilities (to the validation stage) are presented to further validate observations after the collocation and to possibly improve the collocation result iteratively. The third and last processing step, the global gravity field modelling, treats the SH analysis and possibilities to combine the solution with SGGMs to derive the final EGGM. Here, special emphasis is given to properties of the SH analysis on geographic grids, the block-diagonal SH analysis and the so-called kite solution as combination strategy with SGGMs. Eventually, validation strategies for the resulting EGGM are presented along with further possibilities to feedback the final model to the validation stage to iteratively improve the reduction, and, with that, also the final EGGM. Comparing the SGDTv1 model (as result of the GRAV-D test case) to other EGGMs by means of GNSS/levelling reveals that the experimental model performs equally or better than its competitors (in areas with observation coverage), which underlines the correct functionality of the integrated approach.

Contents

Abbreviations

1D	1-Dimensional
2D	2-Dimensional
3D	3-Dimensional
AVISO	Archiving, Validation and Interpretation of Satellite Oceanographic data
CPU	Central Processing Unit
d/o	Degree and Order
EGGM	Extended Global Gravity field Model (extended with ground gravity observations)
EGM2008	Earth Gravitational Model 2008
FSYNTH3D	Fast SYNTHesis method in the 3D space
GGM	Global Gravity field Model
GNSS	Global Navigation Satellite System
GOCE	Gravity field and steady-state Ocean Circulation Explorer
GOCO	Gravity Observation Combination
GOCO06s	Satellite-only model of GOCO group (version 6)
GRAV-D	Gravity for the Redefinition of the American Vertical Datum
GRS80	Geodetic Reference System 1980
HPC	High Performance Computing
HWHM	Half Width at Half Maximum
IAG	International Association of Geodesy
IAPG	Institute of Astronomical and Physical Geodesy
ICGEM	International Centre for Global Earth Models
IHRF	International Height Reference Frame
I/O	Input/Output
LSA	Least-Squares Adjustment
LSC	Least-Squares Collocation
NEQS	Normal EQuation System
NGA	National Geospatial-Intelligence Agency
NGS	National Geodetic Survey
PC	Personal Computer
PE-LSC	Partition-Enhanced Least-Squares Collocation
RAM	Random Access Memory
RCR	Remove-Compute-Restore

RMS	Root Mean Square
SATOP	SAtellite-TOPography model
SGDTv1	Satellite-GravD-Topography model Version 1
SGGM	Satellite-only Global Gravity field Model
SH	Spherical Harmonic or Spheroidal Harmonic
T/P	Topex/Poseidon
US	United States
XGM2019	eXperimental Gravity field Model 2019, up to d/o 719
XGM2019e	eXperimental Gravity field Model 2019 extended with topographic information

Chapter 1

Introduction

1.1 Motivation and Scope

The development of a consistent and accurate description of the Earth's gravity field through Global Gravity field Models (GGMs) is crucial for numerous applications in geosciences and, hence, one of the main tasks of physical geodesy. As important example, the International Association of Geodesy (IAG) has declared the definition and realization of an International (physical) Height Reference System (IHRS) as one of its main objectives. Since physical heights are strongly connected to the Earth's (exterior) gravity field, the IAG recognizes GGMs as one of the central aspects for the establishment of an International Height Reference Frame (IHRF, *Ihde et al.*, 2017). However, not only from GGMs physical heights can be obtained, but also any other gravity field functional, such as the potential, accelerations, deflections of the vertical or gradients. Thus, GGMs are very flexible in their application and may serve as input/constraint to any model influenced by gravity, which basically includes all mass-related dynamic physical processes.

Due to the importance of GGMs, significant effort has already been put into enhancing theories and methodologies behind global gravity field modelling, resulting in a variety of different approaches. While the developed methods may strongly vary, the use of spherical resp. spheroidal harmonics (both abbreviated with SHs in this thesis, see Sec. 2.1.2 and *Heiskanen & Moritz*, 1967) for the representation of the final GGM emerged as de-facto standard in geodesy over the last few decades (see *ICGEM*). Dealing with GGMs therefore means dealing with SHs and thus, also with the SH analysis (see Sec. 2.1.2). Here, it is imperative to notice that modern GGMs can be subdivided into two classes: (1) GGMs solely relying on satellite gravity field observations (henceforth denoted as SGGMs), and (2) GGMs extending SGGMs by including additional ground gravity field observations (henceforth called EGGMs). While SGGMs provide very accurate and globally homogeneous gravity field data, their resolution is inevitably limited due to the high-frequency damping caused by the upward continuation to satellite altitude of the gravity signal (see Sec. 2.1.2). Thus, increasing the resolution to enable a higher point-wise accuracy near or on the Earth's surface can only be achieved by means of EGGMs. Due to the high computational demand of directly performing a SH analysis on scattered high-resolution ground gravity datasets the ground gravity datasets have to be preprocessed for EGGMs. Hence, EGGM modelling through SHs can always be considered as an at least two-step approach, where the first step consists of the preparation of the input dataset, and the second step is the SH analysis itself. This two-step structure forms the common basis of most of the EGGM modelling approaches, which mainly differ in how the first step is performed. Usually, it is covered by regional gravity field modelling methods. Regional methods cover, among others, least squares collocation (LSC, *Krarup*, 1969; *Moritz*, 1980; *Forsberg & Tscherning*, 1981, and Sec. 2.3), Fourier transform, wavelets (*Chambodut et al.*, 2005; *Holschneider et al.*, 2003; *Schmidt et al.*, 2005), radial basis functions (e.g, *Marchenko*, 1998; *Schmidt et al.*, 2007; *Eicker*, 2008; *Lieb*, 2017) or Slepian functions (e.g., *Simons*,

2010). Although all those methods are based on different ideas, they all try to find an estimate for the generally ill-posed problem of regional gravity field modelling by introducing some additional assumptions or constraints to 'regularize' the underlying system. Up until now, this regional modelling step is mostly performed independently of the subsequent SH analysis, since gravity data is often just available in an already preprocessed state provided by certain countries or measurement campaigns. This circumstance is not least caused by the fact that raw gravity field observations are often still considered proprietary. So, all currently available GGMs may be considered as reliant on a patchwork of different regional models (presumably also derived through different methods) compiled into a global dataset. Obviously, such a patchwork is predestined to introduce errors and artefacts due to possible inconsistencies between the different regional models. Also, when working with preprocessed gravity field data, it is not possible to introduce any synergies between the SH analysis and the regional gravity field modelling by interconnecting both steps more strongly. As a consequence, *Ihde et al.* (2017) declared current EGGMs (e.g., EGM2008, *Pavlis et al.*, 2012) as inappropriate for the inclusion into the IHRF, mostly due to the unknown reliability of those models.

Understanding these shortcomings in the current state of global gravity field modelling directly leads to the definition of the purpose of this work, namely the study on how to generally improve future EGGMs. The question of how to improve EGGMs can conceptually be split into three main objectives:

O-1 *Improving existing state-of-the-art (regional and global) gravity field modelling methods*
To tackle this objective, for regional gravity field modelling, this thesis builds upon the established and widely used method of LSC, introducing several modifications that allows an efficient, consistent and large-scale application (see publications P-1, P-2 and Chaps. 3, 4). In the scope of the combined global gravity field modelling, the weighted least-squares-adjustment approach (see *Pail et al.*, 2018) is chosen as a starting point, improving the weighting strategy and extending it for very high-resolutions (see P-3 whereas in the course of Chap. 5 some novel assessments on this topic are presented which may shed a new light on it).

O-2 *Creating synergies among the different processing steps by presenting a unified approach*
The final aim of this thesis shall be the presentation of an integrated gravity modelling approach that reaches from in-situ observations to the final high-resolution EGGM, integrating data validation (Chap. 3) as well as regional (Chap. 4) and global gravity field modelling (Chap. 5) into a unified framework and workflow. The presented workflow is highly automatable with the ability to forward error information and feedback results between processing steps, thus producing synergies throughout the integrated approach (cf. Sec. 1.2).

O-3 *Identifying the theoretical limits and possible caveats of current (regional and global) gravity field modelling approaches*
To lead the way for future investigations, the limiting factors of the respective methods used within O-1 and O-2 are discussed by also presenting possible solutions. Additionally, some important caveats of the methods are highlighted.

1.2 Structure

This thesis is of cumulative nature, meaning that major parts are already published in peer-reviewed scientific journals. To create the link to the objectives defined in Sec. 1.1, the publications mainly discuss the individual methodological enhancements (O-1), while the thesis itself mainly aims to present the framework in which

these methods can be merged together to form a unified approach (O-2). Three publications with first authorship (named P-1, P-2, P-3), covering the majority of (O-1), are collected and reprinted in appendix A and contain a declaration of own contribution. Two further related publications with co-authorship (P-4, P-5), highlighting additional synergies, are not reprinted, but can be found in the Journal of Geodesy.

Since this thesis shall be an autonomous work, the content of the underlying publications is summarized for the core findings and integrated in the aforementioned holistic processing framework (O-2). In addition, if applicable, theoretical and practical limits and/or caveats are highlighted for the individual processing steps, providing a scientific added value to the work already performed within the publications (O-3). To keep the structure of this thesis as intuitive as possible, the three main chapters are aligned with the three main processing steps of the framework which luckily also coincide with the three main publications (P-1, P-2, P-3). Fig. 1.1 provides an overview of the proposed holistic framework and simultaneously describes the basic structure. To illustrate and validate the holistic framework a test case is included throughout this work where GRAV-D airborne gravity observations are included into a global model (see Sec. 1.3).

To give the reader the appropriate context the fundamental physical and mathematical backgrounds are recapitulated in Chap. 2. This encompasses the basics of potential theory (Sec. 2.1.1) as well as an introduction to the method of LSC (Sec. 2.3). As part of potential theory, special emphasis is given to the spherical resp. spheroidal harmonic transformation (Sec. 2.1.2) since it is elementary to GGM modelling (Secs. 2.1.3, 2.2 and 2.2.1). For LSC, next to describing the remove-compute-restore technique (Sec. 2.3.2) the focus is laid on the statistical interpretation of the gravity field by using homogeneous-isotropic covariance functions based on degree variances (Sec. 2.3.3).

In chapter 3, the proposed methods of gravity data validation and reduction are explained as summary of P-1. Thereby, the compilation of the combined satellite-topography EGGM (SATOP) is descibed, which forms the basis for gravity data reduction (Sec. 3.1). To allow a fast evaluation of the SATOP model on the actual gravity data locations a fast 3D synthesis method (FSYNTH3D) is introduced (Sec. 3.2). Using the FSYNTH3D method, reductions for the gravity observations can be efficiently calculated leading to reduced (residual) gravity observations. This residuals can then be evaluated in a statistical way (e.g., through thresholds, empirical covariance functions or by low-pass filtering) to eliminated outliers and/or remove offsets or long-wavelenght effects (Sec. 3.3). The final aim of this chapter is to derive validated, reduced gravity observations from raw gravity observations (see Fig. 1.1, yellow box).

Starting from scattered, but already validated and reduced gravity observations, a gridding of the data can be performed according to the first step of the two step EGGM modelling approach defined in Sec. 1.1. This task is described within chapter 4 (see Fig. 1.1, orange box), recapitulating the content of publication P-2 (PE-LSC): in the first stage, the possibly oversampled observations are thinned out to reduce the overall number of observation and, hence, also reduce the computational effort for later steps (Sec. 4.1). Then, a spherical approximation to the coordinates is applied to remove the latitude-dependency of the covariance function derived by spherical degree variances (Sec. 4.2). Based on the observation with the modified coordinates, an isotropic empirical 2D-covariance function is obtained, which can then be directly correlated to degree variances (Sec. 4.3). Having estimated the degree variances, a least squares collocation of the observations can finally be performed to acquire a regular observation grid (Sec. 4.6). Within LSC, special emphasis is given achieving an optimized computation time by introducing a partitioning of the collocation area (Sec. 4.4). Simultaneously, a localization of the LSC kernel is introduced to minimize the fringe effects caused by the partitioning (Sec. 4.5).

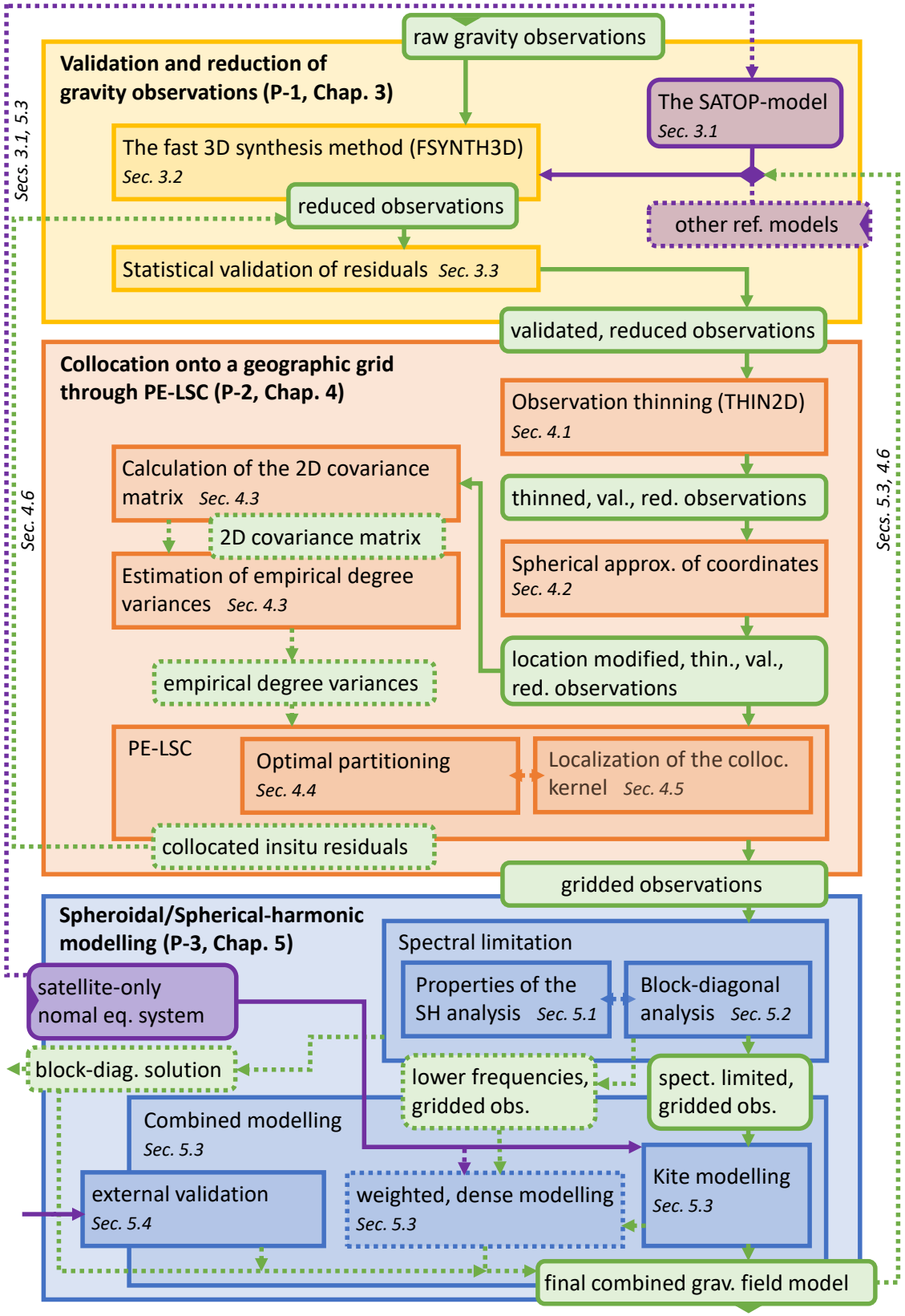


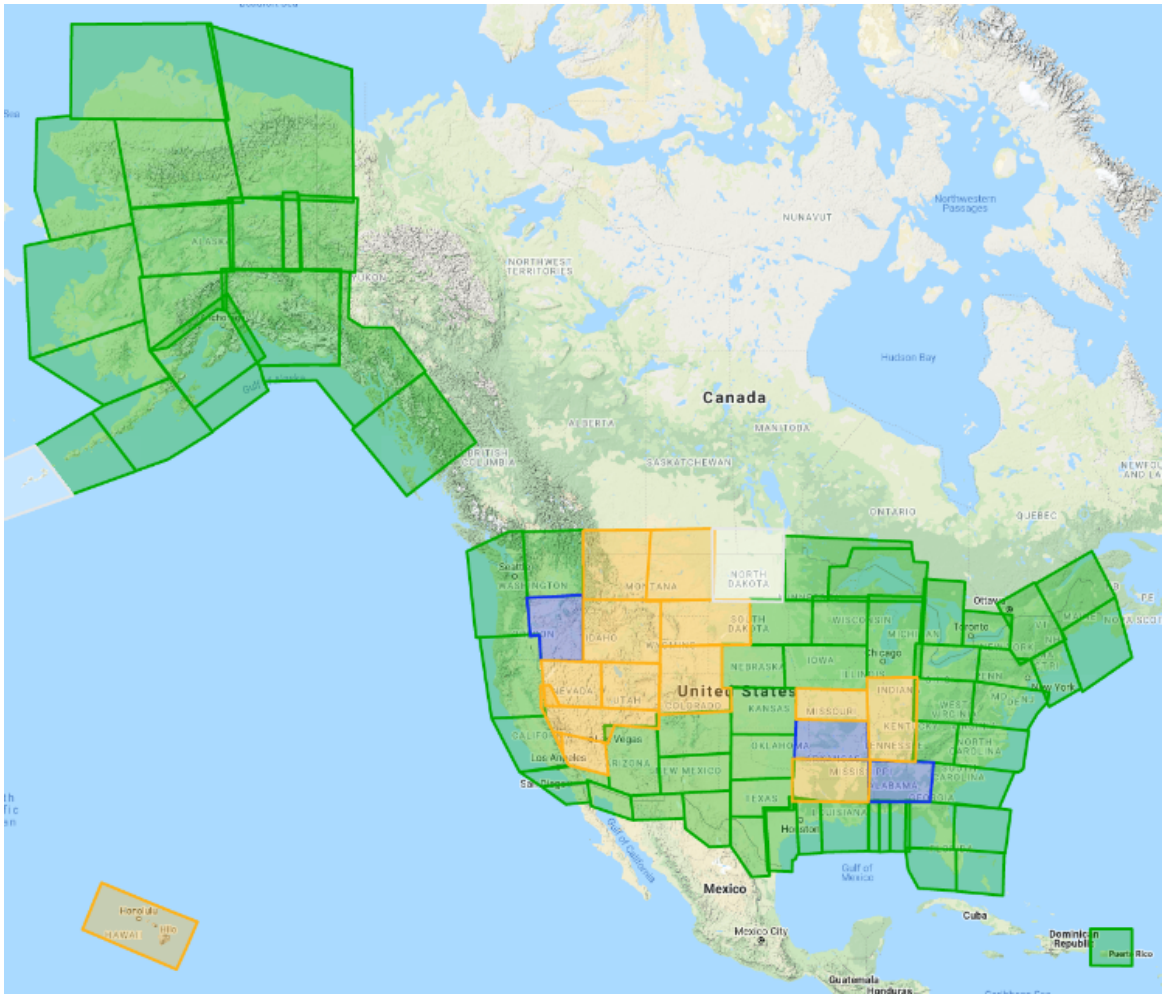
Fig. 1.1 – Overview over the holistic processing chain and, simultaneously, the structure of this thesis. The three main chapters (3, 4, 5) correspond to one of the rectangular main blocks resp. processing steps: Validation and reduction (Chap. 3, yellow block), Localized gravity field modelling (Chap. 4), orange block), Global gravity field modelling (Chap. 5, blue block). The finer structure of the processing is provided through the smaller rectangular blocks within the main blocks. Data is visualized in rounded boxes - green boxes highlight data derived from observations and purple boxes emphasize data which is based on auxiliary information (e.g., other gravity field models). Dotted boxes and arrows signalize optional or alternative processing steps resp. data products.

Finally, when gridded observations are available, an EGGM can be derived. Chapter 5 (see Fig. 1.1, blue box) summarizes the respective workflow from publication P-3 (at least partially). Firstly, the geographic input grid needs to be spectrally limited to the target resolution. To achieve this, the behavior of the SH transformation on geographic grids is investigated (Sec. 5.1) and a block-diagonal SH-analysis is performed, leading to a block-diagonal solution of the gridded observations (Sec. 5.2). The lower frequency part of this block-diagonal solution can then be used to synthesize a lower resolution geographic grid. Subsequently, this lower resolution grid can be combined with a satellite-only gravity field model through a weighted dense least squares adjustment approach. The needed accuracy information (for the weights) of the lower resolution grid can be obtained empirically by comparing it to a so-called kite solution. In a final step, the outcome of the weighted least squares adjustment, i.e., the lower frequency combined model, can be merged together with higher frequency part of the block-diagonal solution to compose the final model (Sec. 5.3). However, it has been shown during the implementation that for the integrated approach it is not even necessary to calculate the weighted combination (which has several disadvantages) because a kite solution is sufficient (Secs. 5.3, 5.4). Concluding this thesis, an outlook to the future of high-resolution global gravity field modeling is provided in chapter 6.

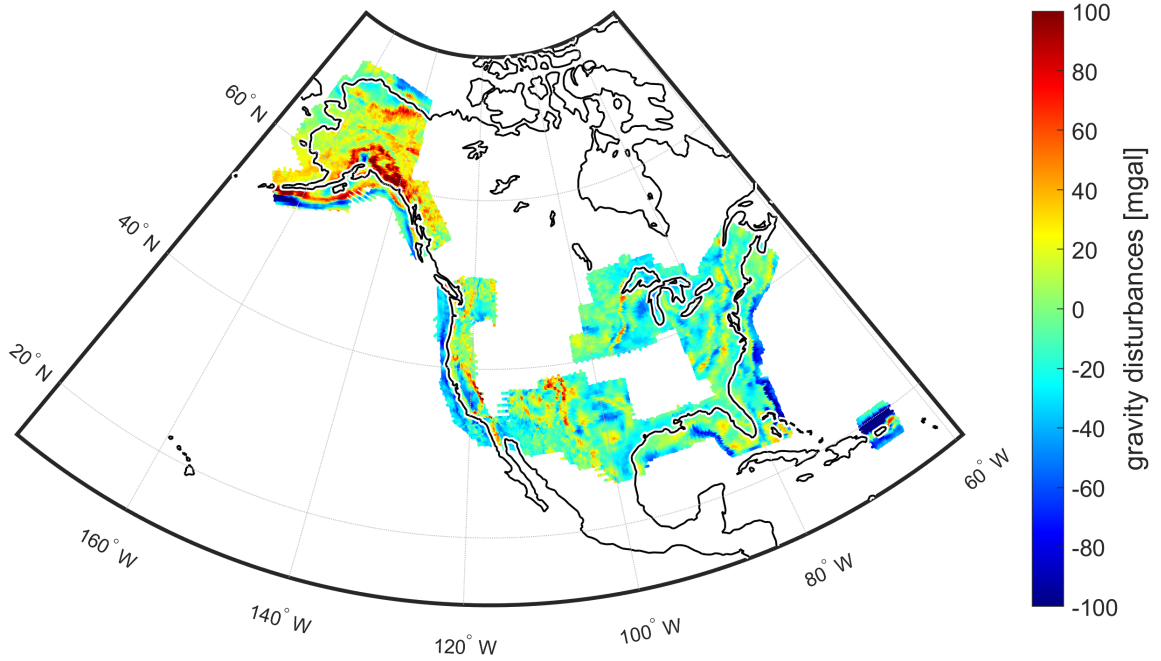
1.3 The GRAV-D test case

To showcase the possibilities of the presented approach a test scenario is included within this thesis. The presented scenario is also meant to accompany the introduction of the different processing steps to provide a concrete context and to give a better illustration. Specifically, the scenario consists of the integration of airborne gravity observations from the GRAV-D project into a SH global model. The GRAV-D project is initiated and managed by the US National Geodetic Survey (NGS) with the primary goal to provide a high-resolution "snapshot" of gravity in the US. As such, GRAV-D consists of an extensive continent-wide collection of airborne gravity observations which are organized in regionally confined blocks, which are processed independently to a certain degree. The data acquisition in the project is designed to enable a target resolution of about 10 km, hence, the nominal distance of adjacent flight tracks is also defined to be 10 km. It is also noteworthy, the campaign (also within specific block) is conducted by a variety of different instruments and airplanes, flown at different heights, resulting in a collection of observations which can be considered variable in accuracy, spatial resolution and spectral content. For more details, the reader is referred to the GRAV-D publications found on their dedicated website (https://geodesy.noaa.gov/GRAV-D/index.shtml). An overview of the covered area and available blocks is given in Fig. 1.2a.

For the test case, data from all blocks where the processing has been finished is taken (cf. Fig. 1.2a, all green blocks, except Kansas). The data is provided by courtesy of NGS in an already preprocessed state (corrected for offsets and high-frequency noise) where the along-track sampling is already reduced (to about 1 km). Additionally, the observations are already reduced to gravity disturbances by removing the normal field of the reference ellipsoid (GRS80 in this case, see *Moritz*, 2000). All in all, the complete dataset consists of about 1.2 million observations which are located on an average altitude of about 6km (see Fig. 1.2b). Although the presented test case has mainly illustrative character in the scope of this thesis, there will also be an actual application for the obtained results. Namely, NGS defined the goal to optimally integrate the GRAV-D observations into a global model with the spatial target resolution of 10 km since they recognize the SH representation as the most versatile for 3rd party usage. For instance, NGS itself wants to use the GRAV-D dataset to validate its terrestrial gravity data source against it. Obviously, for this application a downward



(a)



(b)

Fig. 1.2 – The GRAV-D dataset. (a) Overview of the current processing state of the individual blocks (as of May 2021). Green blocks: data available. Blue blocks: data being processed. Yellow blocks: data acquisition in progress. White blocks: data acquisition planned. Map taken from the GRAV-D website (https://geodesy.noaa.gov/GRAV-D/index.shtml) on 05/05/2021. (b) Visualization of the in-situ observations in terms of gravity disturbances.

continuation of the GRAV-D observations to the terrestrial measurement sites is necessary. Performing a downward continuation is straightforward when having a SH model, which makes EGGMs optimally suited for this task (although, there exist other methods which also have this capability perhaps with some limitations, e.g., LSC, see Sec. 2.3).

Chapter 2

Basic theory on gravity field modelling

This chapter introduces the mathematical and physical principles needed by the various processing steps. Since the shown theory is commonly used in geodesy, for most parts only a brief summary is provided, collecting the most important formulas. An in-depth treatment of the addressed theory can be found, e.g., in the comprehensive work of *Heiskanen & Moritz* (1967) or *Moritz* (1980).

2.1 Mathematical description of the gravity field

2.1.1 Potential theory

In this work, Newtonian law of gravity is used as the fundament for gravity field modelling. This means that relativistic effects are not taken into account for the modelling part (although these may have been considered in the course of the acquisition of the measurements themselves, e.g., within satellite gravity missions). However, in case of the Earth's comparatively weak gravity, the direct perturbations of Newtonian law due to relativity can be neglected regarding the targeted accuracies and application scenarios of the resulting gravity field models. For a sufficiently small test mass P located at $\underline{x}_P$ the attracting gravitational acceleration regarding a body Q reads:

$$\underline{a}\left(\underline{x}_P\right) = \iiint_{\mathcal{X}_Q} G \overbrace{\rho\left(\underline{x}_Q\right)}^{\rho_Q} \frac{\overbrace{\underline{x}_Q - \underline{x}_P}^{\underline{r}_{PQ}}}{\underbrace{\left|\underline{x}_Q - \underline{x}_P\right|^3}_{r_{PQ}}} \, d\mathcal{X}_Q = G \iiint_{\mathcal{X}_Q} \rho_Q \frac{\underline{e}_{PQ}}{r_{PQ}^2} \, d\mathcal{X}_Q. \tag{2.1}$$

Here, $\mathcal{X}_Q$ denotes the volume of the body Q, $\rho\left(\underline{x}_Q\right)$ the density at $\underline{x}_Q$ and G the gravitational constant. It can easily be shown (see, e.g., *Heiskanen & Moritz*, 1967) that $\underline{a}$ has zero curl, i.e. $\nabla \times \underline{a} = 0$. This also implies that a scalar potential $\mathcal{V}\left(\underline{x}_P\right)$ to $\underline{a}\left(\underline{x}_P\right)$ exists, such that $\underline{a}\left(\underline{x}_P\right) =: \nabla \mathcal{V}\left(\underline{x}_P\right)$. By integrating the components of Eq. 2.1 one finds:

$$\mathcal{V}\left(\underline{x}_P\right) = \iiint_{\mathcal{X}_Q} \frac{G\rho\left(\underline{x}_Q\right)}{\left|\underline{x}_Q - \underline{x}_P\right|} \, d\mathcal{X}_Q = G \iiint_{\mathcal{X}_Q} \frac{\rho_Q}{r_{PQ}} \, d\mathcal{X}_Q. \tag{2.2}$$

Thus, the gravitational acceleration $\underline{a}$ as a vector field can be expressed more conveniently through a scalar field, i.e., through the respecitve potential $\mathcal{V}$. As an additional crucial property of the gravitational potential, it can be shown, that the Laplacian $\Delta\mathcal{V} := \nabla\nabla\mathcal{V}$ of the potential $\mathcal{V}$ is zero in the absence of masses. To see this,

one may use the fact, that the divergence operator $\nabla\underline{a}$ can be derived from the divergence theorem on $\underline{a}$ in its differential form (see, e.g., *Bronstein et al.*, 2015):

$$\begin{aligned}
\Delta\mathcal{V} = \nabla\underline{a}\,(\underline{x}_P) &= \frac{\lim\limits_{\mathcal{X}_Q\to\underline{x}_P}\left(\oiint_{\partial\mathcal{X}_Q}\underline{a}\,d\big[\partial\underline{\mathcal{X}}_Q\big]\right)}{d\mathcal{X}_Q} = \\
&= \oiint_{\partial\mathcal{X}_Q}\frac{\lim\limits_{\mathcal{X}_Q\to\underline{x}_P}(\underline{a})}{d\mathcal{X}_Q}\,d\big[\partial\underline{\mathcal{X}}_Q\big] = \\
&= \oiint_{\partial\mathcal{X}_Q} G\rho\,(\underline{x}_P)\lim_{\mathcal{X}_Q\to\underline{x}_P}\left(\iiint_{\mathcal{X}_Q}\frac{\underline{x}_Q-\underline{x}_P}{\left|\underline{x}_Q-\underline{x}_P\right|^3}\frac{d\mathcal{X}_Q}{d\mathcal{X}_Q}\right)d\big[\partial\underline{\mathcal{X}}_Q\big] = \\
&= \oiint_{\Omega} G\rho\,(\underline{x}_P)\left(\frac{\underline{e}_{P'P}}{\delta r^2_{P'P}}\right)\,\big[-\underline{e}_{P'P}\,\delta r^2_{P'P}\,d\Omega\big] = \\
&= \oiint_{\Omega} -G\rho\,(\underline{x}_P)\,d\Omega = -4\pi G\rho\,(\underline{x}_P)\,.
\end{aligned} \tag{2.3}$$

In this equation $[\partial\mathcal{X}_Q]$ denotes the boundary (i.e., surface) of the volume $\mathcal{X}_Q$ (of the body Q) and $d\left[\partial\underline{\mathcal{X}}_Q\right] = d\left[\partial\mathcal{X}_Q\right]\underline{n}$ a directed surface element (regarding the normal direction $\underline{n}$). Since $\rho\,(\underline{x}_P)$ becomes zero in empty space, also the Laplacian of $\mathcal{V}$ is zero. A function $\mathcal{V}$ for which $\Delta\mathcal{V} = 0$ holds true is called harmonic and the stated harmonicity constraint defines the class of Laplacian differential equations. For Laplacian differential equations it is known that a solution (that is, the function values for $\mathcal{V}$) for a harmonic space (i.e., a space where the function is harmonic) can be obtained by knowledge of the function values on the boundary of this space (by means of a boundary value problem). In the actual case of gravity field modelling, when assuming that spatially confined bodies (e.g., the Earth) are defining the boundary to the empty (i.e., harmonic) space, the gravitational potential at every position in this empty space can be determined by knowing the gravitational potential just on the boundary (e.g., the surface of the body). This harmonicity property forms the basis for the definition of the so-called spherical resp. spheroidal harmonic representation as a solution to the Laplacian differential equation regarding spherical resp. spheroidal coordinates (see Sec. 2.1.2).

2.1.2 Spherical resp. spheroidal harmonic representation

Spherical synthesis. In most applications the gravity field of a spherical or spheroidal-shaped body is of interest (as, e.g., the Earth or other planets or larger bodies). For such bodies, it is convenient to describe the gravity field regarding coordinates that follow their shape. Considering a spheroidal body (such as the Earth) an appropriate coordinate system would be the spherical coordinates in a first approximation and the spheroidal (i.e., ellipsoidal) coordinates in a closer approximation. A solution of the Laplacian differential equation $\Delta\mathcal{V} = 0$ can then be sought regarding the respective coordinate system. A detailed solution path for both coordinate systems is presented, e.g., in *Heiskanen & Moritz* (1967). The solution in spherical coordinates (θ, λ, r) is given by the so called (solid) spherical harmonic synthesis, defined as:

$$\mathcal{V}\,(\theta,\lambda,r) = \sum_{n=0}^{n_{max}}\left(\frac{R}{r}\right)^{n+1}\sum_{m=0}^{n}\bar{P}_{nm}\,(\cos\theta)\,(C^s_{nm}\cos m\lambda + S^s_{nm}\sin m\lambda) \tag{2.4}$$

where θ is the spherical (geocentric) co-latitude, λ the longitude, r the spherical (geocentric) radius and R some (constant) reference radius. $\bar{P}_{nm}$ denotes the so-called (fully normalized) associated Legendre polynomials

(see, e.g., *Fukushima*, 2012a). In the stated form, Eq. 2.4 defines the so-called spherical harmonic domain, which is given by the index n, called the degree and the index m, called the order of a specific element (called coefficient) C/S^s_{nm} in this domain. Theoretically, the series in Eq. 2.4 has to be expanded up to infinity, i.e., $n_{max} = \infty$. However, in practical applications, the maximum degree and order (d/o) n_{max} is set to a finite integer value which corresponds to a limitation of the maximum frequency (or minimum wavelength) that is covered by the synthesis. In practice, the harmonic coefficients C/S^s_{nm} are usually given in a normalized fashion c/s_{nm}, such that:

$$c/s^s_{nm} \coloneqq \frac{GM}{R} C/S^s_{nm}. \tag{2.5}$$

with GM as the product of total mass M and gravitation constant G so that the coefficient of the central term of the expansion $c_{00} = 1$. Such normalized coefficients c/s_{nm} are usually referred to as Stokes coefficients. From Eq. 2.4 one can also already see the numerical instabilities induced by the radial component $(R/r)^{n+1}$. As an example, when considering the flattening $f \approx 1/300$ of the Earth, the radial ratio for a point located on the equator and a point located on the pole is given by $1 - f \approx 300/299$. This means that the ratio of the radial terms for both points is given by $(300/299)^n$, which becomes as large as $1 \cdot 10^3, 7 \cdot 10^7, 5 \cdot 10^{15}$ for $n = 2160, 5400, 10800$. Since the arithmetic precision of modern computers is usually limited to $\sim 10^{16}$ (double precision) this exponential amplification quickly becomes uncontrollable. Hence, currently, as a rule of thumb for the application on the Earth's surface, spherical harmonics should not be used higher than d/o 5400 (especially within adjustments, see also Sec. 2.1.2 and 5.2).

Spheroidal synthesis. As alternative to the spherical expansion also a spheroidal expansion of the harmonic potential (regarding spheroidal coordinates) is possible (see *Heiskanen & Moritz*, 1967). For simplicity, in this thesis, the use of the spheroidal expansion is limited to spheroidal surfaces, on which it has an especially simple form:

$$\mathcal{V}(\beta, \lambda, h_{ell} = 0) = \sum_{n=0}^{n_{max}} \sum_{m=0}^{n} \bar{P}_{nm}(\sin\beta) \left(C^e_{nm} \cos m\lambda + S^e_{nm} \sin m\lambda\right). \tag{2.6}$$

Comparing Eq. 2.4 with Eq. 2.6, it can be seen that they are quite similar despite the fact that in Eq. 2.6 the radial component is missing and that the spherical colatitude θ is substituted by the parametric latitude β (connected by the flattening f through $\cot\theta = (1 - f)\tan\beta$). As mentioned, Eq. 2.6 applies for potential values on the surface of the respective spheroid ($h_{ell} = 0$) and the corresponding coefficients C^e_{nm} and S^e_{nm} differ from the spherical counterparts in case that the flattening is not zero (in the special case of $f = 0$ Eq. 2.4 is equivalent to Eq. 2.6). Spheroidal coefficients C/S^e_{nm} and spherical coefficients C/S^s_{nm} are related through the transform described in *Jekeli* (1981):

$$C/S^s_{nm} = \sum_{n=m}^{n_{max}} t^{e\to s}_{nm}\, C/S^e_{nm}, \quad C/S^e_{nm} = \sum_{n=m}^{n_{max}} t^{s\to e}_{nm}\, C/S^s_{nm}. \tag{2.7}$$

The required transformation coefficients $t^{e\to s}_{nm}$, $t^{s\to e}_{nm}$ can be calculated in a numerically stable way (see *Jekeli*, 1981) and, thus, also the transformation itself is stable. Noteworthy, only harmonic coefficients of a certain order m are related through this transform, reflecting the fact that the effect of oblateness is limited to the latitudinal direction. Hence, using the spheroidal harmonic expansion (Eq. 2.6) as well as the transformation to spherical harmonics (Eq. 2.7), the previously mentioned problem of the numerical instabilty of the spherical harmonic expansion on the Earth's surface can be prevented (e.g., within adjustments).

Please note that in this thesis the term spheroid (resp. spheroidal) is used to denote an ellipsoid of revolution (i.e., with two identical principal axes) to distinguish from arbitrary ellipsoids (with three arbitrary principal axes). This seems necessary since next to spheroidal harmonics also (3-axis) ellipsoidal harmonics (see, e.g., *Romain & Jean-Pierre*, 2001) exist, which are not used within this thesis (because the Earth can be well approximated through a spheroid). Further, since many statements in this thesis are valid for spherical harmonics and spheroidal harmonics alike the abbreviation SH(s) will be used to simultaneously address both.

Orthogonal linear operator. For later considerations, it is useful to interpret the harmonic expansion (Eq. 2.4 or Eq. 2.6) as a linear operator $\mathrm{Y}_{\mathcal{H}}^{\mathcal{X}}$ that transforms from the spectral domain $\mathcal{H}$ to the spatial domain $\mathcal{X}$:

$$\begin{aligned}
&f_{\mathcal{X}} = \mathrm{Y}_{\mathcal{X}}^{\mathcal{H}}\, f_{\mathcal{H}} \\
&\mathcal{X} \coloneqq \left\{(\theta,\lambda,r)\,|\theta \in \mathbb{R} \bmod \pi, \lambda \in \mathbb{R} \bmod 2\pi,\ r \in \mathbb{R}^{+}\right\} \\
&\mathcal{H} \coloneqq \{(n,m)\,|n \in \mathbb{N}, m \in \mathbb{Z}\} \\
&f_{\mathcal{X}} = f_{\mathcal{X}}\,(\theta,\lambda,r) \coloneqq \mathcal{V}\,(\theta,\lambda,r) \\
&f_{\mathcal{H}} = f_{\mathcal{H}}\,(n,m) \coloneqq \frac{GM}{R}\begin{cases} c_{nm} & \Leftrightarrow m \geq 0 \\ s_{nm} & \Leftrightarrow m < 0 \end{cases} \\
&\mathrm{Y}_{\mathcal{X}}^{\mathcal{H}} = \mathrm{Y}\left((\theta,\lambda,r) \in \mathcal{X},\,(n,m) \in \mathcal{H}\right) \coloneqq \left(\frac{R}{r}\right)^{n+1} \bar{P}_{n|m|}\,(\cos\theta)\begin{cases} \cos m\lambda & \Leftrightarrow m \geq 0 \\ \sin|m|\lambda & \Leftrightarrow m < 0 \end{cases}
\end{aligned} \tag{2.8}$$

For this linear operator $\mathrm{Y}_{\mathcal{X}}^{\mathcal{H}}$ it is important to notice that spatial base functions $\mathrm{Y}_{nm}\,(\theta,\lambda,r=R)$ are orthogonal regarding the integration over the sphere (resp. spheroid), i.e.:

$$\iint_{\Omega} \mathrm{Y}_{nm}\,(\theta,\lambda)\,\mathrm{Y}_{n'm'}\,(\theta,\lambda)\ d\Omega = 4\pi\delta\,(n,n')\,\delta\,(m,m')\,. \tag{2.9}$$

Analysis. Using this property, Eq. 2.4, resp. Eq. 2.8 can be inverted, yielding the so-called analysis formula for spherical resp. spheroidal (surface) harmonics:

$$\begin{aligned}
\iint_{\Omega} \mathrm{Y}_{nm}\,(\theta,\lambda)\,f_{\mathcal{X}}\,(\theta,\lambda)\ d\Omega &= \iint_{\Omega} \mathrm{Y}_{nm}\,(\theta,\lambda)\left[\sum_{n'=0}^{n_{max}}\sum_{m'=0}^{n'} Y_{n'm'}\,(\theta,\lambda)\,f^{\mathcal{H}}\,(n,m)\right] d\Omega = \\
&= \sum_{n'=0}^{n_{\infty}}\sum_{m'=0}^{n'} \overbrace{\left[\iint_{\Omega} \mathrm{Y}_{nm}\,(\theta,\lambda)\,Y_{n'm'}\,(\theta,\lambda)\ d\Omega\right]}^{4\pi\delta(n,n')\delta(m,m')} f_{\mathcal{H}}\,(n,m) = 4\pi f_{\mathcal{H}}\,(n,m) \\
\Rightarrow\quad f_{\mathcal{H}}\,(n,m) &= \frac{1}{4\pi}\iint_{\Omega} \mathrm{Y}_{nm}\,(\theta,\lambda)\,f_{\mathcal{X}}\,(\theta,\lambda)\ d\Omega.
\end{aligned} \tag{2.10}$$

Thus, by knowledge of the continuous potential field on the body's surface, harmonic coefficients can be derived independently of each other due to the given orthogonality. However, the orthogonality (Eq. 2.9) does in general not hold in the application-related discrete scenario, where, e.g., potential field values are just provided through some sort of sampling. Hence, in the realistic discrete case, the analysis usually needs more attention and generally leads to a formulation through least squares adjustment (LSA, see 2.2).

2.1.3 Disturbing potential and derived quantities

Since the gravitational potential field $\mathcal{V}$ of large bodies (as the Earth) is usually quite homogeneous on large scales due to the isostasy driven by its own masses, it is reasonable to remove a reference potential $\mathcal{V}'$ to obtain a so-called disturbing potential field $\mathcal{T}$:

$$\mathcal{T} \coloneqq \mathcal{V} - \mathcal{V}'. \tag{2.11}$$

Normal gravity. For the Earth, a well-approximating reference potential $\mathcal{V}'$ is found by modelling the gravitational potential of a so-called (physical) reference ellipsoid resp. spheroid (see, e.g., *Heiskanen & Moritz*, 1967). In this context, the obtained reference field is usually denoted as normal field, since the direction of the gradient of this potential field (plus the centrifugal potential) is defined to be aligned with the normal direction of the spheroidal surface of the corresponding reference ellipsoid. A (physical) reference spheroid E of a body is usually described by (1) its geometric form, e.g., through the semi-major axis a_E and the flattening f_E, (2) GM_E the product of the total mass M_E and gravitational constant G, and (3) the angular velocity ω_E, that is $\mathcal{V}'_E = \mathcal{V}'\left(a_E, f_E, GM_E, \omega_E\right)$. Ultimately, since $\mathcal{V}'_E$ is harmonic, it can also be expressed through a spherical harmonic expansion, only concerning the even zonal ($m = 0$) coefficients $c^E_{2n0}\left(a_E, f_E, GM_E, \omega_E\right)$ due to rotational and equatorial symmetry:

$$\mathcal{V}'_E\left(\theta, r\right) = \frac{GM}{R} \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{2n+1} \bar{P}_{2n0}\left(\cos\theta\right)\, c^E_{2n0} \tag{2.12}$$

For details about the calculation of the coefficients c^E_{2n0} the reader is referred to *Heiskanen & Moritz* (1967). It shall be noted that in this considerations the centrifugal potential of rotating bodies, which is not harmonic, is assumed to be already removed beforehand (by using ω_E). Since the harmonic synthesis is a linear operation, the disturbing potential $\mathcal{T}$ can be expressed directly via Stokes coefficient differences $\Delta f_{\mathcal{H}}$ (cf. Eq. 2.8):

$$\mathcal{T} = \mathcal{V} - \mathcal{V}' = \mathrm{Y}^{\mathcal{H}}_{\mathcal{X}}\, f_{\mathcal{H}} - \mathrm{Y}^{\mathcal{H}}_{\mathcal{X}}\, f_{\mathcal{H}}{}' = \mathrm{Y}^{\mathcal{H}}_{\mathcal{X}} \overbrace{\left(f_{\mathcal{H}} - f_{\mathcal{H}}{}'\right)}^{\Delta f_{\mathcal{H}}} =: \mathrm{Y}^{\mathcal{H}}_{\mathcal{X}}\, \Delta f_{\mathcal{H}} \tag{2.13}$$

Disturbing gravity field quantities. Based on the disturbing potential $\mathcal{T}$, other disturbing gravity field functionals $\delta f_?$ can be derived by means of derivatives of $\mathcal{T}$, such as, e.g., the gravity disturbance, the deflection of the vertical, or the gravity gradient. When using derivates along the axis of the coordinate system of the harmonic expansion (e.g. spherical coordinates), the functionals (in the respective approximation) can be expressed as:

$$\begin{aligned} \delta f_? = \frac{\partial^{i+j+k}\mathcal{T}\left(\theta, \lambda, r\right)}{\partial\theta^i \partial\lambda^j \partial r^k} &= \frac{\partial^{i+j+k}\mathrm{Y}^{\mathcal{H}}_{\mathcal{X}}}{\partial\theta^i \partial\lambda^j \partial r^k}\, \Delta f_{\mathcal{H}} \\ \frac{\partial^{i+j+k}\mathrm{Y}^{\mathcal{H}}_{\mathcal{X}}}{\partial\theta^i \partial\lambda^j \partial r^k} &= \frac{\partial^i \bar{P}_{n|m|}\left(\cos\theta\right)}{\partial\theta^i}\, \frac{\partial^j \cos/\sin|m|\lambda}{\partial\lambda^j}\, \frac{\partial^k \left(\frac{R}{r}\right)^{n+1}}{\partial r^k} \end{aligned} \tag{2.14}$$

Hence, it is discernible that all those functionals $\delta f_?$ can be derived from the same set of disturbing Stokes coefficients $\Delta f_{\mathcal{H}}$ by just modifying the linear operator $\mathrm{Y}^{\mathcal{H}}_{\mathcal{X}}$ accordingly (for the computation of the derivatives of the Legendre polynomials, see, e.g., *Fukushima*, 2012c). This property makes the Stokes coefficients $\Delta f_{\mathcal{H}}$ universally applicable, which is probably the main reason why they became the de facto standard in geodesy for

representing/defining gravity fields. Among all the different feasible gravity field functionals (see, e.g., *Meissl*, 1971), in this work, special emphasis is given to the spherically approximated gravity disturbance $\delta g \approx -\partial \mathcal{T}/\partial r$ since it is used as the primary observable to model the GRAV-D dataset:

$$\begin{aligned} \delta g \approx -\frac{\partial \mathcal{T}}{\partial r} = \frac{\partial \mathrm{Y}_{\mathcal{X}}^{\mathcal{H}}}{\partial r}\,\Delta f_{\mathcal{H}} = -\frac{n+1}{r}\mathrm{Y}_{\mathcal{X}}^{\mathcal{H}}\,\Delta f_{\mathcal{H}} =: \mathrm{Y}_{\delta g}^{\mathcal{H}}\Delta f_{\mathcal{H}} = \\ = \sum_{n=0}^{n_{max}} -\frac{n+1}{R}\left(\frac{R}{r}\right)^{n+2}\sum_{m=0}^{n}\bar{P}_{nm}\left(\cos\theta\right)\left(C_{nm}^{s}\cos m\lambda + S_{nm}^{s}\sin m\lambda\right) \end{aligned} \tag{2.15}$$

Anomalous gravity field quantities. Note that next to the disturbing gravity field quantities, in geodesy also the so-called anomalous gravity field quantities are historically of importance, since in earlier days the ellipsoidal height of the observation sites was generally unknown. Thus, relying on physical heights, the geoid height resp. height anomaly (as difference to the ellipsoidal height) had to be co-estimated which led to the formulation of the so-called fundamental equation of physical geodesy and, eventually, to the definition of the so-called gravity anomaly Δg as counterpart to the gravity disturbance. However, since the theory behind anomalous gravity field quantities is of minor importance for this work, it is omitted at this point and the reader is referred to *Heiskanen & Moritz* (1967) for a detailed explanation. Ultimately, anomalous quantities can be expressed similarly to disturbing items by means of Stokes coefficients by modifying the linear operator $\mathrm{Y}_{\mathcal{H}}^{\mathcal{X}}$ (see Eq. 2.15, e.g., $\mathrm{Y}_{\Delta g}^{\mathcal{H}} := -\frac{n-1}{r}\mathrm{Y}_{\mathcal{X}}^{\mathcal{H}}$).

Block-mean quantities. Block-mean quantities $\delta\bar{f}_{?}$ regarding spherical or geographic surface elements $(\Delta\theta, \Delta\lambda)$ may be defined as

$$\begin{aligned} \delta\bar{f}_{?}\left(\theta_0,\lambda_0,r\right) = \int_{\theta_0}^{\theta_0+\Delta\theta}\int_{\lambda_0}^{\lambda_0+\Delta\lambda}\left[\frac{\partial^{i+j+k}\mathcal{T}\left(\theta,\lambda,r\right)}{\partial\theta^i\partial\lambda^j\partial r^k}\right]\partial\theta\,\partial\lambda = \overbrace{\int_{\theta_0}^{\theta_0+\Delta\theta}\int_{\lambda_0}^{\lambda_0+\Delta\lambda}\left[\frac{\partial^{i+j+k}\mathrm{Y}_{\mathcal{X}}^{\mathcal{H}}}{\partial\theta^i\partial\lambda^j\partial r^k}\right]\partial\theta\,\partial\lambda}^{=:\bar{\mathrm{Y}}_{?}^{\mathcal{H}}}\,\Delta f_{\mathcal{H}} \\ \bar{\mathrm{Y}}_{?}^{\mathcal{H}} = \int_{\theta_0}^{\theta_0+\Delta\theta}\left[\frac{\partial^i\bar{P}_{n|m|}\left(\cos\theta\right)}{\partial\theta^i}\right]\partial\theta\int_{\lambda_0}^{\lambda_0+\Delta\lambda}\left[\frac{\partial^j\cos/\sin|m|\lambda}{\partial\lambda^j}\right]\partial\lambda\,\left[\frac{\partial^k\left(\frac{R}{r}\right)^{n+1}}{\partial r^k}\right]. \end{aligned} \tag{2.16}$$

Here, the ? indicates that possibly different gravity field functionals (as disturbances, anomalies, gradients, etc.) may be addressed by the block-mean operation (cf. Eq. 2.14). As seen, for the block-mean quantities the integrals of the associated Legendre polynomials $\bar{P}_{n|m|}$ are needed which introduces some additional effort regarding their computation (see, e.g., *Fukushima*, 2014 and *Fukushima*, 2012b). On the other hand, obtaining the integrals of the trigonometric functions in longitudinal direction is straightforward since they remain basically unchanged (except of a scaling and a phase shift). This circumstance will be important when deriving orthogonality properties (see Sec. 2.2.1).

2.2 Least squares adjustment

As highlighted in the previous section (2.1.2), a determination of Stokes coefficients is theoretically possible through knowledge of the potential field given continuously on a boundary surface of the body of interest. It has been shown that in the continuous spherical resp. spheroidal case the spatial basis functions

$\mathrm{Y}_{nm}(\theta, \lambda, r = 0)$ are orthogonal (see Eq. 2.9) and Stokes coefficients can be derived independently of each other by means of integration (see Eq. 2.10). In practical applications, though, it can neither be assumed that the potential field is known in a continuous manner, nor that the given boundary surface is spherical resp. spheroidal. Thus, instead of having a continuous operator $\mathrm{Y}_{?}^{nm}(\theta, \lambda, r = 0)$ an arbitrary discrete operator (i.e., matrix)

$$\mathrm{A}_{?}^{\mathcal{H}} = \mathrm{A}_{?}(\underline{\boldsymbol{x}}, \underline{\boldsymbol{h}}) \coloneqq \mathrm{Y}_{?}((\theta_i, \lambda_i, r_i) =: \underline{\boldsymbol{x}} \subset \mathcal{X},\ (n_k, m_k) =: \underline{\boldsymbol{h}} \subset \mathcal{H}) \tag{2.17}$$

is applied. Having a finite set $\underline{\boldsymbol{h}}$ of Stokes coefficients $\underline{f} \coloneqq \Delta f_{\mathcal{H}}(\underline{\boldsymbol{h}})$ describing some gravity field functionals $\underline{f}_{?} \coloneqq \Delta f_{?}(\underline{\boldsymbol{x}})$ on a finite set $\underline{\boldsymbol{x}}$ of spatial locations, the following linear relation can be established (according to Eq. 2.10):

$$\underline{f}_{?} = \mathrm{A}_{?}^{\mathcal{H}}\, \underline{f}_{\mathcal{H}}. \tag{2.18}$$

This linear equation system can now be used to formulate a least squares adjustment (LSA) problem (see *Niemeier*, 2008) when interpreting $\underline{f}_{?} \approx \hat{\underline{l}}_{?} = \underline{l}_{?} + \hat{\underline{v}}_{?}$ as observation $\underline{l}_{?}$ with corrections $\hat{\underline{v}}_{?}$:

$$\underline{l}_{?} + \hat{\underline{v}}_{?} = \mathrm{A}_{?}^{\mathcal{H}}\, \hat{\underline{f}}. \tag{2.19}$$

Without loss of generality, when the problem is well-posed, i.e., the equation system is overdetermined, a statistically optimal solution can be found through the minimization of $\left(\underline{l}_{?} - \mathrm{A}_{?}^{\mathcal{H}}\, \hat{\underline{f}}\right)^2$ (assuming normally-distributed items and no prior information). The solution to this minimization problem leads to the well-known formula of LSA (see *Niemeier*, 2008):

$$\hat{\underline{f}} = (\underbrace{\mathrm{A}_{?}^{\mathcal{H}\prime} \overbrace{\mathrm{C}_{ll}^{-1}}^{=:\mathrm{P}_{ll}} \mathrm{A}_{?}^{\mathcal{H}}}_{=:\mathrm{N}=\mathrm{C}_{\mathcal{H}\mathcal{H}}^{-1}})^{-1} \overbrace{\mathrm{A}_{?}^{\mathcal{H}\prime} \mathrm{C}_{ll}^{-1} \underline{l}_{?}}^{\underline{q}} = \mathrm{N}^{-1}\, \underline{q}. \tag{2.20}$$

In Eq. 2.20 C_{ll} denotes the covariances matrix of the observations, $\underline{l}_{?}$ and P_{ll} the respective weighting matrix. N is the so-called normal equation matrix and its inverse $\mathrm{C}_{\mathcal{H}\mathcal{H}}$ the covariance matrix of the sought unknowns $\hat{\underline{f}}$. Note that for the sake of simplicity in this thesis it is generally assumed that $\mathrm{Q}_{ll} = \mathrm{C}_{ll}$, i.e., that the a-posteriori covariance C_{ll} is the same as the a-priori covariance Q_{ll}.

Normal equation stacking. Having several uncorrelated normal equation systems (short NEQSs) $\mathrm{N}_i\, \hat{\underline{f}}_i = \underline{q}_i$ from several different sets of observations $\underline{l}_i$ an optimally combined NEQS can be derived by summing up (stacking) the individual equation systems:

$$\overbrace{\sum_i \mathrm{N}_i}^{\mathrm{N}_\Sigma}\, \hat{\underline{f}}_\Sigma = \overbrace{\sum_i \underline{q}_i}^{\underline{q}_\Sigma}, \quad \mathrm{N}_\Sigma\, \hat{\underline{f}}_\Sigma = \underline{q}_\Sigma. \tag{2.21}$$

The technique of stacking normal equations is useful, since in some cases observation sets are just provided by means of normal equations. This has the advantage that the user does not have to be concerned about the possibly difficult functional models behind the equations (e.g., when combining with satellite models).

2.2.1 Orthogonalities within normal equation systems

Depending on A and P, when building the normal equation matrix $\mathrm{N} = \mathrm{A}'\mathrm{PA}$ some entries may turn zero due to orthogonalities between the occurring basis function (i.e., columns of matrix A, that is $\underline{a}_{nm}$). For instance, as seen in Sec. 2.1.2, Eq. 2.9, in the continuous case, when $\mathrm{A} = \mathrm{Y}$ (and $\mathrm{P} = \sin\theta$, i.e., the area weight) all basis functions are orthogonal, and, thus, N would become diagonal and the LSA problem would decay to a simple integration problem again as shown in Eq. 2.10. Though, even if this general orthogonality does not hold in the arbitrary discrete case, with some constraints it is possible to establish at least a partial orthogonality between the basis functions (cf. *Colombo*, 1981):

m-block orthogonality. Within the spherical harmonic synthesis (Eq. 2.4) it is easy to see that in longitudinal direction regarding m the formula equals a discrete Fourier transform (by swapping the summations):

$$f_? \left(\theta, \lambda, h\right) = \sum_{m=0}^{n_{max}} \overbrace{\left(\sum_{n=m}^{n_{max}} C^?_{nm}\, q_{nm}\left(h\right) \bar{P}^?_{nm}\left(\cos\theta\right) \right)}^{=:C_m(\theta,h)} \cos m\lambda + \overbrace{\left(\sum_{n=m}^{n_{max}} S_{nm}\, q_{nm}\left(h\right) \bar{P}^?_{nm}\left(\cos\theta\right) \right)}^{=:S_m(\theta,h)} \sin m\lambda =$$
$$= \sum_{m=0}^{n_{max}} C_m\left(\theta, h\right) \cos m\lambda + S_m\left(\theta, h\right) \sin m\lambda \tag{2.22}$$

In Eq. 2.22 the ? indicates once again the possibly different expressions for the respective parts depending on different functionals. $q_{nm}\left(h\right)$ is used to arbitrarily describe the possible height dependency which, in the general case, may depend on n and m (referring to the Legendre Polynomials of the second kind in the general spheroidal harmonic case, see, e.g., *Fukushima*, 2013). In the following it is assumed to have (1) samples on a regular grid $\underline{\theta} \times \underline{\lambda}$ which is (2) complete and equally spaced in longitudinal direction, i.e.

$$\underline{\lambda} = \left(\lambda_i\right) = \left(\lambda_0 + \frac{i-1}{n_i} 2\pi\right) \tag{2.23}$$

where n_i is the number of samples in longitudinal direction and λ_0 an arbitrary shift. It is further demanded (3) that $h = h(\theta)$, meaning that the height of the sample locations is invariant within the same latitude, and, thus, $C/S_m\left(\theta, h\right) \equiv C/S_m\left(\theta\right)$. Finally, when considering covariances C_{ll} for the observations, it is also demanded (4) that $\mathrm{C}_{ll} = (\mathrm{cov}(\theta_i))$, i.e., C_{ll} (and, thus, also P_{ll}) is independent of the longitude. Under these conditions (1-3) Eq. 2.22 describes a regular discrete Fourier transform from which is known to retain the orthogonality between the m basis functions (since the sample points cover all of the so-called roots of unity of the complex form of the trigonometric functions). So, when building the normal equation matrix, considering also condition (4), matrix elements from distinct orders will become 0. When ordering the coefficients m-wise, a block-diagonal matrix emerges with the block size corresponding to the number of occurring degrees n in the respective order m, i.e. $(n_{max} - m)$. It is important to notice that gravity field functionals obtained through derivation or integration (cf., Eqs. 2.14, 2.16) retain the basic structure as indicated above in Eq. 2.22. Hence, fulfilling conditions (1-4), the m-block orthogonality takes effect for all those derived functionals (for an illustration see Fig. 5.5a in Sec. 5.3).

Equatorial symmetry. Next to the m-block orthogonality, it is also possible to exploit the symmetry properties of the associated Legendre polynomials. From $\bar{P}_{n|m|}$ it is known to be symmetric in case $(n - m)$ is even and

antisymmetric in case $(n-m)$ is odd. Parameterizing $\bar{P}_{n|m|}$ with $\cos\theta$ resp. $\sin\beta$ transforms the symmetry/antisymmetry into equatorial symmetry/antisymmetry. Thus, if one has a symmetric setup of sample locations $\underline{x}$ (of samples of the same functional) regarding the equator, that is $(\theta_i, \lambda_i, h_i) \in \underline{x} \Leftrightarrow (\pi - \theta_i, \lambda_i, h_i) \in \underline{x}$, then two basis functions $\underline{a}_{n_i m_i}$, $\underline{a}_{n_k m_k}$ are orthogonal in case $(n_i + n_k - m_i - m_k)$ is odd. So, also the respective element N_{ik} in the normal equation N becomes 0:

$$
\begin{aligned}
\mathrm{N}_{ik} &= \left\langle \underline{a}_{n_i m_i}, \underline{a}_{n_k m_k} \right\rangle = \\
&= \sum_j \left[\bar{P}^{?}_{n_i|m_i|}(\theta_j) \ldots \right] \left[\bar{P}^{?}_{n_k|m_k|}(\theta_j) \ldots \right] = \\
&= \sum_{(j:\ \theta_j \le \pi/2)} \left[\bar{P}^{?}_{n_i|m_i|}(\theta_j) \ldots \right] \left[\bar{P}^{?}_{n_k|m_k|}(\theta_j) \ldots \right] + \sum_{(j:\ \theta_j > \pi/2)} \left[\bar{P}^{?}_{n_i|m_i|}(\theta_j) \ldots \right] \left[\bar{P}^{?}_{n_k|m_k|}(\theta_j) \ldots \right] \equiv \\
&\equiv \sum_{(j:\ \theta_j \le \pi/2)} \Big[\bar{P}^{?}_{n_i|m_i|}(\theta_j)\, \bar{P}^{?}_{n_k|m_k|}(\theta_j) + \underbrace{\bar{P}^{?}_{n_i|m_i|}(-\theta_j)\, \bar{P}^{?}_{n_k|m_k|}(-\theta_j)}_{=-(\cdots) \ \Leftrightarrow\ (n_i+n_k-m_i-m_k) \bmod 2 = 1} \Big] \ldots = \\
&= 0 \ \Leftrightarrow\ (n_i + n_k - m_i - m_k) \bmod 2 = 1
\end{aligned}
\tag{2.24}
$$

For sake of simplicity, C_{ll} is omitted in this derivation. However, if C_{ll} is also symmetric regarding the equator, it can be shown that N_{ik} remains 0 in case of odd parity. When exploiting this additional symmetry feature together with the m-block orthogonality, the diagonal m-blocks halve their size, resulting in so-called mp-blocks where the 'p' stands for parity.

Gauss-Legendre quadrature. As pointed out, e.g., in *Sneeuw* (1994), with an appropriate choice of sampling locations (nodes) $\underline{\theta}^G = (\theta_j^G)$ and weights $\mathrm{P}^{\mathrm{G}} \coloneqq \mathrm{diag}(\underline{w}^G)$ it is possible to retain the orthogonality of the associated Legendre polynomials in the discrete case ($\mathrm{diag}(...)$ denotes the vector to diagonal matrix transform). The method corresponds to the so-called Gauss-Legendre quadrature from which it is known to resemble an exact integration of polynomials of degree $2N-1$ in the interval $[-1, 1]$, where N is the number of nodes. Hence, choosing at least $N = n_{max} + 1$ nodes, all polynomials $\bar{P}^{?}_{n_i|m_i|}\bar{P}^{?}_{n_k|m_k|}$ can be integrated exactly:

$$
\int_{-1}^{1} \bar{P}^{?}_{n_i|m_i|}(x)\bar{P}^{?}_{n_k|m_k|}(x)dx = \sum_{j=1}^{n_{max}+1} w_j\, \bar{P}^{?}_{n_i|m_i|}(\cos\theta_j)\bar{P}^{?}_{n_k|m_k|}(\cos\theta_j) = 0 \ \Leftrightarrow\ i \neq k. \tag{2.25}
$$

The main difficulty when using this method is the determination of the $n+1$ nodes and weights. As defined by the approach, the nodes θ_j have to be located at the $n+1$ roots of the Legendre polynomial P_{n+1} and the weights w_j can then be calculated by, e.g. (see, *Sneeuw*, 1994):

$$
w_j = \frac{2\left(1 - (\cos\theta_j)^2\right)}{\left[(n+1)P_n(\cos\theta_j)\right]^2} \tag{2.26}
$$

For finding the roots θ_i several efficient and accurate methods exist (up to machine precision, e.g., *Bogaert*, 2014). Using this so-called Gauss-grid (together with the weights and the m-block orthogonality) leads again to a diagonal structure of the normal equation matrix N and corresponds again to a (weighted) simple quadrature. While this approach is, from the perspective of the computational efficiency, far superior to all other solutions acquired through LSA, there are two major limitations: (1) the observations have to be given on the Gauss-grid which is usually not the case for pre-compiled datasets, and (2) the applied weights are solely synthetic and

do not have to resemble the actual accuracies of the observations. However, while the weights are purely synthetic it has been shown that they resemble a $\sin\theta$ weighting in good approximation (see *Sneeuw*, 1994, converging to it with $N \to \infty$) which corresponds to an equal-area weighting (as in the quadrature case through $d\Omega$, see Eq. 2.10).

2.3 Least squares collocation

This section provides a brief overview of the LSC theory and introduces the notation. Thereto, a somewhat different access to LSC is presented with the aim to highlight similarities and disparities to least-squares adjustment. This approach is neither completely new nor mathematically fully rigorous, but might support readers who are more familiar with LSA than LSC. Advanced readers may also find it interesting to get a glimpse at LSC from a slightly different perspective. For a more complete and elaborated treatment of this topic the reader is referred to *Moritz* (1980).

2.3.1 The LSC method

The measurement as random variable. LSC is a statistically optimal method for estimating (predicting) quantities based on their stochastic properties. Since it is a statistical technique, all treated quantities within the scope of LSC have a probabilistic character and can be modelled mathematically through random variables. Consequently, before explaining the method itself, to build the basis for later, the structure and possible interpretation of the random variables which are used must be described first. For LSC, it is convenient to define the random variable V as

$$\begin{aligned} V \coloneqq & V^{\mathcal{X}} + V^{\mathcal{M}} \quad \text{with} \\ V^{\mathcal{X}} : & \, \mathcal{X} \to \mathbb{R} \, , \\ V^{\mathcal{M}} : & \, \mathcal{M} \to \mathbb{R}. \end{aligned} \tag{2.27}$$

as the central quantity. In words, V consists of a random variable $V^{\mathcal{X}}$ with spatial reference $\mathcal{X}$ (see Sec. 2.1.2) and a random variable $V^{\mathcal{M}}$ with state dependency $\mathcal{M}$. One can think of V as a measurement (e.g., through an instrument) of a signal $V^{\mathcal{X}}$ (e.g., gravity field item) with (additive) noise $V^{\mathcal{M}}$ (e.g., instrument errors). In the sense of a random variable, $(\mathcal{X}, \mathcal{M})$ builds the so-called sample space of V, where in LSC a single element (i.e., sample) $\boldsymbol{x} \in \mathcal{X}$ is denoted as location (or coordinates) and $\boldsymbol{m} \in \mathcal{M}$ as (measurement) state of a concrete event $v \coloneqq V(\boldsymbol{x}, \boldsymbol{m})$. An event v represents an actual realization of a "measurement" (assuming a fixed sample $(\boldsymbol{x}, \boldsymbol{m})$) and is henceforth called observation (i.e., the outcome of a measurement). For random variables, samples are defined to be non-deterministic. Consequently, the state $\boldsymbol{m}$ is interpreted as purely random entity on which V is reliant on (e.g., some unobserved environmental variables influencing the measurement such as temperature, pressure, humidity, time, etc.). In classical LSC also the location $\boldsymbol{x}$ is treated as random regarding the statistical behavior of $V^{\mathcal{X}}$ (e.g., when using homogeneous isotropic covariances functions, all locations are treated as a sample with identical statistical properties, see, e.g., *Moritz*, 1980), although $\boldsymbol{x}$ usually has a deterministic character (since the locations of measurements are generally assumed to be known).

Normally distributed random vectors. In the general application, in LSC one does not just have one random variable, but an entire (ordered) set of "measurements", forming a so-called random vector $\underline{V} \coloneqq (V_i)$

(the underscore indicates multivariate items, i.e., vectors). Accordingly, an event of the random vector $\underline{V}$ can be interpreted as a set of observations $\underline{v} = \underline{V}(\underline{\boldsymbol{x}}, \underline{\boldsymbol{m}})$ depending on a set $\underline{\boldsymbol{x}}$ of locations and a set $\underline{\boldsymbol{m}}$ of states. Here, for the sake of simplicity, $\underline{V}$ shall be a vector in a finite-dimensional Euclidean space (though, in principle, an interpretation of $\underline{V}$ as a random process regarding $\mathcal{X}$ would also be legit). Thus, in the following, all items related to $\underline{V}$ shall also be understood as ordinary vectors resp. matrices in a Euclidean sense. For LSC it is premised that $\underline{V} \sim \mathcal{N}(\dot{\underline{v}}, \mathrm{C}_{VV})$, i.e., $\underline{V}$ is normally distributed with expectation $\dot{\underline{v}} = \mathbb{E}(\underline{V})$ and covariance $\mathrm{C}_{VV} = \mathrm{C}_{VV}^{\mathcal{X}} + \mathrm{C}_{VV}^{M}[+\mathrm{C}_{VV}^{\mathcal{X}M} + \mathrm{C}_{VV}^{M\mathcal{X}}]$. Here, square brackets are used to indicate terms which are usually omitted in practice. $\mathrm{C}_{VV}^{\mathcal{X}}$ is frequently called the signal (or error) covariance of $\underline{V}^{\mathcal{X}}, \mathrm{C}_{VV}^{\mathcal{M}}$ the (measurement) noise covariance of $\underline{V}^{\mathcal{M}}$ [and $\mathrm{C}_{VV}^{\mathcal{X}\mathcal{M}}$ the cross-covariance between signal and noise].

Objective and prerequisites. Within the context of LSC (other than in LSA) it is assumed that $C_{VV}^{\mathcal{X}}$, $C_{VV}^{\mathcal{M}}$ [and the cross-covariance $C_{VV}^{\mathcal{X}\mathcal{M}} = \mathrm{cov}\left(\underline{V}^{\mathcal{X}}, \underline{V}^{\mathcal{M}}\right)$] are known, and the expectation $\dot{\underline{v}}$ is only approximated for a subset $\underline{L} \subseteq \underline{V}$ through given observations $\underline{l}(\underline{\boldsymbol{x}}_L, \underline{\boldsymbol{m}}_L) \subseteq \underline{v}$, ($\underline{\boldsymbol{x}}_L \subseteq \underline{\boldsymbol{x}}$ are called observation points). Under these conditions, the aim of LSC is defined by finding a statistically optimal (linear) estimate $\hat{\underline{s}}$ for the expectation $\dot{\underline{s}}$ of a chosen (arbitrary) subset of the signal $\underline{S}^{\mathcal{X}} \subseteq \underline{V}^{\mathcal{X}}$ under the premise of the given observations $\underline{l}$. Note that in this formulation both sets, the signal part $\underline{L}^{\mathcal{X}}$ of the observations as well as the signal part $\underline{S}^{\mathcal{X}}$ of estimates are modeled to origin from the common signal $\underline{V}^{\mathcal{X}}$, and, furthermore, $\underline{L}^{\mathcal{X}}$ and $\underline{S}^{\mathcal{X}}$ do not necessarily have to be disjoint. This interpretation seems to be advantageous since dealing with parts of a common covariance (matrix) is in our view easier than dealing with individual covariances (especially in the scope of covariance propagation, e.g., Eq. 2.37).

LSC in terms of the least squares principle. Based on this setup, the LSC solution for $\hat{\underline{s}}$ is typically found by minimizing the diagonal of the estimate's covariance $C_{\hat{S}\hat{S}}$ (called the minimum variance principle, see *Moritz*, 1980). However, since in this work it is intended to draw comparisons to LSA, a different approach is presented which nonetheless results in the same final formulation for LSC. For this alternative derivation, to be compatible with LSA, it is assumed in the following that the stated estimation problem could also be solved through classical LSA (by means of the least squares principle, see *Moritz*, 1980). This requirement basically implies that a relation of the form:

$$\underline{L}^{\mathcal{X}} = \mathrm{A}_L^S \, \underline{S}^{\mathcal{X}} \tag{2.28}$$

shall (at least theoretically) exist, meaning that the signal $\underline{L}^{\mathcal{X}}$ within the observations is linearly dependent on the estimate's signal $\underline{S}^{\mathcal{X}}$. Be aware that this is a premise that does not exist in the classical derivation of LSC. The impact of this assumption will be briefly discussed at the end of this section and in Sec. 2.3.3. Having A_L^S at hand, it is straightforward to derive the well-known formula for the least-squares solution (Eq. 2.29). Although, in this formula A_L^S is generally not explicitly known in the scope of LSC, it is possible to substitute the a posteriori covariances $\mathrm{C}_{\hat{S}\hat{S}}$ resp. $\mathrm{C}_{\hat{S}L}$ (by means of covariance propagation) so that all occurrences of A_L^S vanish, finally yielding the LSC estimator $\widetilde{\mathrm{A}}_S^L$:

$$\begin{aligned} \hat{\underline{s}} = & \underbrace{\overbrace{\left(\mathrm{A}_L^{S\prime}\mathrm{C}_{LL}^{-1}\mathrm{A}_L^S\right)^{-1}}^{\mathrm{C}_{\hat{S}\hat{S}}}\mathrm{A}_L^{S\prime}}_{\mathrm{C}_{\hat{S}L}}\mathrm{C}_{LL}^{-1}\underline{l} = \\ = & \mathrm{C}_{SL}\mathrm{C}_{LL}^{-1}\underline{l} = \widetilde{\mathrm{A}}_L^S l, \quad \text{with} \quad \widetilde{\mathrm{A}}_L^S = \mathrm{C}_{SL}\mathrm{C}_{LL}^{-1}. \end{aligned} \tag{2.29}$$

Obviously, by having direct knowledge of the covariances $\mathrm{C}_{\hat{S}L}$ and C_{LL} the LSC estimate can be obtained in the same way as the respective LSA estimate. While $\mathrm{C}_{LL} \coloneqq \mathrm{C}\left(L, L\right)$ is known as per definition, $\mathrm{C}_{\hat{S}L}$ must be chosen appropriately. With the choice of $\mathrm{C}_{\hat{S}L}$ one can explicitly steer the optimal covariances the estimate $\underline{\hat{s}}$ shall have. Without loss of generality, in LSC one usually follows the definition $\mathrm{C}_{\hat{S}\hat{S}} \coloneqq \mathrm{C}^{\mathcal{X}}_{SS}$, meaning that one defines the estimate's optimal covariance $\mathrm{C}_{\hat{S}\hat{S}}$ to be the given a-priori signal covariance $\mathrm{C}^{\mathcal{X}}_{SS} \coloneqq \mathrm{C}^{\mathcal{X}}_{VV}\left(S, S\right)$ (without the noise component). Using this definition, $\mathrm{C}_{\hat{S}L}$ is obtained by:

$$\begin{aligned}\mathrm{C}_{\hat{S}L} \coloneqq \mathrm{C}_{SL} &= \operatorname{cov}\left(S^{\mathcal{X}}, L = L^{\mathcal{X}} + L^{\mathcal{M}}\right) = \\ &= \operatorname{cov}\left(S^{\mathcal{X}}, L = L^{\mathcal{X}}\right) + \operatorname{cov}\left(S^{\mathcal{X}}, L = L^{\mathcal{M}}\right) = \mathrm{C}^{\mathcal{X}}_{SL}\left[+\mathrm{C}^{\mathcal{X}M}_{SL}\right].\end{aligned} \tag{2.30}$$

Thus, the adapted, but still most general, LSC estimate reads:

$$\underline{\hat{s}} = \widetilde{A}^{L}_{S}\underline{l} = \left(C^{\mathcal{X}}_{SL}\left[+C^{\mathcal{X}\mathcal{M}}_{SL}\right]\right)\left(C^{\mathcal{X}}_{LL} + C^{\mathcal{M}}_{LL}\left[+C^{\mathcal{X}\mathcal{M}}_{LL}+C^{\mathcal{M}\mathcal{X}}_{LL}\right]\right)^{-1}\underline{l}. \tag{2.31}$$

As indicated, in many applications it is assumed that the cross-covariances $C^{\mathcal{X}\mathcal{M}}_{VV}$ are zero (i.e., that there is no location-dependent measurement noise).

Difference to the minimum variance principle. Note that within this derivation neither $\underline{l}$ nor $\underline{\hat{s}}$ must necessarily be centered, that is $\underline{\dot{v}} = 0$, as it is the case in the classical derivation given by Moritz (1980). However, since this derivation follows the same rules as LSA, also the same limitations apply, i.e., (1) A^{S}_{L} must (at least theoretically) exist and (2) C_{LL} and $\mathrm{A}^{S\prime}_{L}\mathrm{C}^{-1}_{LL}\mathrm{A}^{S}_{L}$ must be invertible. In other words, this means $\underline{S}^{\mathcal{X}}$ must be unambiguously derivable from $\underline{L}^{\mathcal{X}}$ and vice versa, although a result might still be obtained by LSC even if this is not the case (see also Sec. 2.3.3). As an example, predicting non-reduced gravity observations $\underline{\hat{s}}$ on a regional scale may produce an unrealistic result when the long-wavelength components within $\underline{l}$ are almost linear dependent and, hence, cannot be propagated correctly to the estimate $\underline{\hat{s}}$. In case of gravity field modelling, this can be especially problematic when performing up-/downward continuations or when changing the functional. As such, the requirement of centered observations can be interpreted as equivalence to the requirement of an invertibility of the forward model, while the latter is less restrictive in the sense that it does not forbid uncentered random vectors a priori. As a counterexample, predicting unreduced gravity observations by LSC on a global scale will most likely yield a good result even if the global mean in $\underline{l}$ does not vanish (and hence contradicts the premise of centered random vectors). This is easily discernible by the fact that a corresponding LSA would also succeed.

2.3.2 LSC with a-priori reduction

As highlighted in the previous section, signal parts within $\underline{S}^{\mathcal{X}}$ that are not recoverable from $\underline{L}^{\mathcal{X}}$ should be avoided. This can be achieved by reducing such signal components using a-priori information before the collocation and by restoring them afterwards (known as remove-compute-restore technique, RCR, see *Forsberg & Tscherning*, 1981). To reuse the previous notation, $\underline{V}^{\mathcal{X}}$ can be reinterpreted as the reduction itself and $\underline{V}^{\mathcal{M}}$ as the measurement to be reduced (see P-4). With this definition the reduced random vector $\Delta\underline{V}$ is defined by:

$$\Delta\underline{V} = \underline{V}^{\mathcal{M}} - \underline{V}^{\mathcal{X}} \tag{2.32}$$

It is assumed that realizations $\underline{v}^{\mathcal{X}}$ (called reductions) for all $\underline{V}^{\mathcal{X}}$ and realizations $\underline{l}^{\mathcal{M}}$ (called observations) for all $L^{\mathcal{M}}$ exist. Having $\Delta\underline{l}$ instead of $\underline{l}$, $\Delta\underline{\hat{s}}$ instead of $\underline{\hat{s}}$ and $\mathrm{C}^{\mathcal{X}}_{VV}$ as the error covariance (instead of the

signal covariance), the estimator $\widetilde{\mathrm{A}}_S^L$ of the previous section remains basically unchanged (only difference: the sign of $\mathrm{C}_{VV}^{\mathcal{X}\mathcal{M}}$ changes). Together with the restore step, the modified LSC equation, based on residuals, reads:

$$\begin{aligned} \Delta\underline{\hat{s}} &= \widetilde{\mathrm{A}}_S^L \Delta\underline{l} \rightarrow \\ \underline{\hat{s}} &= \underline{s}^{\mathcal{X}} + \Delta\underline{\hat{s}} = \underline{s}^{\mathcal{X}} + \widetilde{\mathrm{A}}_S^L \left(\underline{l}^{\mathcal{M}} - \underline{l}^{\mathcal{X}}\right) \end{aligned} \tag{2.33}$$

with

$$\widetilde{\mathrm{A}}_S^L = \left(\mathrm{C}_{SL}^{\mathcal{X}} \left[-\mathrm{C}_{SL}^{\mathcal{X}\mathcal{M}}\right]\right) \left(\mathrm{C}_{LL}^{\mathcal{X}} + \mathrm{C}_{LL}^{\mathcal{M}} \left[-\mathrm{C}_{LL}^{\mathcal{X}\mathcal{M}} - \mathrm{C}_{LL}^{\mathcal{M}\mathcal{X}}\right]\right)^{-1}. \tag{2.34}$$

A respective rigorous covariance propagation yields the covariance for the restored estimate $\underline{\hat{s}}$:

$$\mathrm{C}_{\hat{S}\hat{S}} = \mathrm{C}_{SS}^{\mathcal{X}} - \mathrm{C}_{\hat{S}L}\mathrm{C}_{LL}^{-1}\mathrm{C}'_{\hat{S}L} \tag{2.35}$$

This formulation of LSC including a RCR procedure (short RCR-LSC) is widely used (*Forsberg & Tscherning*, 1981; *Kühtreiber*, 2002) as it allows to remove the non-recoverable signal parts, such that the remaining resolvable parts can be expressed sufficiently well by the implicit forward model within LSC. Of course, this is only possible if appropriate a-priori information of the non-recoverable signal part is available. In gravity field modelling this complementary information can be derived from satellite-based global gravity field models and topographic models, respectively.

2.3.3 The covariance function

Covariance propagation. As shown above, LSC and LSA can be derived in a similar way and share the same basic properties. However, they differ in obtaining the cross-covariance $\mathrm{C}_{\hat{S}L}$: while in LSA the a posteriori covariances are calculated through the propagation of the a-priori covariances (i.e.,C_{LL}), in LSC the a posteriori cross-covariances are assumed to be known. With this, the LSC approach holds the advantage that the assumed knowledge circumvents possible instabilities in the calculation of $\mathrm{C}_{\hat{S}L}$ and, hence, allows a solution even if the estimates are linearly dependent. A further advantage is that the forward model A_L^S (cf. Eq. 2.28) does not have to be explicitly known. This can be helpful since in real-world scenarios the connection between $\underline{L}^{\mathcal{X}}$ and $\underline{S}^{\mathcal{X}}$ is possibly only indirectly given over a (concealed) domain $\mathcal{H}$, resulting in a linear relationship of the form

$$\begin{pmatrix} \underline{L}^{\mathcal{X}} \\ \underline{S}^{\mathcal{X}} \end{pmatrix} = \begin{pmatrix} \mathrm{A}_L^{\mathcal{H}} \\ \mathrm{A}_S^{\mathcal{H}} \end{pmatrix} \underline{H}. \tag{2.36}$$

In case the covariance information C_{HH} to the random vector $\underline{H}$ is known, $\mathrm{C}_{VV}^{\mathcal{X}}$ can easily be obtained by covariance propagation:

$$\mathrm{C}_{VV}^{\mathcal{X}} = \mathrm{A}_V^{\mathcal{H}}\mathrm{C}_{\mathcal{H}\mathcal{H}}^{\mathcal{X}}\mathrm{A}_V^{\mathcal{H}\prime}. \tag{2.37}$$

Accordingly, all covariances needed for LSC can be calculated without explicit knowledge of A_L^S. This directly implies an important difference to LSA which is that in LSC the result $\hat{s}_i$ for one estimate is usually independent of the choice of all other estimates $\underline{s}$. This is obvious, as the calculation of $\mathrm{C}_{VV}^{\mathcal{X}}$ (Eq. 2.37) and thus $\mathrm{C}_{SL}^{\mathcal{X}}$ can be performed element-wise, while the calculation of $\mathrm{C}_{\hat{S}L}$ within LSA (Eq. 2.29) requires an adequate choice of

a whole set of estimates $\underline{s}$. In LSA, this adequate set $\underline{s}$ has to be chosen in such a way that the corresponding forward model A_L^S explains the actual observation $\underline{l}$ (see Eq. 2.28 where the problem of underfitting or overfitting may be introduced). In LSC this requirement is generally substituted through the prior knowledge of $\mathrm{C}_{VV}^{\mathcal{X}}$ which is believed to already cover all available functional coherences. However, as explained in Sec. 2.3.1, by using such a-priori covariances one may risk that LSC (in contrast to LSA) succeeds even when $\hat{\underline{s}}$ is not unambiguously derivable by $\underline{l}$ according to the forward model, hence yielding unrealistic results which might not be easy to uncover as such.

Homogeneous-isotropic covariances. In gravity field modelling, the domain $\mathcal{H}$ can be considered as the spherical resp. spheroidal harmonic domain (with $\underline{f}_{\mathcal{H}}$ the set of spherical harmonic coefficients, being a realization of $\underline{H}$, see Sec. 2.1.2). In the classical application $\mathrm{C}_{\mathcal{H}\mathcal{H}}^{\mathcal{X}}$ is modeled by so-called degree variances

$$\mathcal{C}_n \coloneqq \sum_m \left(C_{nm}^2 + S_{nm}^2\right) \tag{2.38}$$

yielding a homogeneous-isotropic covariance function $\mathcal{C}_{\mathcal{T}\mathcal{T}}$ (here in terms of potential values, that is, $V^{\mathcal{X}} \coloneqq \mathcal{T}$, see Sec. 2.1.3)

$$\begin{aligned} \mathcal{C}_{\mathcal{T}\mathcal{T}}(\psi, u) &\coloneqq \iiint_{\mathcal{X}} \left[\frac{1}{2\pi} \int_{\alpha=0} \mathcal{T}(\theta, \lambda, r)\ \mathcal{T}(\theta', \lambda', r')\, d\alpha\right] d\mathcal{X} \Big/ \iiint_{\mathcal{X}} d\mathcal{X} \\ &= \sum_n u^{n+1} P_n(\cos\psi)\ \mathcal{C}_n \\ &\approx \mathrm{cov}(\mathcal{T}, \mathcal{T}') \end{aligned} \tag{2.39}$$

with

$$\cos\psi = \cos\theta\cos\theta' + \sin\theta\sin\theta\cos(\lambda' - \lambda), \quad \sin\alpha = \frac{\sin\theta'\sin(\lambda' - \lambda)}{\sin\psi}, \quad u \coloneqq \frac{R^2}{rr'} \tag{2.40}$$

in the domain $\mathcal{X}$. In Eq. 2.39 the inner integral over α can be interpreted as the averaging operation over all points which have a spherical distance ψ from the actual evaluation point. As such, the inner integral establishes the isotropy property of the resulting covariance function. The outer volume integral over $\mathcal{X}$, on the other hand, averages over all locations and, thus, constitutes the homogeneity of $C(\psi, u)$. As Eq. 2.39 shows, using a homogeneous-isotropic covariance function can be interpreted as the attempt of deriving covariances in an empirical manner from actual observations $\mathcal{T}$ (or from given models in terms of C_{nm}, S_{nm}) by making use of simplified assumptions (i.e., the homogeneity and isotropy). The proof of the relation given in Eq. 2.39 can be found, e.g., in *Heiskanen & Moritz* (1967) (without the r- resp. u-component). While in Eq. 2.39 the covariance is modeled regarding potential values $\mathcal{T}$, the covariance and cross-covariance of other gravity field functionals can also be deduced from it by calculating the respective derivatives along the spatial coordinates (in the same fashion as shown in Sec. 2.1.3, by means of Eq. 2.40). As an example, and since important for the actual application presented in this work (for GRAV-D), the derivation of the covariance function in terms of (spherical) gravity disturbances $\mathcal{C}_{\delta g \delta g}$ is shown. As highlighted Sec. 2.1.3, (spherical) gravity disturbances δg are obtained through the radial derivative of the disturbing potential $\mathcal{T}$, that is $\delta g \coloneqq \partial\mathcal{T}/\partial r$. Consequently,

the propagation of the resp. covariances $\mathcal{C}_{\delta g \delta g} \approx \operatorname{cov}(\delta g, \delta g')$ requires the derivative of $\mathcal{C}_{\mathcal{T}\mathcal{T}}$ along r and r':

$$\begin{aligned}\mathcal{C}_{\delta g \delta g}(\psi, u) = \frac{\partial^2 \mathcal{C}_{\mathcal{T}\mathcal{T}}}{\partial r \partial r'} = \sum_n \frac{\partial \Big(\overbrace{\frac{R^2}{rr'}}^{=u} \Big)^{n+1}}{\partial r \partial r'} P_n(\cos\psi)\, \mathcal{C}_n = \sum_n \frac{1}{R^2} \Big(\frac{R^2}{rr'}\Big)^{n+2} P_n(\cos\psi)\, \mathcal{C}_n = \\ = \sum_n \frac{u^{n+2}}{R^2} P_n(\cos\psi)\, \mathcal{C}_n = \frac{u}{R^2} \mathcal{C}_{\mathcal{T}\mathcal{T}}.\end{aligned} \tag{2.41}$$

While, as mentioned, $\mathcal{C}(\psi, u)$ can easily be modeled through given Stokes coefficients, there are also other possibilities of deriving it when no fitting global model is given. For instance, one can use a simple parametric model to design a fitting degree variance curve (as, e.g., done by *Tscherning & Rapp*, 1974). The method presented in Sec. 4.3 can be considered similar, although more sophisticated. In modern approaches also the full covariance information from global gravity field models is sometimes introduced (see P-4 or *Pail et al.*, 2010) which results in arbitrary covariance functions in $\mathcal{X}$. A number of other methods exist which yield non-homogeneous and non-isotropic covariance functions. However, since these methods will not be used within this thesis the reader is referred to literature (see, e.g., *Darbeheshti*, 2009).

Chapter 3

Gravity data reduction and validation

Within this chapter the workflow of gravity data reduction and validation is presented (cf. publication P-1). Reducing gravity data means removing known parts of the gravity field from given observations to reduce their variance, resulting in residual observations, i.e., the difference between measured and known gravity. Working with residual gravity field observations is useful since nowadays certain parts of the gravity field signal are already known sufficiently well through complementary sources. As an example, the long wavelength part of the gravity signal is already determined very accurately on a global scale through SGGMs up to a wavelength of about ≈100 km (≈d/o 200, see *Kvas et al.*, 2019). The very short wavelength signal, on the other hand, can be approximated to a certain degree by topographic forward models (*Hirt et al.*, 2017). Because gravity datasets are usually limited in their spatial and spectral resolution, using such complementary reductions is imperative to prevent the introduction of aliasing errors in the final model. Thus, in a first step the SATOP reduction model (Sec. 3.1) is introduced, which shall form the basis for all gravity data reductions in this thesis. Since applying reductions pointwise on large scattered datasets is computationally expensive, the use of a fast 3D synthesis approach based on interpolation (FSYNTH3D, see Sec. 3.2) is proposed. Based on the acquired residuals (through subtraction of the reduction model) some validations of the gravity datasets can be performed (e.g., inspecting the long-wavelength residuum or the overall variance of individual datasets). The insights of this validation can then be used to identify outliers or remove unwanted offsets and/or other long-wavelength signals (see Sec. 3.3) to obtain a validated and reduced gravity dataset as input for the subsequent gridding (see Chap. 4).

3.1 The SATOP model

Outline. For a reduction model, to be suited as complementary gravity data and validation source, it is beneficial to be (1) "as good as possible" and (2) "as independent as possible" of the gravity data to reduce. For the reduction and validation of ground gravity data, this means that preexisting EGGMs drop out as possible candidates, since they already strongly rely on the ground gravity data itself (see *ICGEM*) and, hence, would not be independent of it. Thus, an EGGM is needed that is not built upon ground gravity data. In theory, a SGGM would be suited, but, due to its limited resolution it cannot be considered as "as good as possible". Therefore, the only other gravity information available, beside SGGMs and EGGMs, is forward modelled gravity from topography. So, an EGGM combining both, satellite information and forward modelled topographic information would still be independent of ground gravity data and simultaneously would contain all complementary data sources which are (globally) available. Hence, such a model would obviously meet both requirements stated above. Such a model does not yet exist (at least not on *ICGEM*), and so, it has to be created from scratch.

Since this reduction model will be a combination of (SA)tellite and (TOP)ography data it will be denoted as the SATOP model henceforth. As starting point for the compilation of SATOPv2 (an updated version of the original SATOP model as presented in P-1), the appropriate SGGM and topographic model on which the new reduction model is based has to be chosen. For the satellite model, the latest GOCO version (GOCO06s, see *Kvas et al.*, 2019) is selected, as it the most recent static SGGM with the highest possible resolution, combining satellite data from various missions. Additionally, the complete NEQS is available for GOCO06s, which allows for a more rigorous combination with the topographic model. For the topographic model, also one of the most recent ones, EARTH2014 (see *Hirt & Rexer*, 2015 and *Rexer*, 2017), is chosen since it includes the newest global topography data and is available in the highest spectral resolution (d/o 5480). In a statistical sense, the optimal combination of both models could be achieved when combining (stacking) the NEQSs (cf. Eq. 2.21):

$$\underline{x}_S = (\mathrm{N}_S + \mathrm{N}_G)^{-1} \, (\underline{q}_S + \underline{q}_T) \tag{3.1}$$

with N_S and N_T as the normal equation matrix of the satellite resp. topographic model, and $\underline{q}_S$, $\underline{q}_T$ the respective right-hand sides of the NEQSs. However, for EARTH2014 no NEQS or other statistical information is available, only the coefficients themselves. Hence, statistical information in form of degree error variances has to be derived empirically by comparing the coefficients to another EGGM. In case of SATOPv2, EARTH2014 is compared to XGM2019 for the estimation of the empirical degree error variances for EARTH2014. Having the error information for each coefficient in form of degree error variances $\underline{\mathcal{C}}_T \coloneqq \left(\mathcal{C}_n^T\right)$ (see Sec. 2.3.3), a simplified (diagonal) NEQS for the topographic model can be obtained:

$$\mathrm{N}_T = \mathrm{diag}(\underline{\mathcal{C}}_T^{-1}), \quad \underline{q}_T = \mathrm{diag}(\underline{\mathcal{C}}_T^{-1}) \, \underline{x}_T \tag{3.2}$$

where $\underline{x}_T$ denotes the original coefficients of EARTH2014. Having this simplified NEQS, the combination as described by Eq. 3.1 can be calculated since all items are now available. For additional details about the processing and performance of SATOP (which is in principle identical to SATOPv2 except for the use of GOCO05s instead of the more recent GOCO06s model) the reader is referred to the publication P-1.

Performance. When inspecting the empirical degree variances of SATOP, one can assess its performance on a global average degree-wise (see Fig. 3.1). Summing up, SATOP globally reduces nearly 100% of the gravity signal up to d/o 200 (≈100 km), with a slight degradation over the poles due to the polar gap in the GOCE satellite mission (see *Kvas et al.*, 2019). Above d/o 200, the signal reduction is decreased to about 60% due to the uncertainties in topographic forward modelling, with the largest error contribution of about 6 mGal below d/o 720 (≈30 km). It shall be pointed out, that the whole processing is performed in the spheroidal harmonic domain which holds the advantage to reduce the correlations between coefficients of adjacent degrees (see *Jekeli*, 1988), and thus supports the assumed diagonality of the topographic NEQS. Additionally, hav-

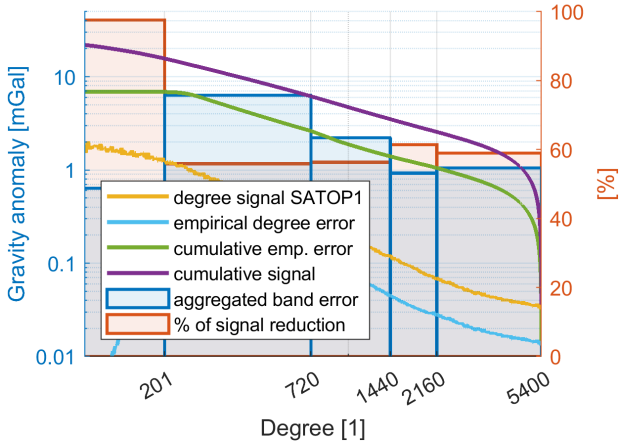


Fig. 3.1 – Performance of the SATOP model in terms of gravity anomalies (from P-1, regarding spheroidal harmonics). Yellow: degree signal/amplitude. Blue: empirical error. Green: aggregated (cumulative) error from max. degree downwards. Green: aggregated signal power. Blue box: aggregated error within spectral bands. Orange box: signal reduction within spectral bands.

ing spheroidal harmonic coefficients allows for the same reason the cutting-out of spectral bands without introducing cut-off effects (near the spheroidal surface). This is advantagous, since the available gravity data to reduce might have been already spectrally limited in a preceding processing step. Hence, in order to properly reduce such datasets also a spectrally limited version of SATOP should be used (see Sec. 3.2).

3.2 FSYNTH3D - A fast 3D synthesis method

Outline. The evaluation of the SH synthesis formula (see Eq. 2.4) on scattered locations is computationally expensive, especially for high-resolution EGGMs such as SATOP, since it requires the seperate evaluation of Legendre polynomials for each location. Hence, when having large scattered gravity datasets, a faster method would be favorable for an efficient calculation of the reduction values. Such a method does not have to be as numerically accurate as the rigorous synthesis, it is sufficient that the calculation accuracy is well below the accuracy of the gravity observations to reduce. For instance, when assuming airborne gravity observations (as within GRAV-D) with an estimated accuracy of 3 mGal, having reduction values with an accuracy of, e.g., below 1 mGal can be considered sufficient. To achieve this, an interpolation approach from a regular 3D geographic grid is chosen. The resolution of this regular grid has to be adapted according to the spectral content of the model to represent. This can be achieved by considering the sampling theorem for spherical harmonics, which is in conformity (at least as an upper bound) to the Nyquist-Shannon sampling theorem (see also Sec. 5.1) for the Fourier transform: to avoid aliasing, one needs at least $2l_{max}+1$ sample points in longitudinal direction and $l_{max}+1$ sample points in latitudinal direction, where l_{max} denotes the maximum d/o of the EGGM. No theorem exists for the sampling in the vertical direction. Hence, the sampling is chosen empirically by investigating the maximum interpolation error. Further details on this matter can be found in publication P-1. After the 3D grid has been defined, an appropriate interpolation method can be chosen according to the accuracy requirements (e.g., linear, cubic, spline, etc.). It has been shown that the rigorous result can be approximated to a very high degree (a few µGal) when using a sophisticated interpolation method (such as spline interpolation). However, such a method is usually computationally more expensive (especially in terms of memory requirements) and the high accuracy is not needed in the actual application. Thus, a simpler method like linear or cubic might be preferrable since it has been shown that the accuracy requirements can still be met (see P-1). Since the SATOP model may be split up into several spectral bands (cf. boxes in Fig. 3.1), also separate 3D grids may be calculated according to them. This grants the possibility to reduce the SATOP model from the observations exclusively within certain spectral bands, which may be useful for the validation/classification of the datasets (see Sec. 3.3). However, since for n bands n 3D grids need to be stored, this technique is highly demanding with respect to memory, which practically limits the number of bands.

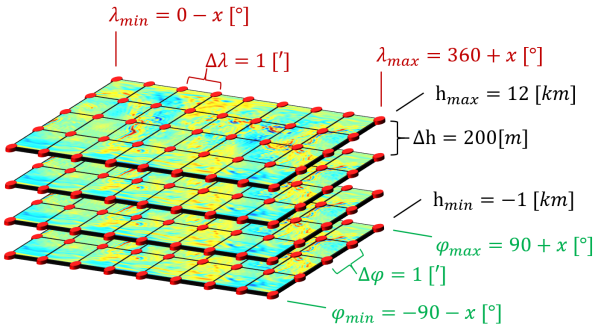


Fig. 3.2 – Illustration of the essential parameters of the regular 3D geographic grid for the GRAV-D application.

Application. In preparation to future applications, the 3D grid for interpolation is chosen to have global coverage (although, for the GRAV-D test case a regional coverage would have been sufficient). The grid resolution

is set to 1' (2 km on equator) to safely cover the spectral content of the SATOP model, which at d/o 5480 corresponds to roughly 2' resolution on the equator. The spacing in height direction is defined to be 200 m starting from a negative (ellipsoidal) altitude of -1 km up until 12 km. The sampling value of 200 m is chosen empirically to safely avoid aliasing. The selected altitude range shall cover all typical ground gravity measurement sites (terrestrial and airborne). Noteworthy, to avoid fringe effects on the poles and the datum border, the grid is extended beyond the poles/datum border for several values (to also allow higher-order interpolation methods to work correctly). A visualization of the grid is given in Fig. 3.2. As a methodical novelty in the test case (not yet contained in P-1), the grid values are based on (disturbing) potential values (instead of gravity disturbances directly). The motivation for this modification is the fact that it is possible (without numerical problems) to derive the most important gravity field functionals (up to second derivative) by applying a numerical differentiation to the interpolated values. Since the total global 3D grid size is rather large ($\approx$700 GB in double precision, for 6 bands), it is favorable to store only one functional from which all others can be derived. Thus, the functional of the disturbing potential is the optimal candidate, since all other functionals just need to be differentiated (not integrated) from it, which is a numerically stable operation (at least for low differentiation orders).

Outlook and limitations. The 3D synthesis method has been proven to function without problems. The constant evaluation quality over the entire interpolation area can be seen as the main advantage of this method. This stands in contrast to the common alternative evaluation strategy in which Taylor-series expansion is used for the height component (see, e.g., *Hirt*, 2012). There, the evaluation quality diminishes with increasing distance to the expansion point (i.e., height). However, the proposed 3D interpolation approach has the disadvantage that the storage requirements are usually larger than for Taylor-series expansion. Further, the 3D grid as presented here is not yet optimized for minimum storage: since Euclidean distances within geographic grids are distorted towards the poles, the target resoltion of 2 km in longitudinal direction is just realized on or near the equator for a 1' spacing. With higher latitudes, the 1' spacing represents smaller distances in longitudinal direction (due to the meridional convergence), and, hence, the spacing in latitudinal direction could be increased in order to save storage. Theoretically, the storage requirements could be reduced to $\sim$63.7% ($2/\pi$) when exploiting this circumstance optimally. Additional storage could be saved for 3D grids of lower spectral bands because there the Nyquist spacing would be larger (according to the maximum d/o of the appropriate band). However, considering different spacings for different latitudes (and different bands) strongly increases the complexity of the implementation of such a method. Such a method would thus also be more prone to programming errors, and so, it was decided to not consider it for the initial reference implementation.

3.3 Validation

Outline. By inspecting the residuals after subtracting the reduction model (e.g. SATOP) a cross-comparison between the model and the observations is possible. Especially, when some properties of the model and/or the observations are known, some conclusions can be drawn towards whether the residuals are reasonable or not. If not, some further action may be required either to rectify the respective residuals or to exclude them from further processing. Since the application scenarios may strongly vary (e.g., terrestrial vs. airborne observations, mountains vs. plains, preprocessed vs. raw measurements), finding the most probable cause for the discrepancies (and, thus, the optimal proceeding) can hardly be automatized, and, hence, has to be figured out manually for each individual case. For certain preprocessed datasets, it may also be reasonable to

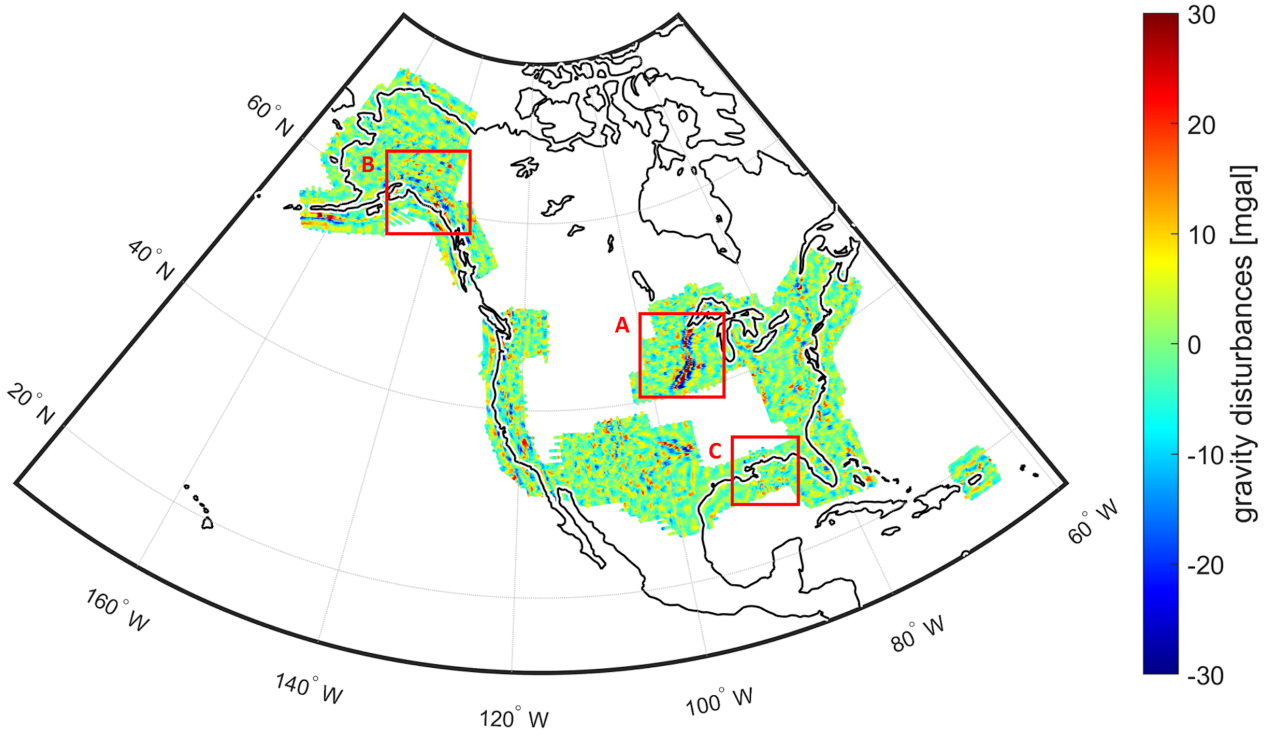
investigate the reduction performance with respect to specific spectral bands since the data may have already been spectrally limited beforehand. For such datasets, it may be unfavorable to reduce the whole spectrum of the SATOP model, as the energy in the spectral bands which are not covered by the datasets is directly added as error, and, thus, increases the overall variance (see P-1).

Application. In case of reducing the SATOP model, through knowledge of its performance (see Fig. 4.4b), one may conclude that long wavelength patterns (> 100 km) in the residuals are most likely caused by the observations, since SATOP should be very accurate in this range. This fact is especially helpful if one suspects offsets or other large-scale errors in the datasets. For instance, one may apply an appropriate Gaussian smoother to low-pass-filter the residuals to the frequency domain where SATOP is accurate in order to quantify the dataset error in this lower spectrum. Ultimately, when an error is identified, it can directly be corrected by subtracting the low-pass filtered residuals from the dataset (see, e.g., Fig. 3.3b). As an example, this strategy is applied in the AntGrav project dataset-wise to rectify offsets and other long-wavelength effects in older and already preprocessed datasets. For these preprocessed datasets reducing the lower bands of SATOP has proven imparative (as mentioned above) since in some cases a low-pass filtering was obviously already applied during the processing (cf. P-1).

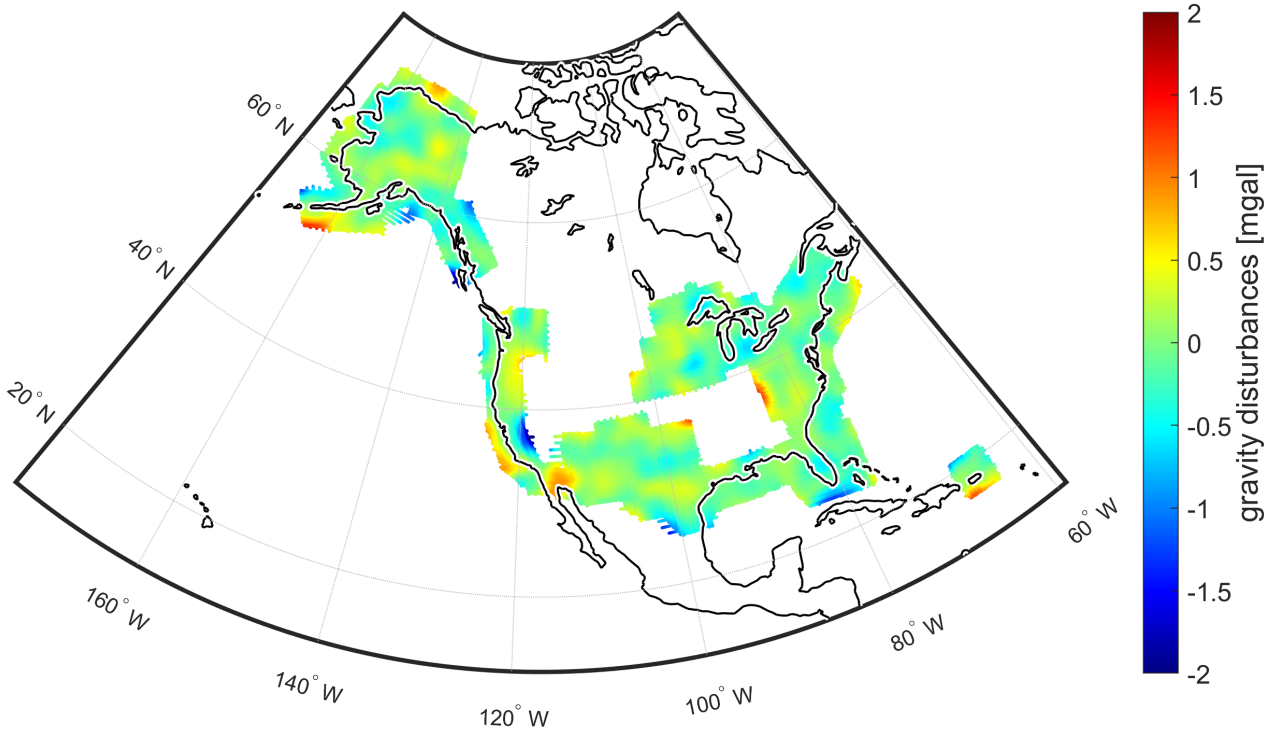
In the actual GRAV-D test case, all observations are based on airborne gravity measurements, and thus, the heterogeneity between datasets is not as prominent as it is in the AntGrav project (see P-1). Especially offsets and drifts are already corrected in the preprocessing by comparing it to an EGGM (EGM2008 in this case, see, e.g., *GRAV-D Team*, 2013). Hence, when inspecting the residuals in Fig. 3.3a and Fig. 3.3b, no obvious long-wavelength differences are present. Nonetheless, even though the longer wavelengths seem to be unproblematic in case of GRAV-D, larger differences can be identified on shorter scales. This is to be expected when keeping in mind the performance of the SATOP model which strongly deteriorates above the satellite resolution ($\sim$100 km). However, it can be seen (Fig. 3.3a) that the deviations not evenly distributed in space, meaning that there are regions with larger as well as smaller residuals. Also this circumstance is not really surprising since the spatial variations may probably also be largely explained by the properties of the topographic model contained in SATOP. Still, larger residuals may also indicate problems/errors in the specific datasets. To explain some potential different reasons for locally larger residuals, three regions (A, B and C, see Fig. 3.3a) are selected where differences are more prominent:

A This area south of Lake Superior, covering parts of the states of Minnesota, Wisconsin, Iowa and Nebraska (see Fig. 3.4a). Here, the topography is rather flat (<500 m). Thus, a direct error from the topographic model due to improper density assumptions can be excluded. However, topographic models do not account for density anomalies in the Earth's interior. Due to the longer wavelengths of the residuals (but also not too long, starting above satellite resolution) in this region, it is probable that they are caused by some anomalous features in the crust. For the same reason (the longer wavelength character), errors in the observations are also highly unlikely since the error would have to influence all flight tracks systematically. Thus, the final conclusion for this region is that the residuals describe the actual signal and can be used as they are given.

B In the southern part of Alaska a strong topographic signal is present due to the pronounced mountain ranges (Alaska range, Saint Alias Mountains, see Fig. 3.4b). Those are fold mountains caused by several faults in this area which are parts of the northernmost section of the pacific ring of fire. Thus, next to the topographic signal itself, also a stronger signal from underground anomalies is to be expected here. In the residuals, exactly this is discernible; similar as in area A, also here the signal is dominated by longer wavelength features which can be attributed again to crustal anomalies which are not modelled by the



(a)



(b)

Fig. 3.3 – Residuals between GRAV-D and SATOPv2. (a) Residuals, using SATOPv2 up to the max. resolution (d/o 5480). Areas for further investigations are highlighted through red boxes. (b) Same residuals as in Fig. 3.3a but low-pass filtered with a Gaussian smoother with a 150 km HWHM (half width at half maximum distance, cf. P-1 and Fig. 4.12b).

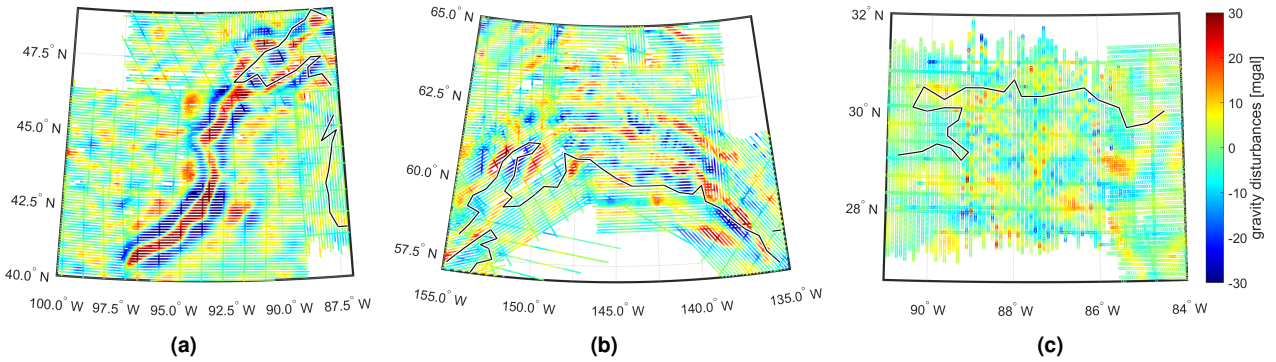


Fig. 3.4 – Three zones of investigation of the GRAV-D residuals (regarding SATOPv2, see also Fig. 3.3a). (a) Area south of Lake Superior. (b) Area in the southern part of Alaska. (c) Coastal zone between Louisiana and Florida.

topographic model within SATOP. However, in contrast to area A, also some residual signals from the direct topographic signal are to be expected here, mainly because varying densities in the topographic model are neglected. Such residuals are expected to be of a more short-scale nature and partially correlated to the topography itself. In fact, when closer inspecting Fig. 3.4b, one can see that also higher-frequency signals are included in the residuals. Theoretically, it is also more likely that higher-frequency signal is related to some errors in the datasets (since often only apparent in one track), but, since in this case there is a physical explanation, this is not assumed here. Thus, also in this area the residuals can be explained sufficiently well.

C Covering the coastal zones of Louisiana, Mississippi, Alabama and Florida, area C is located in an exceptionally flat region (< 50 m elevation). However, when looking at the residuals (Fig. 3.4c), higher-frequency signal is visible. In particular, three things are remarkable: firstly, in this region, one would physically not expect so much short wavelength signal. Secondly, this signals show little correlations between adjacent flight tracks. Thirdly, high frequency signal seems to be spatially confined to the central part of the area. In fact, the suspicious region coincides with the GRAV-D datasets CS1 and CS8. Hence, it seems obvious that there may be some problems with the mentioned datasets. However, at this point, this is just a suspicion, and thus, the datasets are kept unchanged until the final validation (after the collocation see Chap. 4).

Outlook and Limitations. In the scope of the complete processing chain, the validations done at this point can be seen as preliminary. After the collocation of the data (see Sec. 4.6), more accurate assessments which make use of the result of the collocation, are possible. However, the early validations at this stage are still necessary to identify problems early on, especially those which would have a negative impact on the collocation method such as the mentioned long-wavelength residuals or large outliers. The identification of the latter is not explicitly discussed in this section, though such an identification method can easily be implemented by defining some reasonable boundaries regarding the residuals. To reflect the findings of the validation of the GRAV-D data, it can be summarized that (1) no long-wavelength residuals resp. offsets can be identified in the datasets which would hamper the collocation, and (2) there might be some problems with the datasets CS1 and CS8.

Generally speaking, the validation based on a-priori residuals can be seen as just one kind of quality assessment of observations which is, as an advantage, generically applicable, but, as a disadvantage, cannot account for individual properties of certain observation types. For instance, the possibility for airborne observations to acquire further quality indications by performing, e.g., a crossover analysis (see, e.g., *GRAV-D Team*, 2013) or

time series analysis (see P-5 or *Childers et al.*, 1999) should be mentioned. For other observation types other dedicated analysis methods may exist, which makes it difficult to handle such validation techniques individually at this stage. Hence, in the course of the processing chain it is assumed that such observation-related validations have already been performed beforehand as part of the data acquisition itself. In fact, for GRAV-D, some data analysis is already performed for each data block during the processing of the raw measurements (see, e.g., *GRAV-D Team*, 2015).

Chapter 4

Localized gravity field processing

At this stage of the processing it is assumed to have already reduced and validated gravity field observations as input (see Chapter 3). For the reduced observations it is additionally assumed that especially the long-wavelength part of the signal has been removed reliably. This is important since for the localized modelling in this chapter it is necessary for the correlation between observations to also be local, i.e., spatially limited. The aim of the localized processing is to estimate gravity field functionals based on the given observations on predefined locations which are chosen to be optimal for a subsequent global modelling (e.g., on a regular geographic or Gaussian grid). To accomplish this, the method of least squares collocation is used (cf. Sec. 2.3) and improved to overcome several problems of the classical formulation in order to allow for an application on a large scale or even globally in a realistic timeframe.

This improvements comprise as a first step an optimal spatial (2D) data thinning of the observations regarding a specified target resolution based on the evaluation of local convex hulls (Sec. 4.1). This is important, since the main computational effort can be attributed to the actual observation density. Subsequently, the observation geometry is adjusted to correct the long wavelength pattern in the spherical covariance function introduced by the Earth's flattening (Sec. 4.2). Based on the rectified geometry, regionally adjusted degree variances are estimated by using an empirical 2D covariance function as input for a non-linear regression model (Sec. 4.3). Having optimally fitted empiricial degree variances, (homogeneous-isotropic) covariances can be deduced, and, based on these, the collocation can be performed. To dramatically decrease the computational requirements of the collocation, an optimized and automated partitioning approach is implemented based on a 3D divide-and-conquer approach (Sec. 4.4). Eventually, to avoid drift-aways in the estimates and to minimize fringe effects introduced by the partitioning, the kernel function is localized according to the detected correlation length within the residual observations (Sec. 4.5). All mentioned improvements which will be summarized in more detail throughout this chapter are collected to the so-called (P)artition-(E)nhanced-LSC approach, which is also described in publication P-2. Instead of 'regional' gravity field processing this chapter is intentionally named 'localized' gravity field processing to emphasize that, eventually, the improved methods are not limited to a regional scale anymore, but instead are characterized by the need of a locally dominant covariance function.

4.1 Optimal thinning of observations

Outline. Evaluating the formula for LSC (cf. Eq. 2.34) requires matrix multiplication and inversion (i.e., Cholesky factorization) which have a computational complexity t_c (i.e., computing time) in the order of at least

$t_c \sim \mathcal{O}\left(n_{obs}^3\right)$ where n_{obs} is the number of observations. Assuming a constant distance d_{obs} between observations on a regular rectangular 2D grid with edge length d_0, n_{obs} can be approximated by

$$n_{obs} \approx \frac{a_{obs}}{d_{obs}^2} = d_0^2\,\rho_{obs}^2 \tag{4.1}$$

where a_{obs} denotes the covered area and $\rho_{obs} \coloneqq 1/d_{obs}$ the (1D) observation density. Inserting Eq. 4.1 in the formula for the computational complexity t_c one finds:

$$t_c \sim \mathcal{O}\left(n_{obs}^3\right) \approx \mathcal{O}\left(\frac{a_{obs}^3}{d_{obs}^6}\right) = \mathcal{O}\left(d_0^6\,\rho_{obs}^6\right) \tag{4.2}$$

Thus, one can see that the observation density ρ_{obs} is the driving factor for the complexity next to the extents d_0 of the collocation. Thus, when halving the observation density ρ_{obs}, the computation time can be reduced by a factor $(1/2)^6 = 1/64$. This means, that if the original collocation takes one hour, the collocation with the halved density would finish in less than one minute, or, as another example, by increasing the observation distance d_{obs} by just 12%, the computation time can be halved. Hence, for an optimized calculation time of the collocation, it is of utmost importance to not include more observations than necessary according to the targeted spatial resp. spectral resolution. An appropriate thinning of the initial set of observations becomes thus essential at this point.

Another critical reason for thinning the observations before the collocation is the fact that the spatial covariances $\underline{c}_i^{\mathcal{X}} \coloneqq \mathrm{cov}^{\mathcal{X}}\left(L_i, \underline{L}\right)$ of the signal parts of two measurements $L_1^{\mathcal{X}}\left(\boldsymbol{x}_1\right)$ and $L_2^{\mathcal{X}}\left(\boldsymbol{x}_2\right)$ that are located too close to each other, i.e., $\boldsymbol{x}_1 \approx \boldsymbol{x}_2$, become nearly identical when using arbitrarily smooth location-dependent covariance functions (see Secs. 2.3.1, 2.3.3):

$$\lim_{\boldsymbol{x}_1 \to \boldsymbol{x}_2} \mathrm{cov}\left(L_1^{\mathcal{X}}, \underline{L}^{\mathcal{X}}\right) \coloneqq \lim_{\boldsymbol{x}_1 \to \boldsymbol{x}_2} \mathrm{cov}^{\mathcal{X}}\left(\boldsymbol{x}_1, \underline{\boldsymbol{x}}_L\right) = \mathrm{cov}^{\mathcal{X}}\left(\boldsymbol{x}_2, \underline{\boldsymbol{x}}\right) \quad \Rightarrow \quad \underline{c}_1^{\mathcal{X}} \approx \underline{c}_2^{\mathcal{X}} \;\Leftrightarrow\; \boldsymbol{x}_1 \approx \boldsymbol{x}_2. \tag{4.3}$$

This implies that the two rows resp. columns $\underline{c}_1^{\mathcal{X}}$, $\underline{c}_2^{\mathcal{X}}$ of the covariance matrix $C_{LL}^{\mathcal{X}}$ are nearly colinear, which deteriorates the numerical stability. Especially when in addition the measurment noise $C^{\mathcal{M}}$ is low and, hence, does not substantially contribute to the regularity of $C_{LL} = C_{LL}^{\mathcal{X}} + C_{LL}^{\mathcal{M}}$ (cf. Eqs. 2.31, 2.34) singularities in $C_{LL}^{\mathcal{X}}$ are problematic since the needed inverse C_{LL}^{-1} is not well defined. Thus, preventing observations that are too close to each other through thinning does not only speed up the computation, but also increases the solution's numerical stability.

As the need for observation thinning is obvious, an appropriate method is required for this task. Before the method itself can be defined, one has to assess what a 'good' result should look like after the thinning. For this matter, requirements can be determined that define 'good' results:

(1) For any position in space, a good result shall not create gaps that have a larger distance to surrounding observation points than a certain predefined distance d_{th}.

(2) A good result should not shrink the overall area covered by observations.

(3) In a good result, points should be thinned out evenly, i.e., the thinning should preferably create equally sized distances between adjacent observations.

While these requirements seem intuitive at a first glance, a lack of precision can be asserted after a closer inspection. For instance, in requirement (1) the term gap is not defined. From a logical point of view, a gap is an empty space that is surrounded by something, in this case observations. Crucial for this interpretation is the term 'surrounded', which implies that the empty space defining the gap must not be connected to the

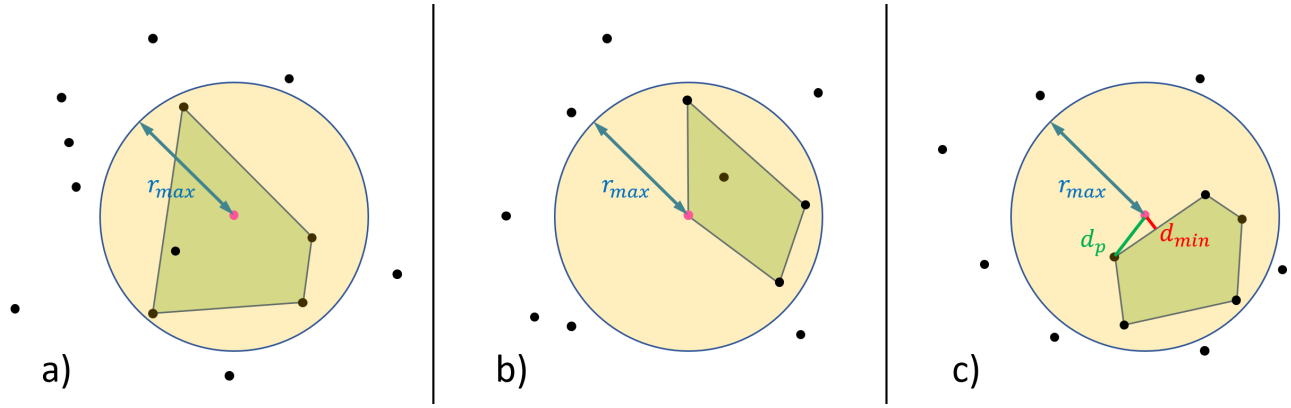


Fig. 4.1 – Illustration of the thinning algorithm's operating principle based on local convex hulls for one evaluation point (magenta, see Sec. 4.1). (a) Evaluation point inside local convex hull. (b) Evaluation point outside local convex hull. (c) Evaluation point close to local convex hull.

'outer' space, i.e., the empty space outside the coverage of the observations. A coverage in terms of a point cloud can be understood mathematically intuitively as its convex hull, which can easily be obtained by, e.g., a Delaunay triangulation (see, e.g., *Preparata & Shamos*, 1985). Thus, when calculating a local convex hull (i.e., a convex hull defined by the set of points which are within a radius of r_{max} regarding an evaluation point), assertions can be made on whether the evaluation point is in a 'gap' and whether the 'gap' gets too large when deleting it (cf. Fig. 4.1). When choosing $r_{max} = d_{th}/2$ and:

- the evaluation point is not part of the local convex hull, the point can safely be deleted according to requirement (1) and obviously also according to requirement (2) since the coverage, defined by the convex hull, is not altered then (see Fig. 4.1a)
- the evaluation point is part of the local convex hull, the point cannot be deleted according to requirement (1) since the new gap would potentially become larger than d_{th} (see Fig. 4.1b). In this case it is also possible that the evaluation point is part of the global convex hull, and, hence, must not be deleted according to requirement (2).
- the evaluation point is part of the local convex hull, but does not contribute notably to it, the point can also be deleted (see Fig. 4.1c). This 'unsharpening' is important for degenerated (1D) cases where points would otherwise not be deleted when located nearly on a straight line (as it is the case, e.g., in along-track direction in airborne gravimetry). As a measure for 'not contributing notably', the distance ratio $q := d_{min}/d_p$ of the closest distance to the convex hull d_{min} and the distance to the next point d_p has to be smaller than a predefined ratio q_{th}, i.e., $q < q_{th}$ (cf. Fig. 4.1c)
- the evaluation point is very close to another point (within a distance r_{min}), the evaluation point is deleted independently of the local convex hull to not endanger the numerical stability of the LSC.

Having this set of rules, an algorithm can be implemented which randomly inspects every observation point and its neighborhood (up to a distance of r_{max}) according to the properties needed by the rules to judge whether the point can be deleted or not. Such an algorithm can be seen as a so-called greedy algorithm, which tries to locally delete as many observation points as possible without concern the global picture. This makes the algorithm very efficient and, e.g., realizable through a divide-and-conquer approach. However, while an algorithm of this form is able to fulfill requirements (1) and (2) (since they are based on local properties), it fails on requirement (3) to thin out points preferably evenly. Establishing equally sized distances within the result is problematic since it would require some sort of global optimization which would become computationally too

complex to handle for large datasets. By using a workaround though, evenness can at least be approximated by a local (i.e., greedy and, thus, efficient) approach. Specifically, it is possible to perform several iterations of thinning by using different, increasing search radii r^i_{max}. In this way, initially only the closest points get deleted and then, incrementally, also more distant ones. This avoids having too small and too large distances in the thinned result. Using n iterations of the efficient greedy algorithm the computational complexity for this approach is still manageable.

The algorithm can be additionally refined by considering datasets individually: for instance, one can choose a smaller thinning distance d_{th} for points of the same dataset than for points of distinct datasets. This leads to intra-dataset thinning with subsequent inter-dataset thinning. The advantage of this approach is, that it can often be assumed that observations of the same dataset have a higher correlation to each other, and, thus, can be deleted more generously than observations of different datasets which are assumed to be more independent of each other.

Application. As in nearly all airborne observation campaigns, also within GRAV-D the observation point density is highly anisotropic and aligned in along-track direction where the spatial sampling is much higher than in across-track. While the across-track sampling ($\sim$10 km) is chosen to reflect the target resolution of the whole campaign, the along-track sampling is defined by the sample frequency of the instrument and the flight velocity of the airplane(s). Thus, the spatial along-track sampling may also vary according to the used instruments and aircrafts. Additionally, airborne observations show a time-dependent correlation which is amplified by the applied temporal (Guassian) low-pass filter needed to reduce the high-frequency noise. In order not to miss any information due to the thinning, it is decided to stay well below the target sampling distance of 10 km and stick to d_{th} =5 km. Going below a 5 km sampling is assumed to be unreasonable since (1) the applied Gaussian low-pass filter to the observations prevents higher-frequency signal (and obviously noise) and (2) the upward continuation of the gravity signal on mean flight altitude ($\sim$6 km) also strongly attenuates shorter wavelength signal (at least over plains). In the course of thinning airborne observations it seems reasonable to mainly only thin out in along-track direction while trying to retain observations of different tracks, e.g., when having cross-overs. To accomplish this, each flight track is interpreted as an individual dataset and a very small (800m) inter-dataset thinning distance is used. This way, only very close cross-over observations get deleted (to retain the regularity of $\mathrm{C}^{\mathcal{X}}_{LL}$). The GRAV-D thinning result of a small excerpt is shown in Fig. 4.2. It can be seen that (1) the algorithm always retains edge points, i.e., points that are part of the convex hull, (2) thins out evenly, and (3) keeps most adjacent points of different flight tracks.

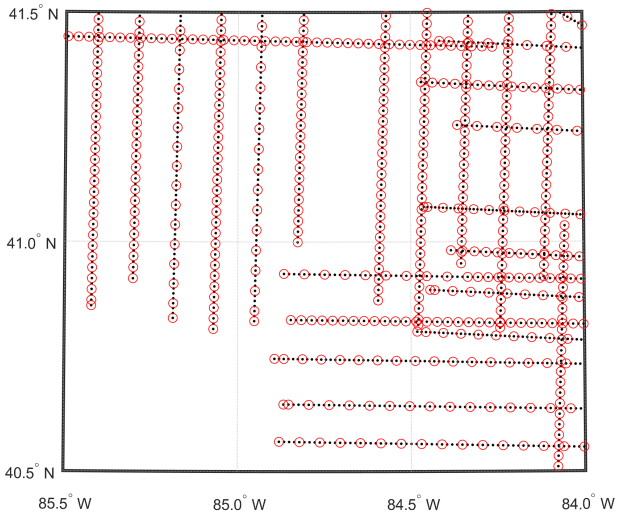


Fig. 4.2 – Illustration of the thinning result of the GRAV-D observations. Black: original observation points. Red circle: Remaining observation points after thinning.

Outlook and limitations. For the application to observations which show a two-dimensional distribution, the presented method works as expected. Especially, when not only having airborne observations but a whole collection of different datasets covering one area, the algorithm can demonstrate its strength by the distinction of intra- resp. inter-dataset thinning. However, there are also some limitations: for instance, as seen in Fig. 4.2, through the iterative thinning it is possible that in the worst case the result contains gaps in the

range of d_{th} to $d_{th}/2$ which contradicts the requirement (3) of an even distribution to some extent. This occurs because sometimes an unfortunate subsampling was hit in the first iterations and propagates in such a way that in the final iteration the distance $d_{th}/2$ to the next observation is either scarcely exceeded or undershot. Generally, the distance d_{th} is always an upper bound for gaps for the algorithm and the resulting average gap distance is statistically somewhere below d_{th} near $3/4\,d_{th}$. Consequently, if one is more interested to hit the gap size d_{th}, on average one should choose $4/3\,d_{th}$ as parameter for the thinning. Another current limitation is the fact that the algorithm is just implemented for the two-dimensional case and not for an arbitrary three-dimensional point distribution. Theoretically, the method of using local convex hulls can easily be extended to the n-dimensional case. However, more dimensions mean more possibilities for degenerated convex hull cases of which one would want to take care of in an actual implementation. This would increases the overall complexity of the algorithm. All in all, the presented algorithm is an attempt to solve the general thinning problem, and has its limitations. For specific thinning problems such as thinning flight tracks dedicated methods may exists which may yield even better results while being much simpler at the same time (e.g., applying a plain subsampling to the observation time series). However, the aim of this work is to present a generic approach that covers as many applications as possible and not just a dedicated one for a limited number of scenarios.

4.2 Handling the oblateness

Outline. In LSC, when using homogeneous-isotropic covariance functions derived from spherical degree variances, the homogeneity and isotropy is related to a spherical reference surface and does not hold for spheroidal reference surfaces as it is the case for the Earth. For example, when calculating signal variances of gravity disturbances δg_i from a spherical homogeneous-isotropic covariance function (cf. Eq. 2.41)

$$\operatorname{var}(\delta g_i) \coloneqq \operatorname{cov}(\delta g_i, \delta g_i) \coloneqq \mathcal{C}_{\delta g \delta g}\left(\Psi_{ii} = 0, u_{ii} = \frac{R^2}{r_i^2}\right) = \mathcal{C}_{\delta g \delta g}(r_i) = \sum_n \frac{1}{R^2}\left(\frac{R^2}{r_i^2}\right)^{n+2} \mathcal{C}_n \tag{4.4}$$

the obtained variance just depends on the geocentric radius r_i of the evaluation point. Hence, when staying on a sphere where $r_i = const.$ variances are homogeneous. However, when replacing the sphere with a spheroid, then $r_i = r(\theta_i)$, and, thus, the variance $\mathcal{C}_{\delta g \delta g}(\theta_i)$ is dependent on the latitude and is not homogeneous anymore. From a physical perspective though, it cannot be justified that the signal variance of the Earth's gravity field has a general dependency on the latitude. Hence, this effect can be interpreted as a purely methodical deficiency when using spherical homogeneous-isotropic covariance functions. The mathematically correct way of circumventing this effect would be to migrate to a spheroidal frame on the basis of spheroidal harmonics. However, within spheroidal harmonics the height component gets much more difficult to handle and the use of such spheroidal covariance functions is not common in geodesy. So, to preserve the commonly used spherical formulation which also has the advantage of being easy to use (especially when deriving covariances resp. cross-covariances to other functionals), another strategy must be applied to avoid the latitudinal dependency. Since this effect is geometry related as explained, one could simply try to alter the geometry to again make r_i latitude-independent. This is possible through a spherical substitution of the spheroid, by defining modified coordinates $\breve{\boldsymbol{x}}_i$ as:

$$\breve{\boldsymbol{x}}_i = \left(\bar{R} + h_i^{ell}\right)\boldsymbol{e}_i\,, \quad \boldsymbol{e}_i \coloneqq \frac{\boldsymbol{x}_i}{|\boldsymbol{x}_i|}. \tag{4.5}$$

Here, $\boldsymbol{x}_i$ denotes the original (geocentric) coordinates of an evaluation point i, $\boldsymbol{e}_i$ the normalized unit vector in direction of $\boldsymbol{x}_i$ and h_i^{ell} the ellipsoidal-geographic height. Through this modification, the ellipsoidal heights are directly projected onto a sphere with reference radius $\bar{R}$ which might be chosen, e.g., as mean geocentric radius of the spheroidal surface in the collocation area. Considering the new geometry of observations that lie within a sufficiently small perimeter, it is found that their relative distances remain widely preserved, since locally, a spheroid (with small eccentricity) can always be closely approximated by an appropriate sphere. The larger-scale deviations in the geometry through this modifications are of minor concern here since it is assumed that correlations on larger distances can be neglected within the collocation when using reduced gravity field items (see also Secs. 4.3, 4.5). Though, by introducing this modification, the u-coordinate of the homogeneous-isotropic covariance function $\mathcal{C}(\Psi, u)$ is altered, and, thus, also the degree variances $\mathcal{C}_n$ have to be computed using the modified geometry (see Sec. 4.3). A similar strategy is also applied in P-4.

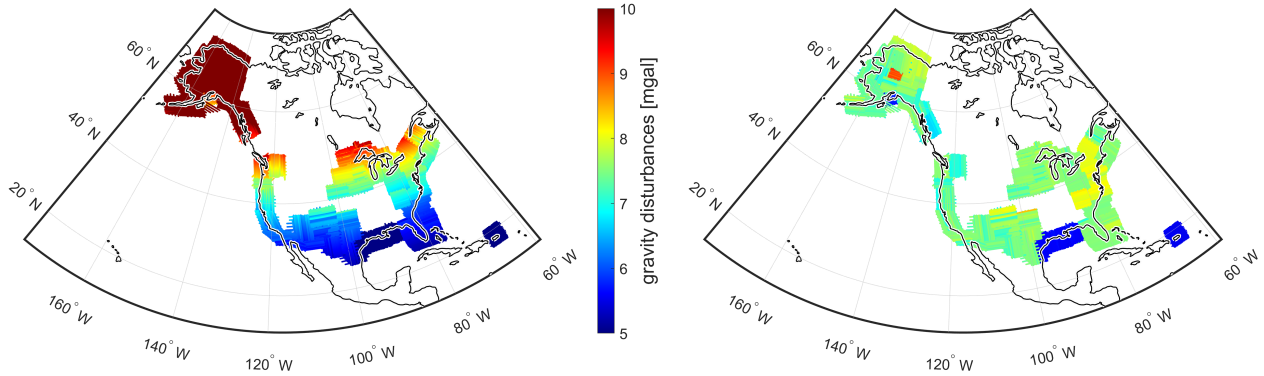


Fig. 4.3 – A-priori signal energy (i.e., square root of the variance entries in $C_{ll}^{\mathcal{X}}$) modelled by the estimated degree variances (see Fig. 4.4). (a) Calculated using the unmodified geometry. (b) Calculated using the spherically modified geometry.

Application. Since the extents of the GRAV-D dataset are large (continental scale) and cover about 50° in latitudinal direction, the influence of the oblateness to the covariance function is substantial. This can be seen when calculating the a-priori signal variances $C_{LL}^{\mathcal{X}}$ by means of $\mathcal{C}_{\delta g \delta g}$ (Eq. 2.41) for the GRAV-D observation sites (Fig. 4.3, using the estimated degree variances from Sec. 4.3): when using the original geometry the modeled signal energy (square root of variance) can become as large as 160 mGal in the northernmost areas and as small as 3.6 mGal on the southernmost areas through the varying geocentric radii (Fig. 4.3a). Trying to perform a collocation in those areas with the given covariance function would either result in a much too rough result (in the north) or in a somewhat too smooth result in (in the south). In contrast, when using the modified geometry, the estimated signal energy is everywhere in a close range from 6 to 9 mGal and reflects mainly the flight altitudes within the specific blocks. It is expected that in this case the collocation yields a desirable result (at least on average). Hence, it is seen that it is essential to apply this modification when dealing with such large collocation areas when trying to use one single consistent homogeneous-isotropic covariance function.

Comments. In the actual application, this 'workaround' is working as expected and is able to efficiently remove the effect of oblateness in the covariance function. In fact, the degree variances related to the modified geometry are closely related to spheroidal harmonics (see Sec. 2.1.2). To see this, one can assume that in a simplified case the gravity field is evaluated on the spheroid, that is, $h_i^{ell} = 0$ and $V^{\mathcal{X}}$ can be evaluated in terms

of spheroidal harmonics by means of Eq. 2.6. Through comparison with the spherical case (Eqs. 2.4, 2.38, 2.39) it becomes obvious that homogeneous-isotropic (surface) covariances from spheroidal harmonics can be derived in the same fashion as their spherical counterparts. The only difference is that the geocentric co-latitude has to be substituted by the parametric co-latitude which, however, has only a very minor influence on the spherical distance ψ between adjacent evaluation points. Eventually, this means that using spheroidal harmonic homogeneous-isotropic covariances rigorously basically corresponds to using the spherical harmonic ones with the modified geometry (at least on or near the spheroidal surface). Only when introducing a height component the two cases would start to diverge. But also then, if a small eccentricity is assumed, the results would have to be very close to each other for smaller heights since spherical harmonics are always the edge case for spheroidal harmonics (when the eccentricity goes to 0). Only when the height values grow larger, the similarity between the approaches and, thus, the validity of this 'workaround' cannot be guaranteed anymore (e.g., for satellite altitudes).

4.3 Derivation of the empirical covariance function

Outline. One of the central aspects when dealing with LSC is the adequate modelling of the signal covariance $C_{VV}^{\mathcal{X}}$. In many applications, $C_{VV}^{\mathcal{X}}$ is not provided a priori but has to be estimated empirically, e.g., through the observations themselves. By doing so, it is often necessary to introduce simplifications due to the lack of information (i.e., due to the limited number of observations available). In case of gravity field modelling the most common simplification introduced is to assume homogeneity and isotropy within the covariances which leads to a covariance function defined by degree variances. While the theory and some systematical considerations about these covariance functions have already been discussed (Secs. 2.3.3, 4.2), it is still uncertain how the needed degree variances $\mathcal{C}_n$ can be obtained. Theoretically, as mentioned in Sec. 2.3.3, degree variances from a global model could be used. However, there are two limitations that come into play then: firstly, degree variances from a global model always represent a global average which might not coincide with the average over the collocation area. Secondly, because residual observations are used as input, also residual degree variances have to be considered, and obtaining these from global models is not trivial (theoretically, from model differences reflecting the model error, e.g., SATOPv2 - other GGM). Thus, an alternative, more empirical method for the determination of degree variances is sought that is able to reproduce the correlations within the given residual observations. For this task, one usually tries to find a set of degree variances that fits the one-dimensional empirical ψ-covariance function which depends only on the spherical distance ψ only (as visualized in Fig. 4.4f). In literature, it is proposed to either adjust a fully analytical degree variance model (e.g., see *Tscherning & Rapp*, 1974) or to somehow scale preexisting degree variances accordingly (cf. P-4). The general problem here is that estimating degree variances directly from a given set of empirical covariances is a highly instable/singular/non-linear task, especially when just using the mentioned 1D empirical $\psi-$covariance function. However, one usually has a second dimension available for breaking down the empirical covariance function, namely the height dependent component u (cf. Eq. 2.39) which allows

the derivation of a 2D ψ, u-covariance function and establishes a (pseudo) linear relation to degree variances $\underline{c}_{\mathcal{H}}$:

$$\begin{aligned} \mathcal{C}_{(ik)} &\coloneqq \mathcal{C}_{\mathcal{TT}}\left(\psi_i, u_k\right) = \sum_n \overbrace{u_k^{n+1} P_n\left(\cos\psi_i\right)}^{=:q^n_{(ik)}} \mathcal{C}_n = \sum_n q^n_{(ik)}\, \mathcal{C}_n = \left\langle \underline{q}_{(ik)} \coloneqq (q^n_{(ik)}),\, \underline{c}_{\mathcal{H}} \coloneqq (\mathcal{C}_n) \right\rangle = \underline{q}'_{(ik)}\, \underline{c}_{\mathcal{H}} \\ \underline{c}_{\mathcal{TT}} &\coloneqq \left(\mathcal{C}_{(ik)}\right) = \overbrace{\begin{pmatrix} \underline{q}_{(11)} & \cdots & \underline{q}_{(n_i n_k)} \end{pmatrix}'}^{=:\mathrm{Q}^{\mathcal{H}}_{\mathcal{TT}}} \underline{c}_{\mathcal{H}} = \mathrm{Q}^{\mathcal{H}}_{\mathcal{TT}}\, \underline{c}_{\mathcal{H}}. \end{aligned} \tag{4.6}$$

Note that a (pseudo) linear relation of the form $\underline{c}_{\mathcal{TT}} = \mathrm{Q}^{\mathcal{H}}_{\mathcal{TT}}\, \underline{c}_{\mathcal{H}}$ is not only available for the covariances of the disturbing potential $\mathcal{T}$ but also for the (cross-)covariances of all other (linear) functionals $f_?$ of the gravity field (cf. Secs. 2.1.3, 2.3.3). Thus, Eq. 4.6 can easily be extended to the general case, so that

$$\underline{c} \coloneqq \left(\underline{c}_{?_i?_k}\right) = \overbrace{\begin{pmatrix} \mathrm{Q}^{\mathcal{H}}_{?_1?_1} & \cdots & \mathrm{Q}^{\mathcal{H}}_{?_{n_i}?_{n_k}} \end{pmatrix}'}^{=:\mathrm{Q}^{\mathcal{H}}} \underline{c} = \mathrm{Q}^{\mathcal{H}}\, \underline{c}_{\mathcal{H}} \tag{4.7}$$

where $\mathrm{Q}^{\mathcal{H}}_{?_i?_k}$ indicates the propagation matrix from degree variances $\underline{c}_{\mathcal{H}}$ to cross-covariances $\underline{c}_{?_i?_k}$ between functionals $f_{?_i}$ and $f_{?_k}$. Theoretically, when having estimates for the covariances $\underline{c} \approx \mathrm{C}^{\mathcal{X}}_{LL}$ this equation system could be inverted to derive optimal degree variances $\underline{c}_{\mathcal{H}}$ by means of least squares adjustment (Sec. 2.2). However, the fact that degree variances $\underline{c}_{\mathcal{H}}$ are defined to be positive impedes this endeavor since it basically destroys the linearity of the relation (hence, denoted as pseudo-linear). To circumvent the possibility of estimating negative degree variances, it is proposed to use a mapping that prohibits negative values, e.g., into the logarithmic space, that is:

$$\underline{c}_{\mathcal{H}} \coloneqq \exp\left(\underline{x}_{\mathcal{H}}\right) \quad \Rightarrow \quad \underline{c}_{\mathcal{H}} = \mathrm{Q}^{\mathcal{H}} \exp\left(\underline{x}_{\mathcal{H}}\right) \tag{4.8}$$

with $\exp(\dots)$ as the exponential function. Obviously, with this mapping $\underline{x}_{\mathcal{H}}$ can cover whole $\mathbb{R}$ while $\underline{c}_{\mathcal{H}}$ is restricted to $\mathbb{R}^+$. Based on Eq. 4.8 an iterative least squares adjustment approach can be implemented to estimate $\underline{x}_{\mathcal{H}}$ resp. $\underline{c}_{\mathcal{H}}$. As mentioned, trying to use the functional model given by Eq. 4.8 directly usually results in rank deficiencies or instabilities in the NEQS due to the nature of $Q^{\mathcal{H}}$ and the generally limited number of covariance samples available (since observations are usually regionally limited). To avoid this, one can use additional regularization techniques, such as demanding smoothness of the degree variance curve (e.g. by using splines) or by limiting the slopes. More details on these issues can be found in publication P-2. An additionally advantage when using 2D grids of empirical covariance functions is that the calculation of actual covariances between observations, that is the calculation of the covariance matrix $C^{\mathcal{X}}_{VV}$, can be vastly accelerated. This is possible since the calculation can be reduced to a simple interpolation within the 2D grid (e.g., c.f. Fig. 4.4c assuming a certain smoothness). This way, an explicit evaluation of Eq. 2.39, a fairly time consuming task (due to the need of evaluating many Legendre polynomials for every point pair) , is avoided. The respective procedure and the mathematical background for its application on, amongst others, cross-covariance is explained in more detail in publication P-2.

Application - degree variance estimation for the GRAV-D dataset. For the GRAV-D observations, empirical (homogeneous-isotropic) covariances based on gravity disturbances $\mathcal{C}_{\delta g \delta g}$ where estimated onto a discrete $\psi_i \times u_k$ grid which is obtained by binning ψ resp. u into classes $[\psi_i, \psi_{i+1}[$ resp. $[u_k, u_{k+1}[$. For ψ, a range from $]0^\circ,\ 4^\circ]$ was chosen by using 400 bins, and, for u, 30 bins between $\sim [3.7\ \text{km},\ 11.3\ \text{km}]$ in terms of the

so-called equivalent height h_{eq} are used. h_{eq} is a vivid interpretation of u as the corresponding height in case that both points reside on this height. It can be calculated by:

$$h_{eq} \coloneqq R\left(\frac{1}{\sqrt{u}} - 1\right). \tag{4.9}$$

The resulting covariances on the $\psi_i \times u_k$ grid are shown in Fig. 4.4: while some disagreements are visible in certain height bands, a relatively clear systematic pattern is discernible. Based on this grid, empirical degree variances are estimated as previously explained. Concretely, the estimation was performed on a reduced number of B-spline control points using smoothness and slope constraints (see publication P-2). The estimated degree variances are depicted in Fig. 4.4b. From a first check, the estimated degree variances look reasonable since they resemble basically the degree variances of the SATOP model (cf. Fig. 3.1). They can be inspected more closely when applying Eq. 4.7 to calculate the estimated 2D covariance function (cf. Fig. 4.4c) and inspecting the differences to the empirical covariance function (Fig. 4.4e): it can be seen that the overall pattern is removed from the empirical covariance function. However, some larger areas with either a slightly positive or negative trend remain. This can probably be attributed to the fact that it is simply not possible to perfectly describe the vast area covered by the GRAV-D dataset by only one homogeneous-isotropic covariance function, since the statistics may vary locally depending on, e.g., roughness of the terrain, underground structures, flight altitude or speed. Nevertheless, it can be assumed that the obtained result shows the best fit to an average observation in the area. Also, during the estimation, the sample number per bin (Fig. 4.4d) is introduced as individual weight. Consequently, when inspecting the differences (Fig. 4.4e), it is evident that the estimation fits best where the sample density is highest, and all larger disagreements are located where the covariance function is not well determined due to the low sample number (especially visible within certain height bands). This behavior supports the assessment that the estimate is reliable and that the procedure is working as expected. Summarizing the results, one can also inspect the aggregated (over the height component) 1D visualization of the 2D grids (Fig. 4.4f) depending solely on the spherical distance ψ: in this representation the strong agreement between empirical and estimated covariances is even more prominent as the differences are basically zero. An interesting feature are the visible fine ripples in the empirical covariance function. They are obviously induced by resonances with the spatial sampling distances of the GRAV-D observations since the same ripples also occur in the number of samples per bin. This could indicate that the data includes some artefacts (maybe induced by the filtering of the airborne observations) since theoretically the sample number and the obtained covariance should be mostly uncorrelated (at least when the sample number is large anyway).

Remarks and restrictions. In the actual example, the presented method using 2D covariance grids works as expected and without problems. However, it cannot be stated that this will be the case for any given scenario. Particularly, when the collocation area becomes smaller and all observations are close to each other, the estimation problem gets more and more instable and one has to apply an even stronger regularization to still obtain reasonable a result. This leads to another problem, namely that the presented method has to be adjusted manually for each case, e.g., regarding reasonable parameters for the $\psi_i \times u_k$ grid and the design of the whole regularization. Up until now, finding the right setup can be considered as an iterative trial-and-error procedure and a fundamental improvement of this situation is not yet in view. Moreover, and this is a general shortcoming, the use of degree variances themselves for modelling spatial covariances has its limits as seen, e.g., in the differences between the empirical and estimated covariance function (Fig. 4.4e), where some remaining systematics are obvious. As already previously mentioned, this can probably be mainly attributed to the assumed homogeneity (and, thus, also isotropy) which, on a large scale, cannot hold true.

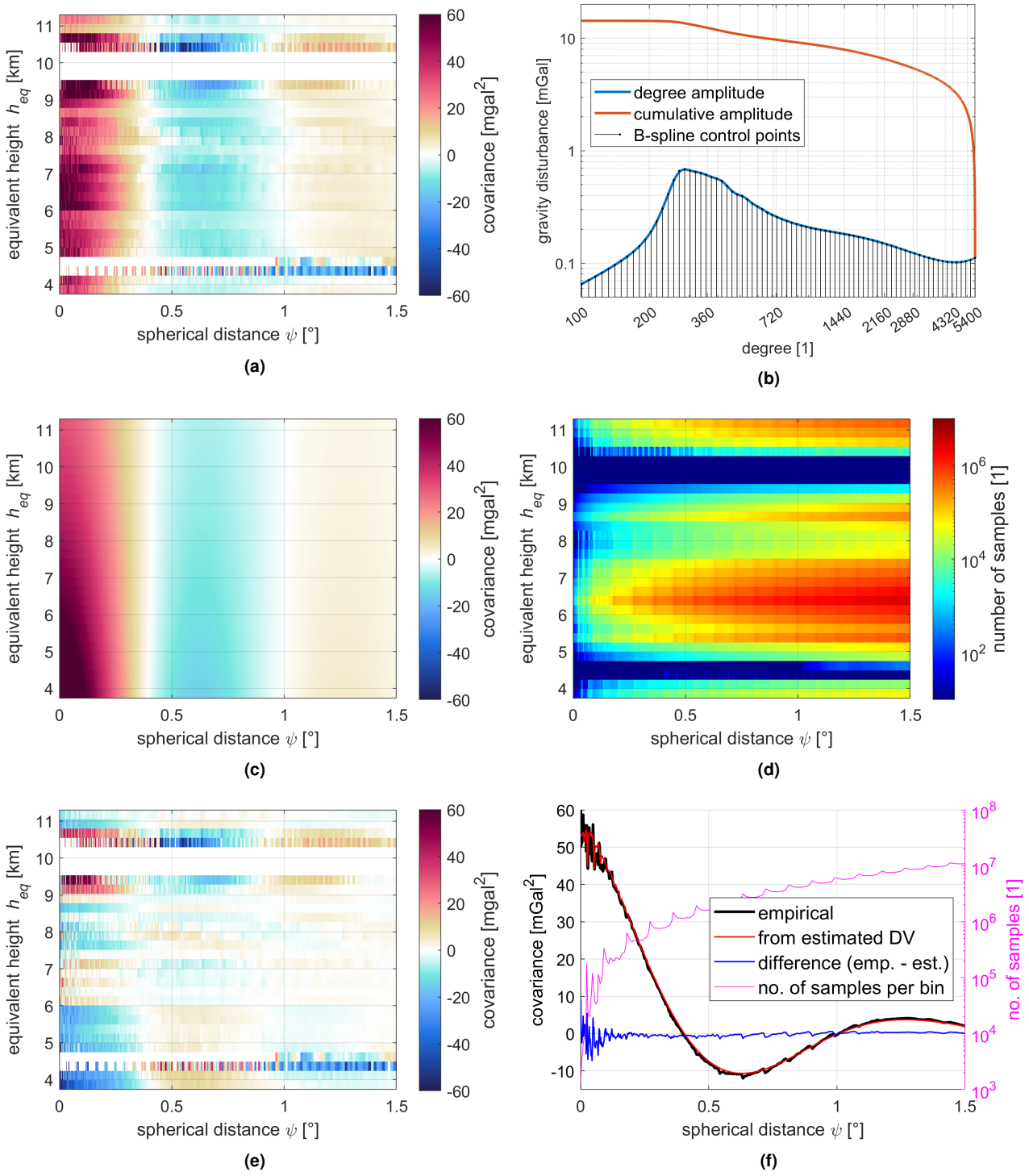


Fig. 4.4 – Empirical homogeneous-isotropic covariance function and derived estimates. (a) 2D empirical homogeneous-isotropic covariance function in terms of gravity disturbances regarding the spherical distance ψ and the equivalent height h_{eq} (see Eq. 4.9). (b) Estimated degree amplitudes (i.e., square root of degree variances) up to d/o 5400 in terms of gravity disturbances (blue line) and cumulative amplitudes (from right, red line). B-spline control points used for the estimation are depicted through vertical black lines. (c) 2D covariance function (cf. Fig. 4.4) derived from the estimated degree variances (Fig. 4.4b). (d) Number of samples per bin for the calculation of the empirical covariance function (Fig. 4.4). (e) Difference between empirical covariance function (Fig. 4.4) and estimated covariance function (Fig. 4.4c). (f) aggregated 1D view on the covariances depending only on the spherical distance ψ obtained by a weighted average (weighted by number of samples, see Fig. 4.4d). Black: empirical covariance function (cf. Fig. 4.4). Red: estimated covariances function (cf. Fig. 4.4c). Blue: difference between empirical and estimated covariance function (cf. Fig. 4.4e). Magenta: number of samples per spherical distance class (cf. Fig. 4.4d).

There exist several theories to introduce non-homogeneous covariance functions which are comprehensivly described, e.g., by *Darbeheshti* (2009). Though, even if such methods exist, there is also an inherent limitation that probably prevents an actual implementation of those theories: when allowing inhomogeneity, one also always introduces a new 'dimension' (or more precisely, two dimensions on the sphere) of parameters which are needed to characterize the local variations. Since those parameters must be additionally estimated from the data, such an estimation would become highly instable (when not assuming a very dense observation coverage) and an 'overfit' is very likely to happen. In fact, when fitting the covariance perfectly to the actual observations, no real advantages from the LSC (working as a filter) can be expected. Thus, due to the lack of better knowledge of the actual covariances (and to prevent overfitting), there is, at least as of now, no real alternative to using degree variances. A possible option to circumvent this deficiency by an iterative procedure will be discussed briefly in Sec. 4.6.

4.4 Automated partitioning

Outline. From a computational perspective, the method of LSC as described in Sec. 2.3 has the same complexity as least squares adjustment. This can be seen, e.g., through Eq. 2.29 where both methods basically require a matrix multiplication and a matrix inversion. As already mentioned in Sec. 4.1, both of these operations have a complexity in the order of $\mathcal{O}(n^3)$ in terms of the estimates resp. observations, and, so, also the overall complexity of the methods is in the same order. A complexity or computation time of $t_c \sim \mathcal{O}(n_{obs}^3)$ implies that one is usually strongly limited regarding the number n_{obs} of observations that can be used as input for the collocation (cf. Fig. 4.5). Also the memory requirements in the order $\mathcal{O}(n^2)$ for keeping fully occupied matrices in the working memory usually poses a hard limit. For instance, assuming a workstation with 128 GB memory (RAM), a square matrix with a dimension of maximum 130.000 can practically be processed (e.g., inverted, see Fig. 4.5). As already pointed out in Sec. 4.1, assuming a constant observation density $\rho_{obs} = const.$, the number of observations increases linearly with respect to the collocation area or quadratically regarding the edge length d_0 (assuming a square area):

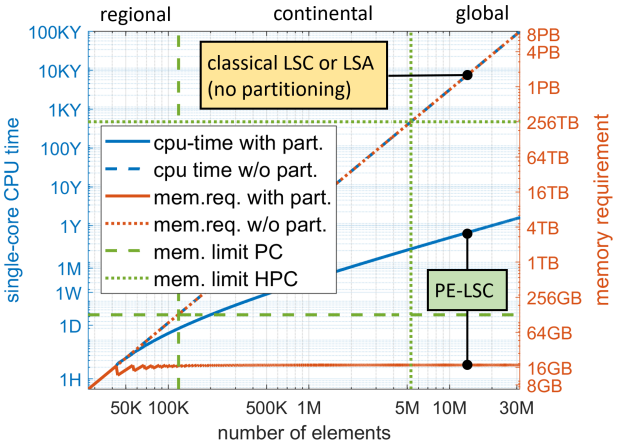


Fig. 4.5 – Comparison between the classical LSC/LSA approach and the partitioned LSC method when assuming a constant linear observation density of $\rho_{obs} = 1\ \mathrm{obs.}/4$ km and a partitioning parameter $r = d = 2.5$ (from P-2). Computation times are scaled according to the single-core performance of an Intel Skylake-SP processor with AVX512 enabled. Solid blue: Computation time of the partitioned LSC method depending on the number of input points (left axis). Dashed blue: computation time of the classical LSC/LSA approach without partitioning (left axis). Solid orange: memory requirements (RAM) when using partitioning (right axis). Dotted orange: memory requirements without partitioning (right axis). Dashed green: memory limit (horizontal line) of a workstation PC with 128 GB RAM and the respective maximum collocation size (unpartitioned). Dotted green: Same for an HPC system with 256 TB of RAM.

$$t_c \approx k_0\, n_{obs}^3 = \overbrace{k_0\, \rho_{obs}^6}^{=:k}\, d_0^6. \tag{4.10}$$

Here, k_0 describes some (unknown) scaling factor accounting for the actual performance of the calculation. Since the computation time t_c increases with the power of 6 regarding the extents d_0 of the area to collocate, at the time of writing, the classical LSC approach is always limited to a local or limited regional extent and cannot be applied in this way on larger areas (e.g., continental scale or even global). According to Eq. 4.10, the only way to reduce the computation time is to somehow decrease d_0 which, obviously, is not possible without introducing some assumptions resp. simplifications. In the actual framework, when using residuals, it seems legit to assume a certain localized correlation of an observation to its surrounding since longer-wavelength correlations should be compensated by the reduction model (more specific, by the precise satellite data within the reduction model). An inspection of the previously derived empirical covariances (see Sec. 4.3 and Fig. 4.4f) should support this assumption, where covariances should converge towards zero with an increasing spherical distance ψ. Exploiting this simplification justifies the estimation of collocated values by just considering observations within a certain distance to the evaluation point. As a consequence, it is possible to partition the collocation area accordingly into smaller independent parts by ensuring that a minimum surrounding, specified by a radius r, is included for every evaluation point (similar as proposed by *Reguzzoni et al.*, 2009). This can be guaranteed by adding a buffer with radius r to every partition (of evaluation points, see Fig. 4.6a). For the determination of the arising complexity, in a simplyfied case it is assumed that the square area with edge length d_0 is partitioned into equally sized smaller squares with edge length d, which leads to a number of

$$m \approx \left(\frac{d_0}{d}\right)^2 \tag{4.11}$$

(assumed identical) partitions. Since for every evaluation partition an extended observation partition with edge length $d_e = d + 2r$ has to be considered, the new overall computation time t_{cp}, using partitioning can be approximated by substituting d_0 by d_e in Eq. 4.10 and multiplying the result by m (because the same effort for every partition is assumed):

$$t_{cp} \approx m\, k\, d_e^6 = \left(\frac{d_0}{d}\right)^2 k\,(d+2r)^6 = \overbrace{k\, d_0^2}^{=:\bar{k}} \frac{(d+2r)^6}{d^2} = \bar{k}\,\frac{(d+2r)^6}{d^2}. \tag{4.12}$$

Since in this equation the partition edge length d is not yet predefined, it can be chosen in such a way that the expression $t_{cp} = t_{cp}(d)$ becomes minimal. By setting the derivative to zero one finds that the minimum is found when $d = r$ (see publication P-2) and so the optimized computation time t_{cpo} becomes:

$$t_{cpo} \approx k\, d_0^2 3^6\, r^4 \quad \Rightarrow \quad \frac{t_c po}{t_c} = \frac{k\, d_0^2\, 3^6\, r^4}{k\, d_0^6} = 9\left(\frac{3r}{d_0}\right)^4, \tag{4.13}$$

with t_{cpo}/t_c as the computation time reduction when applying partitioning. It is seen that (1) the complexity becomes linear regarding the collocation area d_0^2, and that (2) the computation time reduction is strongly dependent on the chosen buffer width r and, thus, r should be chosen as small as possible. Fig. 4.5 shows a comparison between the classical and partitioned approach regarding runtime and memory requirements by means of a realistic scenario (taken from publication P-2). As conclusion it can be stated that through the partitioning (1) the memory requirements usually become unproblematic, and (2) even a collocation on a global scale becomes feasible. As already pointed out, the optimal choice of r depends on the actual correlation length of the observations/residuals and will be discussed further in Sec. 4.5.

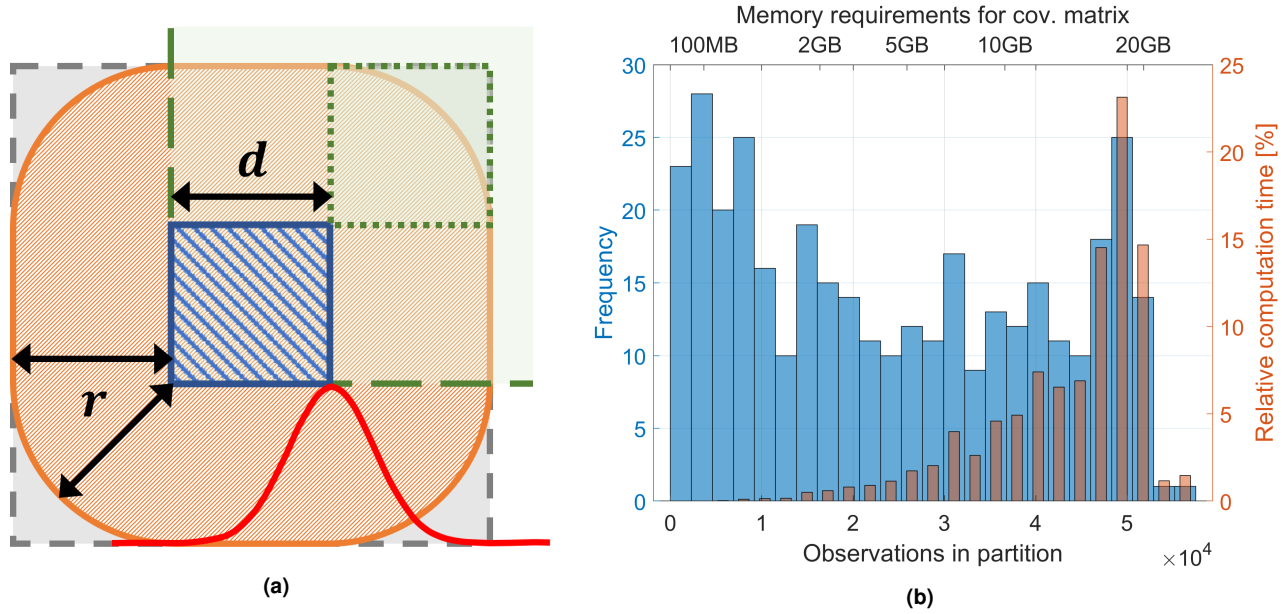
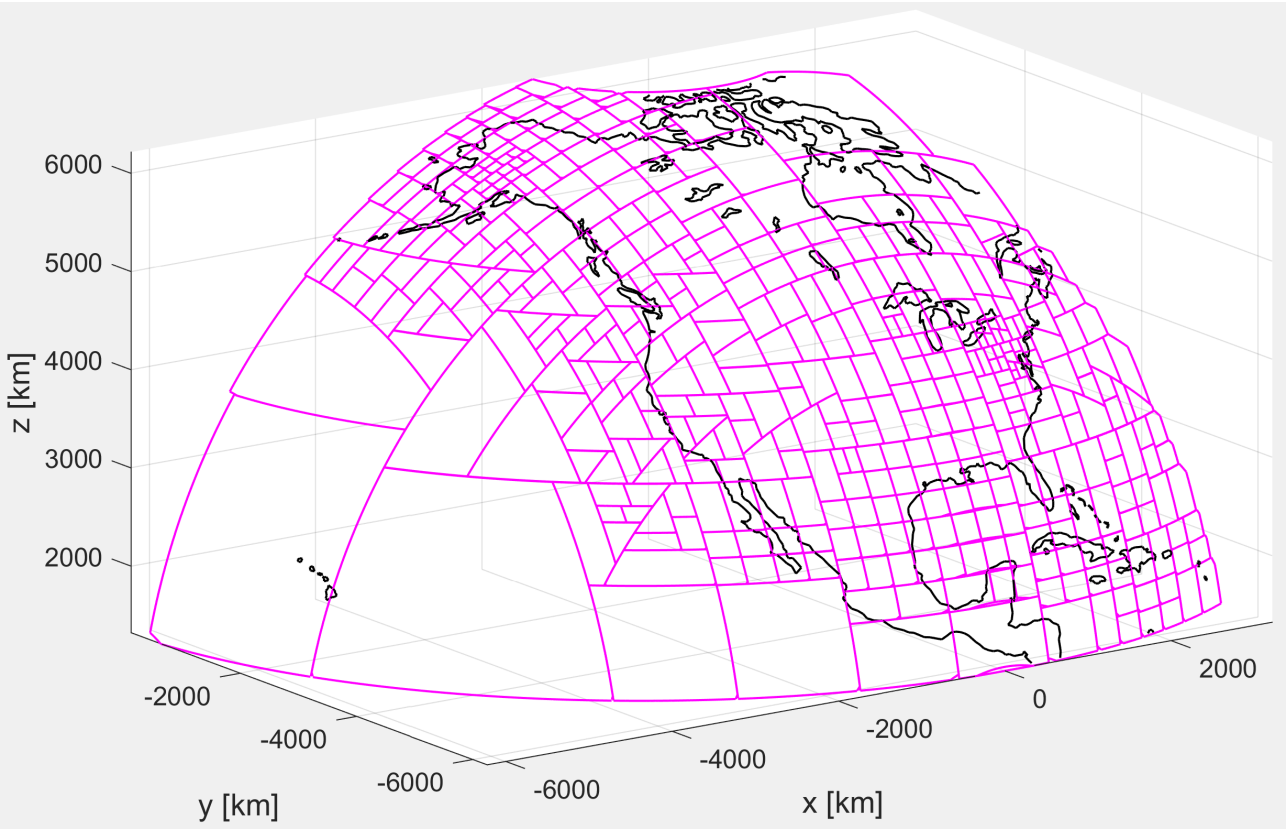
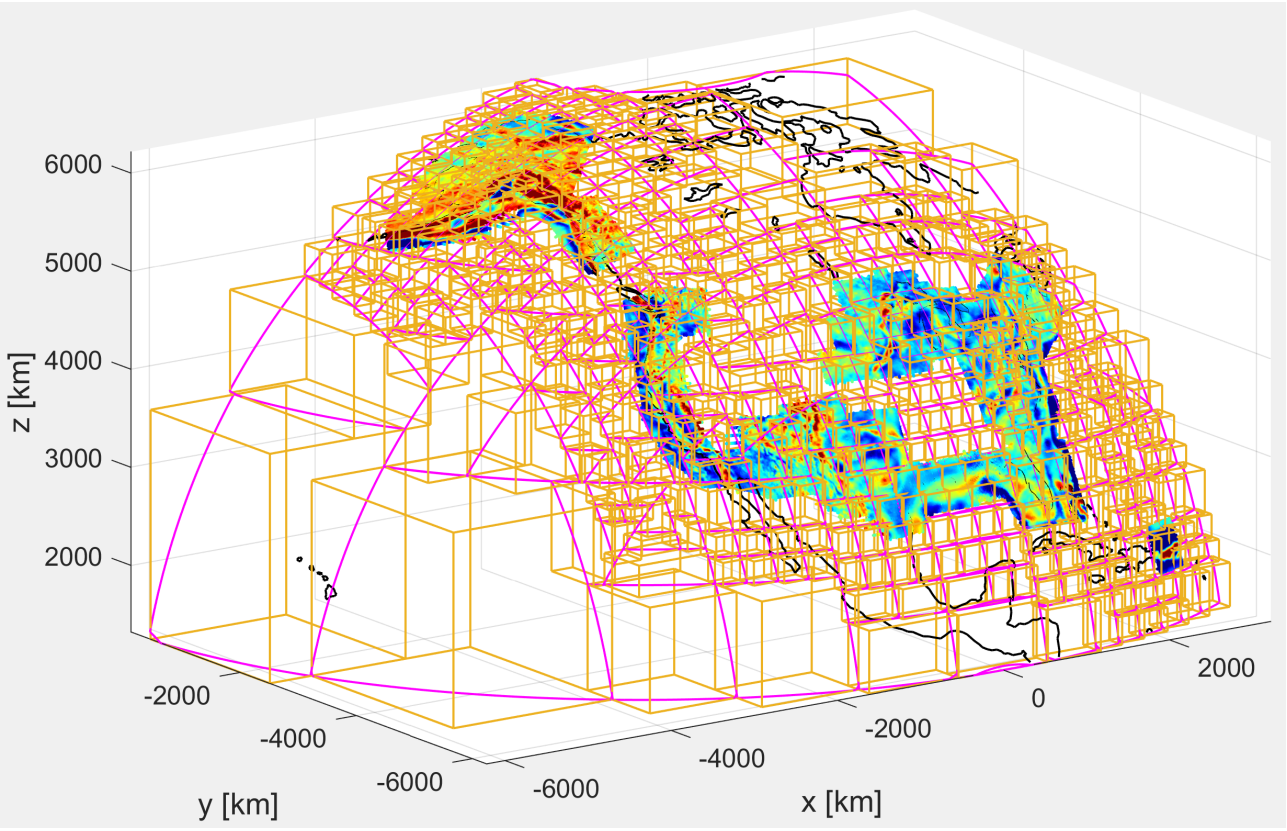


Fig. 4.6 – (a) Simplified (2D) visualization of the partitioning approach. Hatched blue area: core area of one partition where evaluation points may be located in (simplified as 2D square area with edge length d). Hatched orange area: extended partition wherein observations are sought for the evaluation points in the inner partition. The inner partition is extended with a buffer of width r in every direction to ensure a minimum surrounding of radius r around each evaluation point. Hatched green area: a neighboring partition that overlaps the actual partition through its buffer (indicated with dashed green line). (b) Blue bars: Frequency distribution of the partition sizes (in terms of number of observations per partition) created by the automated partitioning algorithm for the GRAV-D dataset (see Fig. 4.7). Red bars: Relative aggregate computation time for all partitions in the respective partition size bins.

Implementation - Application. The presented idea of partitioning enables the needed speed-up to perform a large-scale collocation (as in case of GRAV-D) within a realistic timespan and with reasonable computational resources. However, to be eventually applicable on a large scale an automated partitioning approach is needed, since for large areas the number of required partitions and consequently, the effort to set them up manually, increases. Especially if some areas feature a high observation density, individual treatment may be needed to ensure to stay, e.g., within the memory limit of the system. A simple but efficient approach that accomplishes the partitioning is found in a 3D divide-and-conquer algorithm. The idea behind it is to recursively divide the collocation area along the coordinate plane with the largest extent until a certain maximum partition size falls below a certain threshold or until the number of points per (extended) partition is small enough. With this procedure, also the mentioned special treatment of areas with high observation densities can be handled easily. The algorithm operates based on 3D geocentric coordinates, which holds the advantage of functioning without projections, and not being affected by distortions and all the difficulties connected with it. For collocating the GRAV-D dataset, the resulting partitioning (without buffers) is depicted in Fig. 4.7a. The different partition forms arise through the intersection of the 3D partitions with the Earth's surface as shown in Fig. 4.7b. As already mentioned, it can be recognized that the algorithm adjusts the partition size depending on the point density in the respective areas to keep the memory limits. This is also seen when inspecting the statistics of the partitioning in terms of a frequency distribution regarding the number of observations per partition (Fig. 4.6b). For the used computing system, the memory limit for C_{LL} was set to 20GB, and it is seen that also the largest partitions are limited to this size ($\sim$50k elements). The peak at 50k elements in the frequency distribution shows that probably some partitions where further divided to keep this limit (by sacrificing the optimality in the sense of the ideal partition size). All in all, for GRAV-D, the algorithm produces 360 (non-empty) partitions (when using $d = r = 4°$, see Sec. 4.5). The computation time for the least squares collocation of these



(a)



(b)

Fig. 4.7 – Result of the automated partitioning of the GRAV-D collocation area when using $r = d = 4^\circ \approx 900$ km. (a) intersections of the 3D partitions with the Earth's surface (magenta lines, approximated through a spheroid). (b) Same intersections (magenta lines) but with the 3D partitions (yellow lines) and observations visualized.

360 partitions constitutes about 1.5 days on a workstation with 28 cores (on an Intel Haswell architecture). In the example of GRAV-D, the gain in terms of computation time through the partitioning can be quantified by a factor of $\sim$7 (days instead of weeks), and more importantly in terms of memory requirements by a factor of $\sim$60 (20GB instead of 1.2TB). In Fig. 4.6b it can also be seen that about 60% of the computation time is needed for the partitions containing more than 45k observations. This shows once again (as already derived in Eq. 4.2 and Eq. 4.12) the need to reduce the partition size as much as possible which can either be done by (1) thinning out more observations (cf. Sec. 4.1) or (2) reducing the needed buffer radius r (which is possible, e.g., when having a better reduction model, see Sec. 4.5).

4.5 Localizing the collocation kernel

Outline. The partitioning as applied in Sec. 4.4 assumes that the collocations within the different partitions can be performed independently of each other. Obviously, this assumption is in general not valid since the LSC estimator (i.e., kernel) $\widetilde{\mathrm{A}}_L^S = \mathrm{C}_{SL}\mathrm{C}_{LL}^{-1}$ (see Eq. 2.29) requires matrix inversion of C_{LL} which correlates all observations to a certain extent. However, from the perspective of the covariance function which generates C_{LL} (see Sec. 4.3, Fig. 4.4f) there is a certain decorrelation with the spherical distance ψ discernible when assuming previously reduced observations. Beyond a certain distance $r \approx 2\sin(\psi_{max}/2)$ in most cases it can be expected that no significant correlation between observations remains (when having applied a reduction model with good performance in the long wavelengths). Under this assumption it is counterintuitive that a distant observation can have a significant influence on a local estimate. However, it has been shown that this can be the case when an estimate has no observations in its surroundings. Then, the spatial structure of the estimator expands and includes a much wider area (which is comprehensible since no local observations are available to support the estimate). This behavior is generally undesirable when trying to retain a local influence of observations to the estimates since through the limited buffer width of the partitioning in this case one would still omit a certain number of non-zero elements in the original estimator. To prevent this, it is proposed to limit this influence by applying a distance dependent weighting to the estimator to forbid non-zero elements beyond the buffer width r (e.g., c.f. Fig. 4.6a). As pointed out in more detail in publication P-2 several weighting functions might be suitable for this task. In the actual implementation though, it is decided to use a slightly modified Gaussian bell function $w^G(\psi_{ik})$ since it is (1) smooth and (2) shows a relatively sharp transition at a specified spherical distance ψ_{max}:

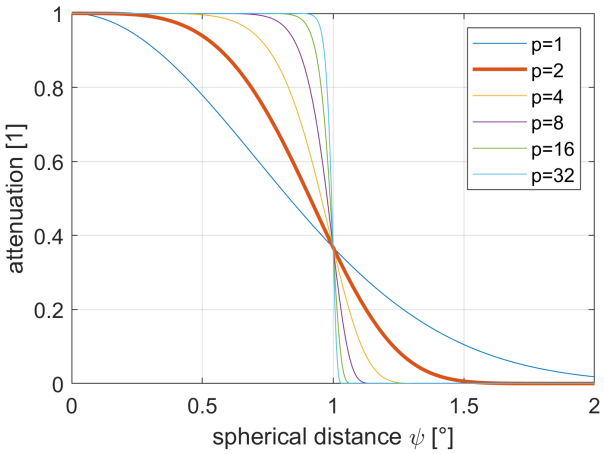


Fig. 4.8 – The (spherical) distance dependent attenuation function w^G (Eq. 4.14) for $\psi_{max} = 1$ and different values for the sharpness p.

$$w_{ik}^G = w^G(\psi_{ik}) \coloneqq \exp\left(-\left(\frac{1-\cos\psi_{ik}}{1-\cos\psi_{max}}\right)^p\right). \tag{4.14}$$

The parameter p can be used to control the sharpness of the transition. A higher value corresponds to a sharper transition (see Fig. 4.8). As default a value of $p = 2$ is chosen which represents a good compromise between smoothness and an adequate attenuation. Applying this attenuation function to the elements $\widetilde{a}_{ik}$ of

the estimator $\widetilde{\mathrm{A}}_L^S$ the elements $\breve{a}_{ik}$ of the modified estimator $\breve{\mathrm{A}}_L^S$ can simply be obtained by an element-wise multiplication:

$$\breve{a}_{ik} = w_{ik}^G \, \widetilde{a}_{ik}. \tag{4.15}$$

Eventually, this modification has the effect of tieing estimates in far-off locations (i.e., locations with no observations in their neighborhood) towards zero, which only demands a very small sacrifice in terms of optimality while widely reducing the possibility of fringe effects between different partitions. On the other hand, in locations with close-by observations, the modification should have no influence at all due to the already local structure of the original estimator there. To quantify the possible difference between original and modified estimator a kernel disparity measure Δ_i^{ker} is defined by:

$$\Delta_i^{ker} = \sqrt{\frac{\sum_k \left(\widetilde{a}_{ik} - \breve{a}_{ik}\right)^2}{\sum_k \widetilde{a}_{ik}^2}}. \tag{4.16}$$

This measure Δ_i^{ker} is normalized in such a way that it becomes 0 when the modified kernel equals the original, and 1 if there is no similarity at all between original and modified kernel.

Application. Before the kernel localization can be applied to the GRAV-D collocation, an appropriate spherical transition distance ψ_{max} has to be chosen as parameter for the attenuation function. The choice of this distance should depend on the applied covariance function (which is taken as a hint of how the actual estimator may behave). As a rule of thumb, ψ_{max} should be large enough to preserve the main features of the covariance function. In case of the reduced GRAV-D observations, $\psi_{max} = 2^\circ$ is chosen which is located near the second minimum of the covariance function (see Fig. 4.9a). Having fixed the value for ψ_{max}, also the maximum (spherical) distance can be derived beyond which the localized estimator can be assumed to be zero. For small values for ψ_{max} (e.g., below 5°) it has been shown that a value of $5/3\,\psi_{max}$ is a good compromise since the attenuation at this distance is already larger than 99.95% (cf. Fig. 4.9a). Thus, in accordance with Sec. 4.4, a value of $5/3\,\psi_{max}$ might also be a good choice for a minimal acceptable buffer distance r in this case. However, in the actual GRAV-D case, a buffer distance of $r = 4^\circ$ is chosen for the partitioning instead, since the computational limits were uncritical and, furthermore, an even higher consistency between partitions can be expected from a larger buffer distance (see P-2 for comparisons between different buffer widths).

To present the influence of the kernel localization to the estimates, a smaller region of the GRAV-D collocation area is chosen (same region as in Fig. 4.2 located between Michigan, Indiana and Ohio). This region was chosen because it contains area with and without observation coverage. As previously pointed out, the localization has quite a different behavior in the respective cases. To demonstrate this, the estimator is explicitly examined in two sample locations, one positioned in the area covered with observations (point A, see Fig. 4.9b) and the other one located in the observation-free area (point B, see Fig. 4.9c). While the estimator for point A remains widely unchanged through the localization ($\Delta_A^{ker} = 0.0004\%$) the estimator for point B is strongly modified ($\Delta_B^{ker} = 40\%$). This can also be seen when comparing Fig. 4.9c and Fig. 4.9d: in the localized case (Fig. 4.9d) all components beyond the buffer distance r are widely suppressed, while in the original case they reach far beyond the buffer width r. Eventually, this non-suppressed components make the occurrence of fringe effects between partitions very likely, as shown, e.g., in Figs. 4.10d, 4.10e at longitude $\sim$89°W where a partition boundary is located. The influence of the kernel localization is presented in a more complete picture in Fig. 4.9e when inspecting the kernel disparity Δ^{ker} as defined by Eq. 4.16: it is evident that the estimates are preserved to a very high degree in all locations close to existing observations and just differ in areas with-

out observations. Although this kernel disparity measure is still independent of the actual observations, it is a good indicator of where the actual estimates will differ from each other. This is confirmed by inspecting the difference of the estimates obtained through the localized and the original kernel (Fig. 4.10d). Here, larger differences only occur where the kernel disparity is high, i.e., in locations without observation coverage. There, the localization has the effect to constrain the estimates towards zero (if no observations lie within a spherical distance of r around the evaluation point the estimate becomes effectively zero, see Fig. 4.10f). Even though this tie implies larger differences to the original solution, the formal error as obtained through covariance propagation using $\check{\mathrm{A}}_L^S$ increases only slightly. This is seen when examining the increase of the formal error in the modified solution (since the original LSC estimator minimizes the formal error any, modification to it must actually lead to an increase of the formal error, see Figs. 4.10b and 4.10c): although in the observation-free areas the formal error increases the most, the increase is still everywhere below 2% of the original error (or less than $\sim$100 µGal in total). This sacrifice of optimality is seen as uncritical in the actual application where observations are assumed to have an accuracy of about 3 mGal.

Limitations and discussion. As empirically shown for the GRAV-D collocation, the kernel localization effectively reduces fringe effects (e.g., compare Figs. 4.10e and 4.10f). However, this statement cannot be easily extended to the general case, because the behavior of the covariance function depends on (1) the quality of the measurements resp. reduction model, and (2) on the given functionals (e.g., gravity disturbances, potential values, gravity gradients, etc.). If the assumed locality of the covariance function is disturbed for some reason, the correct behavior of this method cannot be assumed anymore. This means that the localization may yield suboptimal results. For a general application the potentially variable behavior of the covariance function has to be taken into account, which means that it may be neccesary to apply different localizations (e.g., through different attenuation functions) depending on the actual case. For instance, when estimating potential values from gravity disturbances, a wider attenuation function may be neccesary than when estimating gravity gradients from gravity disturbances (due to the different eigenvalues present in the construction of the covariance functions). As of now, such an individual treatment of the localization is not implemented simply because it was not really necessary for the application cases (GRAV-D and publication P-2). Additionally, one further limitation of the current implementation is that the applied attenuation function is only dependent on the spherical distance and not on the height component (cf. Sec. 4.3). Further improvements (even if they are not significant) may be achieved when considering a 2D attenuation function. Due to the attenuation of shorter wavelengths with higher altitudes it is generally expected that the (spherical distance dependent) attenuation function needs to get wider (i.e., ψ_{max} should be increased, see Eq. 4.14). Eventually, it should be noted that even when applying the localization perfectly, it cannot be guaranteed that fringe effects between partitions are fully avoided. This is apparent since in different partitions different covariance matrices C_{LL} containing different observations are inverted. Since the result of the inversion generally depends on all elements of the matrix (that is, all observations), inverted covariance matrices may fundamentally differ from each other even if just one element (observation) is altered. Nonetheless, a certain insensitivity regarding far-off observations is expected when inverting covariance matrices from locally dominated covariance functions (see publication P-2 for a more detailed treatment of this topic). Ideally, one could calculate every estimate independently by regarding its individual surrounding which would introduce the need to invert the respective covariance matrix for every estimate individually. By doing so, fringe effects could be completely avoided since no partitions would exist then in the first place. However, as pointed out in Sec. 4.4, by doing so, the optimal computation time would be missed by a significant margin (cf. Eq. 4.12 by setting $d = 0$) since an explicit matrix inversion would be necessary for every single estimate. Hence, this idea has a purely theoretic character.

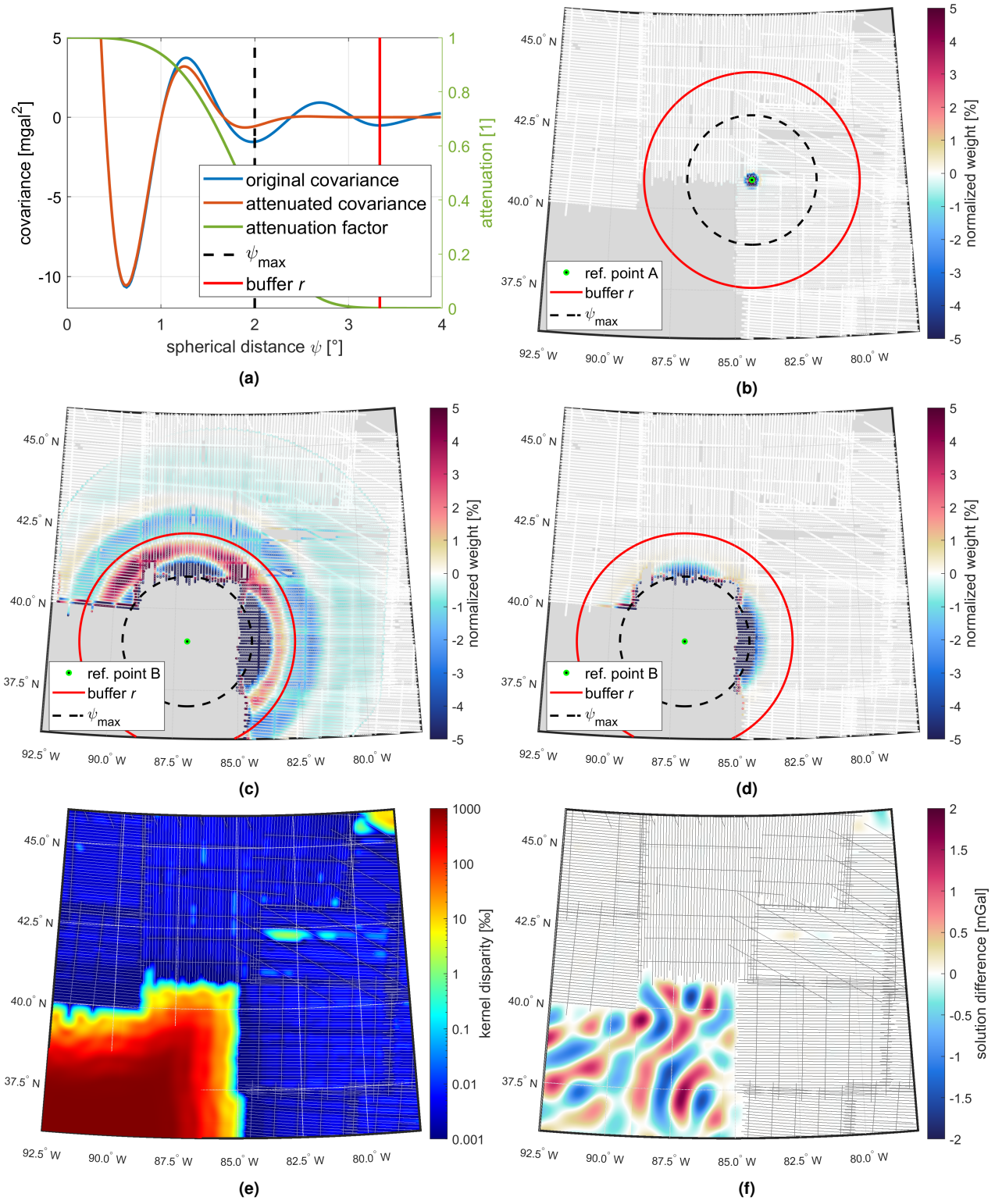


Fig. 4.9 – Influence of the kernel localization on the LSC solution. (a) The exemplary influence of the attenuation function w^G to the covariance function (cf. Eq. 4.14 and Figs. 4.4f, 4.8). Blue: original covariance function. Orange: attenuated covariance function. Green: attenuation function. Dashed black: spherical distance psi_{max} as parameter for the attenuation function. Red: optimal buffer distance r chosen as $r = 5/3\,\psi_{max}$ (cf. Sec. 4.4). (b) The estimator visualized for one sample point (marked as green dot) located in an area covered by observations (point A). The distances psi_{max} and r are again visualized as dashed black resp. red circles (as in Fig. 4.9a). In this case, the original estimator and the attenuated one cannot be distinguished since the kernel disparity in this point is $\Delta_A^{ker} = 0.0004\%$. (c) The original estimator for a sample point located in area without observations (point B). (d) The attenuated estimator for point B ($\Delta_B^{ker} = 40\%$). (e) The kernel disparity measure as defined by Eq. 4.16 visualized for every point (logarithmic scale). (f) The difference in the solutions (attenuated-original) caused by the kernel localization.

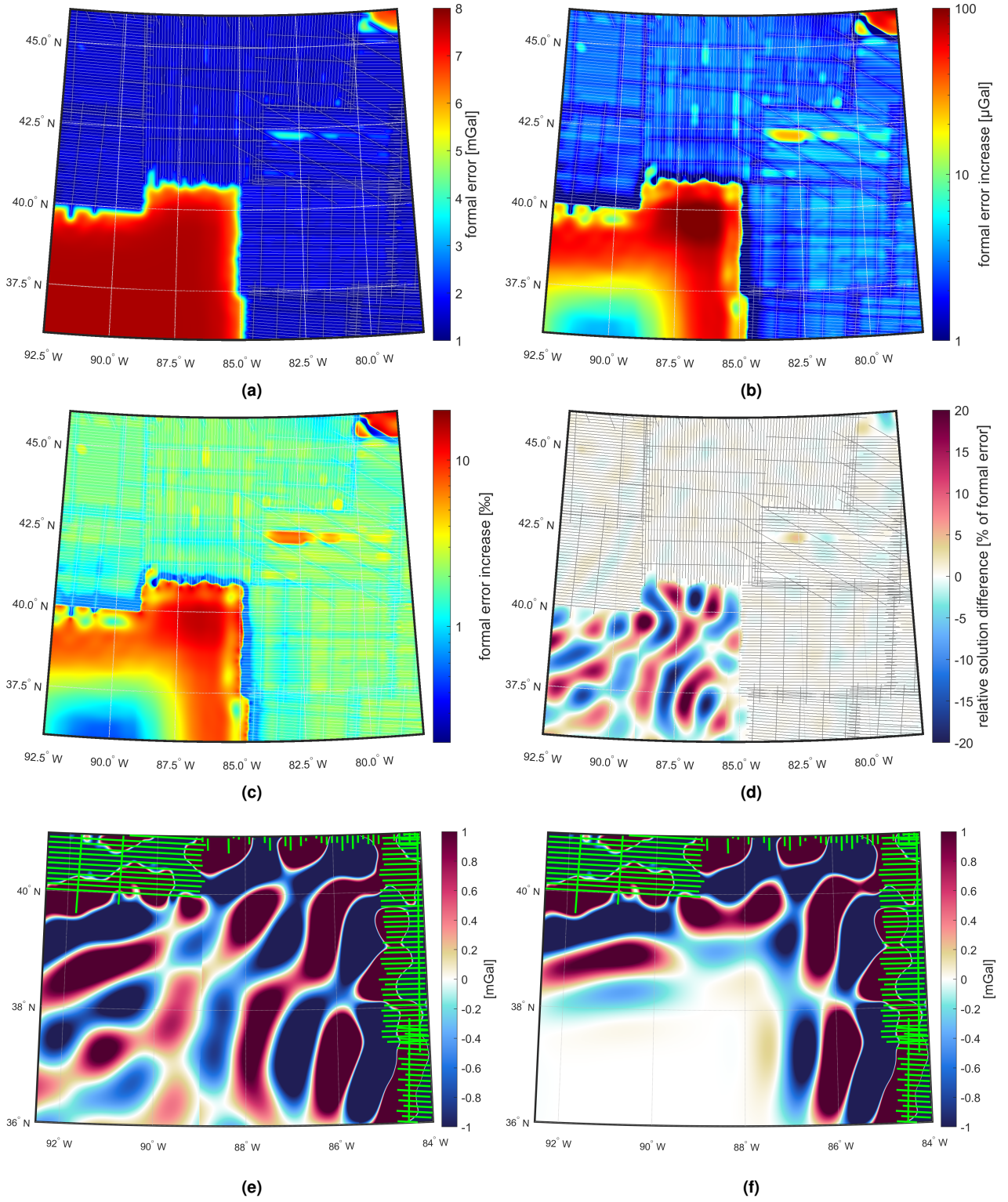


Fig. 4.10 – Influence of the kernel localization on the formal error of the LSC solution. (a) The formal error (cf. Eq. 2.35, i.e., square root of diagonal) obtained by the original kernel. (b) Increase of the formal error due to the modification of the estimator. (c) Relative increase of the formal error regarding the original formal error (Fig. 4.10a). (d) Relative solution difference (attenuated-original, cf. Fig. 4.10d) regarding the original formal error (Fig. 4.10a). (e) LSC solution using the original kernel. (f) LSC solution using the localized kernel.

4.6 PE-LSC results and validation

Outline. With the modifications presented in the previous sections, it is now possible to apply the PE-LSC method to the available (reduced) observations $\Delta\underline{l}$ to collocate them onto the specified target grid. However, through the collocation it is not only possible to estimate the target grid, but also to calculate estimates on the locations of the observations (using the observation's functionals). In this way it is possible to use the collocation as a filter to derive filtered observations $\Delta\underline{\hat{l}}$. Assuming the validity of the covariance function, the filtered observations $\Delta\underline{\hat{l}}$ represent an optimal estimate for the observation's signal $\Delta\underline{l}^{\mathcal{X}}$ (without the noise component $\underline{l}^{\mathcal{M}}$). Having the original observations $\Delta\underline{l}$ and the estimates for the signal $\Delta\underline{\hat{l}}^{\mathcal{X}}$ also the estimated observation's noise resp. error $\Delta\underline{\hat{l}}^{\mathcal{M}}$ can be deduced by using the relation given by Eq. 2.27:

$$\Delta\underline{\hat{l}}^{\mathcal{M}} = \Delta\underline{l} - \Delta\underline{\hat{l}}^{\mathcal{X}} = \Delta\underline{l} - \mathrm{C}_{LL}^{\mathcal{X}} \left(\mathrm{C}_{LL}^{\mathcal{X}} + \mathrm{C}_{LL}^{\mathcal{M}}\right)^{-1} \Delta\underline{l} = \left(\mathrm{I} - \mathrm{C}_{LL}^{\mathcal{X}} \left(\mathrm{C}_{LL}^{\mathcal{X}} + \mathrm{C}_{LL}^{\mathcal{M}}\right)^{-1}\right) \Delta\underline{l}. \tag{4.17}$$

By inspecting the estimated observations' error $\Delta\underline{\hat{l}}^{\mathcal{M}}$ and by comparing it to the assumed error specified by $\mathrm{C}_{LL}^{\mathcal{M}}$ one can deduce whether the estimated noise is realistic or not. This can be achieved, e.g., by comparing the estimated error $\Delta\underline{\hat{l}}^{\mathcal{M}}$ to the formal error $\Delta\underline{\sigma}_L^{\mathcal{M}} = \sqrt{\mathrm{diag}\left(\mathrm{C}_{LL}^{\mathcal{M}}\right)}$; if an estimated error $\Delta\hat{l}_i^{\mathcal{M}}$ is k times larger than its expected (formal) error $\sigma_i^{\mathcal{M}}$, the respective observation can be classified as suspicious (i.e., as outlier) since the probability that this occurs according to the assumed (centered) normal distribution (characterized by $\sigma_i^{\mathcal{M}}$) grows smaller and smaller with increasing k. As an example, when choosing $3\,\sigma_i^{\mathcal{M}}$ as threshold for the outlier detection there is only a probability of $\sim 0.27\%$ that this deviation is regular (in the sense of the normal distribution). When further increasing the threshold to $4\,\sigma_i^{\mathcal{M}}$ this probability drops further to 0.0063%. To directly compare observation errors to their expected error the normalized error $\Delta\grave{l}_i^{\mathcal{M}}$ is introduced as

$$\Delta\grave{l}_i^{\mathcal{M}} = \frac{\Delta\hat{l}_i^{\mathcal{M}}}{\sigma_i^{\mathcal{M}}}. \tag{4.18}$$

When inspecting $\Delta\grave{l}_i^{\mathcal{M}}$, outliers are easily identifiable, since $\Delta\grave{l}_i^{\mathcal{M}}$ can be compared directly to the σ-factor k. To identify outliers safely, choosing $k \approx 4$ may be reasonable according to the probabilities given above. An observation l_i is then defined as outlier if $\Delta\grave{l}_i^{\mathcal{M}} > k$.

Usually, an initial (a priori) observation noise $\mathrm{C}_{LL}^{\mathcal{M}}$ is provided when performing the collocation. Mostly, this observation noise is assumed to be uncorrelated and the variances are chosen according to the specified instrument accuracies σ_i^0 (that is $\mathrm{C}_{LL}^{\mathcal{M}}$ is diagonal with $(\sigma_i^0)^2$ as diagonal entries). While this noise assessment might be sufficient for a first iteration, having now $\Delta\underline{\hat{l}}^{\mathcal{M}}$ it is possible to calculate a better suited data-driven estimate for the standard deviation σ_i^0, denoted now with σ_i^e, which can be acquired through local averaging (e.g., by using a Gaussian kernel):

$$\sigma_i^e = \sqrt{\frac{\sum_k w_{ik}^G \left(\Delta\hat{l}_k^{\mathcal{M}}\right)^2}{\sum_k w_{ik}^G}}, \quad w_{ik}^G \coloneqq \exp\left(-\frac{1-\cos\psi_{ik}}{2(1-\cos\psi_0)}\right). \tag{4.19}$$

With ψ_0 one can adjust the (spherical) width (i.e, sigma) of the Gaussian averaging kernel. The derived empirical standard deviation σ^e describes a local estimate for the misfit between the actual observations and their assumed behavior given by C_{LL}. Assuming C_{LL} is realistic, σ^e actually characterizes the spatial error pattern of the observations. By inspecting this spatial error pattern, further assessments on the observations can be made (beyond the outlier detection as explained above). For instance, if a faulty instrument or pro-

cessing strategy was applied in a certain region, this may be quite easy to identify through an increased value for σ_i^e in this area. Vice versa, when expecting the a-priori accuracies σ_i^0 to be correct, an increased resp. lowered value for σ_i^e may indicate a misfit due to C_{LL} through an under- resp. overfit in the respective area (e.g. because of an unmodelled but varying behavior of the actual covariance function in plains or mountain ranges).

Application. For the succeeding calculation of a global gravity field model, the GRAV-D observations are collocated onto a regular 1' grid of (spherical) gravity disturbances. To minimize the influence of up- resp. downward continuation, a dedicated spheroid E_{GD} whose surface is located on mean flight altitude of the observations is chosen as reference surface. To achieve this, E_{GD} is created by taking the semi-minor axis b_{G80} and semi-major axis a_{G80} of the GRS80 spheroid (see *Moritz*, 2000) and increasing it by 6 km (mean flight altitude), so that $b_{GD} = b_{G80} + 6$ km and $a_{GD} = a_{G80} + 6$ km. The a-priori observation noise $\mathrm{C}_{LL}^{\mathcal{M}}$ is assumed to be diagonal with an expected constant instrument accuracy of $\sigma^0 = 4$ mGal (according to crossover error analyses, see, e.g., *GRAV-D Team*, 2017). As signal covariance $\mathrm{C}_{VV}^{\mathcal{X}}$ the covariance function derived from the estimated degree variances (see Sec. 4.3) is used. Also the modification of the geometry, the partitioning and the kernel localization are applied as explained in Secs. 4.2, 4.4 and 4.2. As pointed out above, for validation purpose, also gravity disturbances at the observation locations are estimated.

The result of the collocation (using $\sigma^0 = 4mGal$) in terms of collocated residuals $\Delta\underline{\hat{s}}$ (cf. Eq. 2.33) is shown in Fig. 4.11a and the respective formal (a-posteriori) error in Fig. 4.11b (according to Eq. 2.35). By inspecting the collocated residuals (Fig. 4.11a) themselves it is not easy to make any assessments of the quality of the result. However, with naked eye, it can at least be recognized that the residuals look as expected (no severe outliers) and that through the kernel localization the result is efficiently tied to zero in absence of observations. When looking at the formal errors (Fig. 4.11b), mainly the different flight altitudes in different patches are apparent which manifest themselve through increasing error with an increased altitude. Also, if the distance between adjacent flight tracks is increased for some reason, the formal error increases accordingly. Despite this minor variations, according to the applied model (σ^0 and $\mathrm{C}_{LL}^{\mathcal{X}}$) the estimated formal error is relatively homogeneous with values mostly between 1-2 mGal in areas with observation coverage. Since a more detailed evaluation is not possible by inspecting the collocated residuals $\Delta\underline{\hat{s}}$ resp. formal errors, the previously defined estimated observation errors $\Delta\underline{\hat{l}}^{\mathcal{M}}$ are additionally examined (Figs. 4.12a and 4.12b). When looking at the observation errors in terms of normalized errors $\Delta\grave{l}^{\mathcal{M}}$ (Fig. 4.12a, Eq. 4.18) and the thereby derived empirical standard deviations σ^e (Fig. 4.12b, Eq. 4.19), several statements can be made:

1. generally, it can be seen that the increased observation errors are accumulated in areas with rougher topography (i.e., in mountainous areas in the western part of US, see also Figs. 4.14b,e,h,k). As already discussed partially in Sec. 3.3, the causes behind this may be manyfold, and, thus, it might not be easy to backtrack a single source. One reason might be the not fully adequate covariance function for these areas since the applied degree variances (see Sec. 4.3) represent an average over the entire GRAV-D region. However, also problems with the airborne observations (e.g., through filtering out short-wavelength signals in the preprocessing) or problems through the reduction model itself (e.g., higher uncertainties due to the density assumptions made within the topographic model) can cause the increased observation errors (see P-5). Eventually, it can be put on record that rougher terrain is much more challenging for the gravity processing than flat terrain, and that for a reliable backtracking of the error sources more sophisticated evaluations would be needed which cannot be performed in the course of this thesis. Two simple ideas for adjusting inadequate covariance functions are presented below (3) and in the outlook at the end of this section.

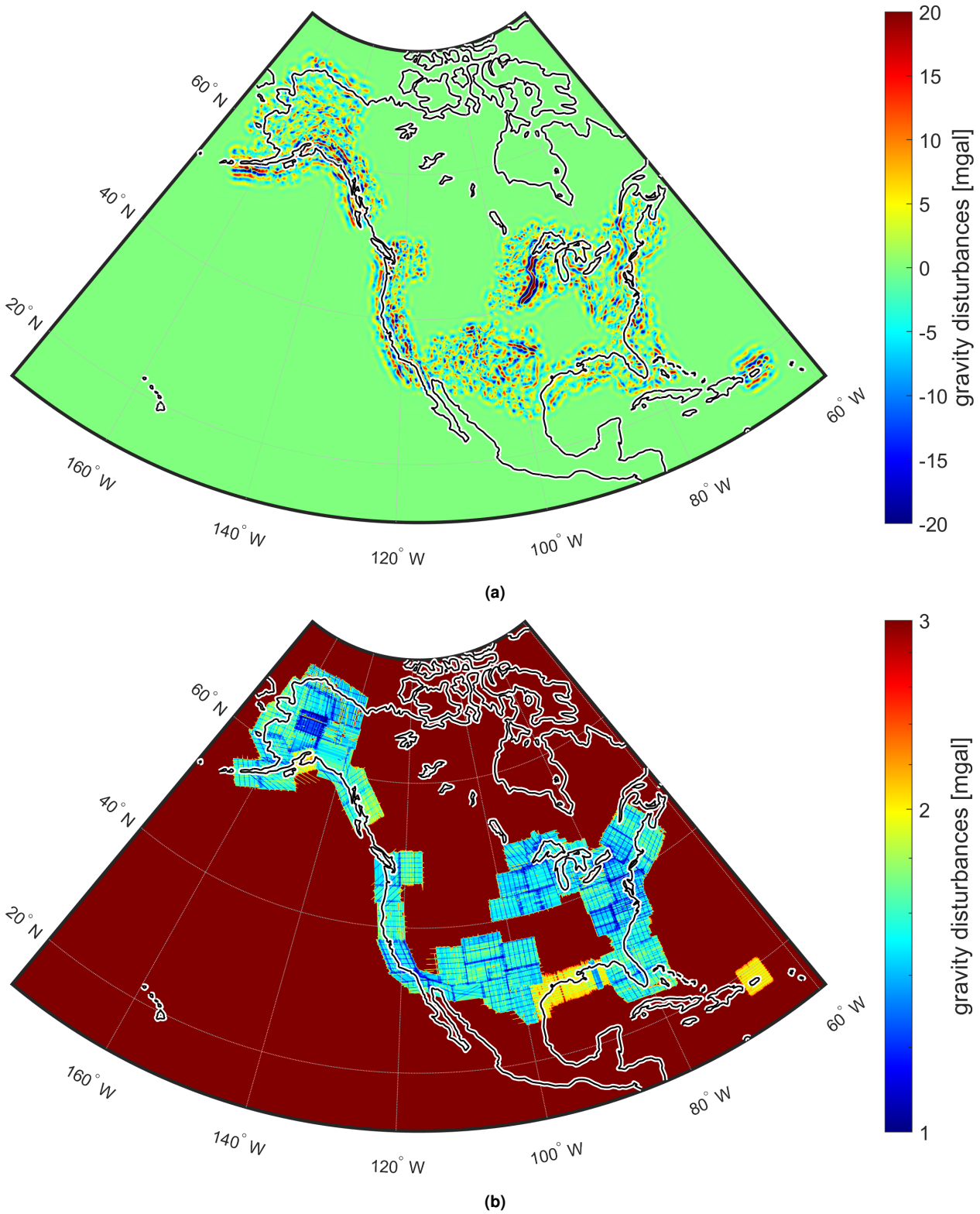


Fig. 4.11 – Result of collocating the GRAV-D dataset using PE-LSC and $\sigma^0 = 4$ mGal. (a) The estimated residual gravity disturbances on a 1' grid on the surface of the E_{GD} spheroid. (b) Estimated formal error to the collocated residuals (Fig. 4.11a, according to Eq. 2.35).

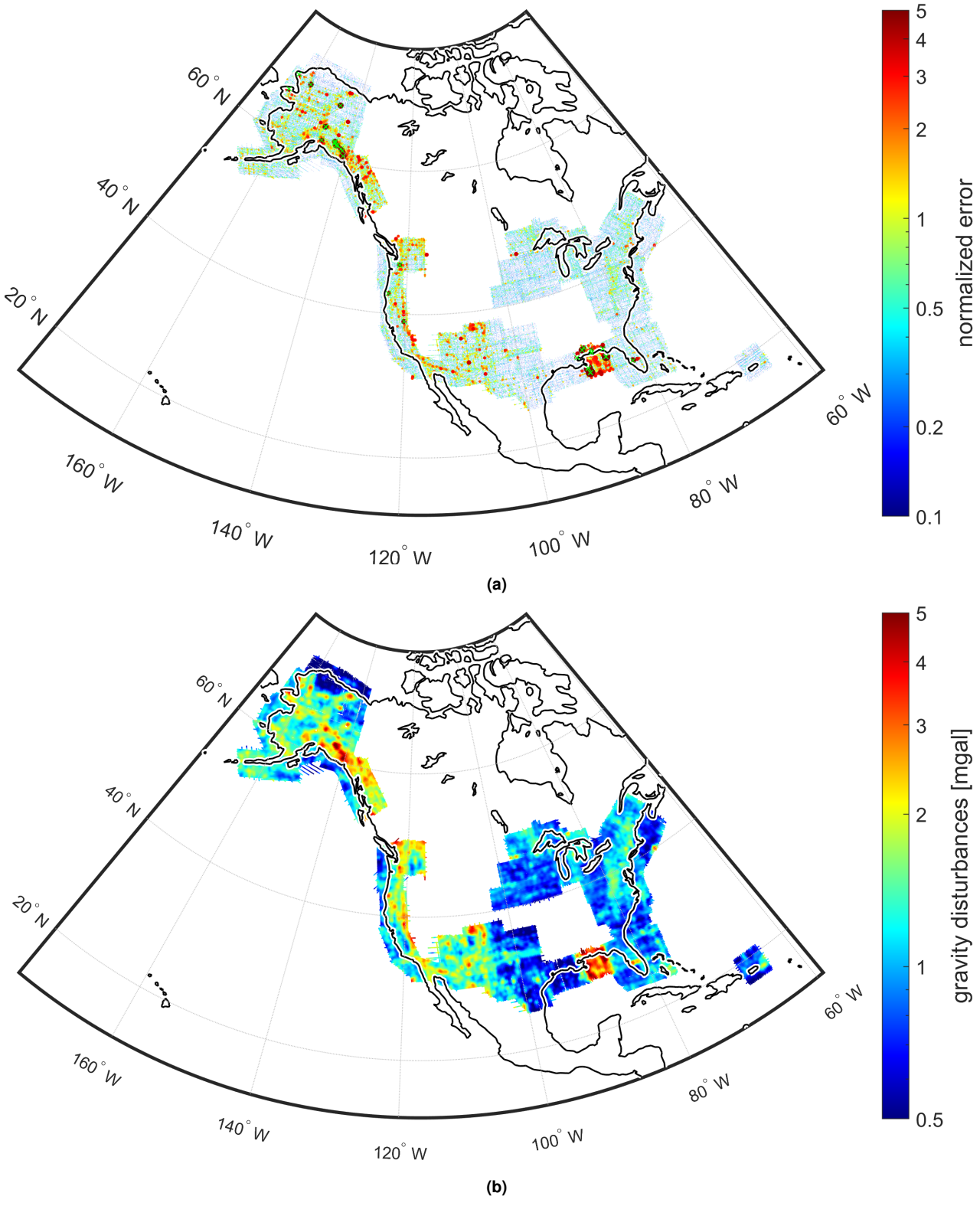


Fig. 4.12 – Post-collocation validation of the GRAV-D observations using PE-LSC and $\sigma^0 = 4$ mGal. (a) Normalized observation errors $\Delta\grave{l}^{\mathcal{M}}$ with $\sigma^0 = 4$ mGal. Assumed outliers with $k > 4$ are additionally marked with green crosses. (b) Emprical standard deviation σ^e (see Eq. 4.19) using a Gaussian averaging kernel with a width of 30 km (HWHM, that is $2\log 2\,\sigma$).

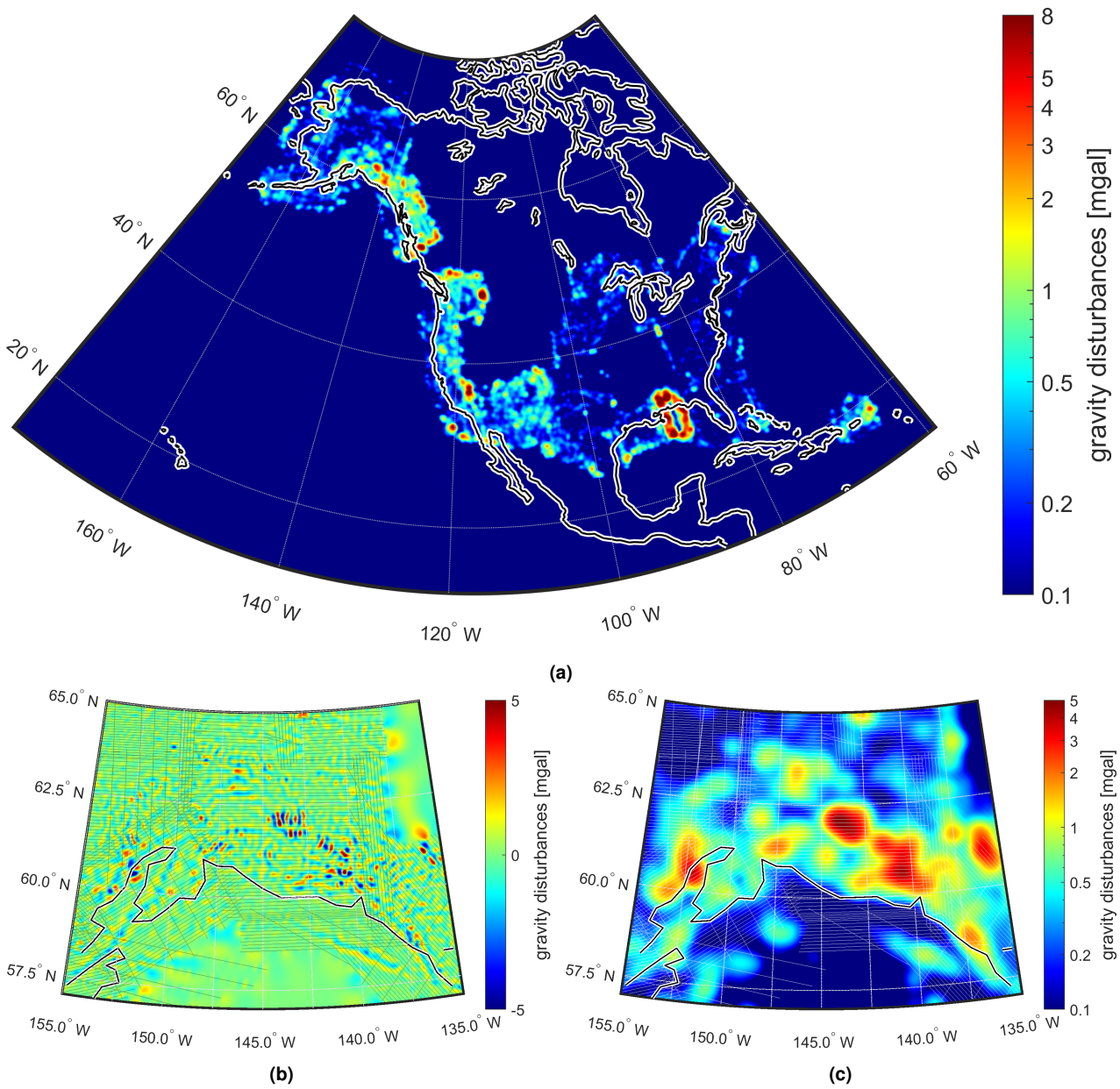


Fig. 4.13 – Difference between the PE-LSC collocation result (i.e., the gridded data, cf. Fig 4.11a) when decreasing the assumed a-priori accuracy σ^0 to 2 mGal (from 4 mGal). (a) Standard deviation of the solution difference (derived by using a Gaussian averaging kernel with a width of $\sim$30 km, cf. Fig. 4.12b and Eq. 4.19). (b) Solution difference in validation zone B (see Fig. 3.3a). (c) Standard deviation of solution difference in validation zone B (excerpt of Fig. 4.13a).

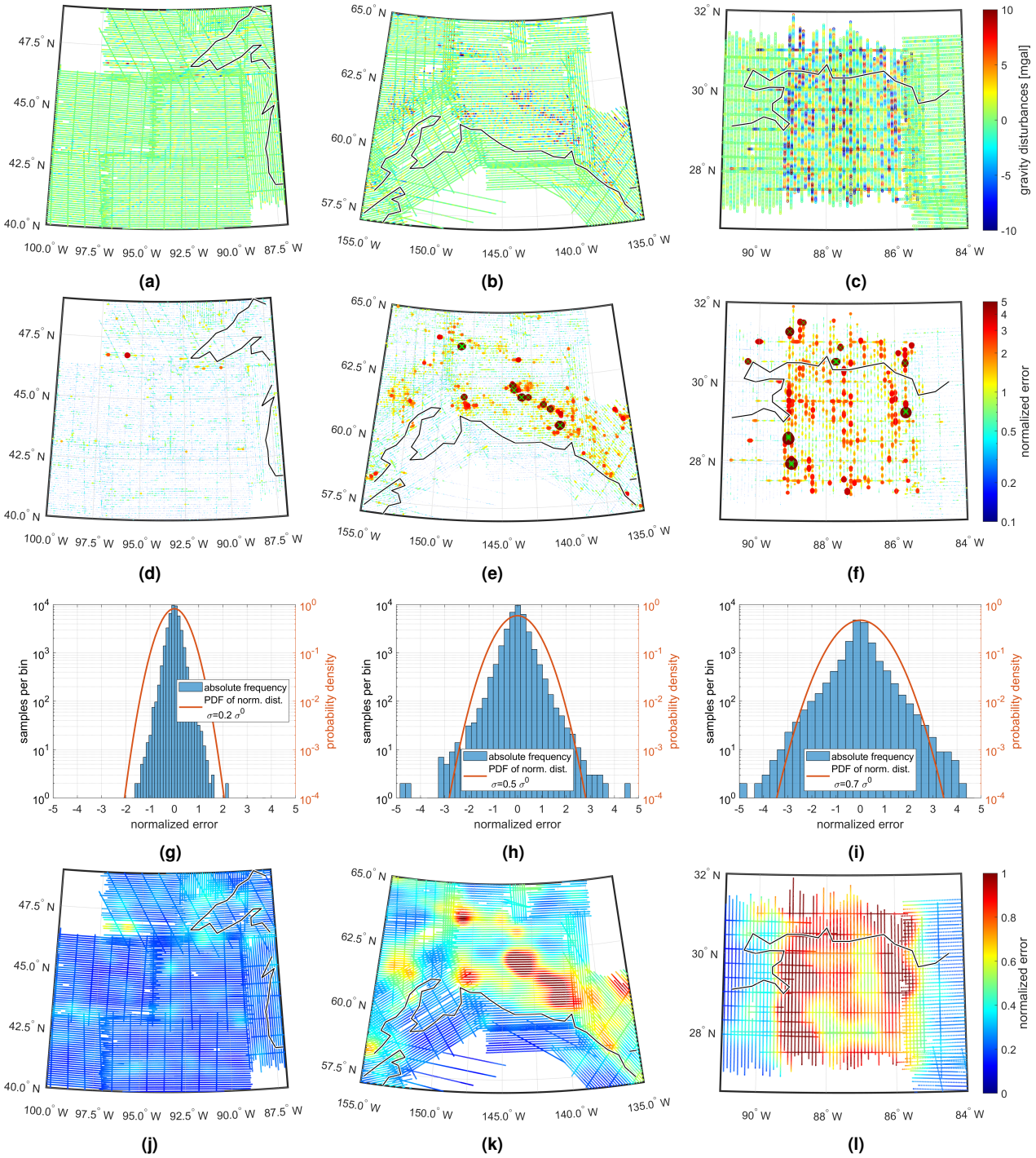


Fig. 4.14 – Post-collocation validation of the three zones defined in Sec. 3.3 (cf. Fig. 3.4). (a-c) Estimated observation errors $\Delta\hat{\underline{l}}^{\mathcal{M}}$. (d-f) Normalized observation errors $\Delta\hat{l}^{\mathcal{M}}$ with $\sigma^0 = 4$ mGal. Assumed outliers with $k > 4$ are additionally marked with green crosses. (g-i) Frequency distribution of the normalized errors in the validation zones, corresponding to Fig. d-f. (j-l) Empirical standard deviation σ^e (Eq. 4.19, same as in Fig. 4.12b, normalized according to $\sigma^0 = 4$ mGal).

2. in the south of US, in the coastal zone between Louisiana and Florida (without rough topography), an area with a highly increased observation error is apparent (cf. Figs. 4.14c,f,i,l). As already noted in the validation of the residuals in Sec. 3.3, it is suspected that the GRAV-D data block CS08 situated here causes some sort of problem (possibly due to the low-pass filter which was applied differently than on other blocks, see *GRAV-D Team*, 2015). The estimated observation error and the empirical standard deviation clearly supports this assumption. Having identified data block CS08 the as source of the problem one could either try to (1) fix the problem in the dataset, (2) decrease the assumed a-priori accuracies for the included observations or, (3) simply omit the data. In general, option (1) is preferable. If this is not feasible though, one should weigh option (2) against option (3) according to the assumed new a-priori accuracy; if the new a-priori accuracy is higher than the model's input accuracy (from the covariance function $\mathrm{C}_{VV}^{\mathcal{X}}$), only a small gain is expected through the inclusion of the data, and, hence, omitting the data in the first place might be reasonable. Since the processing shown within this thesis only serves an explanatory purpose, a backtracking of the error source and a reprocessing is not conducted. Next to data block CS08 also some other individual outliers can be detected throughout the whole GRAV-D dataset (mostly only single points in addition to a single flight track over Florida).

3. over wide areas, especially plains (see also Figs. 4.14a,d,g,j), the observation errors are below the expected accuracy of 4 mGal. Calculating the standard deviation of all observation errors yields a value of about 1.4 mGal. Thus, it can be assumed that the a-priori accuracy (4 mGal) was chosen somewhat too pessimistic, and that lowering σ^0 in a second iteration may produce a better result. With the choice of a decreased a-priori error, the tie towards the actual observations is strengthened, which ultimately allows for stronger regional deviations from the assumed covariance function ($\mathrm{C}_{VV}^{\mathcal{X}}$). This behavior may be helpful in areas where the covariance function does not fit optimally (e.g., over rough terrain, see above), but may also be dangerous because actual observation noise is also more likely to be recognized as signal by the collocation. For testing purposes, a dedicated solution with an adjusted a-priori accuracy of 2 mGal is calculated. The comparison to the 4 mGal solution is depicted in Fig. 4.13. Generally, the differences are comparable with the observation errors, which means that wherever the observation errors are increased also larger differences between both solutions are to be expected. Additionally, changing the a-priori accuracy seems to also introduce differences in the swing-off zones, i.e., observation-free regions at the boundaries to areas with observation coverage, where the signal gets tied toward zeros due to the kernel localization (see Fig. 4.13b).

For a more detailed view on the result, the three zones of investigation with suspicious residuals defined in Sec. 3.3 (see Figs. 3.3a and 3.4) are reevaluated in terms of observation errors in Fig. 4.14. What can be seen is that:

A the underground anomalies prominent in the residuals in zone A vanish completely after collocation (cf. Figs. 4.14a,d,g,j). This can probably be explained by the good agreement of the signal with the expected structure defined by the covariance function.

B the stronger residuals in the mountainous area of zone B are also widely reduced after the collocation (cf. Figs. 4.14b,e,h,k). However, as already mentioned above, the observation errors are considerably larger than in zone A due to the given reasons with several $4\sigma^0$-outliers. While smaller observation errors can be explained through a misfit of the covariance function, $4\sigma^0$-outliers are probably more difficult to explain since they correspond to punctual deviations of more than 16 mGal at an altitude of 6 km. However, since the flight altitude is also around 6 km and the mountain ranges reach up to 5 km there is a much weaker damping due to upward continuation in these zones. Theoretically, this could be a reasonable

explanation for the larger deviations. It is also possible that there is a certain spectral misfit between the observations and the reduction model since the airborne observations are filtered (in terms of a time series) and the reduction model (SATOP) is not. Applying SATOP in a reduced resolution might help to mitigate some of the observed errors in this zone.

C zone C, located over the erroneous data block CS08 (cf. Figs. 4.14c,f,i,l) shows the largest observation errors. The possible reasons have already been explained above. In comparison to the residuals (Fig. 3.4c), the problem is even more prominent in terms of observation errors (Fig. 4.14c).

Iterative refinement strategy and conclusions. It has been shown that PE-LSC works as expected and that with a post-collocation validation at least certain data problems can be identified. However, it has also been shown that the increased observation errors over mountainous areas may indicate a certain deficiency in the modelling of the covariances. This is already discussed partially in Sec. 4.3 where it has been pointed out that the implicit assumption (when using degree variances) of homogeneity and isotropy is only a rough approximation of the actual behavior of the gravity field. Especially when assuming homogeneity (i.e., location independence), a strong simplification is introduced which, e.g., leads to the assumption of identical gravity field behavior over both rough and flat terrain, although this is evidently not the case. As a consequence, since the estimated degree variances (see Sec. 4.3) represent the averaged behavior of the gravity field over the whole collocation area, the derived empirical covariance function is probably too rough for flat country and too smooth for mountainous regions. While the former (flat country) seems to be more or less unproblematic (although it can be assumed that more noise is included in the solution than in the optimal case), the latter (rough terrain) seems to suffer from the too smooth covariance function in this area (since too much signal is then interpreted as noise). This problem cannot be solved by using covariances derived from degree variances. However, the effect of non-homogeneity can at least be partially simulated when applying a workaround to reduce the impact of the. The idea of this workaround is to restart the collocation with the narrowed down residuals given through the observation errors $\Delta\hat{\underline{l}}^{\mathcal{M}}$. The procedure can be summarized as follows (with the index i denoting the i-th iteration):

1. Set the observation errors $\Delta\hat{\underline{l}}^{\mathcal{M}^{(i)}}$ (Eq. 4.17) of the previous iteration to be the new input residuals $\Delta\underline{l}^{(i+1)}$ within the collocation of the next iteration, i.e., $\Delta\hat{\underline{l}}^{\mathcal{M}^{(i)}} \rightarrow \Delta\underline{l}^{(i+1)}$. With this assignment one basically implies that the estimated observation noise still contains residual signal.

2. Disregard all new residuals $\Delta\underline{l}^{(i+1)}$ which show an estimated normalized standard deviation $\sigma^{e(i)}/\sigma^{0(i)}$ (Eq. 4.19) below a certain threshold. This is best done by effectively setting the residuals with a small standard deviation (below a certain threshold) to zero. With that, the new collocation just influences areas with significant misfits. For the transition from selected to omitted residuals, it is proposed to apply some sort of distance-dependent damping function $\underline{w}_d^{(i)}$ to the new residuals in order to avoid sharp (high-frequency) edges. With that, far-off residuals can safely be deleted from the new residuals as they would not influence the result (since zero). This should help to speed up subsequent collocations (since the number of observations is reduced from iteration to iteration). An illustration for the transition function and the remaining input residuals for the collocation is given in Fig. 4.15.

3. Based on the remaining new residuals, re-estimate the empirical covariance function according to Sec. 4.3. By readjusting the covariance function $\mathrm{C}_{VV}^{\mathcal{X}\ (i+1)}$ one essentially simulates the aforementioned non-homogeneity since one now basically uses different covariance functions in different areas (namely in those areas containing the selected residuals). Note that only the selected residuals in the core area (outside the transition zone) are taken for the calculation of the empirical covariance function. A possible

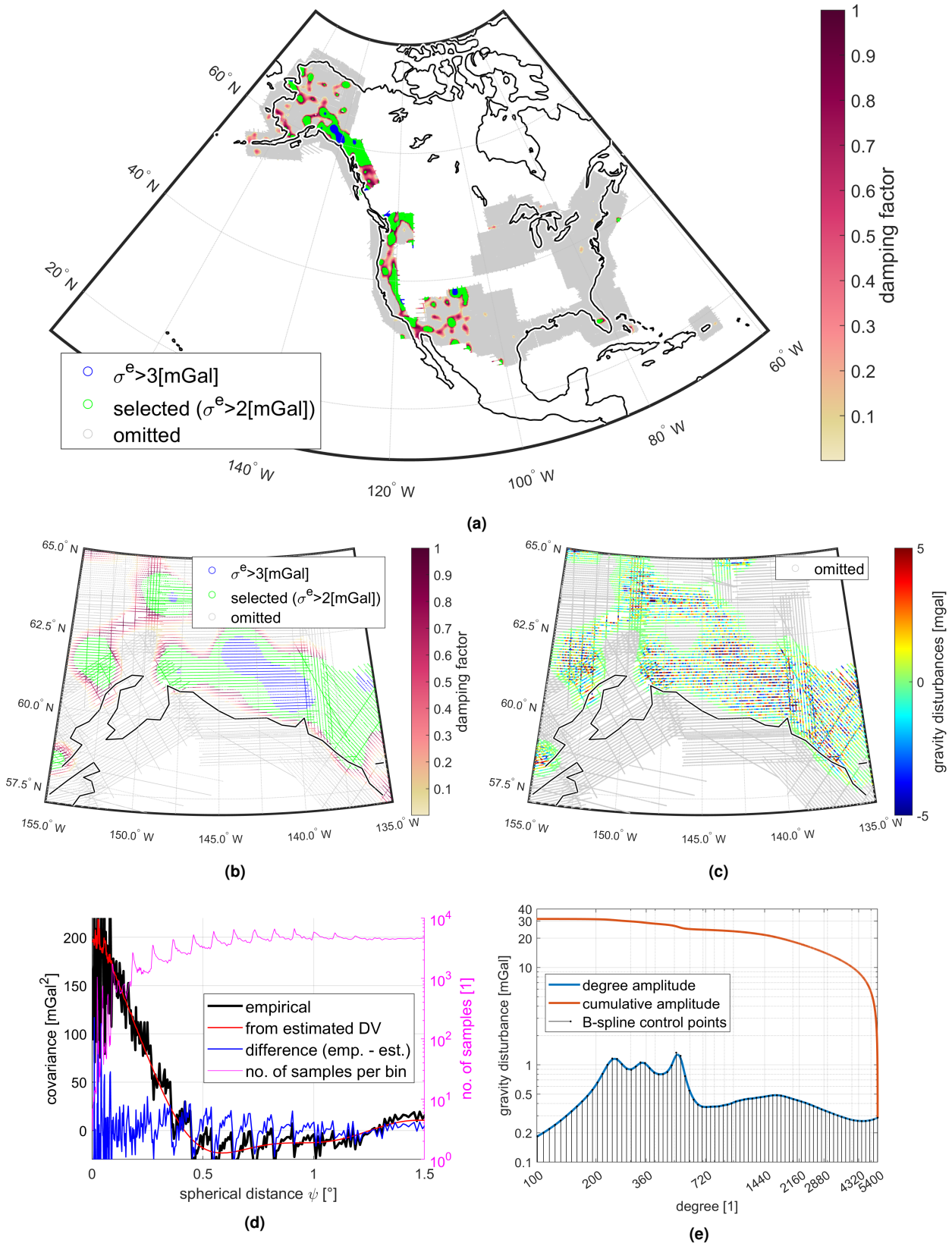


Fig. 4.15 – Preparation of the collocation for the second iteration. (a) Selection of the new residuals (green and blue points) with the used damping function highlighted (red-gray colormap) to model the transition towards the omitted residuals (gray points). Selected are all residuals with an empirical observation error $\sigma^e > 2$ mGal (see Eq. 4.19 using 50 km HWHM). The damping function $\underline{w}_d$ consists of a sine function parameterized through the closest distance to a selected residual. (b) Same as Fig. 4.15a for validation zone B. (c) Resulting new residuals for the next iteration of the collocation with the damping function (see Fig. 4.15b) already applied. Omitted residuals are marked with gray color. (d) Estimated new empricial covariance function (1D) including all residuals where $\sigma_i^e > 3$ mGal (blue points in Fig. 4.15b, cf. Sec. 4.3, Fig. 4.4f). (e) Degree variances estimated from the empirical covariance function (Fig. 4.15d, cf. Sec. 4.3, Fig. 4.4b).

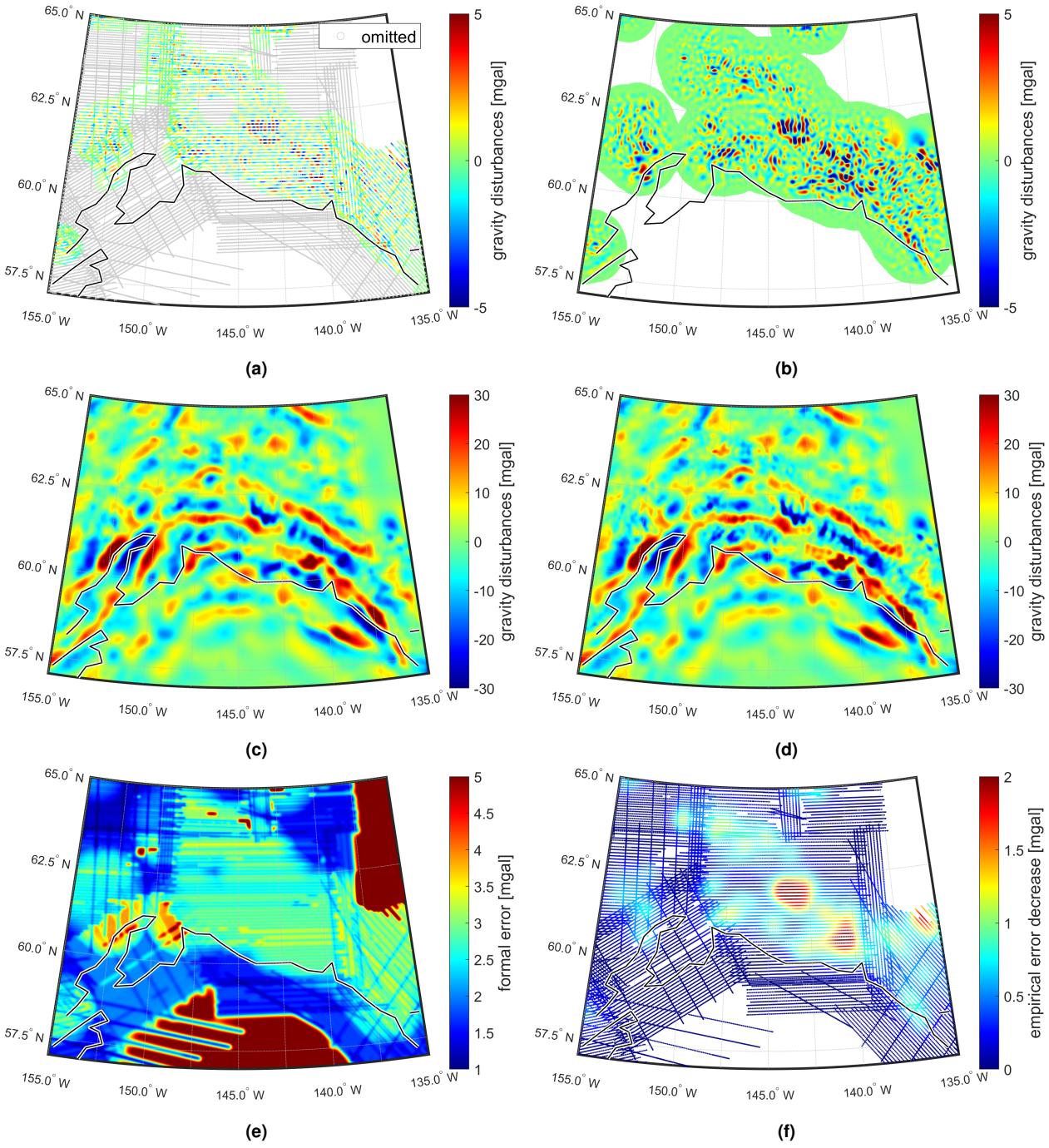


Fig. 4.16 – Results of the second iteration of the collocation shown exemplary for validation zone B. (a) Estimated observation error differences $\Delta\hat{\underline{l}}^{\mathcal{M}(2)}$ to the original observation errors of the first iteration (Fig. 4.14b). This is identical to the difference in the collocation result $\Delta\underline{l}^{(2)}$ on the observation sites. (b) Difference of the collocation result $\Delta\underline{\hat{s}}^{(2)}$ on the target grid. (c) Original result of the collocation $\Delta\underline{\hat{s}}^{(1)}$ on the target grid after the first iteration (same as Fig. 4.11a). (d) Refined result of the collocation $\Delta\underline{\hat{s}}$ on the target grid after the second iteration. (e) Formal error $\Delta\underline{\hat{\sigma}}$ of the refined result. (f) Reduction of the empirical standard deviation through the refined solution $\Delta\hat{\underline{l}}$ in comparison to the original solution $\Delta\hat{\underline{l}}^{(1)}$ $(\sigma^{e(1)} - \sigma^{e})$.

result for a second iteration of the GRAV-D residuals is visualized in Fig. 4.15d and 4.4b. Since the number of residuals is reduced for the estimate of the new degree variances (in comparison to the original estimate for the initial collocation, compare Fig. 4.15d with Fig. 4.4f), the result can be considered less stable resp. overdetermined. However, in the actual example of GRAV-D it seems that the remaining residuals (for the second iteration) still allow the estimation of degree variances sufficiently well.

4. Assess a realistic value for $\sigma^{0^{(i+1)}}$. For that, one may choose the initial value of the first iteration (if considered realistic) or an estimated one from $\Delta\underline{l}^{(i+1)}$.

5. Calculate the new residual estimates $\Delta\underline{\hat{s}}^{(i+1)}$ (cf. Fig. 4.16b) and $\Delta\underline{\hat{l}}^{(i+1)}$ (cf. Fig. 4.16a) from $\Delta\underline{l}^{(i+1)}$, $\mathrm{C}_{VV}^{\mathcal{X}}{}^{(i+1)}$ and $\Delta\underline{l}^{(i+1)}$. The final estimates $\Delta\underline{\hat{s}}$ (cf. Fig. 4.16d) can then be obtained by $\Delta\underline{\hat{s}} = \sum_i \underline{w}_d^{(i)} \odot \Delta\underline{\hat{s}}^{(i)}$ considering the same damping function $\underline{w}_d^{(i)}$ (now applied on the output grid) as used for the input residuals (with $\odot$ denoting element-wise multiplication). Omitting correlations, the final formal error $\Delta\underline{\hat{\sigma}}$ (cf. Fig. 4.16e) can be approximated identically through $\Delta\underline{\hat{\sigma}}^2 = \sum_i \underline{w}_d^{(i)} \odot \Delta\underline{\hat{\sigma}}^{(i)^2}$

6. Inspect the newly obtained observation errors $\Delta\underline{\hat{l}}^{\mathcal{M}^{(i+1)}}$ (e.g., in terms of the empirical standard deviation σ^e, cf. Fig. 4.16f) and assess whether another iteration is necessary. If so, restart the entire procedure starting from step 1.

With this strategy one should be able to at least refine the initial result by a certain degree. It should be noted that some of the proposed methods (in this validation section) are just feasible because in the GRAV-D collocation only gravity disturbances occur. In the general case where a mixture of several gravity field functionals may be present, adjustments on several points are needed. For example, calculating the empirical standard deviation according to Eq. 4.19 would not be possible anymore, since different functionals would obviously yield different values. So, Eq. 4.19 would have to be modified by, e.g., replacing the estimated error $\Delta\hat{l}^{\mathcal{M}}$ by the normalized error $\Delta\grave{l}^{\mathcal{M}}$. In this way, one would get a normalized empirical standard deviation which can then be used for further validations resp. calculations instead (which then would also have to be adjusted). Although a second iteration is calculated for the GRAV-D example, the obtained solution (see Fig. 4.16) is not used in the following since the whole iterative procedure is still considered experimental at the time of writing.

Chapter 5

Global gravity field processing

After having the reduced and validated gravity field observations collocated onto a global regular geographic grid (see Fig. 4.11a), the data is ready for integration into a global model given in the harmonic domain. Publication P-3 illustrates the basic procedure of how to do so on the example of the XGM2019 model. In this chapter, the content of publication P-3 will be summarized and extended by topics which are relevant for the holistic processing approach using again the GRAV-D mission for demonstration. While the presented methodology is not too complex from a mathematical perspective, it is even more complicated in terms of computational complexity and from a numerical point of view. Thus, the major effort in the field of global gravity field processing has been put into an efficient and numerically accurate implementation of the procedures discussed below. Due to the enormous computational requirements of certain parts of the processing, the code had to be optimized for execution on distributed high-performance computing systems, which requires a lot of rethinking of the serial algorithms. Therefore, much could be said about workload balancing, numerical stability, vectorization, efficient I/O concepts, etc. However, such details would probably require a deeper informatic background knowledge from the reader which would go beyond the scope of this thesis that should retain its geodetic character. Therefore, in this chapter, the focus will be strictly kept on the methodological part, mentioning computational details only when necessary. The readers interested in details about the actual implementation are referred to earlier works which cover these topics sufficiently well (e.g., *Zingerle*, 2015, *Fecher*, 2015, *Brockmann*, 2014, *Gruber*, 2001 or *Schuh*, 1996).

This chapter is organized as following: first, the properties of the SH transform on geographic grids are investigated (Sec. 5.1). They are of relevance to understand the interaction between the collocation and the following SH analysis. Based on the gained knowledge, the gridded dataset is analyzed using a block-diagonal spheroidal harmonic approach and evaluated regarding possible analysis errors (Sec. 5.2). This block-diagonal solution can then be used to spectrally limit (and/or downsample/convert) the initial gridded data to a desired format, which can then be used as ground observation basis for a combination with other (satellite) models. In Sec. 5.3, several combination methods are briefly described and compared against each other and evaluated in terms of applicability within the integrated framework to derive the final EGGM. Finally, some strategies are presented for an external validation of the resulting EGGM against 3rd party data and models (Sec. 5.4).

5.1 The SH transform on geographic grids

Outline. For an arbitrary signal (or function) which is sampled on a geographic grid, two main statements regarding its spectral behavior resp. recoverability (in the spherical harmonic domain) can be made: first of

all, the signal (or function) does not necessarily have to be spectrally limited. This implies that the original spectrum (or function) cannot be recovered when having a finite spacing within the regular grid. Moreover, it is known that non-recoverable frequency components can cause aliasing, meaning that these components can bias the estimate of the recoverable components (see example Fig. 5.1, 5.2). In general, there is nothing one can do to avoid this except for using a finer spacing in the first place to make as much of the signal recoverable as possible. Thinking of a classical (flat) discrete Fourier transform (see, e.g., *Bronstein et al.*, 2015) there exists a well-defined threshold frequency (denoted as Nyquist frequency, see *Lüke*, 1999) for the recoverability of the signal solely defined by the spacing (or sampling):

$$m_{max} = \frac{n_s}{2} - 1 = \frac{\pi}{\Delta\lambda} - 1, \tag{5.1}$$

where m_{max} is the Nyquist frequency (i.e., wave number), n_s the number of equidistant samples and $\Delta\lambda$ the spacing between the samples as angular fraction of the unit circle. For instance, when interpreting a geographic grid with a spacing of $\Delta\alpha = 1^\circ$ as flat and performing a discrete 2D Fourier analysis on it, the respective Nyquist wave numbers would be given by $(1/2)\,360^\circ/1^\circ - 1 = 179$ (longitudinal) and $(1/2)\,180^\circ/1^\circ - 1 = 89$ (latitudinal). Performing the discrete 2D Fourier analysis up to the determined Nyquist frequencies would yield a 1:1 equivalent of the signal in the spectral domain, meaning that a subsequent synthesis would fully recover the original grid (since 360*180 sample points are orthogonally projected on 360*180 coefficients). However, and as second statement, the same is not true in case of the SH transformation where the maximum SH order m_{max} (in analogy to the Nyquist wave number) and degree l_{max} is induced by the Fourier transform in longitudinal direction (cf. Eq. 2.22 which would correspond in the stated example to $m_{max} = l_{max} = 179$, identical to the 2D Fourier case). The subsequent latitudinal convolution though differs from the Fourier case since after the transformation through the associated Legendre polynomials only $(l_{max} - m) + 1$ coefficients per order m remain (since $m \leq l$), which sums up to 180*180 coefficients in total. Thus, it is evindent that a SH transform does not retain the 1:1 relationship since a subsequent SH synthesis cannot reproduce the original grid from a information reproducing perspective (actually, it's a 2:1 relationship since the information is halved in the analysis step). This loss of information resp. coefficients can geometrically be explained best through the meridional convergence on the sphere, reflecting the fact that the equiangular sampling in the geographic grid is not equidistant (even if not fully true). Ultimately, this means that the geographic grid is able to hold much more high-frequency signal, especially, in the polar regions (in longitudinal direction) than is recoverable through spherical harmonics limited to a maximum degree l_{max}. Thus, even if the signal on the original grid is not affected by aliasing, the derived spherical harmonics are in general not able to reproduce this signal (e.g., see Fig. 5.2c). While this is generally no problem, the question may arise whether the possible additional signal content affects the spectral estimate up to degree l_{max} (according to Sec. 2.2, see below).

In publication P-3, it was assumed that there may be a negative impact of this extra signal to the estimate, and a special procedure was presented which is able to eliminate non-reproducible signal before the estimation takes places (see Fig. 5.2). This simple method consists of a low-pass filter that is applied in the spatial domain (cf. Fig. 5.2e) which is then reverted after the analysis step in the SH domain through multiplication with the inverse filter coefficients (cf. Fig. 5.2f and *Jekeli*, 1981). In this manner, only the low-pass filtered signal needs to be analyzed, and the effect of unresolvable higher frequency content is supposedly strongly reduced (see P-3). However, a closer evaluation of the impact of the additional signal has shown that retaining it does not affect the final estimate (see Fig. 5.2f, compare blue with dashed orange line). This circumstance can be explained best by the fact that the SH analysis (on a regular grid) consists in a first step of a Fourier transform (cf. *Sneeuw*, 1994 or *Colombo*, 1981) which is able to orthogonally decompose the longitudinal information

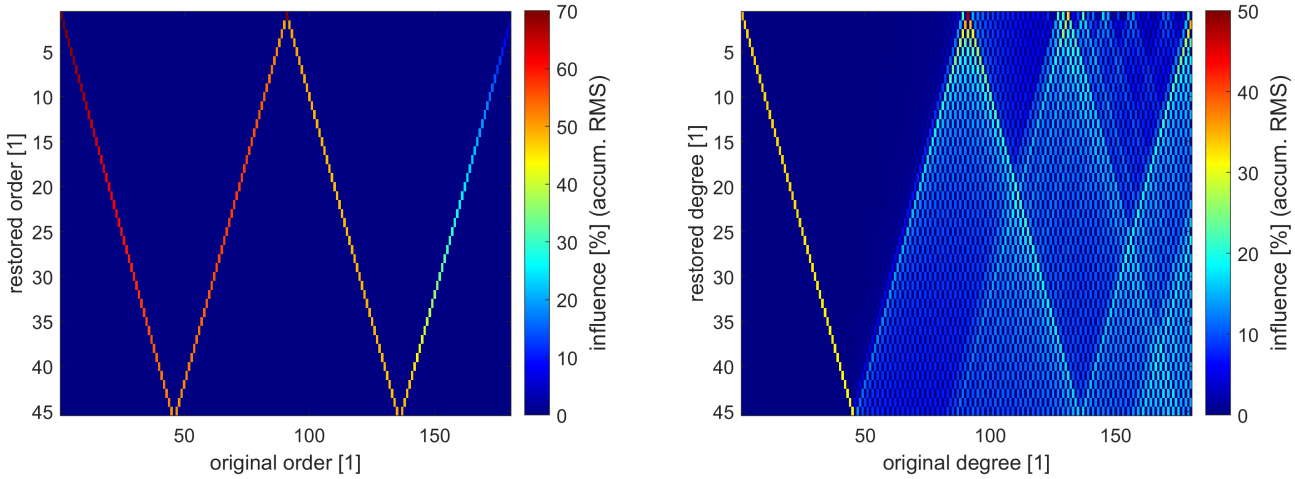


Fig. 5.1 – Aliasing behavior of the SH transform on geographic grids. In the example, a SH signal up to d/o 180 is assumed which is synthesized onto a 4°x4° geographic grid (90x45 sample points) estimated (restored) up to the grid's 'Nyquist' d/o of 44. (a) Aliasing (mapping) of the signal within the original orders to the signal within the restored orders. Shown are accumulated RMS values (over all degrees of a specific order) of the least squares estimator normalized regarding the restored orders (row-wise). The aliasing structure is in principle identical with the structure expected from the Fourier transform (showing the folding frequencies). Though, the aliasing differs from the Fourier case due to the attenuation of the effect towards higher orders. Hence, lower orders are less influenced by aliasing. (b) Aliasing of the signal within the original degrees to the signal within the restored degrees. The aliasing structure of the associated Legendre polynomials is much more complicated than the structure regarding the trigonometric functions (no unique folding frequencies discernible). Due to the equatorial symmetry in the example, even and odd degrees are still uncorrelated. In contrast to the order-aliasing, all degrees are affected by aliasing more or less in the same order of magnitude.

without any loss of information. In the succeeding convolution with the associated Legendre polynomials, the Fourier coefficients of higher orders (containing the additional signal) are simply damped towards the poles (cf. Fig. 5.3b). Nevertheless, even if the shown procedure does not improve the estimate as initially expected, it also does not negatively affect the result (see Fig. 5.2f).

Conclusions. At the bottom line, it can be concluded that the only way to really improve the estimate is to perform the analysis on a grid that is fine enough to not be influenced by aliasing in the first place. Additionally, to avoid spectral leakage (see Fig. 5.3a) the SH analysis should be performed up to the highest d/o that is possible to estimate on the respecitve grid (i.e., up to the 'Nyquist' d/o, see above). In case that the signal on the geographic grid was derived by collocation as presented in the previous chapter (4), it can be assumed that the estimated signal follows the spectral behavior defined by the used degree variances (see, e.g., Fig. 4.4b). Thus, when adjusting the grid resolution to the expected maximum signal wavelength one should be able to widely avoid any kind of aliasing in the SH analysis step (which does not mean that aliasing, possibly introduced through the collocation itself, can be avoided). Performing a SH analysis on a fine geographic grid up to the 'Nyquist' degree l_{max} is a computationally very complex task when not exploiting the orthogonality properties as explained in Sec. 2.2.1 (see Fig. 4.5). In fact, one of the main reasons why the collocation is needed in the first place is to obtain the data on the regular grid which forms the basis for all orthogonality properties of SHs, and, hence, also constitutes the foundation for all speed-up approaches. The utilization of these properties leads to a block-diagonal structure or even diagonal structure (in the Gauss-Legendre case) which can be evaluated very efficiently (the calculation can also be parallelized relatively easily). However, as already explained in Sec. 2.2.1, the constraints for using the block-diagonality do not allow to introduce individual point weights which would be available from the collocation. While this may seem problematic, it is

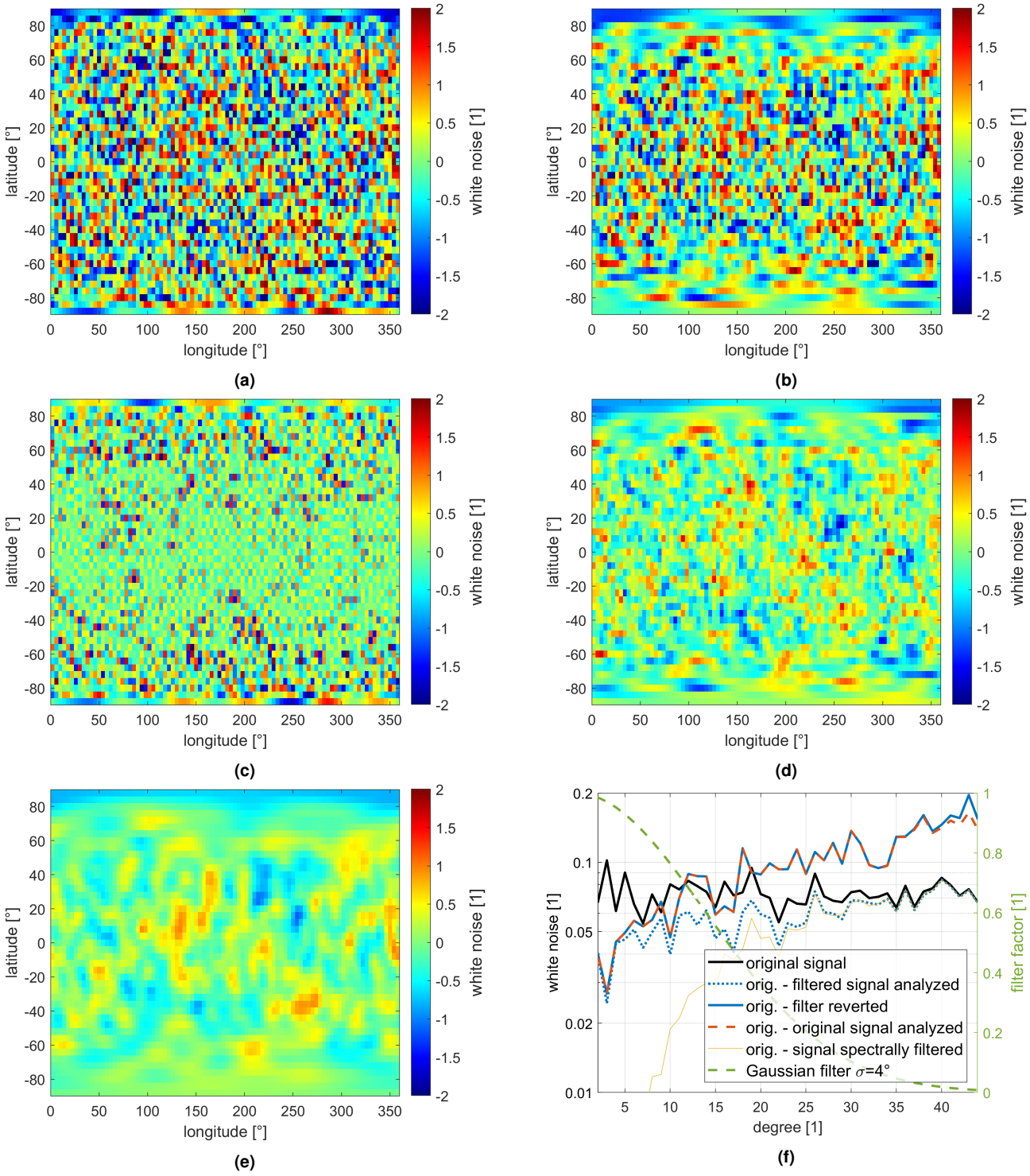


Fig. 5.2 – White-noise aliasing on a 4°x4° grid when introducing signal up to d/o 179 (cf. Fig. 5.1). (a) Reference signal (white noise) synthesized up to d/o 180 on a 4°x4° grid. (b) Recovered (estimated) signal, i.e., signal of Fig. 5.2a analyzed up to d/o 44 and synthesized back on the 4°x4° grid. (c) Difference between reference signal (Fig. 5.2a) and recovered signal (Fig. 5.2b). This represents the possible extra signal in geographic grids which is not recoverable through an SH analysis (cf. Sec. 5.1). (d) Reference signal synthesized up to d/o 44. This would be the optimal solution if no aliasing were present. (e) Reference signal (Fig. 5.2a) filtered with a Gaussian filter with $\sigma = 4°$. (f) Various signals in terms of degree amplitudes (square root of degree variances). solid black line: reference white noise signal (theoretically available up to d/o 180, cf. Fig. 5.2a and 5.2d). blue dotted line: difference between the analyzed filtered signal (Fig. 5.2e) and reference signal. solid blue line: difference between the filter-reverted signal to the reference signal. Dashed orange line: difference between analyzed reference signal and reference signal. Solid yellow line: reference signal filtered in the spectral domain. Solid green line: SH Gaussian filter factors ($\sigma = 4°$, for the calculation details see *Jekeli*, 1981).

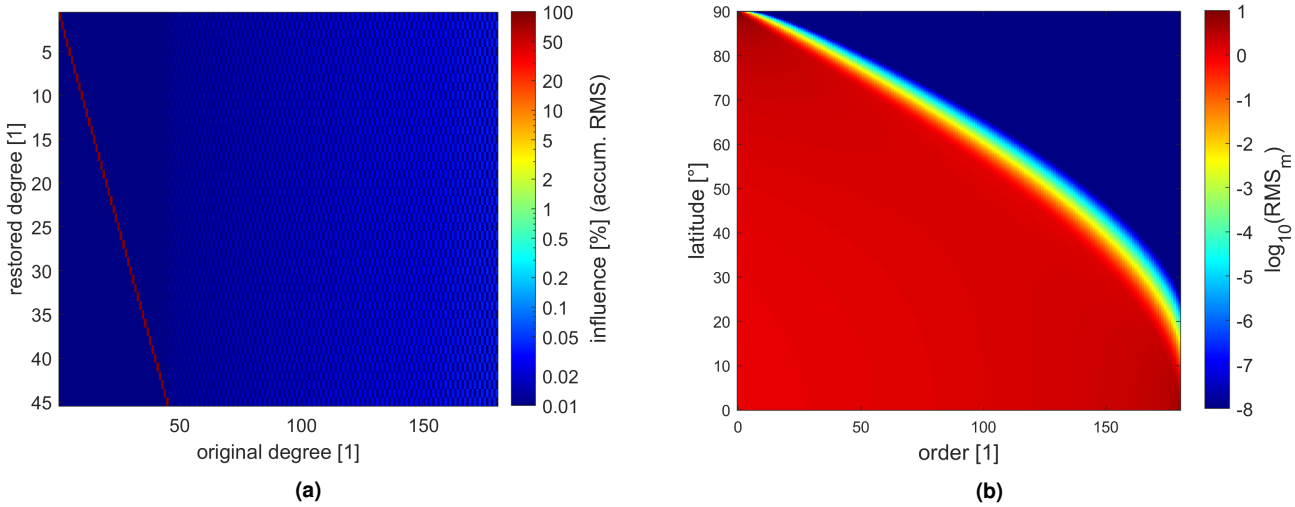


Fig. 5.3 – (a) Spectral leakage of the SH transform on geographic grids. In the example, a SH signal up to d/o 179 is assumed which is synthesized onto a 1°x1° geographic grid (360x180 sample points, thus d/o 179 is the grid's 'Nyquist' frequency) estimated (restored) up to d/o of 44 (note the difference to the aliasing setup in Fig. 5.1). Shown are accumulated RMS values (over all orders of a specific degree) of the least squares estimator normalized regarding the restored degrees (row-wise). In contrast to the aliasing error, the spectral leakage error is much weaker in amplitude (cf. Fig. 5.1a). Because orders are uncorrelated in the block-diagonal case, spectral leakage occurs just between degrees. For the same reason spectral leakage does not occur at all on Gaussian grids. (b) Damping of higher orders through the associated Legendre polynomials towards the poles. Shown are accumulated RMS values (over all degrees of a specific order) of associated Legendre polynomials for different latitudes.

not in the underlying case of a simple analysis where no combination with other data is performed, since a weighting would just influence the estimate if inconsistencies in the data are present (if overdetermined). It is maybe noteworthy that the aforementioned extra signal towards the poles (see Fig. 5.2c) represents such an overdetermination. If weighting is introduced, the Fourier transform will lose its orthogonality properties, which is why the previously made conclusion about the extra signal not influencing the result is then obsolete. As mentioned though, if data on the grid is obtained from collocation, it can be assumed that no significant extra signal is present in the data (because presumably already spectrally limited according to the degree variances, see Sec. 5.2). Hence, in case of collocation, the weighting (even when assuming full covariance information) cannot influence the estimate since no contradictions exist in the data according to the functional model of the analysis.

Spatial resolution of SHs. An interesting finding when evaluating the SH transform on geographic grids is that the actual spatial resolution of a signal which is provided up to the grid's 'Nyquist' d/o is actually lower than the grid's maximum spacing (at the equator). To see this, one has to compare the available information (i. e., the number of SH coefficients) with the area on which they apply. In the 2D Fourier case, for instance, when assuming a normalized area of $2\pi^2$ sr (steradiant, to resemble a 'flat' geographic grid), one would come up with a 2D resolution (when having $2m \cdot m$ coefficients) of $(\pi/m)^2$ sr which again is nothing else than the original resolution of the gird if the Fourier transform is performed up to the Nyquist wavenumber (because then, m is simultaneously the number of samples in latitudinal direction) . This is obvious since a 2D Fourier transform is able to exactly reconstruct the original grid. In the spherical case however, the same normalized grid has an area of 4π sr and one ends up with a spherical 2D resolution (when having m^2 coefficients) of $(2/\sqrt{\pi}\,\pi/m)^2$ sr. Hence, the shortest distance which can be expressed through SHs estimated from a grid (up to its 'Nyquist' d/o) is $2/\sqrt{\pi} \approx 1.12838$ times larger than the maximum distance occurring in the same grid (in latitudinal direction and on the equator in longitudinal direction). Hence, even on the equator, an arbitrary

(i.e., spectrally not limited) signal cannot be reconstructed through the SH transform. This is seen, e.g., in Fig. 5.2c where non-reconstructable signal is even present at the equator. At this point, one could argue that the expressible resolution of SHs limited to a certain max. d/o may spatially vary depending on the position (e.g., depending on the latitude so that the resolution on the pole could be lower/higher than on the equator). However, this cannot be the case since SHs can be spatially rotated without loss of information (see, e.g., *Heiskanen & Moritz*, 1967). This implies that the resolution of SHs, limited to a certain max d/o, has to be homogeneous-isotropic. As a consequence, when analyzing for instance a 1' grid, the spatial resolution of the obtained SH model is reduced to ∼1'8" (to ∼88.6% of the original resolution) at the equator. In other words, the rule of thumb of $180°/(l_{max}+1)$ for estimating the spatial resolution of SH signal limited to d/o l_{max} should rather read $360°/\left(\sqrt{\pi}\,(l_{max}+1)\right)$.

5.2 The (block-)diagonal SH analysis

Outline. After the more theoretical treatment of the SH transform in the previous section (5.1) the actual application of the (block-)diagonal analysis on the collocated observations will be subject of this section. During the course of the holistic processing chain the (block-)diagonal analysis can be used for two main objectives:

1. To get a SH representation of the collocated signal; as explained in the previous (Sec. 5.1), the solution obtained through the (block-)diagonal analysis is unique (when avoiding aliasing, i.e., the grid's resolution is high enough) and is not influenced by the omitted weighting (when spectrally limited, i.e., when no additional signal is present in the grid). Hence, since unique, the obtained solution can also be considered as optimal regarding the collocated signal, and, therefore, this solution can already be used as the final model (if no contradicting model is to be combined with the collocated signal, see second objective below). Note that the SH transform is a linear transform which implies that the SH analysis can be performed based on collocated residuals (without restoring the reductions). This is possible since the reduction model is also available as SH model, which means that the restore step can be shifted to the spectral domain (i.e., the analyzed coefficients of the residuals are simply added to the coefficients of the reduction model).

2. If the collocated signal is to be combined with a contradicting (satellite) model, a weighted combination (using point weights) is preferrable. A contradicting model in this sense denotes a model that possibly introduces observations that are in contradiction to the collocated signal, so that the weighting controls the composition of the combined model (for instance, GOCO06s would not contradict since it is used within the SATOPv2 model, see Sec. 5.3). Since no orthogonalities are preserved in the weighted case, the combination can probably not be performed based on the full resolution of the collocated grid (due to the high computation complexity, cf. Fig. 4.5). Thus, a lower-resolution version of the original grid is needed for this step (see Sec. 5.3). This lower-resolution grid should ideally not be influenced by aliasing and should also not contain unresolvable extra signal which could distort the estimate in the weighted case. This basically implies that the lower-resolution target grid needs to be spectrally limited to the grid's 'Nyquist' d/o (Eq. 5.1). Obviously, this can easily be achieved when the full spectrum of the collocated signal is available (through the block-diagonal analysis), which can then be synthesized to the desired grid by just using the signal up to the respective 'Nyquist' d/o l_{max}.

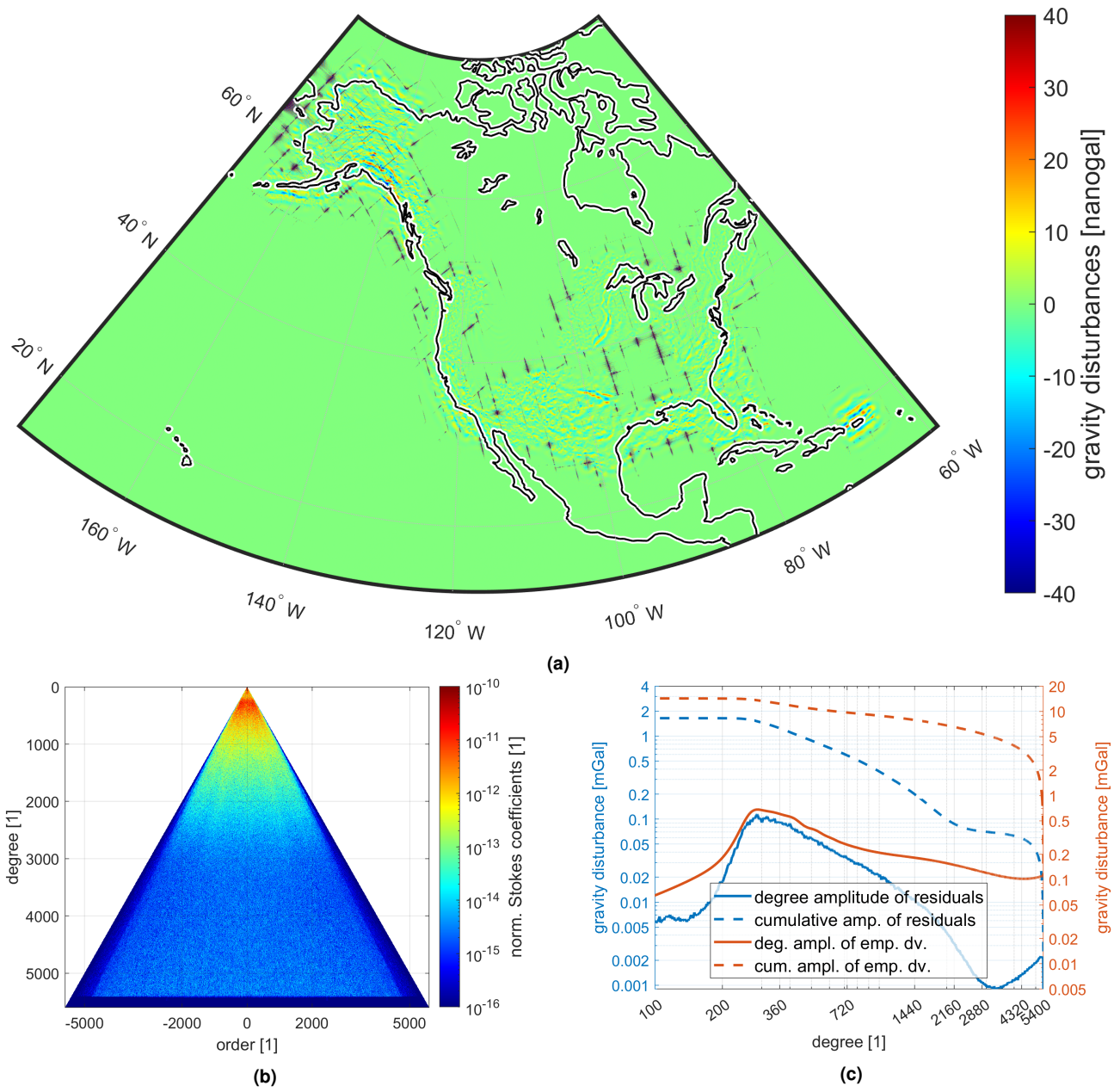


Fig. 5.4 – Result of the block-diagonal analysis of the residual collocation solution (see Fig. 4.11a) up to d/o 10,700. The analysis is performed in the spheroidal harmonic domain using the E_{GD} spheroid (see Sec. 4.6). (a) Difference of the original collocation result and the resynthesized block-diagonal solution. Note that the unit is nanoGal (10^{-6} mGal). (b) Block-diagonal solution (of the collocated residuals) in terms of Stokes coefficients in the spheroidal harmonic domain visualized as coefficient triangle (negative orders correspond to sine-coefficients) up to d/o 5,600 (coefficients above d/o 5,400 strongly converge towards zero). (c) Evaluation of the spectrum in terms of degree amplitudes (i.e., square root of degree variances) regarding gravity disturbances. Blue lines: degree amplitudes of the coefficients of the residual collocation result (Fig. 5.4b, solid) and the respective cumulative amplitude (from right, dashed). Orange lines: in comparison, the degree amplitudes estimated from the empirical covariance function used in the collocation (cf. Fig. 4.4b, solid) and the respective cumulative amplitude (from right, dashed).

Application - the SGDTv1(BD) model. As argued in Chap. 4 the target grid of the GRAV-D collocation is chosen to be 1' ($\sim$2 km) to safely cover the spectral signal content of the whole collocated signal (keeping in mind that GRAV-D was designed for a 10 km sampling). Thus, the resulting grid has a size of 21,600*10,800 sample points and a 'Nyquist' d/o of 10799 (see Fig. 4.11a). To avoid the instable downward-continuation term $(R/r)^{n+1}$ (see Sec. 2.1.2, Eq. 2.4) in the functional model, the analysis is performed in the spheroidal harmonic domain on the surface of the dedicated E_{GD} spheroid (to which the residuals are collocated, see Sec. 4.6). The result of the block-diagonal analysis of the collocated residuals is shown in Fig. 5.4. Several conclusions can be drawn from the result:

1. The difference between the input signal (collocated residuals) and resynthesized block-diagonal solution (Fig. 5.4a) is really small and close to numerics (in the nanoGal ,$10^{-6}mGal$, range). This difference corresponds to the aforementioned additional signal not recoverable by the SH analysis on geographic grids (see, e.g., Fig. 5.2c). Since this signal is almost zero, the assumption that the collocated residuals are spectrally limited is widely confirmed. Inspecting the spatial pattern of the differences reveals that the partitioning within the PE-LSC approach is the main reason for the remaining signal. This is not surprising, because even though the localization (see Sec. 4.5) is able to greatly reduce the fringe effects between partitions, they are not fully eliminated. However, considering observation accuracies in the mGal range, effects in the nanoGal area can be tolerated.

2. Inspecting the Stokes coefficient triangle of the collocated residuals (Fig. 5.4b), one can identify the following two features: firstly, near-sectorial coefficients converge towards zero. This is caused by regional (not global) coverage of the GRAV-D signal. Especially the absence of data in the equatorial region causes this effect, since (near-)sectorial coefficients are dominated by signal within this zone (as supported by Fig. 5.3b). Secondly, there is a strong decay of the signal above d/o 5,400. This can be explained by the fact that by limiting the estimated degree variances in the collocation to d/o 5,400 one basically prohibits any signal above this d/o. Another explanation is found in the airborne observations themselves which were low-pass filtered in the preprocessing (*GRAV-D Team*, 2013), and the SATOPv2 reduction model which also only includes signal up to d/o 5,480. Thus, it is very unlikely that in the input residuals of the collocation much signal can be found above d/o 5,400.

3. From the degree amplitudes of the collocated residuals (Fig. 5.4c, solid blue line) it is seen that the form of the curve resembles to a certain degree the form of the estimated degree amplitudes used as input to the collocation step (solid orange line). This is to be expected since the LSC should statistically 'trim' the original signal to the form of the assumed degree variances. However, there is also a certain mismatch between both curves especially in the shorter wavelengths where the decay of the actual signal is stronger than the decay of the estimated one. Here, one should keep in mind that interpreting the degree amplitudes of the collocated residuals may be dangerous since the GRAV-D dataset only covers a limited region of the Earth while degree amplitudes always refer to a global statistic (this is also the reason for the much weaker signal strengths of the analyzed signal compared to the estimated degree variances).

For numerical reasons (see below) the analysis is performed only up to d/o 10,700 (instead of 10,799). However, since there is basically no signal above d/o 5400 (cf. Fig. 5.4) this should be of minor concern at this point. Due to the absence of signal above d/o 5400, it is decided to limit the final SH model to d/o 5,480 (to the same max. d/o as the SATOPv2 reduction model, see Sec. 3.1). In a final step the spheroidal harmonic coefficients are converted to the spherical harmonic domain and added to coefficients of the SATOPv2 model to form the final model, denoted as SGDTv1(BD) ([S]atellite-[G]RAV[D]-[T]opography model [V]ersion

[1] [B]lock-[D]iagonal). The solution was calculated on the SuperMUC-NG high performance computing system (of the Leibniz Supercomputing Centre) using 16 nodes with 48 cores/nodes in about 15 minutes (solving 10,700*4 NEQSs of varying size). This corresponds to a single-core computation time of about 200 hours.

Remarks and caveats. As shown in Fig. 5.4 and explained above, the SH analysis of the grid performs as expected and yields the desired SH representation. Also, extending the grid resolution and the maximum d/o further would not be problematic from a numerical or computational perspective. Solving block-diagonal systems up to d/o $\sim$100,000 ($\sim$200 m sampling) would still be computationally feasible (or even beyond when having a Gauss-Legendre grid). Yet, there are also some caveats to consider when dealing with the SH analysis and with spherical harmonics in general:

- To be compatible with earlier definitions, the latitudinal spacing was chosen to be equidistant in accordance with the geographic latitude. This implies that associated Legendre polynomials are not orthogonal on this discrete sampling, and, thus, the resulting NEQS is block-diagonal (instead of diagonal when using the Gauss-Legendre spacing). Theoretically, one could also have chosen a (spheroidal) Gauss-Legendre grid as output (target) grid of the collocation which would have yielded a simple diagonal system. However, since the computation time of the block-diagonal solution is still relatively low (see above) using the simpler equidistant spacing is not problematic. On the other hand, using an equidistant spacing in terms of the geographic latitude means that the spacing in terms of the parametric latitude (which is used for the spheroidal harmonic analysis, see Sec. 2.1.2) is not equidistant. A consequence of this is that the block-diagonal normal equation matrix becomes numerically unstable when analyzing up to the grid's 'Nyquist' d/o (since the latitudinal sampling distance is in some regions larger than the average spacing, thus basically violating the 'Nyquist' sampling rule). An effective way to mitigate this problem is to not analyze up to the 'Nyquist' d/o, but to stay safely below, as it is done in the GRAV-D example where the analysis is performed up to d/o 10,700 (instead of 10,799). Obviously, this problem could also be solved by changing the latitudinal grid to either be equidistant regarding the parametric latitude or to follow the Gauss-Legendre rule. Consequently, it can be stated that from a numerical perspective, choosing a grid with equidistant spacing regarding the geographic latitudinal is suboptimal for the SH analysis.

- When using double-precision arithmetic and calculating associated Legendre Polynomials to a higher d/o the numerically stable sectorial standard recursions for calculating fully normalized associated Legendre Polynomials (see, e.g., *Heiskanen & Moritz*, 1967) suffer from numerical underflow. This has the consequence that subsequent elements in the recursion are bound to zero, although they would usually start to grow again in amplitude within the tesseral recursion. To circumvent this problem, *Fukushima* (2012a) proposed to use an extended arithmetic (called the x-numbers) which prevents underflow by manually extending the exponent of the double precision floating point format. Accordingly, in the used implementation, all associated Legendre Polynomials are calculated using this technique (with additional modifications, which specifically optimize for vectorization and parallelization). Note that for the whole procedure also associated Legendre polynomials of the second kind are needed within the transform from spheroidal to spherical harmonics (see *Jekeli*, 1988; *Fukushima*, 2013 and Sec. 2.1.2).

- Finally, and as general remark (already pointed out in Sec. 2.1.2), it can be concluded that spherical harmonics become numerically more and more unstable with higher degrees if upward or downward continuation is involved. This is easily seen through the factor $(R/r)^{n+1}$ (in Eq. 2.4) which attenuates or amplifies coefficients with exponentially increasing strength with higher degrees. Even though this problem is mitigated in the shown analysis procedure by staying on the reference surface, it comes back

into play when using the final model for upward resp. downward continuation. In both cases a very high relative precision between coefficients is demanded so that no artifacts are generated (cf. Sec. 2.1.2). Because of the Earth's topography, downward and upward continuation is always necessary when starting from observations on the surface and when ending up with synthesizing the estimated signal back onto the surface. For instance, in the actual GRAV-D example, upward continuation is performed within the collocation when estimating observations on an altitude of 6 km, and downward continuation would be necessary when synthesizing observations back onto the surface. If the reference height of the collocated grid had been chosen to be 0 km, the situation would be reversed. As explained in Sec. 2.1.2, the effect is still manageable when considering double-precision, a max. d/o of 5,400 and a max. radial difference between pole and equator (plus topography). Though, going even higher with the spectral resolution becomes very dangerous from a numerical perspective. While this problem is strongest when using spherical harmonics, it still exists within spheroidal harmonics (although in a weaker form) since the Earth's oblateness is of no concern then (only the topographic elevation itself). Hence, also when using (3D) spheroidal harmonics one eventually runs into the same numerical troubles at some point (within the associated Legendre polynomials of second kind which would then basically replace the (R/r) term). Thus, when one seeks an ultra-high-resolution general-purpose representation of a body's gravity field the use of spherical resp. spheroidal harmonics is not recommended (at least not inside the bodies sphere/spheroid of convergence, i.e. the smallest sphere/spheroid containing all masses, see, e.g., *Hirt & Kuhn*, 2017, *Bucha et al.*, 2018). However, as, e.g., *Bucha et al.* (2018) states, even if the series is diverging, a high-resolution model (e.g., up to d/o 5400) might still yield a good representation of the surface gravity signal when the model was acquired through a fit to observations located on that surface (although the aforementioned numerical problem still remains).

5.3 Combined SH modelling

Outline. In a general application, the observation-based input grid to the SH analysis (see previous section 5.2) is thought to be widely independent of additional sources such as satellite information. Thus, when assuming that solely ground-only observation within the input grid is available (i.e., the result of the collocation), it may be reasonable to seek an optimal combination between all available data sources (including, e.g., satellite information) for the final SH model. Without loss of generality, such an optimal combination (in a statistical sense) can be acquired in the spectral domain when stacking the NEQSs of all available data sources (see Sec. 2.2, Eq. 2.21). In the most common application, NEQSs containing satellite data (which are usually available from third parties) are stacked to the NEQS obtained from the input grid. For an optimal combination, the correct weighting of the normal equation components is of utmost importance. From the (externally obtained) satellite part it is normally assumed that the provided covariance information is correct. Hence, the main variable for influencing the combination is given by the weighting of the input grid. Depending of what is needed and what is available, several options exists for the choice of the weighting:

- *kite modelling*: the most simple and numerically most efficient way is found by assuming identical and uncorrelated errors for all grid points, such that the resulting NEQS becomes block-diagonal. If seen as advantage, one is also free to use individual weights resp. errors for every latitude (retaining one identical weight for all longitudinal value of one latitude) since the m-block orthogonality induced by the Fourier transform is still preserved (even if the equatorial symmetry might be violated, cf. Sec. 2.2.1). When stacking a block-diagonal NEQS of a higher max. d/o with a fully occupied NEQS of a lower

max. d/o (e.g., from a satellite model) a so-called kite NEQS can be obtained which is partially fully occupied and partially block diagonal (see Fig. 5.5). Such a kite system can be solved very efficiently by applying a reverse Gaussian elimination procedure on the emerging 3x3 super-block matrix which retains the sparsity of the original system (see for details *Zingerle*, 2015; *Gruber*, 2000; *Schuh*, 1996; *Bosch*, 1993). Usually, the solution obtained from a kite-system is already relatively close to the solution of a correctly weighted system if a satellite model is involved. This is because the errors in the satellite models grow very quickly with increasing degree due to the strong noise amplification caused by the downward continuation from the satellite altitude to the reference sphere (see, e.g., publication P-3, Fig. 4, violet dashed line). Due to this steep error growth in the satellite model, the actual errors within the input grid (which normally do not vary too strongly, maybe by a max. factor of 20, see, e.g., *Fecher*, 2015) are of minor concern since the spectral transition band (from satellite data to ground data) would not be altered significantly (cf. publication P-3, Fig. 4, when the yellow line is moved up and down, the intersection degree with the violet dashed line changes only slightly). Hence, in the mentioned application a kite solution can be used very well as a first approximation for a correctly weighted solution. The big advantage of kite systems (next to the numerical efficiency) is that the input grid can be used in full resolution and does not have to be downsampled (see previous Sec. 5.2). Note that a lot of effort has been put into the development of efficient parallelized algorithms for handling the sparse structure of huge kite systems and the appropriate linear algebra through all required processing steps (including, e.g., the inversion to acquire variance information). All details on this subject can be found in *Zingerle* (2015).

- *dense modelling with a priori covariance information*: if the input grid is obtained from a method such as collocation, a relatively realistic error (covariance) information is usually available. Theoretically, these errors can be used when creating the NEQS of the input grid either as simple (diagonal) variance information or even as (fully occupied) covariance information. However, in both cases the resulting normal equation matrix is fully occupied, and, thus, the gained advantages of using a regular geographic input grid are lost again. Even on the largest current HPC systems, the maximum solvable d/o of fully occupied systems is very limited (see, e.g., Fig. 4.5 and publication P-3), which is why such a combination can just be performed on a down-sampled (and spectrally limited) versions of the original input grid. As explained in Sec. 5.2, such a grid can be obtained by a preceding block-diagonal analysis of the original dataset. In previously performed combinations (e.g., XGM2019) it has been shown that above the max. d/o of the satellite model the combined solution is strongly converging to a block-diagonal solution (at least when the input grid it already spectrally limited, see publication P-3, Fig. 4, yellow line). Hence, it is justifiable (without mathematical proof) to extend the spectrally limited solution of the (fully occupied) stacked system with the block-diagonal solution to reconstruct the full spectrum of the original input grid. Intuitively, this can be explained by the fact that the satellite system does not hold information above its max. d/o, and, thus, cannot significantly influence the higher frequency information within the input grid. Another problem when using a spectrally limited input grid is that observations on the input grid are correlated after the spectral limitation (even if they were not before). One could theoretically use the original covariance information (if present), however, this information probably applies only partially to the spectrally limited grid since the modification through the spectral limitation itself is not considered in it. Hence, introducing covariance information from the original grid is not recommended when using a spectrally limited version (beside the fact that handling covariance information of the spatial geographic grids is numerically even more demanding than in the spectral domain due to the 2:1 relation of information, see Sec. 5.1).

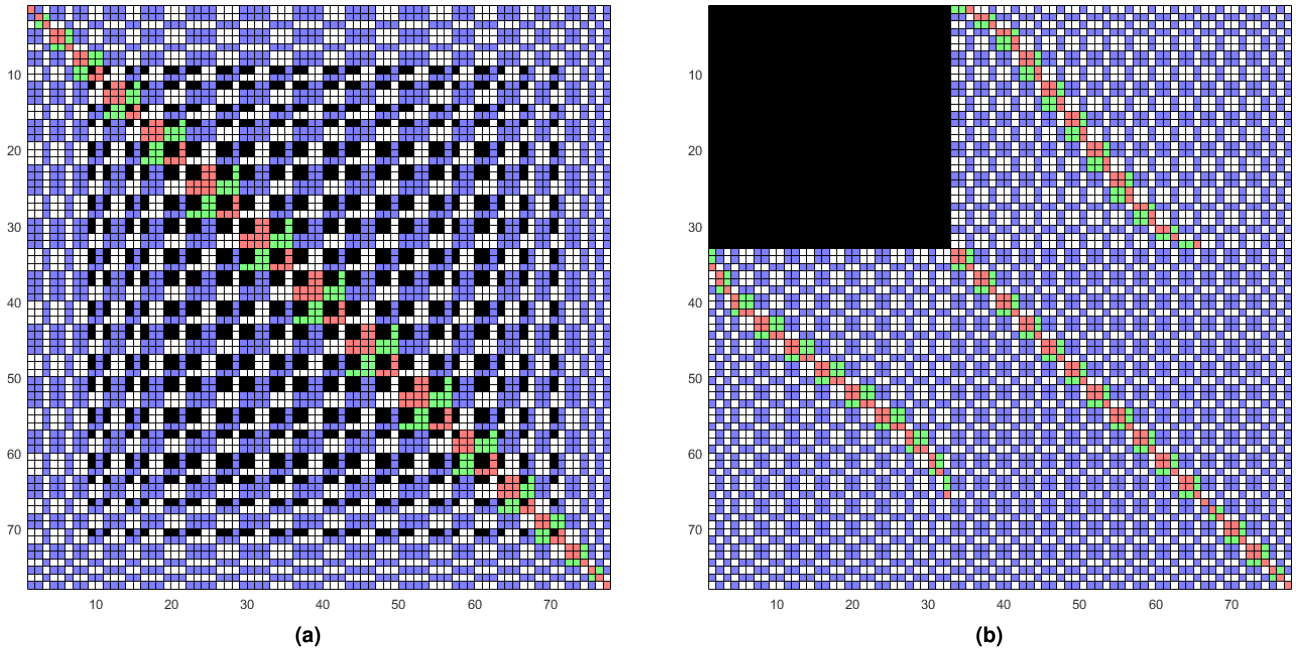


Fig. 5.5 – Illustration of the sparse structure of a stacked normal equation matrix consisting of a fully occupied lower d/o normal equation matrix (up to d/o 5) and a block-diagonal higher d/o normal equation matrix (up to d/o 8). The minimum d/o is set to 2 in the example. Different colors represent different reasons for orthogonality (cf. Sec. 2.2.1). Blue: orthogonal due to m-block orthogonality. Green: orthogonal due to the equatorial symmetry. White: orthogonal due to m-block orthogonality and equatorial symmetry. Red: non-orthogonal (occupied) element of the block-diagonal matrix. Black: element of the fully occupied matrix. (a) Structure of the stacked normal equation matrix when using a block-diagonal sorting of the elements (i.e., Stokes coefficients, sorted by sine-cosine -> order -> odd-even -> degree). Using this sorting, the elements of the fully occupied matrix are scattered over the resulting matrix. (b) Structure of the stacked normal equation matrix when using a so-called kite sorting of the elements (sorted by d/o$\leq$5 $\rightarrow$ order$\leq$5 $\rightarrow$ sine-cosine $\rightarrow$ order $\rightarrow$ odd-even $\rightarrow$ degree). When sorting all elements which have a d/o smaller than that of the fully occupied matrix (5 in this example) to the beginning a so-called kite matrix emerges. The kite matrix is characterized by a fully occupied left upper block and block-diagonal right lower 'tail'. In addition, correlations exist between the fully occupied and the block-diagonal part which are represented by the off-diagonal 'wings' which also show a block-diagonal structure (taken from *Zingerle*, 2015).

- *dense modelling using empirically-derived variances*: when no other error information about the input grid is available, the errors can also be empirically estimated by evaluating the differences in the longer wavelengths between the input grid and the satellite model. An elegant way of comparing the long wavelength differences is found by using the aforementioned kite solution with the same satellite model: obviously, kite solution also the converges to the block-diagonal solution above the max d/o of the satellite model (for the same reason as the fully occupied system does, see above). Thus, when building the difference between block-diagonal and kite solution, only the relevant long wavelength differences remain. From this differences (in the spatial domain) an empirical standard deviation can be obtained in the same fashion as in Sec. 4.6 (Eq. 4.19). While this method holds the advantage that the variances of the input grid can be estimated without any prior knowledge, the obvious disadvantages are that only the error in the long wavelengths can be estimated and that it is not possible to acquire covariance information (which is obviously present when working with spectrally limited grids). Further, all the other previously explained problems when working with fully occupied normal equation also apply in this case. This procedure was used, e.g., in the processing of XGM2019 due to the lack of additional error information for the ground gravity grid (see publication P-3 for further details such as rescaling of the empirical errors).

The main assumption when performing a combination with a satellite model is that the satellite model contains observations which to some extent contradict the observations within the input grid (otherwise the input grid would not be altered through the combination with the satellite model). While this may be true for arbitrary input grids, it is not the case in the shown holistic processing chain. To see this, one has to go back to the very beginning of the processing (chapter 3) to look at how the SATOPv2 model (used for observation reduction) was constructed. In fact, the SATOPv2 model already contains the satellite information to a large extent (since the topographic model is assumed to have a very high error in the lower wavelengths, see publication P-1, Fig. 1a, yellow curve). When reducing the SATOPv2 model from the observations and assuming that no signal is contained in the lower degrees in the collocation (through the empirical degree variances, see Fig. 4.4b), one basically prohibits signal to be estimated into the longer wavelengths. Hence, after restoring the SATOPv2 model, the satellite model in the longer wavelengths is still widely preserved. In fact, since the actual covariance of (static) satellite models is, due to its nature, relatively homogeneous and also isotropic (at least when using combined satellite models such as GOCO06s and when disregarding the polar regions) the collocation is able to emulate the satellite's covariance reasonably well through the used homogeneous-isotropic covariance function induced by the empirical degree variances. Eventually, this means that the collocation itself is already able to perform a weighted combination with the satellite model, so that a combination in the spectral domain should be redundant, and, thus, unnecessary. Additionally, performing the combination in the spatial domain has several advantages: firstly, it is not affected by the disadvantages of the spectral methods mentioned above (i.e., unrealistic weighting, limited resolution, omission of covariances, problem of restoring the original spectrum, etc.). Secondly, due the localization within the PE-LSC approach, the 'combination' in the spatial domain is numerically much more efficient than in the spectral domain where a 'localization' of the problem is not possible. Further, the combination in the spatial domain can probably be considered as more optimal than the combination in the spectral domain since it is performed based on the original observations and not on the already processed gridded data. At this point, one could argue that a combination in the spectral domain is still better since the full covariance information of the satellite model can be introduced there. However, as shown in publication P-4, in the course of the collocation it is also possible to introduce the correctly propagated (into the spatial domain) covariance information from SH models if one finds that assuming homogeneous-isotropic covariances for describing the model's performance is insufficient.

Application - the SGDTv1 model. As argued above, for the gridded GRAV-D dataset it is assumed that it already contains a suitable combination with the GOCO06s satellite model, which is seen as identical or even superior to what a combination in the spectral domain could achieve. Though, in some cases, the collocation is also able to produce erroneous long-wavelength patterns in the result (e.g., if unexpected offsets or similar is present in the observations, see below). In order to check whether the collocation worked as expected and that the gridded GRAV-D dataset does not contain significant differences to the satellite model in the longer wavelengths, a kite solution as explained above is calculated. Within this solution the GOCO06s satellite model is introduced as fully occupied NEQS up to d/o 320, converted into the spheroidal harmonic system referring to the TOPEX/POSEIDON spheroid (short T/P spheroid, see *AVISO*, 1996). The GRAV-D dataset is introduced as 1' grid of gravity disturbances on the surface of the T/P spheroid (by synthesizing the SGDTv1(BD) model to this surface). For the weighting of the grid's observations a latitude-dependent standard deviation of 15 mGal$/\sin\theta$ is used, which should correspond to an error of 3 mGal on the equator regarding a 5' grid (which represents the actual target resolution of GRAV-D). The difference between SGDTv1(BD) and the kite solution is depicted in Fig. 5.6. Fig. 5.6a shows that the differences are mostly well below 1 mGal and only exceed 1 mGal in some few exceptional locations. These differences are remarkably small when considering the fact that GRAV-D observations have an assumed accuracy of 4 mGal. This result basically supports the assumption that the collocation is able to perform an appropriate combination with the GOCO06s satellite

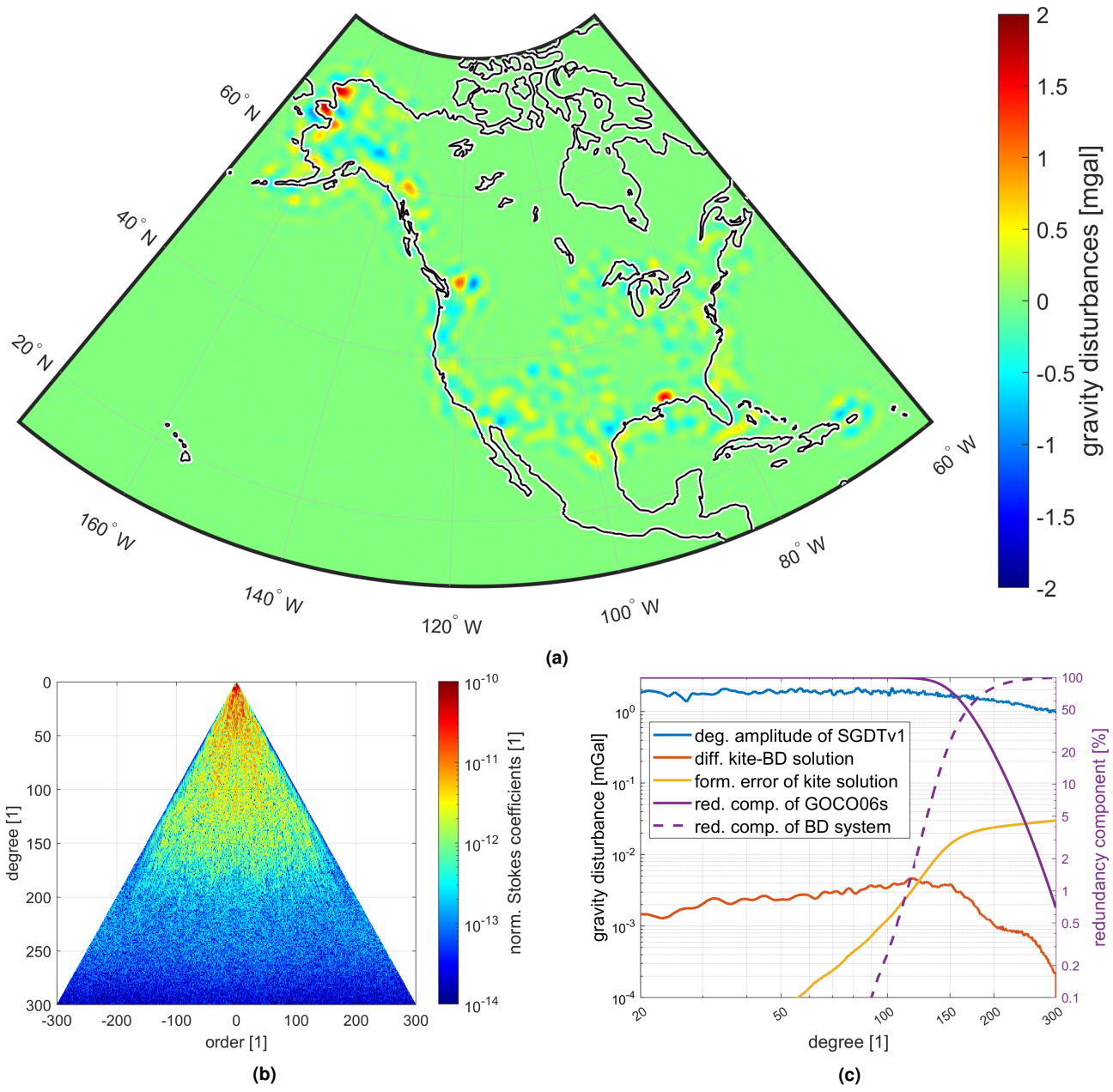


Fig. 5.6 – Result of the kite combination of the collocation result with the GOCO06s satellite model (i.e., the SGDTv1 model, see Sec. 5.3 for the detailed setup). (a) Difference between the original block-diagonal solution (SGDTv1(BD), cf. Sec. 5.2 and Fig. 5.4) and the kite solution (SGDTv1). (b) Same differences as in Fig. 5.6a in terms of Stokes coefficients in the spheroidal harmonic domain visualized as coefficient triangle (negative orders correspond to sine-coefficients) up to d/o 300 (coefficients above d/o 300 strongly converge towards zero). (c) SGDTv1 in terms of degree amplitudes (i.e., square root of degree variances) regarding gravity disturbances. Blue line: full signal of the kite solution (SGDTv1). Orange line: Difference between kite solution and block-diagonal solution (cf. Fig. 5.6a and 5.6b). Yellow line: formal error of the kite solution. Solid violet line: redundancy component of the GOCO06s normal equation part. Dashed violet line: redundancy component of the block-diagonal normal equation part.

model. However, when inspecting the difference further (in terms of degree amplitudes, see Fig. 5.6c), one can see that the majority of the discrepancies (orange line) are spectrally located below the transition band (where the violet lines cross). This implies that those discrepancies can be attributed to the aforementioned erroneous long-wavelength patterns induced by the collocation (since the GOCO06s model should be very accurate in the longer wavelengths, see yellow line). An additional hint that this error is induced by the collocation is provided by the fact that the differences in the spatial domain (Fig. 5.6a) correlate to some degree with the observation errors from the collocation (cf. Fig. 4.12a.) Consequently, by using the kite solution instead of the block-diagonal one, these erroneous patterns can be widely corrected. Thus, it is decided to redeclare the final SGDTv1 model to be the kite solution. Since the differences between block-diagonal solution and kite solution are already very small, calculating an individually weighted (fully occupied) combination is considered unnecessary (see above for the arguments).

Conclusions and outlook. Somewhat surprisingly, it has been shown that it is evidently better (due to the aforementioned reasons) to trust the collocation with the weighted combination between satellite information and ground data. This is in contrast to earlier paradigms where it has always been assumed that for global gravity field modelling the combination with the satellite model has to be performed in the spectral domain (see, e.g., *Pavlis et al.*, 2012, *Fecher*, 2015 or *Pail et al.*, 2018). However, this study indicates that the combination in the critical spectral transition band can also be handled very well (or even better) by the collocation and that the longer wavelengths (which are difficult to catch for PE-LSC) can efficiently be corrected (if even necessary) by a kite combination with a satellite model. Note that for this corrective kite combination it is important to set the weighting of the collocated observations (i.e., the block-diagonal system) high enough so that the shorter wavelengths (which are assumingly estimated correctly through the collocation) are not impaired by this correction which should just affect the longer wavelengths.

The reasons why the collocation estimates signal into the longer wavelengths even if the respective degree variances (covariance function) do not foresee it is not yet conclusively clarified. One explanation may be that long-wavelength signal in the residuals (e.g. an offset) is generally not expressible through a local system, and, hence, causes the error. Another explanation is that perhaps every larger deviation from the assumed statistics (e.g., an outlier) causes a certain stress to the system which may manifest itself as a longer-wavelength pattern in the result. In case of PE-LSC it is also possible that the localization (through damping of far-off observations) has a certain influence on the longer wavelengths (although in the AntGrav project, see publication P-2, it has been shown that the effect is more positive than negative since the classical LSC tends to 'overshoot' in observation-free regions). Eventually, it can be stated that this problem is mainly confined to areas where observations do not follow the covariance function. Thus, by checking the observations carefully before and after the collocation (see Secs. 3.3, 4.6), a lot of problems which cause the deviations (such as offsets, long-wavelength residuals and outliers) can be sorted out. For the GRAV-D example, this error can be considered as small (see Fig. 5.6) since the observations have already been already checked for the mentioned problems beforehand and can be assumed free of serious flaws (with a few exceptions).

At this point, the end of the holistic processing chain has been reached with the kite solution (the SGDTv1 model, cf. Fig. 5.6) as final outcome. In addition to the model itself also spatial covariance information (at least for a local surrounding) can be provided (as additional outcome of PE-LSC). The kite solution also provides variances for the estimated coefficients, although they are considered as not very accurate (due to the unrealistic weighting of the block-diagonal system there). Thus, for users it is suggested to use the covariance information from the PE-LSC method. Although the processing is finished, in the scope of the holistic approach one still has the possibility to iteratively restart the whole procedure by replacing the reduction model

from Sec. 3.1 (i.e., the SATOP model) with the final kite solution (the SGDTv1 model in this case). Using the final model should result in much smaller residuals (cf. Sec. 3.3, Fig. 3.3a) which can be treated as suggested in Sec. 4.6. In fact, starting the iteration from this last point of the processing chain is an alternative to starting the iteration already after the collocation. The difference is that if the collocated observations show long-wavelength 'errors', these 'errors' are reintroduced in the residuals if not corrected through the kite model. In case that these long-wavelength errors were already in the initial observation residuals and are just propagated by the collocation, they will more or less cancel out in the residuals of the second iteration (but not when using the kite solution). In case that these errors were artificially produced by the collocation itself, they will be added in the residuals of the second iteration (in contrast, when using the kite solution, they will be corrected). Hence, it depends on the individual circumstances which iteration approach is to prefer (if an iteration is even needed). As shown in Sec. 4.6, in the course of the GRAV-D case study a second (or maybe even third) iteration is perhaps appropriate due to the stronger differences in the covariance function among different terrain types, which cannot be modeled in a single iteration through a single homogeneous-isotropic covariance function. Since for SGDTv1 the modifications through the kite solution are very small, it does not matter if the iteration is performed based on observation errors (after the collocation) or based on a new reduction model (available after the kite combination).

5.4 External validation of SGDTv1

Outline. For completeness (and in brevity), it shall be shown how a final EGGM can be validated through external data sources and how the model can be benchmarked against other competing EGGMs. Generally, there are plenty of possibilities to do so, and it is not feasible to state a complete list of methods here. However, it is at least possible to identify two main classes of methods which are commonly used:

- *model-to-data comparison*: this class of validations aims to compare the model against other reference data (i.e., gravity field observations or data derived from such observations), which may be dependent (if already included in the modelling) or independent of the model. In the simplest case the SH model is synthesized to the observation's functionals, and, the residuals are investigated in the following (e.g., spatially or statistically). For instance, the inspection of the residual observations (see Sec. 3.3) would fall into this category. In more complicated cases, a further processing of the synthesized data and/or the reference data might be necessary (e.g., when performing orbit fits with the model, see for instance *Gruber et al.*, 2011, comparing with drifter data, see *Siegismund*, 2020). All in all, it can be stated that the quality of the final assessments is governed by (1) the quality of reference data, (2) the independence of the reference data to the model, and (3) the sensitivity of the method/functional to the spectral bands of interest.
- *mode-to-model intercomparison*: when reference data (see above) is available in form of another (competing) SH model, a validation (i.e., intercomparison) between both models in the spectral domain is possible. There, coefficient differences are either investigated directly or in a statistical way (mostly in terms of degree variances or similar). In the scope of this thesis such comparisons have already been performed, e.g., in Fig. 5.6b or 5.6c. As in the model to data comparison case, the quality of the final assessments is driven by the (1) quality of the reference model (in the spectral bands of interest), and by (2) the independence to the model to validate (in the spectral bands of interest). While with this kind of spectral validations the global performances (differences) can be quantified very well, it is not possible to draw conclusions about the spatial distribution (patterns) of the observed deviations. However, with two

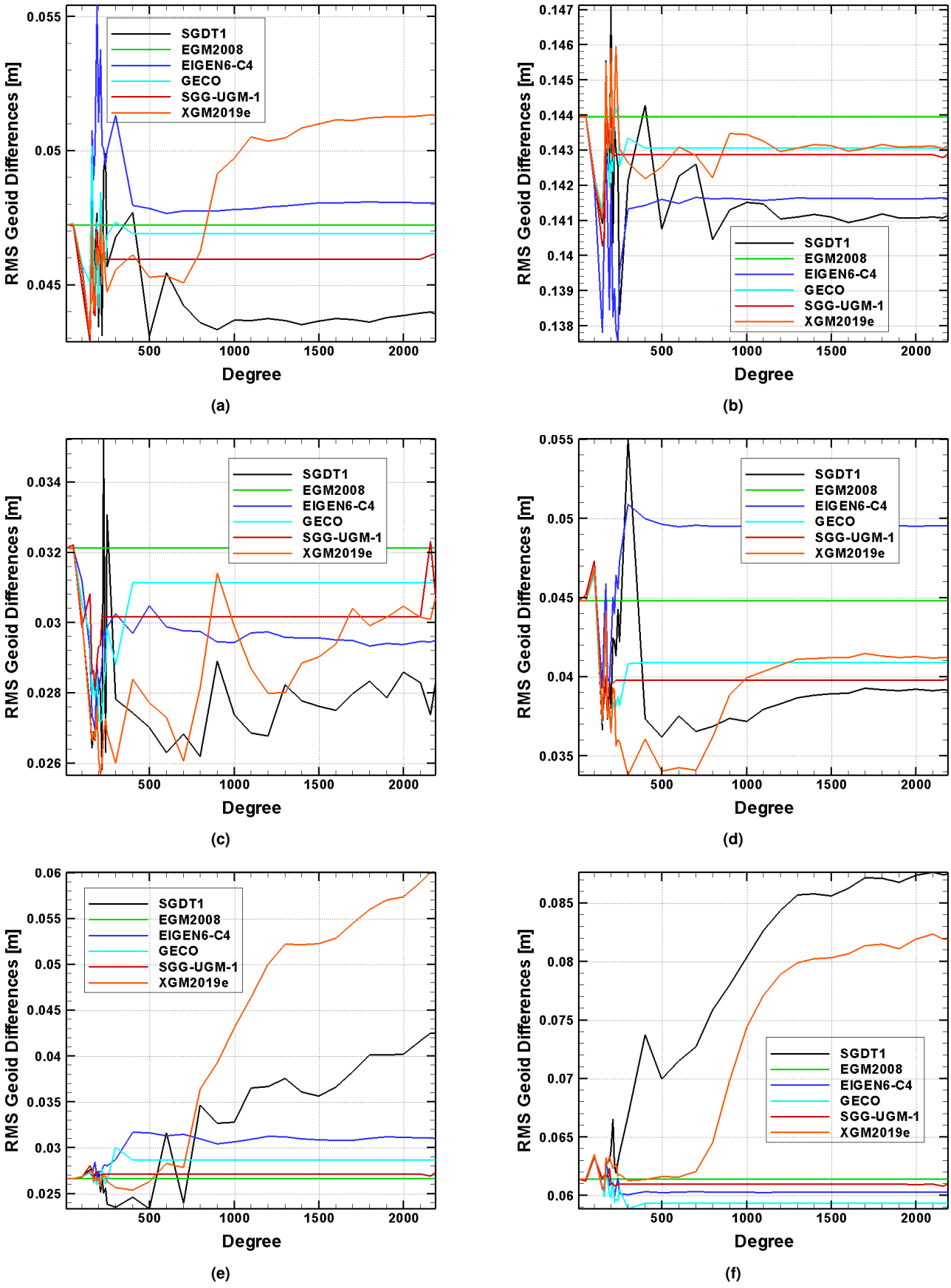


Fig. 5.7 – RMS of GNSS/levelling comparisons for different EGGMs regarding their maximum spectral resolutions in different states of US (see *Gruber & Willberg*, 2019 for more details on the processing and *ICGEM* for details on the different models). (a) RMS for the South Carolina dataset. (b) RMS for the Texas dataset. (c) RMS for the Massachusetts dataset. (d) RMS for the Florida dataset. (e) RMS for the Puerto Rico dataset. (f) RMS for the Arizona dataset.

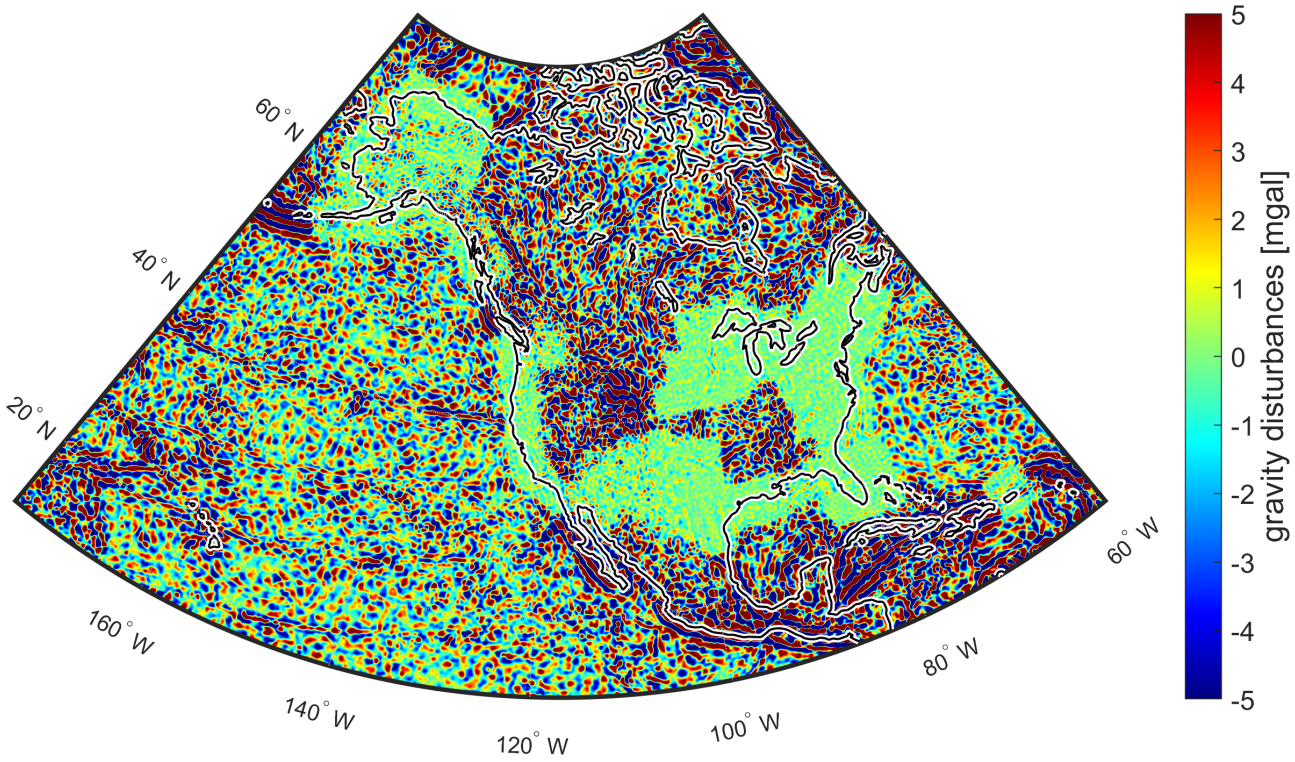


Fig. 5.8 – Difference between the SGTDv1 model and XGM2019 (see publication P-2) up to d/o 719 in terms of gravity disturbances and on an ellipsoidal height of 6 km.

SH models it is always possible to perform a synthesis and to inspect the patterns in the spatial domain (would then again correspond to a model-to-data comparison, see above).

Application. For the external validation of the SGDTv1 model a spectral model-to-model intercomparison seems inappropriate; The performance of SGDTv1 can obviously not compete with other models on a global scale, because ground data is just introduced where GRAV-D observations are available. It is therefore decided to stick to comparisons in the spatial domain. For SGDTv1, two different comparisons are performed:

In a first validation, the SGDTv1 model is used to generate physical heights used to build residuals to available datasets in different US states. The obtained residuals are then compared to other available competing EGGMs (taken from *ICGEM*) up to various maximum spectral resolutions to draw conclusions on the performance of SGDTv1. For an explanation of the detailed procedure, please see *Gruber & Willberg* (2019). The results are depicted in Fig. 5.7. Evidently, SGDTv1 shows a good performance in states which are generously covered by GRAV-D data (with an appropriate buffer, see, e.g., Fig. 5.7a-d) and deteriorates in states which are only partially covered (e.g., without the needed buffer, see Fig. 5.7e) or not covered at all (e.g., Fig. 5.7f). Generally, it shall be noted that this test is disadvantageous for the SGDTv1 model since for the generation of physical heights it is needed to perfom: (1) a field transformation (since the original data is provided in terms of gravity accelerations), and (2) downward continuation to the topography is necessary (since GRAV-D observations are recorded at a mean flight altitude of 6 km). With this, SGDTv1 is at a clear disadvantage compared to other EGGMs which, e.g., contain terrestrial observations (on the topography) or already include physical heights in the modelling. Additionally, the patch-wise coverage of GRAV-D observations is a further handicap since then, the longer wavelengths cannot be determined as stable, which deteriorates the downward continuation as well as the field transformation (which is why a sufficiently larger buffer of observations is necessary). This

probably explains the visible deviations and oscillations of the SGDTv1 model above d/o ∼200 (see, e.g., Fig. 5.7c). Even with these drawbacks though, the model generally seems to outperform its competitors in well-covered states.

In a second examination, SGDTv1 is compared to the XGM2019 model (see publication P-3). Up to d/o 719, XGM2019 is composed of the same satellite model (GOCO06s) and a precompiled ground-gravity dataset provided by courtesy of the NGA. While the processing details and exact data composition of this NGA dataset is unclear, it is known that it contains also GRAV-D data (though not all GRAV-D blocks, because the available dataset was compiled 2017, see *Pail et al.*, 2018). Thus, when comparing XGM2019 to SGDTv1 up to d/o 719, only small-scale differences are expected since similar data sources are included. These are visualized in Fig. 5.8 (up to d/o 719 and on a height of 6 km). It is evident that especially in flat regions (basically eastern part of the US) the model differences are really small (mostly below 1 mGal), and that the differences constitute only max. a few mGal in montaneous regions (still in the range or below the assumed accuracy of GRAV-D) where GRAV-D data is present in both models. This is a strong indication that no severe processing errors have orccured in the processing of either model. Further no long-wavelength residuals (in the range of the GOCO06s model) are discernible, which basically cross-validates both combination methods (LSC+kite vs. weighted SH) and underlines the assumptions made in the previous Sec. 5.3.

Comments. Although the performance of SGDTv1 can compete with other EGGMs in areas with GRAV-D data coverage, the model was neighter designed for maximum performance, nor has was it designed for publication in general. SGDTv1 shall therefore be interpreted as proof of concept for the holistic processing approach. For instance, the final model performance could have been greatly improved (especially in the border regions to areas without GRAV-D data coverage) if instead of the SATOP model a better performing EGGM had been chosen (e.g., XGM2019e). However, this variant was rejected, because otherwise the observations' independence to the reduction models would have been lost. As already mentioned, such dependencies would have been unfavorable in the course of the validation of the observations (Sec. 3.3). At a point where the observations are validated though, it would be a feasible option to replace the SATOP model with another (well performing) EGGM (with the limitation that this EGGM shall contain the same satellite model as used in the context of the final kite combination).

Chapter 6

Conclusions, outlook and final thoughts

Summarizing this thesis, in Chaps. 3-5 it has been shown how an integrated gravity field processing approach can help improve SH-based gravity field models, thereby fulfilling objective O-2 of this work (presentation of a holistic approach, see Sec. 1.1). By presenting the holistic processing approach, several innovations were introduced, starting with the SATOP model (Sec. 3.1), followed by the thinning procedure (Sec. 4.1), the estimation of empirical degree variances (Sec. 4.3), the partitioning/localization through the PE-LSC method (Secs. 4.4, 4.5), and, last but not least, the findings regarding the gridded SH analysis (Secs. 5.1, 5.2) and combination strategies (Sec. 5.3). The sum of these innovations represents the completion of objective O-1 (improvement of methodology). Additionally, in every section in which a method is explained, also an outlook is provided which may serve as a starting point for further improvements or which shows possible limitations of the current implementation. With that, also the last main objective O-2 (identifying limitations) is covered.

On the example of applying the holistic processing approach on the continent-scale GRAV-D dataset also the correct functionality of the proposed methods has been verified: it has been shown that (1) the approach is applicable on a nearly unlimited number of observations while staying within a feasible processing time (if the point density is not too high, cf. Secs. 4.1, 4.4, 5.2), and (2) the approach yields a correct result which which can match or even outperform competing models (resp. modelling techniques, see Sec. 5.3). However, processing a single continent-scale dataset is not the final aim of the presented holistic approach, but only a basis for its actual purpose, namely, a global-scale application on a multitude of datasets with varying properties (e.g., functional, accuracy, density, height, etc.). In fact, by using the localized PE-LSC method (Chap. 4) all the varying datasets can be integrated in a single modelling step into a seamless global grid without limitation to a local region (or a maximum number of observation points). Subsequent to the localized modelling, also the global modelling (Chap. 5) would greatly benefit from a scenario with global data coverage. The global performance of a resulting model (e.g., in terms of degree variances) would strongly improve since it wouldn't rely on the reduction model anywhere. Recognizing that having a global coverage of ground observations would be the ideal case, it has to be noted that at the time of writing (ground) gravity observations are still considered proprietary by many countries and individuals (probably due to economic or military reasons). As long as this situation remains unchanged, even the most advanced processing strategy can only provide a limited improvement of the final result (i.e., the SH model in this case). On the other hand, if a global observation coverage were available, with the presented processing chain it should be straightforward to derive an initial global SH model and to easily maintain it by updating the underlying data base.

While the limitations on the individual parts of the processing have already been discussed in the respective sections, at this point it might be appropriate to address (again) an inherent limitation of the procedure: the use of spherical resp. spheroidal harmonics in general. As already pointed out in Sec. 2.1.2 and 5.2, the SH series

expansion is generally divergent when introducing (1) a height component and (2) high-frequency signal. When dealing with high-resolution gravity field modelling of the Earth, both these issues (different heights and high-frequency signal) inevitably arise due to the Earth's topography, since (1) elevations obviously cause different heights, and (2) vertical slopes, for instance, induce infinitely high frequencies). Eventually, this means that an evaluation (i.e. analysis or synthesis) of the SH model below the sphere/spheroid of convergence (see Sec. 5.2, *Hirt & Kuhn*, 2017 or *Bucha et al.*, 2018) becomes more and more problematic with increasing degree of the signal as well as with a larger downward continuation. As pointed out in Sec. 2.1.2, for the Earth, a max. d/o of $\sim$5400 might be a reasonable compromise (when using double-precision arithmetic) between numerical stability and achievable resolution (note: an additional factor which can further worsen the result, but will not be further discussed in this thesis, are truncation errors of the SH series which also play a prominent role when dealing with high-resolution SHs). A max. d/o of 5400 corresponds to roughly $\sim$4 km spatial resolution. This is well above, e.g., the wavelength of valley-mountain structures of common mountain ridges, and, thus, misses a crucial part of the actual ultra-high-frequency (mostly topographically induced) signal of the Earth's gravity field. As of now, it is not predictable that such structures (with wavelengths down to a few hundred meters) can be resolved using SHs in the near future, since the mentioned divergence effect grows exponentially with respect to the max. degree. This leads to the final question of whether SHs are generally suited as stand-alone representation form for future ultra-high-resolution gravity field models which try to incorporate these wavelengths. The short answer to this question is probably no. SHs can be considered as unsuitable for the application in the ultra-high-resolution regime and below the sphere of convergence because the numerical problems (divergence but also the numerical complexity of the synthesis) greatly outweigh all of the SHs's advantages in the longer wavelengths (e.g., when applied for satellite models). Hence, the shown processing approach (especially, the SH global modelling part, Chap. 5) must be interpreted as a demonstration of what is possible in the context of the SH modelling environment (see *ICGEM*, 2019), but has to be complemented by other tools in the long term when aiming to further increase the resolution. For the general application of SHs, an inherent limit regarding the spatial and spectral resolutions is reached which cannot be overcome through simple methodological improvements. Therefore, a fundamentally different form of representation (other than SHs) is required for describing the ultra-high-frequency features of the gravity field near or on the topography.

In conclusion, the presented holistic SH modelling approach has been proven to work properly up to a certain resultion and is able to integrate heterogeneous observations of the Earth's gravity field accurately and consistently. For the description of shortest wavelengths, however, different modelling approaches are preferrable, which do not rely on reference surfaces other than the body's topography (to avoid downward continuation). Consequently, all currently available spectral methods (like SH transform, Fourier transform, wavelets, etc.) cannot be seen as qualified to form the basis for such models. This basically restricts the search for future alternative approaches to the spatial domain. Spatial methods (except collocation) are characterized by the fact that they try to directly model the masses responsible for the gravity signal. An important class of such spatial methods are radial basis functions (see, e.g., *Marchenko*, 1998, *Schmidt et al.*, 2007) where mass-like 'objects' located at various locations are modelled to describe the gravity signal. While those methods usually do not have numerical problems with downward continuation (or more specifically upward continuation in this case), a crucial, yet unsolved difficulty lies in the choice of a suitable amount of 'objects' and the choice of their corresponding locations. A wrong choice may easily lead to an overfitting (or underfitting) of the given gravity field data, so that the radial basis functions may perfectly fit the observations, but fail to describe the gravity field signal elsewhere. Additionally, as of now, radial basis functions usually do not directly take benefit from topographic information (beside RCR techniques, see, e.g., *Liu et al.*, 2020), which means that they are hardly able to reconstruct gravity field signals induced by a complex topography on small spatial scales

(e.g.,mountains, cliffs, canyons, etc.). In contrast to radial basis functions, the method of topographic forward modelling (see, e.g., *Hirt & Kuhn*, 2014) takes full advantage of topographic information and is able to properly describe the aformentioned small-scale structures which dominate the shorter wavelengths of the Earth's gravity field signal. However, since topographic forward modelling usually uses constant density assumptions, the actual gravity signal can only be roughly approximated.

Recognizing the strengths and the weaknesses of both spatial approaches (radial basis functions and topographic forward modelling), a combination of both methods might be a viable alternative for future ultra-high resolution gravity field models. More specifically, the geometric 'objects' used within radial basis functions might be set as the polyhedra defining the Earth's topography. For these polyherda, individual (vertically constant) 'densities' (i.e., weights) can be estimated which fit the given observations. This way, one could fully benefit from the available topographic information while removing the uncertainties of static density assumptions (of the topographic modelling) and the problem of the choice of amount and location of radial basis functions. To reduce the computational effort, 'densities' (weights) could be estimated in a reduced resolution on an interpolated grid and the topographic data could be evaluated on a multi-resolution grid (e.g., *Hirt et al.*, 2019) depending on the distance to the observation. Further, some additional regularizations could be defined regarding the 'density' (e.g., smoothness) which should stabilize the estimation even in case of a sparse data coverage. In such an approach, the longer wavelengths could still be reduced by means of a SH model (SGGM or EGGM) to spatially decorrelate the problem and to shift the estimation to residual 'densities'. Similar studies linked to this subject already exist in the field of geophysical gravity inversion where one tries to constrain density models to, amongst others, given gravity field observations (see, e.g., *Sambridge & Mosegaard*, 2002). The idea of using voxel models which is related to the idea of using polyhedra has already been examined, for instance, by *Fuchs* (2015). However, the mentioned studies mainly focus on local (3D) applications, they do not aim to provide a general descirption of the gravity field and on a global scale in the first place. Also, the previously proposed strategy is neither implemented nor investigated in any way in the scope of this thesis. So, even if the suggested method should not be applicable for any so far undiscovered reason, this discussion should have revealed that for ultra-higher-resolution applications SHs need a companion to describe the short-wavelength structures of the Earth's gravity field and that there is much room for innovative ideas here which only wait to be investigated...

Bibliography

Aviso (1996) AVISO user handbook: Merged TOPEX/Poseidon products. Techn. Rep. AVI-NT-02-101-CN, 198.

Bogaert I (2014). Iteration-free computation of Gauss–Legendre quadrature nodes and weights. SIAM Journal on Scientific Computing, 36(3), A1008-A1026. doi 10.1137/140954969

Brockmann JM (2014) On High Performance Computing in Geodesy: Applications in Global Gravity Field Determination [Doctoral dissertation, Bonn University]. Bonndoc. https://hdl.handle.net/20.500.11811/5871

Bronstein IN, Semendyayev KAF, Musiol G, Mühlig H (2015) Handbook of mathematics. Springer. doi 10.1007/978-3-662-46221-8

Bosch W (1993) A rigorous least squares combination of low and high degree spherical harmonics. In IAG General Meeting, Beijing, China.

Bucha B, Hirt C, Kuhn M (2018) Divergence-free spherical harmonic gravity field modelling based on the Runge–Krarup theorem: a case study for the Moon. Journal of Geodesy. 93. doi 10.1007/s00190-018-1177-4

Chambodut A, Panet I, Mandea M, Diament M, Holschneider M, Jamet O (2005) Wavelet frames: An alternative to spherical harmonic representation of potential fields. Geophys J Int, 163(3): 875–899. doi 10.1111/j.1365-246X.2005.02754.x

Childers VA, Bell RE, Brozena JM (1999) Airborne gravimetry: An investigation of filtering. Geophysics, 64(1): 61-69. doi 10.1190/1.1444530

Colombo OL (1981). Numerical Methods for Harmonic Analysis on the Sphere. Report No. 310, Department of Geodetic Science, The Ohio State University, Columbus, Ohio.

Darbeheshti N (2009) Modifcation of the least-squares collocation method for non-stationary gravity feld modelling [Doctoral dissertation, Curtin University]. Curtin Theses. http://hdl.handle.net/20.500.11937/2228

Eicker A (2008) Gravity Field Refinement by Radial Basis Functions from In-situ Satellite Data [Doctoral dissertation, Bonn University]. Bonndoc. https://nbn-resolving.org/urn:nbn:de:hbz:5N-13754

Fecher T (2015) Globale kombinierte Schwerefeldmodellierung auf Basis voller Normalgleichungssysteme [Doctoral dissertation, Technical University of University]. MediaTUM. http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:bvb:91-diss-20150713-1238858-1-1

Forsberg R, Tscherning CC (1981) The use of height data in gravity field approximation by collocation. J Geophys Res, 86(B9): 7843-7854. doi 10.1029/JB086iB09p07843

Fuchs MJ (2015) Detection and in-depth assessment of the 2011 Tohoku-Oki earthquake evaluating GOCE gravity gradient data [Doctoral dissertation, Technical University of University]. MediaTUM. http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:bvb:91-diss-20150922-1252237-0-8

Fukushima T (2012a) Numerical computation of spherical harmonics of arbitrary degree and order by extending exponent of floating point numbers. J Geod 86, 271–285. doi 10.1007/s00190-011-0519-2

Fukushima T (2012b) Recursive computation of finite difference of associated Legendre functions. J Geod 86, 745–754. doi 10.1007/s00190-012-0553-8

Fukushima T (2012c) Numerical computation of spherical harmonics of arbitrary degree and order by extending exponent of floating point numbers: II first-, second-, and third-order derivatives. J Geod 86, 1019–1028. doi 10.1007/s00190-012-0561-8

Fukushima T (2013) Recursive computation of oblate spheroidal harmonics of the second kind and their first-, second-, and third-order derivatives. J Geod 87, 303–309. doi 10.1007/s00190-012-0599-7

Fukushima T (2014) Numerical computation of spherical harmonics of arbitrary degree and order by extending exponent of floating point numbers: III integral. Computers & Geosciences, 63, 17-21. doi 10.1016/j.cageo.2013.10.010

GRAV-D Team (2013) GRAV-D General Airborne Gravity Data User Manual. Theresa Damiani and Monica Youngman ed. Version 2. Available September 10, 2021

GRAV-D Team (2015) Block CS08 (Central South 08); GRAV-D Airborne Gravity Data User Manual. Sandra Preaux, Monica Youngman, and Carly Weil ed. Version 1. Available September 10, 2021

GRAV-D Team (2017) Block MS05 (Mountain South 05); GRAV-D Airborne Gravity Data User Manual. Monica A. Youngman and Jeffery A Johnson. BETA Available September 10, 2021

Gruber T (2000) Hochauflösende Schwerefeldbestimmung aus Kombination von terrestrischen Messungen und Satellitendaten über Kugelfunktionen. PhD Thesis. (Scientific Technical Report ; 00/16), Potsdam : GeoForschungsZentrum, 178 p. doi 10.2312/GFZ.b103-00160

Gruber T (2001) High-resolution gravity field modeling with full variance–covariance matrices. Journal of Geodesy 75, 505–514. doi 10.1007/s001900100202

Gruber T, Visser PNAM, Ackermann C, Hosse M (2011) Validation of GOCE gravity field models by means of orbit residuals and geoid comparisons. J Geod, 85(11): 845-860. doi 10.1007/s00190-011-0486-7

Gruber T, Willberg M (2019) Signal and error assessment of GOCE based high resolution gravity feld models. J Geod Sci 9(1):71–86. doi 10.1515/jogs-2019-0008

Heiskanen WA, Moritz H (1967) Physical Geodesy. W.H.Freeman and Company, San Francisco.

Hirt C (2012) Efficient and accurate high-degree spherical harmonic synthesis of gravity field functionals at the Earth's surface using the gradient approach. J Geod 86, 729–744. doi 10.1007/s00190-012-0550-y.

Hirt C, Kuhn M (2014) Band-limited topographic mass distribution generates full-spectrum gravity field: Gravity forward modeling in the spectral and spatial domains revisited, J. Geophys. Res. Solid Earth, 119, 3646– 3661. doi 10.1002/2013JB010900.

Hirt C, Rexer M (2015) Earth2014: 1 arc-min shape, topography, bedrock and ice-sheet models - Available as gridded data and degree-10,800 spherical harmonics. International Journal of Applied Earth Observation and Geoinformation, 39: 103-112. doi 10.1016/j.jag.2015.03.001

Hirt C, Kuhn M (2017) Convergence and divergence in spherical harmonic series of the gravitational field generated by high-resolution planetary topography—A case study for the Moon. J. Geophys. Res. Planets, 122, 1727– 1746. doi 10.1002/2017JE005298

Hirt C, Rexer M, Claessens S, Rummel R (2017) The relation between degree-2160 spectral models of Earth's gravitational and topographic potential: a guide on global correlation measures and their dependency on approximation effects. J Geod 91:1179–1205. doi 10.1007/s00190-017-1016-z

Hirt C, Yang M, Kuhn M, Bucha B, Kurzmann A, Pail R (2019) SRTM2gravity: An Ultrahigh Resolution Global Model of Gravimetric Terrain Corrections. Geophys Res Lett, 46: 4618–4627. doi 10.1029/2019GL082521

Holschneider M, Chambodut A, Mandea M (2003) From global to regional analysis of the magnetic field on the sphere using wavelet frames. Physics of the Earth and Planetary Interiors, 135(2-3): 107-124. doi 10.1016/S0031-9201(02)00210-8

Ince, ES, Barthelmes F, Reißland S, Elger K, Förste C, Flechtner F, Schuh H (2019). ICGEM – 15 years of successful collection and distribution of global gravitational models, associated services and future plans. Earth Syst. Sci. Data, 11, 647–674. doi 10.5194/essd-11-647-2019

Ihde J, Sánchez L, Barzaghi R, Drewes H, Foerste C, Gruber T, Liebsch G, Marti U, Pail R, Sideris M (2017) Definition and Proposed Realization of the International Height Reference System (IHRS). Surv Geophys, 38(3): 549-570. doi 10.1007/s10712-017-9409-3

Jekeli C (1981) Alternative methods to smooth the Earth's gravity field. NASA, Grant No. NGR 36-008-161, OSURF Proj. No. 783210, 48 pp, Dec 1981, N82-22821/4

Jekeli C (1988) The exact transformation between ellipsoidal and spherical harmonic expansions. Manuscr Geod 1988(13):106–113

Krarup T (1969). A Contribution to the Mathematical Foundation of Physical Geodesy. In: Borre K (ed) Mathematical Foundation of Geodesy - Selected Papers of Torben Krarup. Springer, Berlin, 2006. doi 10.1007/3-540-33767-9

Kühtreiber N (2002) High precision geoid determination of Austria using heterogeneous data. In Gravity and Geoid 2002, Proceedings of 3rd meeting of the Int. Gravity and Geoid Commision, Thessaloniki, Greece (pp. 144-149).

Kvas A, Mayer-Gürr T, Krauss S, Brockmann JM, Schubert T, Schuh W-D, Pail R, Gruber T, Jäggi A, Meyer U (2019) The satellite-only gravity field model GOCO06s. GFZ Data Services. doi 10.5880/ICGEM.2019.002

Lieb V (2017) Enhanced regional gravity field modeling from the combination of real data via MRR [Doctoral dissertation, Technical University of University]. MediaTUM. http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:bvb:91-diss-20161222-1325856-1-0

Liu Q, Schmidt M, Sánchez L, Willberg M (2020) Regional gravity field refinement for (quasi-) geoid determination based on spherical radial basis functions in Colorado. J Geod, 94, 99. doi 10.1007/s00190-020-01431-2

Lüke HD (1999) The origins of the sampling theorem. IEEE Communications Magazine, vol. 37, no. 4, pp. 106-108. doi 10.1109/35.755459

Marchenko AN (1998) Parameterization of the Earth's gravity field. Point and line singularities. Lviv Astronomical and Geodetic Society, Lviv, Ukraine.

Meissl P (1971) On the linearization of the geodetic boundary value problem. Report 151, Department of Geodetic Science and Surv, Ohio State University Columbus.

Moritz H (1980) Advanced Physical Geodesy. Herbert Wichmann, Karlsruhe. ISBN 3-87907-106-3

Moritz H (2000) Geodetic Reference System 1980. J Geod, 74: 128-133. doi 10.1007/s001900050278

Niemeier W (2008) Ausgleichsrechnung. De Gruyter. doi 10.1515/9783110206784

Pail R, Reguzzoni M, Sansò F, Kühtreiber N (2010) On the combination of global and local data in collocation theory. Stud Geophys Geod, 54(2): 195-218. doi 10.1007/s11200-010-0010-1

Pail R, Fecher T, Barnes D, Factor JF, Holmes SA, Gruber T, Zingerle P (2018) Short note: the experimental geopotential model XGM2016. J Geod, 92(4): 443-451. doi 10.1007/s00190-017-1070-6

Pavlis NK, Holmes SA, Kenyon SC, Factor JK (2012) The development and evaluation of the Earth Gravitational Model 2008 (EGM2008). J Geophys Res, 117, B04406. doi 10.1029/2011JB008916

Preparata F, Shamos MI (1985) Computational Geometry: An Introduction. Springer-Verlag, Berlin, Heidelberg, ISBN 978-0-387-96131-6

Reguzzoni M, Tselfes N (2009) Optimal multi-step collocation: application to the space-wise approach for GOCE data analysis. J Geod 83, 13–29 (2009). 10.1007/s00190-008-0225-x

Rexer M (2017) Spectral Solutions to the Topographic Potential in the context of High-Resolution Global Gravity Field Modelling [Doctoral dissertation, Technical University of University]. MediaTUM. http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:bvb:91-diss-20170531-1349781-1-7

Romain G, Jean-Pierre B (2001) Ellipsoidal harmonic expansions of the gravitational potential: theory and application. Celestial Mechanics and Dynamical Astronomy, 79(4), 235-275. 10.1023/A:1017555515763

Sambridge M, Mosegaard K (2002) Monte Carlo methods in geophysical inverse problems. Reviews of Geophysics, 40 (3), 1009. doi 10.1029/2000RG000089

Schmidt M, Fabert O, Shum CK (2005) On the estimation of a multi-resolution representation of the gravity field based on spherical harmonics and wavelets. Journal of Geodynamics, 39(5): 512-526. doi 10.1016/j.jog.2005.04.007

Schmidt M, Fengler M, Mayer-Gürr T, Eicker A, Kusche J, Sánchez L, Han S-C (2007) Regional gravity modeling in terms of spherical base functions. J Geod, 81: 17–38. doi 10.1007/s00190-006-0101-5

Schuh, WD (1996) Tailored numerical solution strategies for the global determinations of the Earth's gravity field. Mitteilungen d. Geodat. Inst. d. TU Graz, No. 81, Graz.

Siegismund F (2020) A spectrally consistent globally defined geodetic mean dynamic ocean topography. Journal of Geophysical Research: Oceans, 125, e2019JC016031. doi 10.1029/2019JC016031

Simons FJ (2010) Slepian Functions and Their Use in Signal Estimation and Spectral Analysis. In: Freeden W, Nashed MZ, Sonar T (eds) Handbook of Geomathematics. Springer, Berlin. doi 10.1007/978-3-642-01546-5_30

Sneeuw N (1994) Global spherical harmonic analysis by least-squares and numerical quadrature methods in historical perspective. Geophysical Journal International, 118: 707-716. doi 10.1111/j.1365-246X.1994.tb03995.x

Tscherning CC, Rapp RH (1974) Closed Covariance Expressions for Gravity Anomalies, Geoid Undulations, and Deflections of the Vertical Implied by Anomaly Degree Variance Models. Ohio State University, Department of Geodetic Science and Surveying, OSU Report No. 208

Zingerle P (2015) Hochauflösende Schwerefeldmodellierung auf Basis von dünnbesetzten Systemen [Master thesis, Technical University of University]. MediaTUM. https://mediatum.ub.tum.de/1363992

Acknowledgments

This work wouldn't have been possible without the help of many people. First and foremost, I would like to express my cordial gratitude to my supervisor, Prof. Roland Pail, who has continiously suported me with his knowledge, experience, great ideas and a lot of patience especially in the final stages of the thesis. Knowing that it is not common to have a supervisor who always has an open ear for questions and discussions, I feel very grateful to have been in this privileged position.

Next, I would like to mention all my colleagues at the IAPG; working with you has always been a pleasure. In particular, I would like to highlight Dr. Martin Willberg for all the magnificent collaborations over the past six years and Petro Abrykosov for editing my English. I would also like to give special thanks to Dr. Thomas Gruber and my former colleague Dr. Thomas Fecher for being great tutors and for introducing me to the field of global gravity field modelling.

Further, I'd like to thank all my other co-authors and project partners: Dr. Mirko Scheinert and Theresa Schaller from TU Dresden for supporting me with the Antarctic data and when it came to geophysical questions. Xiaopeng Li and the NGS for providing me the preprocessed GRAV-D data and thereby giving the inspiration to the core idea of this thesis. Xanthi Oikinomidou for her help with the oceanographic validations of the XGM2019e model.

On the institutional side I would like to acknowledge the Leibniz Centre for Supercomputing (LRZ, project pn98be, "EGRAV") for providing me with the needed computational resources and for supporting me with prompt help in numerous technical questions. The German Research Foundation (DFG, grants PA 1543/18-1 and SCHE 1426/24-1 within SPP 1158) for funding the "AntGrav" project and the research on the PE-LSC method. The European Space Agency (ESA, contract 18308/04/NL/MM of the "GOCE-HPF" project) for supporting my studies on the XGM2019e model. The International Graduate School of Science and Engineering (IGSSE, project 13.07, "GOCE4GOSE") of the Technical University of Munich for funding travels and consumables.

Last but not least, I would also like to thank all my family and friends who accompanied me through several difficult moments in the last years and backed me whenever it was needed - I'll always be grateful for that.

Appendix A

Publication P-1: Evaluation of terrestrial and airborne gravity data over Antarctica – a generic approach

Reference

Zingerle P., Pail R., Scheinert M. and Schaller T. (2019) Evaluation of terrestrial and airborne gravity data over Antarctica – a generic approach. Journal of Geodetic Science, 9(1), 29-40. doi 10.1515/jogs-2019-0004

Copyright

Abstract

The AntGrav project, funded by the German Research Foundation (DFG) has the main objective to homogenize and optimize Antarctic gravity field information. Within this project an evaluation procedure is needed to inspect all different kind of gravity field surveys available in Antarctica. In this paper a suitable methodology is proposed. We present an approach for fast 3D gravity point data reduction in different spectral bands. This is achieved through pre-calculating a fine 3D mesh of synthesized gravity functionals over the entirety of the Antarctic continent, for which two different global models are used: the combined satellite model GOCO05s for the longwavelength part, and the topographic model Earth2014 for the shorter wavelengths. To maximize the applicability separate meshes are calculated for different spectral bands in order to specifically reduce a certain band or a selected combination. All meshes are calculated for gravity anomalies as well as gravity disturbances. Utilizing these meshes, synthesized gravity data at arbitrary positions is computed by conventional 3D interpolation methods (e.g. linear, cubic or spline). It is shown that the applied approach can reach a worstcase interpolation error of less than 1 mGal. Evaluation results are presented for the AntGG grid and exemplary for the in-situ measurements of the AGAP and BAS-LAND campaigns. While general properties, large-scale errors and systematic effects can usually be detected, small-scale errors (e.g. of single points) are mostly untraceable due to the uncertainties within the topographic model.

Declaration of own contribution
(PZ: Philipp Zingerle; RP: Roland Pail; MS: Mirko Scheinert; TS: Theresa Schaller)
The study was proposed by RP and MS. Ideas and the study design were developed by PZ with support of RP, MS and TS. PZ calculated the SATOP reduction model, implemented the fast 3D synthesis method and performed evaluations of Antarctic gravity field observations based on various datasets. TS facilitated the validations by preparing the input datasets and by helping to assess their properties. PZ created the figures and formulated the text with suggestions of RP and MS.

The overall own contribution of PZ for *P-I* is estimated at 90 %, which is the average value of the percentage values estimated for the five criteria listed in the table below (Tab. P.1).

Criteria	**Estimated own contribution**
Computation and results	90 %
Ideas and study design	80 %
Analysis and interpretation	90 %
Text	90 %
Figures and tables	100 %
Total	**90 %**

Tab. P.1 – Criteria and estimated contribution share of Philipp Zingerle for *P-1*

Confirmation by the authors

We hereby confirm the correctness of the declaration of own contribution (of 90 % for Philipp Zingerle) for the publication

Zingerle P., Pail R., Scheinert M. and Schaller T. (2019) Evaluation of terrestrial and airborne gravity data over Antarctica – a generic approach. Journal of Geodetic Science, 9(1), 29-40. doi 10.1515/jogs-2019-0004

Roland Pail

Institute of Astronomical and Physical Geodesy, Technical University of Munich, Germany

Signature: Roland Pail Date: 23.09.2021

Mirko Scheinert

Institut für Planetare Geodäsie, Technische Universität Dresden, Germany

Signature: Mirko Scheinert Date: 05. 10. 2021

Theresa Schaller

Institut für Planetare Geodäsie, Technische Universität Dresden, Germany

Signature: Schaller Date: 05. 10. 2021

Research Article **Open Access**

P. Zingerle*, R. Pail, M. Scheinert, and T. Schaller

Evaluation of terrestrial and airborne gravity data over Antarctica – a generic approach

https://doi.org/10.1515/jogs-2019-0004
Received January 25, 2019; accepted April 30, 2019

Abstract: The AntGrav project, funded by the German Research Foundation (DFG) has the main objective to homogenize and optimize Antarctic gravity field information. Within this project an evaluation procedure is needed to inspect all different kind of gravity field surveys available in Antarctica. In this paper a suitable methodology is proposed.

We present an approach for fast 3D gravity point data reduction in different spectral bands. This is achieved through pre-calculating a fine 3D mesh of synthesized gravity functionals over the entirety of the Antarctic continent, for which two different global models are used: the combined satellite model GOCO05s for the long-wavelength part, and the topographic model Earth2014 for the shorter wavelengths. To maximize the applicability separate meshes are calculated for different spectral bands in order to specifically reduce a certain band or a selected combination. All meshes are calculated for gravity anomalies as well as gravity disturbances. Utilizing these meshes, synthesized gravity data at arbitrary positions is computed by conventional 3D interpolation methods (e.g. linear, cubic or spline).

It is shown that the applied approach can reach a worst-case interpolation error of less than 1 mGal. Evaluation results are presented for the AntGG grid and exemplary for the in-situ measurements of the AGAP and BAS-LAND campaigns. While general properties, large-scale errors and systematic effects can usually be detected, small-scale errors (e.g. of single points) are mostly untraceable due to the uncertainties within the topographic model.

Keywords: Antarctica, combined gravity field modelling, fast synthesis method, global geopotential model, gravity field

***Corresponding Author: P. Zingerle:** Institute of Astronomical and Physical Geodesy, Technical University of Munich, Germany, E-mail: zingerle@tum.de

M. Scheinert, T. Schaller: Institut für Planetare Geodäsie, Technische Universität Dresden, Germany

1 Introduction

An accurate and consistent knowledge of the gravity field in Antarctica is crucial for many geophysical applications and thus also for a better understanding of the geological structures of the continent. In order to achieve this objective, Scheinert et al. (2016) presented a first continent-wide collection of gravity field observations within the AntGG grid, containing a majority of all accessible measurements. In concrete, more than 13 million observations collected over the past decades covering about 73% of the continent are included in the AntGG processing. These measurement campaigns have been performed by many different countries using different measurement techniques and analysis methods. Thus, the entire collection of gravity observations is largely heterogeneous regarding:

- **Spatial distribution:** different campaigns may show different spatial distribution patterns. As an example, airborne campaigns usually have a high resolution in the along-track direction, while the cross-track resolution mostly depends on the airborne campaign goals (e.g. grid- vs. star-shaped). Terrestrial observations on the other hand may show no clear pattern at all including larger data gaps or even having just a one-dimensional extent along a profile line.
- **Observation type:** as the earliest campaigns date back to a time before global navigation satellite systems became available, the observed data had to be processed in terms of gravity anomalies using physical heights, as ellipsoidal heights were more difficult to obtain. In contrast, modern airborne campaigns observations are usually processed using gravity disturbances and ellipsoidal heights.
- **Post-processing:** depending on the campaign, the available input datasets may have already been post-processed beforehand, meaning that one may be unable to reconstruct the raw observation data. This implies that inconsistencies made in the post-processing chains of different surveys are difficult to undo, especially since the metadata describing the processing strategies is often missing. The most common post-

processing steps of the observations in the datasets are gridding and reduction to a reference surface (e.g. ellipsoids, offset-ellipsoids).
- **Accuracy:** the observation accuracy obtained by different campaigns can also be heterogeneous, as it depends on instrumentation, transportation of the instruments and the possibly inaccurate postprocessing.
- **Spectral content:** regarding the spectral content of the data one also might observe varying behavior, especially in the higher frequency domain. This can be attributed to multiple reasons. E.g., when dealing with airborne gravity observation one must keep in mind that due to flight dynamics and limitations of the instruments the raw measurement data needs to be lowpass filtered down to several km wavelength. Also, it is possible that during further postprocessing (e.g. gridding, collocation) the data may have already been spectrally limited.

As the aim of the AntGG grid is primarily to represent the available gravity field information, no in-depth evaluation or homogenization of these datasets has been performed so far. To tackle this circumstance and the fact that new gravity data is available by now the German Research Foundation (DFG) funded the so-called AntGrav project as part of which this study is performed. The main purpose of this project is to compute an optimally combined gravity model from satellite and ground data for the Antarctic continent. From this dataset a model of the bedrock topography shall then be derived by means of geophysical inversion.

For the targeted combination of gravity datasets reliable knowledge about the actual accuracy and spectral content of individual measurement campaigns is required as well as the possibility to detect larger outliers and systematic effects. Thus, an adequate evaluation method is needed, which is applicable on the wide range of different survey campaigns available in Antarctica. The method shall therefore meet the following criteria:
- **Consistency:** the method shall be able to inspect all different kinds of campaigns independently of their attributes and with unvarying evaluation quality.
- **Independency:** the evaluation method shall not be correlated to the actual measurements (subject of examination) allowing to make assessments about their correctness.
- **Efficiency:** as millions of datapoints need to be evaluated, the evaluation method shall be fast and efficient w.r.t. computation time and memory consumption.
- **Correctness:** obviously, the evaluation method shall be as correct as possible by itself.

The technique presented in this study will attempt to properly fulfill all these demands. To accomplish this, we make use of:
- **a preexisting gravity field model:** reducing gravity field information from a preexisting model and studying residuals is a feasible (and always applicable) method to examine gravity field observations. Such a gravity field model must be as independent and as correct as possible as well as offer the highest possible resolution. Further, it shall be given in the spectral domain in order to address the consistent adaption on different observation types (e.g. gravity anomalies and disturbances).
- **a fast synthesis method:** having gravity field information in the spectral domain, the transformation to the spatial domain (synthesis) is a time-consuming task when dealing with large scattered point datasets. To reduce computation time, it is proposed to apply a two-step synthesis method, where in a first step gravity field information is pre-calculated on regular geographic 3D grids and then in a second step interpolated on the individual scattered point positions.

In section 2 the used gravity field model SATOP1 is described in more detail, as it is generated specifically for the purpose of evaluation. Section 3 describes the synthesis method with all its characteristics. Subsequently, within section 4 the new method is applied to examine the AntGG grid. Finally, in section 5 examination results for two selected in-situ measurement campaigns are presented.

2 The SATOP1 gravity field model

As explained in section 1, an independent gravity field model shall constitute the basis of the evaluation method. To be truly independent, the model shall not contain any of the observations being evaluated. In the case of Antarctica, this means that no terrestrial gravity field information at all is to be included, because observations over Antarctica are generally sparse (and barely overlapping) and are target of the inspection itself.

Thus, the only available data sources left are satellite models as well as topographic forward models. Both sources show complementary spectral behavior: satellite models are very accurate in the low frequency domain but worsen with increasing resolution due to signal attenua-

tion of the gravity field with altitude. Topographic models, on the other hand, do not include lateral density variations or isostatic effects, which leads to a decreased accuracy especially in the low frequency domain (cf. Fig. 1). Because of these complementary characteristics it is logical to use both data sources in a combined (SA)tellite-(TOP)ographic model (hence called SATOP1).

For the combined model SATOP1, the satellite model GOCO05s (Mayer-Gürr, 2015) and the EARTH2014 (Rexer et al., 2016) topographic model are chosen (up to d/o 5480), as both models include actual data sources and are based on mature processing techniques. As an optimal combination method, a variance-based stacking on normal equation level in the spectral domain is applied, similarly to the model SatGravRET14 (Hirt et al., 2016). This strategy can be interpreted as regularization of the satellite system with the topographic model as a-priori information. Mathematically, the combination can be described as:

$$\underline{x}_{SATOP1} = C_{SATOP1}\left(\underline{q}_{GOCO05s} + \mathrm{diag}\left(\underline{var}_{EARTH}^{-1}\right)\underline{x}_{EARTH}\right) \tag{1}$$

with

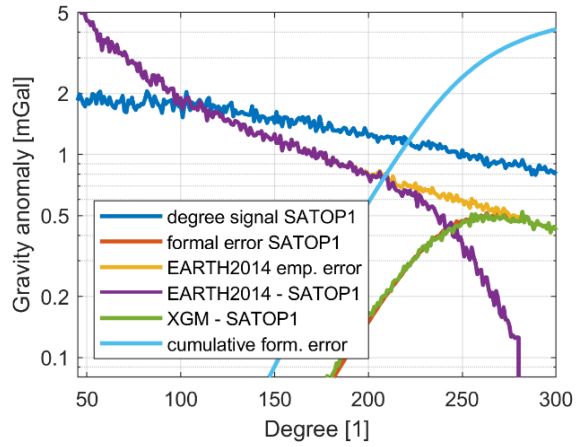
$$C_{SATOP1} = \left[N_{GOCO05s} + \mathrm{diag}\left(\underline{var}_{EARTH}^{-1}\right)\right]^{-1} \tag{2}$$

where $N_{GOCO05s}$ is the normal equation matrix and $\underline{q}_{GOCO05s}$ is the right-hand side of the GOCO05s model, $\underline{x}_{EARTH}$ represents the coefficients and $\underline{var}_{EARTH}$ the variances of the EARTH2014 model, and finally, $\underline{x}_{SATOP1}$ describes the resulting SATOP1 model coefficients and C_{SATOP1} the corresponding covariance matrix.

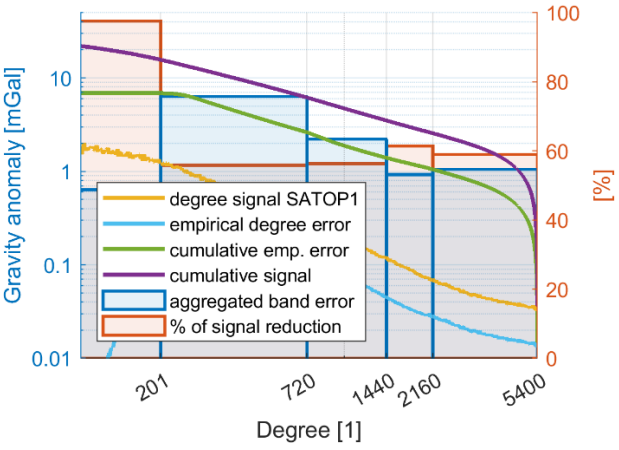
Within the processing of SATOP1, the choice of a realistic variance for EARTH2014 ($\underline{var}_{EARTH}$) is crucial, as it controls the spectral transition from the satellite to the topographic model. Since topographic forward modelling as used in EARTH2014 does not provide any statistical measures for its result, one has to find a different method to derive model variances. For SATOP1 it is assumed that those variances are mostly degree-dependent and hence can be simplified to degree-variances (no order dependency). The degree-variances are then derived empirically from the comparison to another independent model. In this case, XGM2016 (Pail et al., 2018) is used:

$$var_l^{EARTH} \sim \sum_{m=-l}^{l} \frac{\left(c_{lm}^{EARTH} - c_{lm}^{XGM}\right)^2}{2l+1} \tag{3}$$

and c_{lm}^{XGM} denote the spectral coefficients of the EARTH2014 resp. XGM2016 models, with degree l and order m (negative orders indicate sine-coefficients). The spectral characteristics of the models and the empirical variances are visualized in Fig. 1a in form of gravity anomaly degree-RMS.



(a)



(b)

Figure 1: Spectral characteristics of SATOP1 (in spheroidal-harmonics at 4 km altitude): (a) degree-RMS of the satellite band in terms of gravity anomalies. Dark blue: signal of SATOP1, orange: formal error of SATOP1, yellow: empirical error EARTH2014 (for regularization), violet: difference EARTH2014-SATOP1, green: difference XGM2016-SATOP1, light blue: SATOP1 cumulative formal error. (b) SATOP1 error estimate in terms of gravity anomalies. Yellow: signal of SATOP1, light blue: empirical error of SATOP1 (XGM2016-SATOP1), green: SATOP1 cumulative empirical error, violet: cumulative SATOP1 signal (unreduced), dark blue bars: aggregated SATOP1 errors within depicted bands, orange bars: percentage of SATOP1 band signal reduction (refers to right axis).

In contrast to preceding strategies (Hirt et al., 2016) the combination of SATOP1 is performed purely in a spheroidal-harmonic domain adopting the GRS80 reference system (Moritz, 1980, 2000) - all items in Eqs. (1), (2) and (3) are given w.r.t. this domain. To accomplish this,

Jekeli's transform (Jekeli, 1981) is applied in a preceding step to transform all input quantities into the spheroidal-harmonic domain. Performing the degree-variance-based combination within this domain is preferred, since the coefficient smearing (=correlation) between adjacent degrees (as implied by Jekeli's transform) on (or near) spheroidal surfaces is reduced. A second and perhaps more important advantage is, that due to the same reason, spectral band cuttings in the spheroidal domain do not lead to truncation artefacts as e.g. seen in EGM2008 (Pavlis et al. 2012) when omitting its highest degrees (e.g. d/o 2160 upwards).

Thus, the investigation of different spectral bands in the spatial domain (on or near the spheroid) is reasonable. For the following synthesis six (spheroidal) spectral bands ([2-200], [201-359], [360-719], [720-1439], [1440-2159], [2160-5480]) are selected. The first limit is chosen, as it represents an empirical upper boundary for a (nearly) satellite-only solution, the latter limits correspond to spatial resolutions commonly used in global gravity field modelling. To be compatible with the spherical harmonic definition of different gravity field functionals, the specific bands are truncated in the spheroidal harmonic domain and then transformed into the spherical harmonic domain (Jekeli, 1981) for an ordinary synthesis.

Error estimations of the different spectral bands (commission errors) are shown in Fig. 1b: they are derived in the same way as the formal errors of the EARTH2014 model by comparing SATOP1 against XGM2016 extended by EGM2008 from d/o 720 up to 2159. Due to lack of independent comparison data, errors above d/o 2159 are extrapolated. As expected, the largest contribution to the commission error (at a simulated flight altitude of 4 km) occurs in the lowest topographic band (between d/o 201 and 719). While the estimated percentage of signal reduction stays widely stable (at about 60%), the gravity field signal, and thus also the commission error, is attenuated with increasing d/o.

3 The AGRID3D synthesis method

Performing a spherical harmonic synthesis up to d/o 5480 is a CPU-intensive task, especially when dealing with scattered points, as Legendre polynomials need to be recalculated for every single observation. As pointed out in the introduction, a two-step synthesis may be better suited for the purpose of evaluating big data volumes.

Firstly, it exploits the advantage of regular grid synthesis, where Legendre polynomials are only evaluated individually for different latitudes, but not for every single point (Sneeuw 1994). Thus, a geographic grid has to be chosen as a base grid, because polar stereographic grids, for instance, would feature varying latitudes for each grid point.

Secondly, due to the fact that the minimal wavelength is known to be limited (d/o 5480 corresponding to 2' spatial resolution), a 2-times oversampled grid is also known to give good interpolation results, as it fulfills the Nyquist-Shannon sampling theorem (Shannon 1949). Having a pre-calculated grid at hand, the computation time of the synthesis problem scales linearly w.r.t. the number of points being interpolated and is in contrast to the rigorous synthesis (nearly) independent of the maximum d/o (i.e. the number of support points).

Even though other methods exist for the upward continuation of the gravity field besides 3D interpolation (e.g., Rapp (1997), Ivanov et al. (2018)), this method is favourable as it is simple, robust, very fast, and the accuracy of the result is independent of the distance to the surface (cf. subsection 3.2).

3.1 Properties and components of AGRID3D

The first step of the two-step synthesis is the calculation of the regular 3D grid. The dimension of the grid has to be chosen in a way which ensures that every point to be validated is located safely within the grid boundaries. The final AGRID3D properties can be summarized as follows (cf. Fig. 2):

- **Grid extent:** in longitudinal direction the grid ranges from 0° to 360°, in latitudinal direction from -60° to -90°, and in vertical direction from -1 km up to 6 km. On every periodic limit, the grid was extended by 5 cells to avoid increasing interpolation errors near the boundaries.
- **Grid resolution:** in order to ensure that the spherical grid is oversampled a least twice, a latitudinal sampling of 1' is chosen. Adapting to meridian convergence, the longitudinal sampling is reduced to 2' at -60° lat. and 4' from ~ -75.4° southwards. This reduces the grid size and hence saves memory space. For the vertical component, a constant spacing of 200 m is selected empirically.
- **Spectral bands:** to enable also a spectral evaluation, 6 distinct grids are calculated for the different spectral bands as defined in section 2.
- **Gravity field functionals:** as all observation data of the AntGG dataset (Scheinert et al. 2016) are given either in the form of gravity anomalies or gravity distur-

bances, dedicated sets of grids are calculated for these two functionals.

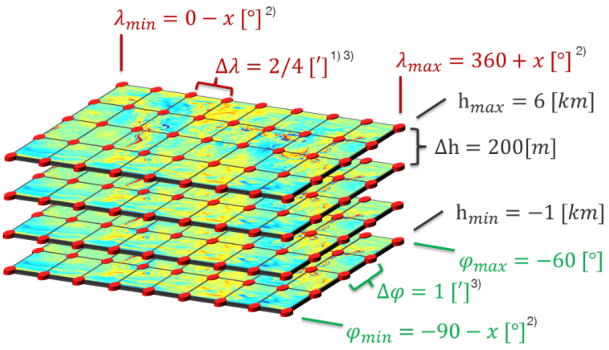


Figure 2: Visualization of the grid properties of the AGRID3D dataset. Notes: 1. Meridian convergence is considered. 2. Limits are extended to simulated periodicity. 3. 2x Nyquist-frequency (for d/o 5400) is guaranteed.

One complete 3D-grid (from $-60°$ to $-90°$) has therefore a memory size of ~ 4.1 GB (in double precision). This marks also the recommended minimum RAM requirement of a system to perform an interpolation based on AGRID3D. Since different grids are needed for the spectral bands (6x) and gravity field functional (2x), the overall size of the complete AGRID3D dataset is ~ 49 GB.

The computation of the grid itself is performed on resources of the Leibniz Computing Center (*LRZ*). For the stable calculation of the Legendre polynomials, methods based on extended numbers (Fukushima, 2012) are used.

3.2 Interpolation performance

Based on the resulting AGRID3D dataset, the gravity values at arbitrary points within the extent of the grid are obtained by interpolation. Within this study, a common (tri-)cubic interpolation approach (Lekien and Marsden, 2005) is used, as it shows good performance (cf. Fig. 4) while keeping memory requirements and calculation costs low. Before those interpolated points can be used as evaluation reference, it must be verified that the interpolation method is accurate and in agreement with a rigorous direct scattered point synthesis.

To prove the correctness of the interpolation method itself (as required for the purpose of evaluation), it is proposed to simulate a worst-case scenario, where all interpolated points are located at the maximum distance centered between the support points (cf. Fig. 3a). This leads to a validation grid which is shifted by half the grid spacing in all three directions relatively to AGRID3D. This validation grid is then evaluated with two different strategies: first, the SATOP1 model is synthesized rigorously to get reference values (using all bands, d/o 2-5480). Second, the interpolation based on AGRID3D is applied to obtain homologous values. Latter values are not only obtained using the cubic approach, but also involving the widely used linear and spline interpolation methods, enabling comparisons between the different techniques. As the first strategy is assumed to be error-free, differences in the values of both methods represent the interpolation error.

Statistical evaluations of those differences for different height layers and interpolation approaches are shown in Fig. 4, and an example for the spatial distribution of the cubic interpolation errors on a near-surface height layer of 100 m is shown in Fig. 3b for the whole spectrum (d/o 2-5480). Generally, a strong correlation of the interpolation error with high-frequency signals (e.g. due to mountains) is evident. This implies that the interpolation error is also strongly correlated with height, since upward continuation of the gravity field leads to an attenuation of higher frequencies and thus reduced interpolation errors. Empirically, this fact can also be seen in the statistics (cf. Fig. 4). Although the spline-interpolation method shows the best interpolation accuracy, the cubic approach is more favorable, as the actual implementation requires less memory and the performance is still within a reasonable limit for the purpose of evaluation (as uncertainties within the SATOP1 model are expected to exceed the interpolation error).

4 AntGG inspection

In a first application, the evaluation method is used to inspect the AntGG (Scheinert et al., 2016) grid. The AntGG dataset consists of a polar stereographic regular surface grid (ice layer including) of gravity anomalies with a spacing of 10 km. The AntGG gravity anomalies are further spectrally limited by an average operator to minimize aliasing effects due to the grid sampling.

4.1 Reducing the SATOP1 model

The SATOP1 model is interpolated on the AntGG grid points (using the cubic AGRID3D interpolation, see section 3) for increasing spectral content (using the 6 bands

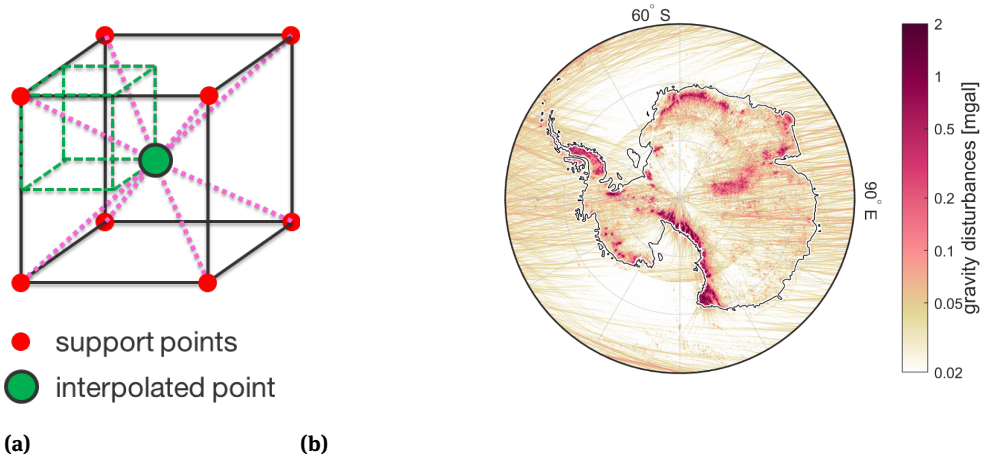


Figure 3: (a) Visualization of the validation concept. Interpolation grid points (shown in green) lie centered between support grid points (shown in red). This is clarified by dashed green lines. The chosen setup leads to equal maximum distances of the interpolation point to all surrounding support points (indicated by dashed purple lines). (b) Interpolation errors (in terms of gravity disturbances) for the cubic interpolation approach on a 100 m height layer for the complete SATOP1 model (including all spectral bands).

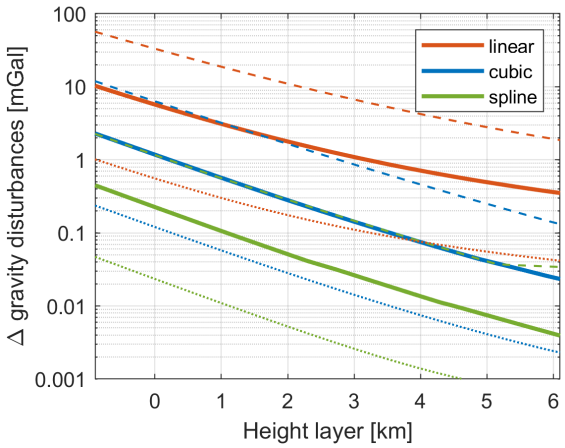


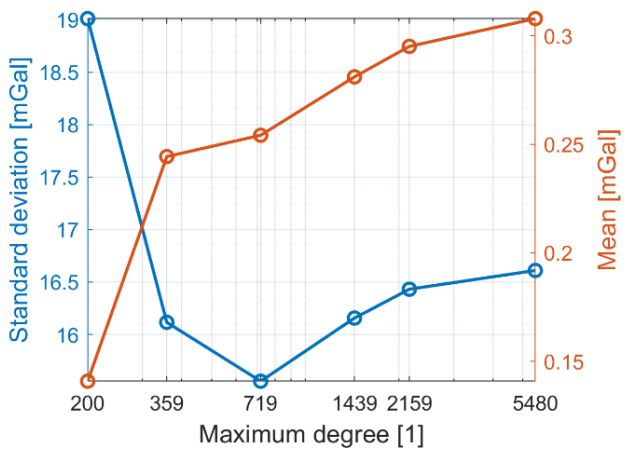
Figure 4: Statistical characteristics of the AGRID3D worst-case scenario errors for different heights and interpolation strategies. Shown are RMS (dotted), absolute maximum (dashed) and absolute 99.9% quantile (continuous) for the linear (red), cubic (blue) and spline (green) interpolation method. Using the cubic approach, on 100 m height, 99.9% of all values are within ±2 mGal, on 2 km all values are within this range.

as defined in section 2). The resulting grids are then subtracted from the original AntGG grid to obtain residuals. As the sources of the gravity data and the SATOP1 model are assumed to be mostly independent, it is reasonable to use these residuals for the purpose of cross-validation: if they are small in a certain region, this is a strong indicator that both datasets are correct (meaning that they are error-free and encompass the same spectral content). Vice versa, when residuals become larger, it should be kept in mind that in principle no statement of the error source can be made. The difference could either result from errors in the real observations, or from inadequacies within the SATOP1 model. However, the analysis of residuals may still help finding suspicious data, while simultaneously reliably identifying trustworthy data. Further, these residuals can also be used to estimate the spectral content of AntGG by inspecting statistical properties for different band reductions (cf. Fig. 5a).

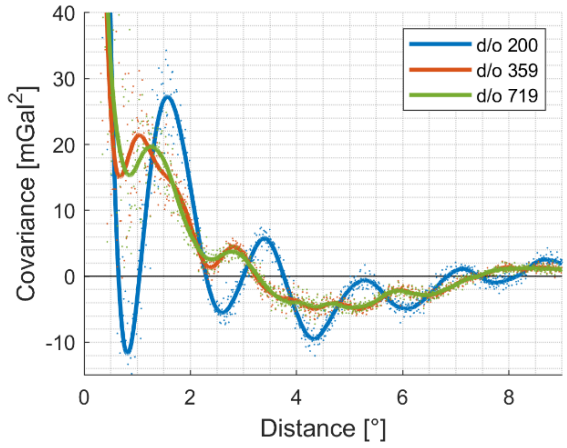
Looking at the remaining standard deviation of the reductions (cf. Fig. 5a), one can identify a minimum when using SATOP1 up to d/o 719. Thus, it is deduced that the spectral content of the AntGG grid is limited at approx. this resolution. In fact, the AntGG grid is low-pass filtered to reduce aliasing effects on the 10 km spaced grid. Hence, the d/o 719 reduction is concluded to fit best to the AntGG data and is therefore chosen for further investigations. The empirical covariances (cf. subsection 5.2) also support this choice, although some longer distance patterns remain due to discrepancies in the models (e.g. isostatic effects or incomplete terrain data in SATOP1, errors in the AntGG grid).

4.2 AntGG validation

Reductions using SATOP1 also include the error of the topographic EARTH2014 model and therefore cannot be directly used to validate other datasets. Nevertheless, SATOP1 also consists of the high-accuracy satellite model part up to about d/o 200 (cf. Fig. 1).

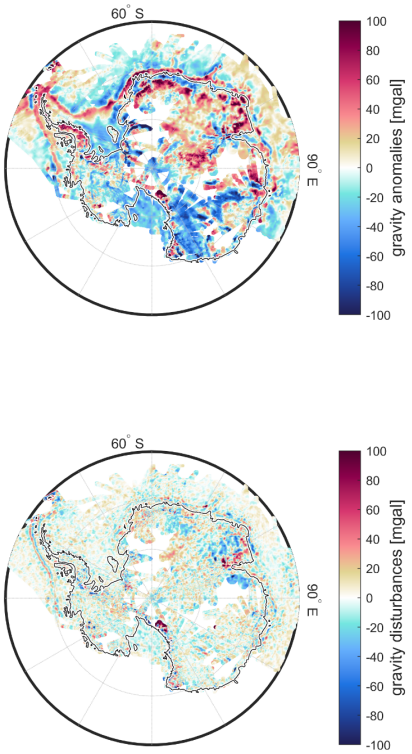


(a)



(b)

Figure 5: Statistical properties of the AntGG-SATOP1 inspection: (a) Standard deviation (blue) and mean (red) for varying maximum degree of the SATOP1 model reduction from the AntGG grid. (b) Empirical covariances depending on the spherical distance for different maximum degree reductions (200-359-719). Colored dots indicate the scattering of the smoothed covariance functions (continuous lines).



(a)

(b)

Figure 6: (a) The AntGG grid as is with white areas where no data is available. (b) Difference to SATOP1, limited to d/o 719.

This issue can be exploited when applying a low-pass filtering to the data which attenuates gravity field information above d/o 200. As the spectral truncation of the signal in the frequency-domain is obviously not (easily) possible (since the data is only available in the spatial domain), Gaussian filtering (Jekeli, 1981) is applied instead to widely reduce the influence of the topographic model.

The spectral strength of the Gaussian low-pass filtering is determined by the choice of the full-width-half-maximum (FWHM) distance (cf. Fig. 7a). For a reliable reduction of signals beyond d/o 200 a distance of 80 km is chosen. For a less reliable but higher resolution result a second variant is calculated with 40 km FWHM. As an example, the 80 km FWHM filtered reduction is shown in Fig. 7b: signals within this plot arise with high certainty from errors in the AntGG datasets.

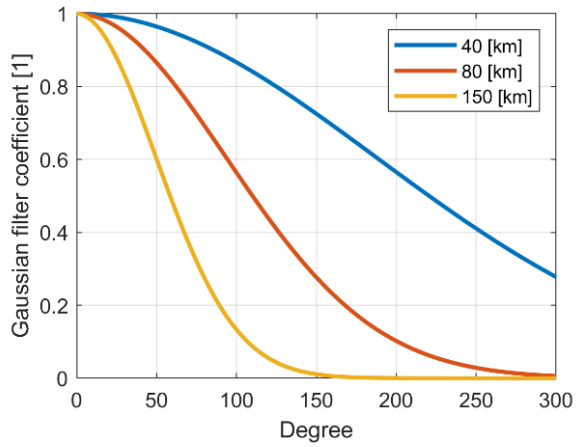
Within the Gaussian filtering, the occurring filter uncertainties near data gaps are considered by weighting down the signal in those regions. The filter uncertainty itself is estimated through comparison of the complete with the incomplete grid filtering result of SATOP1 data.

In a next step the low-pass filtered reductions (Δf) can be further processed to derive a final error measure. To achieve this, local standard deviations ($\bar{s}$) are derived applying weighted integrals once again using a Gaussian kernel (W_G). The appropriate formula can be summarized as (Jekeli, 1981):

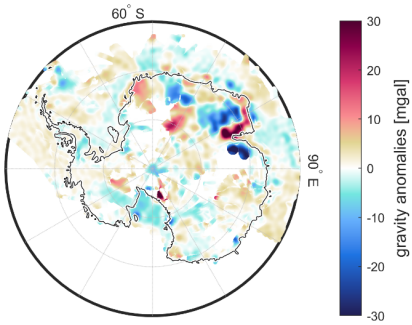
$$\bar{s}\,(\theta,\lambda) := \sqrt{\left(W_G \star \Delta f^2\right)(\theta,\lambda)} = \sqrt{\sum_{i=1}^{n} W_G\left(\psi_i, d_{FWHM}\right)\Delta f_i^2} \tag{4}$$

with the spherical distance:

$$\cos\psi_i = \cos\theta\cos\theta_i + \sin\theta\sin\theta_i\cos\left(\lambda - \lambda_i\right) \tag{5}$$



(a)



(b)

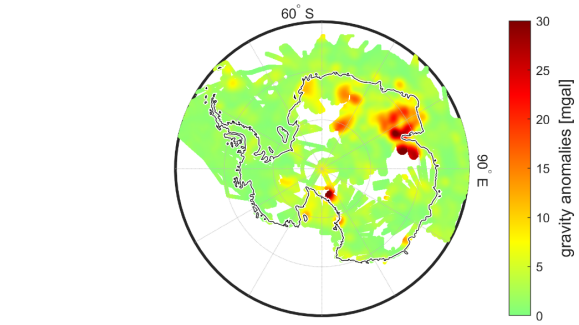
Figure 7: (a) The impact of different FWHM filter distances (40 km blue, 80 km red, 150 km yellow) in the spectral domain, depending on the spherical degree. (b) Difference AntGG-SATOP, up to d/o 719 (cf. Fig. 6b), 80 km FWHM Gaussian filtered.

As a suitable FWHM distance (d_{FWHM}) for the Gaussian kernel the same as for the low-pass filtering is chosen. The final results are shown in Fig. 8a (for 80 km FHWM) and Fig. 8b (for 40 km). The latter may still contain some signals from the topographic model but also delivers a higher resolution than the 80 km version.

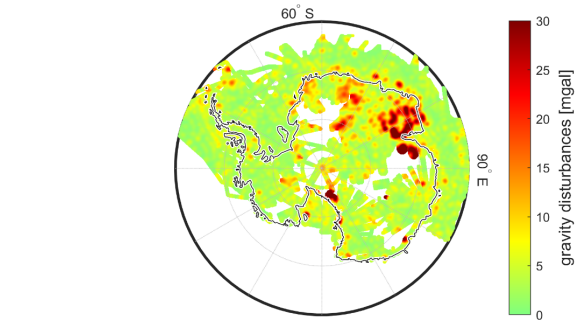
Having these error estimates at hand, one may now take an in-depth look into suspicious regions that show larger deviations by inspecting the underlying gravity observation campaigns (see section 5)

5 In-situ evaluations

As mentioned in section 1, the presented examination strategy is not only meant for application to already combined and gridded datasets such as AntGG, but also to in-



(a)



(b)

Figure 8: AntGG error estimate (standard deviation) based on (a) 80 km resp. (b) 40 km FWHM low-pass filtering.

situ measurements from various gravity campaigns (as the ones used in the AntGG grid). This has the advantage that possible errors/outliers may be identified on observation level and therefore eliminated/corrected without influencing other measurements. To showcase the application, two different campaign types are selected: firstly, a larger recent airborne survey campaign named AGAP (Ferraccioli et al., 2011) and secondly, an older ground measurement campaign of the Antarctic Peninsula, named BAS-LAND (Renner et al. 1985).

5.1 AGAP dataset

Within the Antarctica's Gamburtsev Province (AGAP) project, a comprehensive airborne gravity map of the Gamburtsev Subglacial Mountains was recorded during the 2008/09 field season with a line spacing of 5 km and tie-line interval of ~33 km at a mean altitude of ~4 km. The whole dataset consists of over 2 million datapoints of gravity disturbances.

Performing the same reductions as in subsection 3.1 (but now in-situ, on the actual point of measurement),

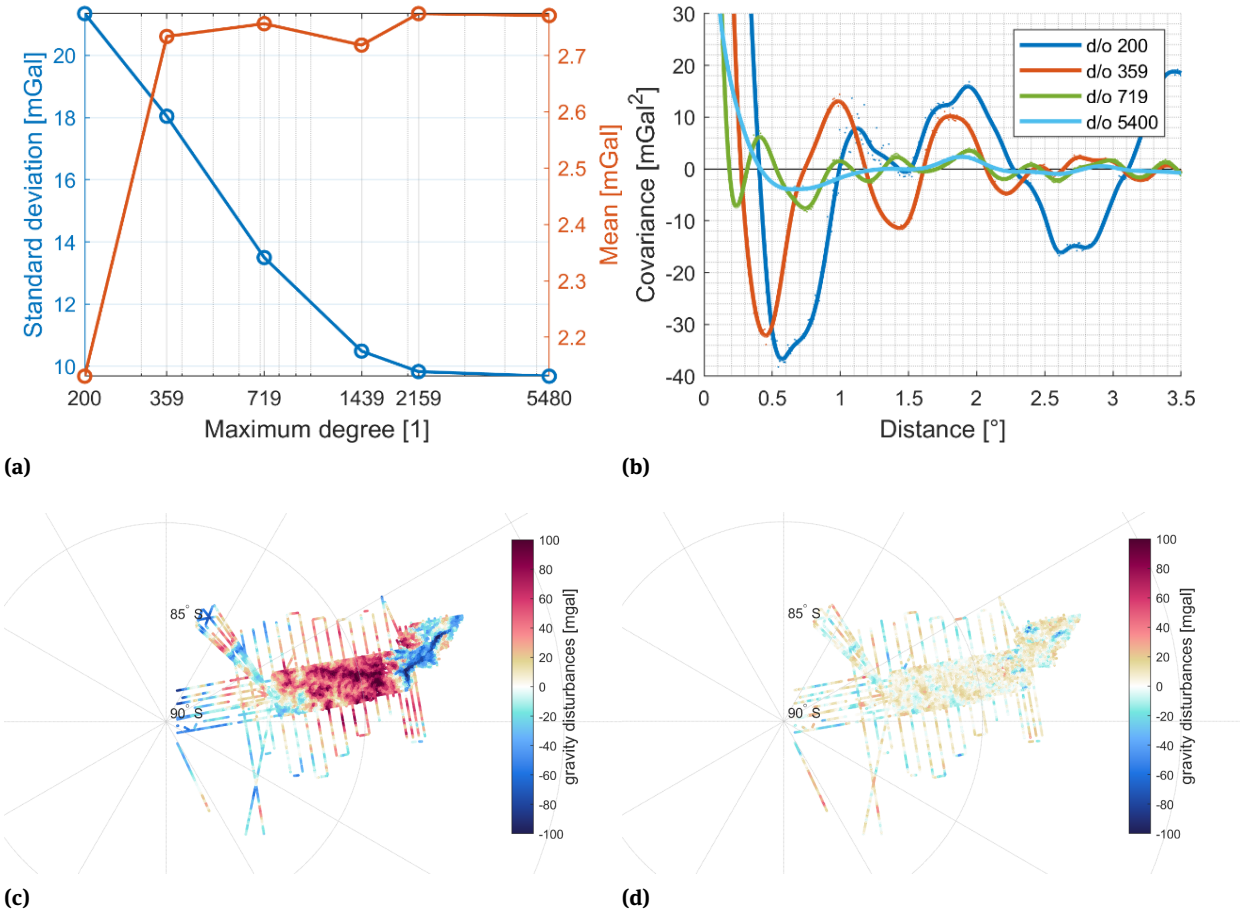


Figure 9: (a) Standard deviation (blue) and mean (red) for varying maximum degree of the SATOP1 model reduction from the AGAP observations. (b) Empirical covariances depending on the spherical distance for different maximum degree reductions (200-359-719-5400). (c) The AGAP observations as provided. (d) Differences to SATOP1 up to d/o 5400.

the results are shown in Figs. 9a-d: Looking at the statistics (Figs. 9a-b), the reduction performance can be considered outstanding while keeping in mind that from d/o 200 upward only topographic forward modelling is used. This also suggests a lack of higher frequency isostasy (beyond d/o 200) in this region (maybe due to the glacial coating). The covariances (Fig. 9b) show an expected behavior: the steepness towards the first minimum increases and the oscillation around 0 decreases with a higher reduction degree.

A visual inspection of the dataset (Figs. 9c-d) reveals no obvious errors or outliers. Also, the detected offset of 2 mGal (Fig. 9a) meets the expectations. Thus, it can be summarized that the AGAP data is most probably very trustworthy. Further, it can be seen (Fig. 9a and visually in 9c) that the spectral content is not significantly limited below d/o 2159 (e.g. due to filtering of airborne data).

It is worth mentioning that due to the use of AGRID3D in combination with the cubic interpolation approach, the reduction of the 2 million observation points can be calculated within a second on a standard PC.

5.2 BAS-LAND

In contrast to the AGAP campaign, the BAS-LAND dataset contains ground measurements of the Antarctic Peninsula (cf. Fig. 10c) from the pre-GNSS era. This leads to an introduction of some new error sources.

Firstly, the geolocalization of points is probably less accurate especially in the vertical direction, as this coordinate refers to an orthometric height. Consequently, the observations are given in terms of gravity anomalies. For the reduction with AGRID3D, a (quasi-)geoid model is therefore needed to convert the physical coordinates

to geometric ones. To maintain consistency, this (quasi-)geoid is calculated correspondingly also from the SATOP1 model, introducing on top of all other effects its own (height-)uncertainty.

Further, terrestrial observations in general include more higher-frequency content than airborne measurements, as no upward continuation is embedded. For the airborne data it can be assumed that due to the common flight altitudes not much signal above d/o 5400 is left. This may not be true for terrestrial measurements, especially in mountainous regions (as the Antarctic Peninsula). So, even if the SATOP1 model were totally correct, there would still be gravity information above d/o 5400 that is not covered by the model.

Another aspect to consider, altough the effect is considered to be minor, is the possible temporal variability of the gravity field signal, as the BAS-LAND data was recorded about 30 years prior to the SATOP1 data. Variable ice-thicknesses may introduce changes in the gravity field above mGal-level, as 25 m ice thickness corresponds to roughly 1 mGal Bouguer gravity (without considering glacial isostatic rebound, variable ice densities etc.).

Finally, the overall consistency of the data may be not as good as from airborne campaigns since e.g. uncertainties due to drift effects of relative gravimeters increase with time and transportation conditions, two influence factors that are generally unknown for many terrestrial datasets (even though Renner et al. 1985 provide some hints for the actual campaign).

When looking at the statistics of the residuals after reducing SATOP1 (Fig. 10a), it can be seen that the standard deviation is twice as high as in the AGAP dataset. This is not surprising regarding all the additional error sources mentioned above. Nevertheless, the behavior of the empirical covariances (Fig. 10b, first minimum positive) implies some systematic effects. These effects can also be found in Fig. 10d as longer wavelength effect (positive in the mountains, negative over the ocean). As there could be different reasons for this phenomenon, no clear assertion can be made in this case. A common cause for such effects in terrestrial observations may be the systematics introduced through the choice of the measurement site (e.g. measuring through valleys, not over crests). The fact that with increasing reduction degree the effect decreases (cf. Fig. 10b) supports this theory (as unilateral measuring can be interpreted as source for aliasing).

Beside the longer-wavelength pattern, a strongly negative measurement track (marked in Fig. 10d) can be identified. As there is no obvious reason in this case (e.g. higher mountains) and the cut in the track is abrupt, a gross error in the observations seems reasonable. The presence of the higher mean value (Fig. 10a) of over 20 mGal is not unusual for a campaign dating back to a time when the tie to global gravity reference systems was not easy to establish (especially in Antarctica).

6 Conclusions

It is shown (through section 3b, 4 and 5) that the presented evaluation approach widely fulfills the demands stated in section 1. Thus, the method is fit for application to further datasets, with the final objective to inspect all datasets included in the AntGG database as well as all other campaign data available. The examples in section 5 were chosen to be most differing in terms of campaign type and therefore represent the majority of the available datasets. Thus, individual campaign examinations will be performed, accordingly.

After the evaluation and correction phase of the data is finished, the SATOP1 reductions can be reused for the remove-compute-restore procedure within the least squares collocation approach when producing the final regional gravity field solution. This is possible since the SATOP1 reductions rely on a spectral model and thus can be adapted consistently to all different kinds of gravity field functionals on arbitrary points in space.

Further, with the SATOP1 model (section 2) two innovations are introduced: firstly, the regularization of the satellite model was performed based on spheroidal harmonic degree variances, minimizing errors due to an implicit downward continuation (towards the poles). Secondly, spectral bands were cut out also in the spheroidal domain in order to avoid truncation effects (in contrast to band limitations in the spherical domain).

In summary, it can be concluded that the overall evaluation performance of the presented approach strongly depends on the dataset at hand – sparsely distributed terrestrial datasets are more difficult to inspect, as they generally feature a lower spatial density while their spectral content is higher (in comparison to airborne observations). This inevitably leads to more aliasing, preventing assessments on small scales (e.g. single points) and possibly also on larger scales (although gross errors may still be detectable, c.f. subsection 5.2). Dense airborne datasets on the other hand are easier to validate for the same reason, even though there is the additional difficulty with the (mostly unknown) low-pass filtering embedded in the data.

Finally, it is important to mention that neither the gravity field model nor the synthesis method presented

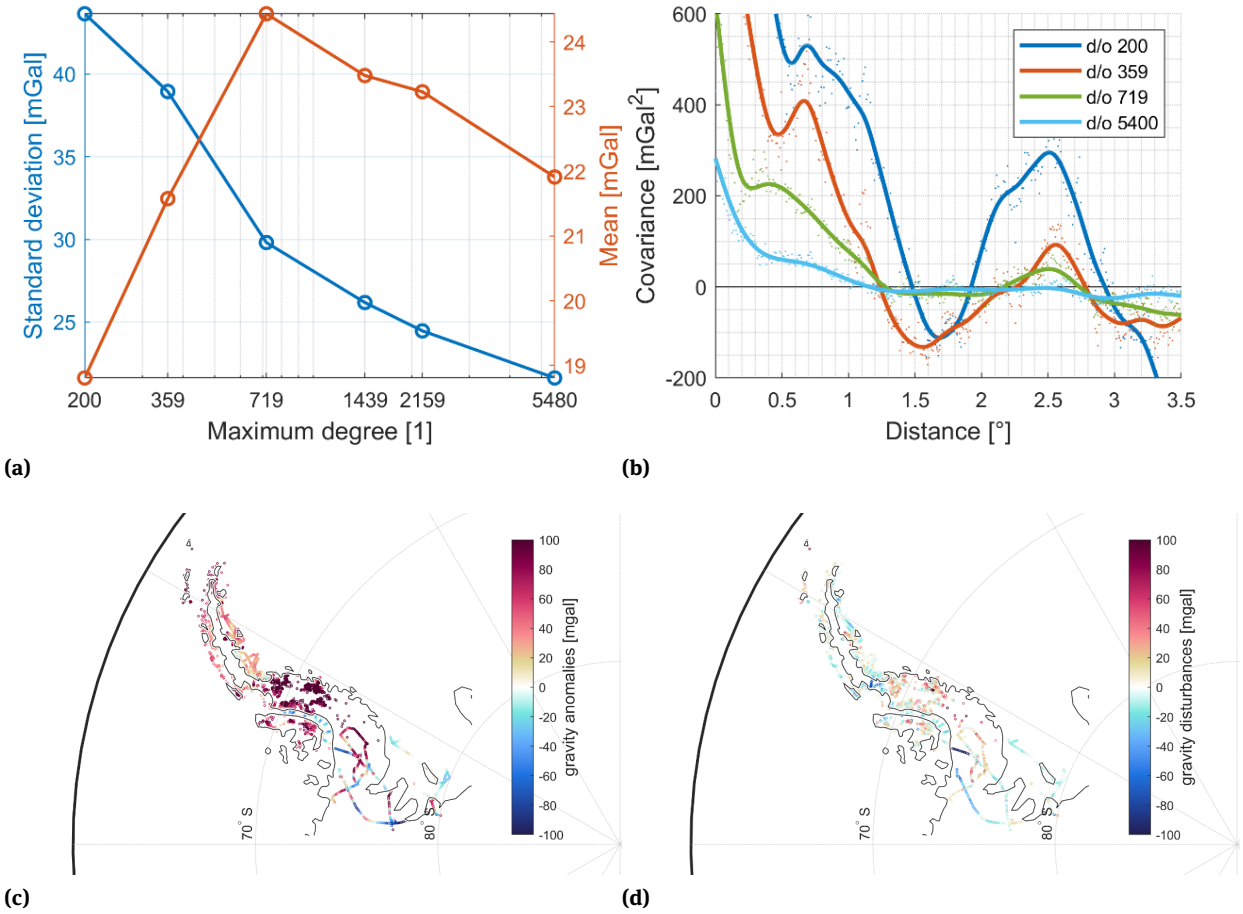


Figure 10: (a) Standard deviation (blue) and mean (red) for varying maximum degree of the SATOP1 model reduction from the BAS-LAND observations. (b) Empirical covariances depending on the spherical distance for different maximum degree reductions (200-359-719-5400). (c) The BAS-LAND observations as provided. (d) Differences to SATOP1 up to d/o 5400, mean value subtracted.

is restricted to the Antarctic region; one is free to use this evaluation strategy globally for nearly all gravity field functionals/observations. The application might be reasonable especially in other sparsely surveyed regions (e.g. Africa or parts of Asia).

Acknowledgements: This work was supported by the Deutsche Forschungsgemeinschaft (DFG) in the framework of the priority programme "Antarctic Research with comparative investigations in Arctic ice areas" by the grants PA 1543/18-1 and SCHE 1426/24-1.

References

Ferraccioli F., Finn C. A., Jordan T. A., Bell R. E., Anderson L. M. and Damaske D, 2011, East Antarctic rifting triggers uplift of the Gamburtsev Mountains, *Nature*, *479*, 388-392. doi:10.1038/nature10566

Fukushima T., 2012, Numerical computation of spherical harmonics of arbitrary degree and order by extending exponent of floating point numbers, *J. Geod.*, *86*, 271-285. doi:10.1007/s00190-011-0519-2

Hirt C., Rexer M., Scheinert M., Pail R., Claessens S. and Holmes S., 2016, A new degree-2190 (10 km resolution) gravity field model for Antarctica developed from GRACE, GOCE and Bedmap2 data., *J. Geod.*, *90*, 105-127.

Ivanov K. G., Pavlis N. K. and Petrushev P., 2018, Precise and efficient evaluation of gravimetric quantities at arbitrarily scattered points in space, *J. Geod.*, *92*, 779-796. doi:10.1007/s00190-017-1094-y

Jekeli C., 1981, *Alternative methods to smooth the Earth's gravity field,* retrieved from http://adsabs.harvard.edu/abs/1981amse.book

Lekien F. and Marsden J., 2005, Tricubic interpolation in three dimensions, *Int. J. Numer. Methods Eng.*, *63*, 455-471. doi:10.1002/nme.1296

Mayer-Gürr T., 2015, The combined satellite gravity field model GOCO05s, *EGU General Assembly Conference Abstracts*, *17*, p. 12364.

Moritz H., 1980, Geodetic reference system 1980, *Bulletin Geodesique, 54*, 395-405. doi:10.1007/BF02521480

Moritz H., 2000, Geodetic Reference System 1980, *J. Geod., 74*, 128. doi:10.1007/s001900050278

Pail, R. Fecher T., Barnes D., Factor J. F., Holmes S. A., Gruber T. and Zingerle P., 2018, Short note: the experimental geopotential model XGM2016, *J. Geod., 92*, 443. doi:10.1007/s00190-017-1070-6

Rapp R. H., 1997, Use of potential coefficient models for geoid undulation determinations using a spherical harmonic representation of the height anomaly/geoid undulation difference, *J. Geod., 71*, 282-289. doi:10.1007/s001900050096

Rexer M., Hirt, C. Claessens, S. and Tenzer R., 2016, Layer-Based Modelling of the Earth's Gravitational Potential up to 10-km Scale in Spherical Harmonics in Spherical and Ellipsoidal Approximation, Surv. Geophys., *37*, 1035. doi:10.1007/s10712-016-9382-2

Scheinert M., Ferraccioli,F., Schwabe J., Bell R., Studinger M., Damaske D., Jokat W., Aleshkova N., Jordan T., Leitchenkov Blankenship G. D. D., Damiani T. M., Young D., Cochran J. R., Richter T. D., 2016, New Antarctic gravity anomaly grid for enhanced geodetic and geophysical studies in Antarctica, *Geophys. Res. Lett., 43*, 600-610.

Shannon C. E., 1949, Communication in the Presence of Noise, *IEEE Proceedings, 37*, 10-21. doi:10.1109/JPROC.1998.659497

Sneeuw N., 1994, Global spherical harmonic analysis by least-squares and numerical quadrature methods in historical perspective, *Geophys. J. Int., 118*, 707-716.

Publication P-2: A partition-enhanced least-squares collocation approach (PE-LSC)

Reference
Zingerle P., Pail R., Willberg M. and Scheinert M. (2021) A partition-enhanced least-squares collocation approach (PE-LSC). J Geod 95, 94. doi 10.1007/s00190-021-01540-6

Abstract
We present a partition-enhanced least-squares collocation (PE-LSC) which comprises several modifications to the classical LSC method. It is our goal to circumvent various problems of the practical application of LSC. While these investigations are focused on the modeling of the exterior gravity field the elaborated methods can also be used in other applications. One of the main drawbacks and current limitations of LSC is its high computational cost which grows cubically with the number of observation points. A common way to mitigate this problem is to tile the target area into sub-regions and solve each tile individually. This procedure assumes a certain locality of the LSC kernel functions which is generally not given and, therefore, results in fringe effects. To avoid this, it is proposed to localize the LSC kernels such that locality is preserved, and the estimated variances are not notably increased in comparison with the classical LSC method. Using global covariance models involves the calculation of a large number of Legendre polynomials which is usually a time-consuming task. Hence, to accelerate the creation of the covariance matrices, as an intermediate step we pre-calculate the covariance function on a two-dimensional grid of isotropic coordinates. Based on this grid, and under the assumption that the covariances are sufficiently smooth, the final covariance matrices are then obtained by a simple and fast interpolation algorithm. Applying the generalized multi-variate chain rule, also cross-covariance matrices among arbitrary linear spherical harmonic functionals can be obtained by this technique. Together with some further minor alterations these modifications are implemented in the PE-LSC method. The new PE-LSC is tested using selected data sets in Antarctica where altogether more than 800,000 observations are available for processing. In this case, PE-LSC yields a speed-up of computation time by a factor of about 55 (i.e., the computation needs only hours instead of weeks) in comparison with the classical unpartitioned LSC. Likewise, the memory requirement is reduced by a factor of about 360 (i.e., allocating memory in the order of GB instead of TB).

Declaration of own contribution

(PZ: Philipp Zingerle; RP: Roland Pail; MW: Martin Willberg; MS: Mirko Scheinert)

The idea of improving the LSC method to support partitioning and the application on a large-scale was introduced by PZ. The study design was created by PZ with support of RP, MW and MS. All new methods were developed and implemented by PZ. Validations were performed by PZ using Antarctic gravity field datasets provided by MS. Additionally, the new methods were independently validated by PZ against an established implementation of the classical LSC approach with the help of MW. PZ designed and generated the figures and wrote the text with recommendations of RP, MW and MS.

The overall own contribution of PZ for *P-2* is estimated at 94 %, which is the average value of the percentage values estimated for the five criteria listed in the table below (Tab. P.2).

Criteria	**Estimated own contribution**
Computation and results	100 %
Ideas and study design	90 %
Analysis and interpretation	90 %
Text	90 %
Figures and tables	100 %
Total	**94 %**

Tab. P.2 – Criteria and estimated contribution share of Philipp Zingerle for *P-2*

Confirmation by the authors

We hereby confirm the correctness of the declaration of own contribution (of 94 % for Philipp Zingerle) for the publication

Zingerle P., Pail R., Willberg M. and Scheinert M. (2021) A partition-enhanced least-squares collocation approach (PE-LSC). J Geod 95, 94. doi 10.1007/s00190-021-01540-6

Roland Pail

Institute of Astronomical and Physical Geodesy, Technical University of Munich, Germany

Signature: ..... Date: 23.09.2021

Martin Willberg

Institute of Astronomical and Physical Geodesy, Technical University of Munich, Germany

Signature: ..... Date: 27.09.2021

Mirko Scheinert

Institut für Planetare Geodäsie, Technische Universität Dresden, Germany

Signature: ..... Date: 05.10.2021

Journal of Geodesy (2021) 95:94
https://doi.org/10.1007/s00190-021-01540-6

ORIGINAL ARTICLE

A partition-enhanced least-squares collocation approach (PE-LSC)

P. Zingerle[1] · R. Pail[1] · M. Willberg[1] · M. Scheinert[2]

Received: 27 November 2020 / Accepted: 28 June 2021 / Published online: 31 July 2021

Abstract
We present a partition-enhanced least-squares collocation (PE-LSC) which comprises several modifications to the classical LSC method. It is our goal to circumvent various problems of the practical application of LSC. While these investigations are focused on the modeling of the exterior gravity field the elaborated methods can also be used in other applications. One of the main drawbacks and current limitations of LSC is its high computational cost which grows cubically with the number of observation points. A common way to mitigate this problem is to tile the target area into sub-regions and solve each tile individually. This procedure assumes a certain locality of the LSC kernel functions which is generally not given and, therefore, results in fringe effects. To avoid this, it is proposed to localize the LSC kernels such that locality is preserved, and the estimated variances are not notably increased in comparison with the classical LSC method. Using global covariance models involves the calculation of a large number of Legendre polynomials which is usually a time-consuming task. Hence, to accelerate the creation of the covariance matrices, as an intermediate step we pre-calculate the covariance function on a two-dimensional grid of isotropic coordinates. Based on this grid, and under the assumption that the covariances are sufficiently smooth, the final covariance matrices are then obtained by a simple and fast interpolation algorithm. Applying the generalized multi-variate chain rule, also cross-covariance matrices among arbitrary linear spherical harmonic functionals can be obtained by this technique. Together with some further minor alterations these modifications are implemented in the PE-LSC method. The new PE-LSC is tested using selected data sets in Antarctica where altogether more than 800,000 observations are available for processing. In this case, PE-LSC yields a speed-up of computation time by a factor of about 55 (i.e., the computation needs only hours instead of weeks) in comparison with the classical unpartitioned LSC. Likewise, the memory requirement is reduced by a factor of about 360 (i.e., allocating memory in the order of GB instead of TB).

Keywords Gravity field · Least squares collocation (LSC) · Covariance function · Data combination · Prediction · Antarctica

1 Introduction

Least-squares collocation (LSC) is a commonly used technique in local or regional gravity field modeling (Moritz 1980; Krarup 1969). This method is very popular because it allows to predict gravity field-related quantities at arbitrary locations. At the same time, the desired functionals can be inferred from different input functionals (e.g., geoid heights from gravity anomalies). Eventually, LSC yields an optimal solution in the stochastic sense as it minimizes the prediction error using a priori covariance information of the observation and estimation points. From this perspective, LSC also allows to consider the full variance–covariance information and an appropriate propagation in the sense of a Gaussian process.

There are also some disadvantages of LSC. The computational cost is proportional to the cubic power of the number of observation points, while memory requirements are proportional to the quadratic power. This limits the practical application of the method to a certain maximum number of observation points. Even if this number can be reasonably handled, in the general application, some sort of tiling strategy (Reguzzoni and Tselfes 2009) must be applied to comply with computational limits. However, this strategy introduces the problem of fringe effects caused by the forced

✉ P. Zingerle
zingerle@tum.de

[1] Institute of Astronomical and Physical Geodesy, Technical University of Munich, Munich, Germany

[2] Institut Für Planetare Geodäsie, Technische Universität Dresden, Dresden, Germany

locality assumption. Another disadvantage of regional LSC is that it requires the input values at the observation points to have a zero mean. In practice, there may be several reasons why a residual (nonzero) mean exists, e.g., due to datum offsets, systematic effects induced by the measurement set-up, errors in the observations or reduction models or just due to the spatial limitation of the region under consideration. In general, it is not trivial to backtrack the source of the residual mean value and, therefore, it is often not justified to preventively remove it as it is usually done. In any case, a residual mean (or unexpected long-wavelength signal) causes a severe problem to the classical LSC, which ultimately culminates in unnatural oscillations and long-wavelength compensation attempts.

In the field of gravity field modeling, the (homogeneous-isotropic) covariance function used in LSC is commonly derived from spherical harmonic degree variances using covariance propagation (Moritz 1980). Since this calculation step involves the evaluation of Legendre polynomials for spherical distances of every pair of data points (i.e., observation and estimation points), the computation of covariance matrices becomes computationally demanding. Especially when a tiling approach is applied and identical point pairs must be evaluated several times (due to the overlap between tiles), the efficiency of the covariance matrix calculation deteriorates.

Tackling these challenges, a number of practical modifications to the original LSC approach was realized resulting in the so-called partition-enhanced least-squares collocation (short PE-LSC). Among these modifications, there are several innovative aspects such as a 2D gridding approach for a fast set-up of the covariances, the use of exponential estimators, and the filtering of the transfer function to achieve improved localization characteristics. From a practical point of view, the most important aspect is a significant acceleration of the run-time, which is primarily achieved by an optimal partitioning strategy of the study area. While the investigations are focused on the geodetic modeling of the gravity field, the usability of the presented methods may be extended to any harmonic functional defined on spherical or near-spherical surfaces (e.g., spheroids with small eccentricities).

The presented investigations are motivated by the IAG Subcommission 2.4f "Gravity and Geoid in Antarctica" (AntGG) where we aim to compile a refined grid of terrestrial gravity data in Antarctica as a major update to the data set published by Scheinert et al. (2016). For this, an optimum combination of a satellite-based global model (accounting for long-wavelength signal parts) and the terrestrial (ground-based or airborne) data is sought for. From the perspective of data processing, this paper is a logical succession to the study by Zingerle et al. (2019) who discussed data reduction and validation making use of a high-resolution gravity field model based on satellite and topography data. While the present paper focusses on the treatment of methodological aspects, resulting final products with respect to the regional gravity field in Antarctica will be subject to a separate paper.

Section 2 recapitulates the basics of the theory of LSC and introduces the notation. In Sect. 3, we explain in detail the different modifications to finally end up with PE-LSC. In Sect. 4, we examine and validate these modifications. All examples and validations presented in Sects. 3 and 4 are based on selected data sets from the AntGG project. Finally, Sect. 5 concludes the discussion by investigating the computational effort when using PE-LSC in real-world scenarios and gives a brief outlook to further research on the topic of LSC.

2 Theory and notation

This section provides a brief overview of the LSC theory and specifies the notation which will be used in this publication. Since we present only the very basics, the reader is referred to Moritz (1980) for a more complete and elaborated treatment of this topic.

2.1 The LSC method

LSC is a statistically optimal method for estimating (predicting) quantities $\underline{s}$ (the underscore below items tags vectors) from observations $\underline{l}$ based on their stochastic relation:

$$\underline{\hat{s}} = C_{SL}\left(C_{LL} + C_{NN}\right)^{-1}\underline{l} = \tilde{A}_S^L\underline{l}, \quad \text{with} \quad \tilde{A}_S^L := C_{SL}\left(C_{LL} + C_{NN}\right)^{-1}. \tag{1}$$

Here, $\underline{\hat{s}}$ denotes the estimates for $\underline{s}$ as result of the collocation, whereby C_{SL} is the cross-covariance matrix between the estimates and observation, C_{LL} the covariance matrix of the observations and C_{NN} additive noise covariance matrix. $\tilde{A}_S^L$ is defined to be the so-called estimator or kernel of the collocation. It is useful to interpret C_{SL} and C_{LL} to origin from a common (signal) covariance matrixC_{VV}, belonging to the common 'signal' $\underline{v}$:

$$\underline{v} := \begin{pmatrix} \underline{s} \\ \underline{l} \end{pmatrix} \leftrightarrow C_{VV} := \left[\mathrm{cov}\left(v_i, v_k\right)\right] = \begin{pmatrix} C_{LL} & C_{LS} \\ C_{SL} & C_{SS} \end{pmatrix}. \tag{2}$$

In the context of this publication, the observations $\underline{l}$ are measurements of the Earth's gravity field and $\underline{\hat{s}}$ are estimates of the gravity field signal on possibly other locations and in terms of other gravity field functionals.

2.2 LSC with a priori reduction

One major limitation of LSC is that it requires the signal $\underline{v}$ to be centered, meaning that the expectation of $\underline{v}$ shall be zero.

This implies that collocating a signal that is not centered probably yields an unfavorable result. For instance, if one tries to collocate gravity observations $\underline{l}$ directly on a regional scale, $\underline{l}$ is generally not centered since it still contains longer-wavelength signal components. Thus, it is advantageous to remove such signal parts beforehand. This can be achieved through prior information $\underline{v}_0$ of the signal which can be used to reduce the actual signal $\underline{v}$ before the collocation and restore it afterward (known as remove-compute-restore technique, RCR, see Forsberg and Tscherning 1981). To reuse the previous notation, $\underline{v}$ can be replaced by $\Delta\underline{v}$ through the relation:

$$\Delta\underline{v} := \underline{v} - \underline{v}_0 = \begin{pmatrix} \Delta\underline{s} \\ \Delta\underline{l} \end{pmatrix} \tag{3}$$

In this scenario, it is effective to substitute C_{VV} from Sect. 2.1 with the error covariance $\mathrm{C}^e_{VV} := \left[\mathrm{cov}\left(\Delta v_i, \Delta v_k\right)\right]$ of $\underline{v}_0$ (which is the same as the signal covariance of $\Delta\underline{v}$). Consequently, $\underline{v}$ is assumed to be error free in this context (despite the noise covariance C_{NN}, see Willberg et al. 2019). Together with the restore step, the modified LSC equation, based on residuals, reads:

$$\begin{aligned} &\Delta\underline{\hat{s}} = \tilde{A}^L_S \Delta\underline{l} \rightarrow \\ &\underline{\hat{s}} = \underline{s}_0 + \Delta\underline{\hat{s}} = \underline{s}_0 + \tilde{A}^L_S\left(\underline{l} - \underline{l}_0\right) \end{aligned} \tag{4}$$

with

$$\tilde{A}^L_S = C^e_{SL}\left(C^e_{LL} + C_{NN}\right)^{-1}. \tag{5}$$

An appropriate covariance propagation yields the error covariance for the restored estimate $\underline{\hat{s}}$:

$$C^e_{\hat{S}\hat{S}} = C^e_{SS} - C^e_{SL}\left(C^e_{LL} + C_{NN}\right)^{-1} C^{e\prime}_{SL}. \tag{6}$$

In gravity field modeling, the needed a priori information can be derived, e.g., from satellite-based global gravity field models and topographic models, respectively.

2.3 The covariance function

One crucial point for LSC is finding an adequate, ideally analytical, expression for the needed covariances C^e_{VV}. In gravity field modeling, it is useful to exploit the harmonic character of the gravity field in the absence of masses (in the exterior space). Doing so leads to the spherical harmonic representation $\underline{h}$ (i.e., Stokes coefficients, see Moritz 1980) of the gravity field, with the relation A^H_V to the spatial gravity signal $\underline{v}$ in form of:

$$\underline{v} = A^H_V \underline{h} \tag{7}$$

Assuming the covariance C^e_{HH} related to $\underline{h}$ in the spherical harmonic domain is known, the covariance C^e_{VV} in the spatial domain can be obtained by covariance propagation:

$$C^e_{VV} = A^H_V C^e_{HH} A^{H\prime}_V. \tag{8}$$

In the classical application C^e_{HH} is modeled by degree variances yielding a homogeneous-isotropic covariance function in the spatial domain (e.g., see Tscherning and Rapp 1974). In modern approaches also the full covariance information from global gravity field models is sometimes used (see Willberg et al. 2019), resulting in arbitrary, but still harmonic, covariance functions.

3 Methodology

Several modifications of the classical approach are introduced to increase the numerical efficiency and stability of RCR-LSC (cf. Sect. 2.2). In most of the strategies, the basic assumption is that a single observation Δl_i has a certain localized influence on the estimates $\Delta\underline{\hat{s}}$. Generally speaking, assuming a localized influence is justified if the covariances $\mathrm{cov}\left(\Delta l_i, \Delta s_k\right)$ become sufficiently small beyond a certain distance from the observation point. As arbitrary covariance functions do not necessarily show this behavior, their application must be restricted to the class of locally dominated covariance functions. For the ease of use this class is further constrained to (locally dominated) homogeneous-isotropic covariance functions. Since this paper deals with functionals that are harmonic and defined on or close to a sphere, the most natural and most general way of describing this class of covariance functions is by using spherical harmonic degree variances (cf. Moritz 1980 and Sect. 3.1). Consequently, they will be used to construct all covariances throughout the rest of this paper. In order to obtain the desired local character of the covariance function, it is required that within the RCR-LSC, the long-wavelength components are reduced beforehand (e.g., see Zingerle et al. 2019). Since the presented methods could theoretically also be applied to a non-reduced LSC (even if not recommended), in the following, each subsection describes a specific modification of the classical (RCR-)LSC method. Together, they form the basis for the PE-LSC approach which is evaluated in Sect. 4.

3.1 Accelerated covariance calculation

As explained in Sect. 2.3, the calculation of the covariance C^e_{VV} from spherical harmonics theoretically requires the calculation of the transformation matrix A^H_V and the evaluation of the (matrix) product $A^H_V C^e_{HH} A^{H\prime}_V$. For the special case of degree variances, this calculation can be simplified to the homogeneous-isotropic form (see Moritz 1980):

$$
\begin{aligned}
c_{ik} &:= \operatorname{cov}\left(\Delta v_i, \Delta v_k\right) = c\left(t_{ik}, u_{ik}\right) \\
&= \sum_{n=0}^{n_{\max}} u_{ik}^{l+1} P_n\left(t_{ik}\right) c_H^n := \left\langle \underline{q}_{ik}^H, \underline{c}_H \right\rangle
\end{aligned} \tag{9}
$$

where

$$
\begin{aligned}
t_{ik} &:= \langle \boldsymbol{e}_i, \boldsymbol{e}_k \rangle \left[= \cos\theta_{ik}\right], \quad \boldsymbol{e}_i := \frac{\boldsymbol{x}_i}{r_i}, \\
u_{ik} &:= \overline{u}_i \overline{u}_k, \quad \overline{u}_i := \frac{R}{r_i}, \quad r_i := \left\|\boldsymbol{x}_i\right\|.
\end{aligned} \tag{10}
$$

Herein, c_H^n denotes the (error) degree variances and P_n the Legendre polynomial of degree n, and R is the radius of the reference sphere chosen for the degree variances. $n_{\max}$ is the maximum degree that is considered in the modeling of c_{ik}. $\boldsymbol{x}_i$ describes the location in form of geocentric coordinates a single signal element v_i refers to. The parameterization via t (cosine of spherical distance θ) and u (product of length ratios to reference radius) is chosen so that a covariance matrix may be efficiently derived by outer (Cartesian) products regarding $\Delta\underline{v} \times \Delta\underline{v}$. For a better readability in figures (see Fig. 1), t is substituted by θ and u by the so-called equivalent height h_{eq} which is defined by

$$
h_{eq} := R\left(\frac{1}{\sqrt{u}} - 1\right). \tag{11}
$$

h_{eq} describes the actual height of a pair of points in case that both points are located at the same height. The evaluation of the simplified Eq. 9 is still time-consuming since it involves the calculation of numerous Legendre polynomials for every element of the resulting covariance matrices. Therefore,

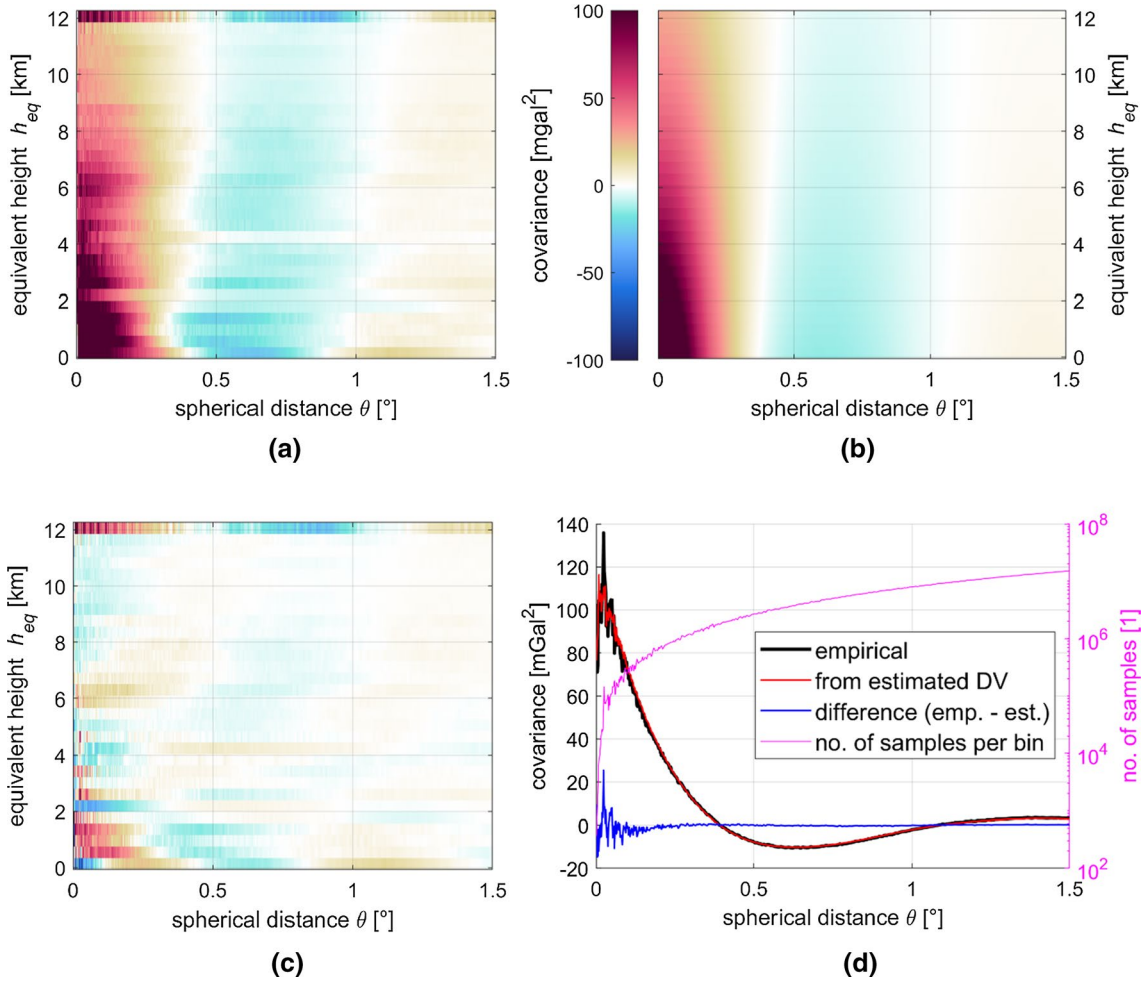


Fig. 1 Covariances in terms of gravity disturbances in the Ω_G domain. **a** Empirical covariances computed from reduced Antarctic gravity data (see Sect. 4). **b** Estimated covariance function derived from estimated degree variances (see Fig. 2). **c** Difference between empirical and estimated covariances. **d** The same covariances as in figure **a-c**, averaged over dimension u. Black: empirical covariance. Red: estimated covariances. Blue: difference between empirical and estimated covariances. Magenta: number of samples per bin for the calculation of the empirical covariances. Results shown are calculated by applying the spherical approximation to the coordinates (see Sect. 3.3)

to avoid the need to do this calculation for every element, it is proposed to introduce an intermediate regular grid $\Omega_G = \underline{t}_G \times \underline{u}_G$ on which the covariance values $\underline{c}_G = c(\Omega_G)$ are pre-calculated. Assuming a certain smoothness of the covariance function, the final elements are then simply obtained by applying a (gridded) two-dimensional interpolator (i.e., the interpolation matrix) I^G to the pre-calculated values $\underline{c}_G$:

$$c_{ik} = I^G\big(\Omega_G \rightarrow (t_{ik}, u_{ik})\big)\underline{c}_G \tag{12}$$

This interpolation can be evaluated very efficiently for large numbers of point pairs avoiding the need of evaluating sums of Legendre polynomials every time. Naturally, the grid vectors $\underline{t}_G$ and $\underline{u}_G$ should be chosen such that all occurring t_{ik} and u_{ik} are within the limits of the appropriate grid vector. The sampling of the grid vectors shall also be adapted to the smoothness of the intermediate covariance function $c(t, u)$ and the overall accuracy requirements for the calculation of the covariance functions. The accuracy can also be controlled by the appropriate choice of the interpolation method (e.g., linear, cubic, spline, etc.).

It is even possible to calculate arbitrary derivatives of the covariance function regarding local frame coordinates (e.g., gravity disturbances, gradients, deflection of vertical, etc.) based on a generalized chain rule. Since this generalized chain rule (also known as multivariate version of Faà di Bruno's formula, see Hardy 2006) is fairly complicated to state and even more complicated to prove, the reader is referred to the appropriate literature (e.g., Hardy 2006). In summary, the generalized chain rule consists of a linear combination of products of partial derivatives. For the commonly used local spherical east-north-up (ENU) frame, the partial derivatives are given in appendix (A.1). Since the number of summands within the generalized chain rule increases rapidly with the order of the derivative, the computation is practically limited to lower orders (i.e., below ten). In practice, this is only a minor limitation as higher-order derivatives are rarely used (at least in the scope of gravity field modeling).

3.2 Estimation of degree variances

Having a set of reduced observations $\Delta\underline{l}$, one can estimate an empirical covariance function $\tilde{\underline{c}}_G$ on the regular grid Ω_G by binning the individual covariance estimates $\tilde{c}_{ik} := \Delta l_i \Delta l_k$ into the 2D classes defined by $\underline{t}_G \times \underline{u}_G$ (see Fig. 1a). In this context, $\Delta\underline{l}$ is assumed to be centered.

Theoretically, by inverting the linear relation of Eq. 9 it would be possible to derive empirical degree variances $\underline{c}_H$ from $\tilde{\underline{c}}_G$. In practice, there are two major obstacles that prevent us from performing this inversion: firstly, the degree variances $\underline{c}_H$ are defined to be only positive; hence, the relation is not linear as Eq. 9 could suggest. Secondly, in general this inversion is highly instable due to the high dynamic range of the estimates and partially high insensitivity of $\underline{c}_H$ to $\tilde{\underline{c}}_G$. Nevertheless, a solution is possible when altering the functional model and including a reasonable regularization: to force positive values and to reduce the dynamic range, it is proposed to translate the estimates into the logarithmic domain (hence losing linearity). In order to tackle the instability issue, it is further proposed to introduce cubic basis splines (B-splines, e.g., de Boor 1978) as representation for the degree variances $\underline{c}_H$ assuming a certain smoothness of the degree variance curve. Obviously, to acquire smoothness the number of B-spline control points must be chosen to be significantly lower than the number of degree variances to estimate. Modifying Eq. 9 accordingly yields the nonlinear relation (with Q_G^H being the matrix extension of $\underline{q}_{ik}^H$ to all grid locations of Ω_G):

$$\begin{aligned} \underline{c}_G &:= Q_G^H \underline{c}_H =: Q_G^H \exp\left(B^{SP}\underline{x}_{SP}\right), \\ \frac{\partial \underline{c}_G}{\partial \underline{x}_{SP}} &=: A_G^{SP} = Q_G^H \mathrm{diag}\big(\exp\left(B^{SP}\underline{x}_{SP}\right)\big)B^{SP} \end{aligned} \tag{13}$$

where B^{SP} denotes the linear B-spline synthesis matrix (i.e., the matrix containing the spline basis functions, transforming from the spline function space to the logarithmic degree variance space). B^{SP} might be obtained efficiently by the Cox de Boor recursion formula (see de Boor 1978). The vector $\underline{x}_{SP}$ contains the appropriate spline parameters to estimate. $\exp(\dots)$ denotes the element-wise exponential function and $\mathrm{diag}(\dots)$ indicates the vector to diagonal square matrix transform. Using the linearization A_G^{SP}, $\underline{x}_{SP}$ can be obtained by an iterative LSA approach

$$\begin{aligned} \Delta\underline{x}_{SP} = &\left(A_G^{SP\prime} C_{GG}^{-1} A_G^{SP} + P_{REG}\right)^{-1} \\ &\left(A_G^{SP\prime}\left(\underline{c}_G - Q_G^H \exp\left(B^{SP}\tilde{\underline{x}}_{SP}\right)\right) + \underline{q}_{REG}\right) \end{aligned} \tag{14}$$

starting with some initial guess $\tilde{\underline{x}}_{SP}$ for $\underline{x}_{SP}$ (and iteratively improving $\tilde{\underline{x}}_{SP}$ by $\Delta\underline{x}_{SP}$). C_{GG} denotes the covariance of $\underline{c}_G$ which can also be derived empirically or can simply be approximated as, e.g., diagonal matrix with the diagonals (variances) as inverse of the number of samples of the appropriate bin. P_{REG} and $\underline{q}_{REG}$ are regularization terms that may help to further improve the result. As an example, one may add additional smoothing constraints (e.g., second derivative set to be zero), add a tie to a baseline value (e.g., zero) or force the slope to be flat (e.g., first derivative set to be zero). The appropriate weights to the regularization terms may be derived empirically such that the iterative LSA converges and yields a good fit as well as a realistic result. Figure 2 (in combination with Fig. 1) exemplarily shows the functionality of this method.

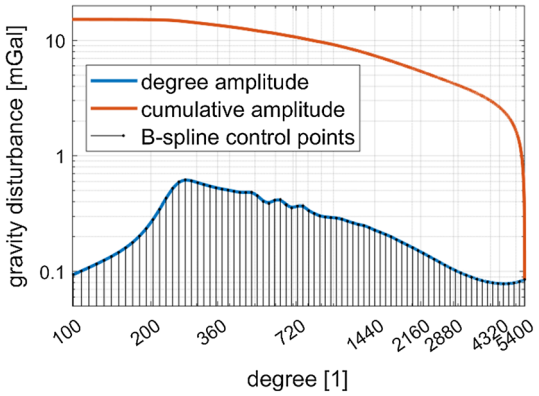


Fig. 2 Estimated degree amplitudes (i.e., square root of degree variances $\underline{c}_H$) from empirical covariances ($\underline{\tilde{c}}_G$, cf. Eq. 13, Fig. 1a) in terms of gravity disturbances. Blue: degree amplitude. Red: cumulative amplitude, i.e., the aggregated total power from the maximum degree downward (from right). Black: control points of the B-spline used for estimating the degree variances (see Eq. 13, equally spaced in the logarithmic domain)

3.3 Spherical approximations

In the following, it is assumed that the reduced signal $\Delta\underline{v}$ describes some harmonic functional in the exterior of a spheroidal-shaped body (e.g., Earth's gravity or magnetic field). When the region Ω_V of $\Delta\underline{v}$ is large and covers a certain latitude range, the use of covariance functions derived by spherical harmonic degree variances introduces artificial latitude-dependent systematics. This is obvious as the quantity u in Eq. 10 depends on the observation's (body-centric) radii which vice versa depend on the latitude if $\Delta\underline{v}$ resides near the surface (see Fig. 3). Apparently, such a latitude-dependent influence cannot be physically explained, especially when the signal $\Delta\underline{v}$ was reduced beforehand for the long-wavelength signal part. Theoretically, these systematics can be avoided by using spheroidal harmonic degree variances instead of spherical harmonic degree variances, as the height-dependent item u would then vanish for the signal on the surface of the spheroid (see Moritz 1980). However, spheroidal harmonic degree variances become difficult to handle when the height is not zero, as it involves the parametric latitude and Legendre polynomials of the second kind. Therefore, they are rarely used in LSC. Hence, a different approach is proposed where once again the locality assumptions are considered: in a local (or regional) setting the curvature of a spheroid (with small eccentricity) can be approximated by a sphere with the curvature radius $\overline{R}$, averaged over the region Ω_V (see also Willberg 2020). Consequently, the oblateness in the geometry can be eliminated by a spherical modification of the coordinates:

$$\breve{\boldsymbol{x}}_i := \left(\overline{R} + h_i^{ell}\right)\boldsymbol{e}_i \tag{15}$$

h_i^{ell} denotes the ellipsoidal height of point $\boldsymbol{x}_i$. Evidently, the modified geometry in $\breve{\boldsymbol{x}}_i$ preserves local relations of adjacent points to a large extent while removing the latitude dependency. When applying this geometry, also the empirical degree variances (see Sect. 3.2) should be derived using the modified coordinates (cf. Eq. 10) in order to preserve consistency. Consequently, results shown in Figs. 1 and 2 are also calculated by using the modified geometry.

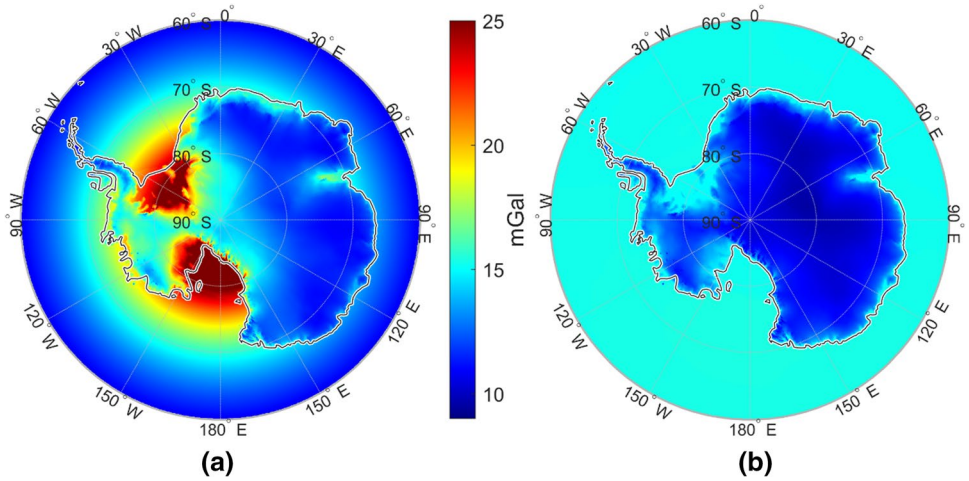


Fig. 3 A priori standard deviations (i.e., square root of variance entries in C_{SS}^{Ω}, see, e.g., Eq. 9) calculated from spherical degree variances (cf. Fig. 2) evaluated on the surface of Antarctica (using a polar-stereographic projection). **a** Standard deviations calculated using the original coordinates. **b** Same standard deviations calculated using the modified coordinates (Eq. 15). As the geocentric radii of the original coordinate decrease toward the pole due to the oblateness of the Earth, the a priori variances systematically increase (cf. Fig. 1b). Variations in (**b**) are solely correlated to varying surface heights (topography)

3.4 Localization of the LSC estimator

Even though the covariance $\mathrm{cov}(\Delta l_i, \Delta s_k)$ between an estimate's signal Δs_k and observation's signal Δl_i is zero (or close to zero), the appropriate element $\tilde{a}_{ik}$ in the LSC estimator $\tilde{A}_S^L$ does not necessarily have to be zero (or even close to zero). This is counterintuitive because one would assume that the estimate does not depend on observations that are statistically uncorrelated. Nevertheless, due to the inversion of $C_{LL}^e + C_{NN}$ (cf. Eq. 5) rather large nonzero elements may pop up especially when the system has a high sensitivity to small changes. This emerging correlation among observations over rather large distances is considered to be undesirable as it destroys all previously made locality assumptions. Consequently, this behavior is mainly responsible for fringe effects when partitioning the region Ω to accelerate the collocation (see Sect. 4.2). Making a non-critical compromise in terms of optimality (cf. Sect. 4.1), this behavior can be avoided when $\tilde{a}_{ik}$ is down-weighted by applying a distance-dependent weighting function w_{ik}. For PE-LSC the distance measure $1 - t_{ik}$ is used for this weighting (see Eq. 16). Alternatively, also the spherical distance $\theta_{ik} := \mathrm{acos}(t_{ik})$ may be chosen as distance measure. The weighting function w_{ik} should be (1) smooth, and (2) strongly attenuate beyond a certain distance $\theta^{\max}$, but (3) should not influence the LSC estimator $\tilde{A}_S^L$ in close proximity (i.e., when θ_{ik} is small). Although there might be many functions that satisfy these requirements, a very simple and effective one is found when applying a slightly modified Gaussian bell curve to t_{ik}

$$w_{ik}^G := \exp\left(-\left(\frac{t_{ik}-1}{\cos(\theta^{\max})-1}\right)^2\right). \tag{16}$$

Figure 4 exemplarily shows the behavior of w_{ik}^G. It should be noticed that using $t_{ik} := \cos(\theta_{ik})$ (instead of θ_{ik}) as parameter in Eq. 16 narrows down the transition width of the function and hence contributes to the requirements (2) and (3) as stated above. While this weighting approach allows for an efficient reduction of fringe effects, it does not guarantee to fully eliminate them (cf. Sect. 4.2 and discussion above).

3.5 Optimal partitioning and corresponding reduction of numerical effort

The main aim of the forced localization introduced in the previous section is to enable the partitioning of the collocation region $\Omega_S \subseteq \Omega_V$ (where the estimates are located in) into smaller subregions Ω_i^P. Dividing Ω_S into independent subregions allows a significant reduction of the overall computation time t_c as well as of memory requirements. To perform this separation correctly every partition must include the necessary surrounding $\delta\Omega_i^P$ (buffer) to Ω_i^P (according to the localization criterion, e.g., $\theta^{\max}$) and subsequently introduces a certain amount of overhead (see Fig. 5a). As this additional overhead is introduced with every partition, the overall overhead grows linearly with the number of partitions and at a certain point becomes larger than the acquired gain. Consequently, when implementing an appropriate algorithm, it is favorable to determine an optimal partition size that allows maximizing the overall gain (i.e., to minimize t_c): from a computational perspective, evaluating the LSC estimator $\tilde{A}_S^L$ requires a matrix inversion as well as matrix multiplication (cf. Eq. 5). In general, matrix inversion as well as multiplication has a computational cost of $\mathcal{O}(n^3) \approx kn^3$, where n is the number of elements of the random vector (that is the number of observation resp. evaluation points) and k is a constant factor depending on the algorithm used. For the following considerations it is assumed that the points are evenly distributed on a sufficiently large 2D surface (region Ω) with a constant (point) density of ρ. Dividing this surface into m identical partitions, each including n points, the needed computation time is given by:

$$t_c = mkn^3 \tag{17}$$

Assuming the subregions Ω_i^P to be squares, their size can be characterized by their edge length d. Having an initial number of partitions m_0 corresponding to an initial edge length d_0 the actual number of partitions can be calculated (in the 2D case) by:

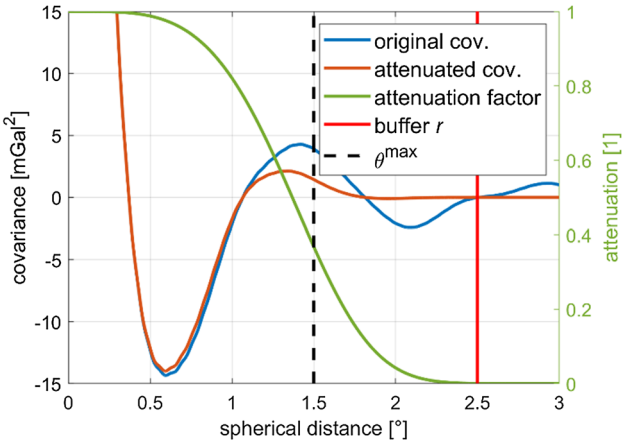


Fig. 4 Modified Gaussian bell attenuation function w_{ik}^G (cf. Eq. 16) and its exemplary influence to the covariance function. Blue: original covariance function in terms of gravity disturbances (cf. Fig. 1d). Red: attenuated covariance function. Green: attenuation function w_{ik}^G ($\theta^{\max} = 1.5°$). Dashed black: attenuation parameter $\theta^{\max}$ (1σ, cf. Fig. 5a). Vertical red: buffer distance $r = 5/3\theta^{\max}$ (cf. Fig. 5a). Covariance function is evaluated on the ellipsoid

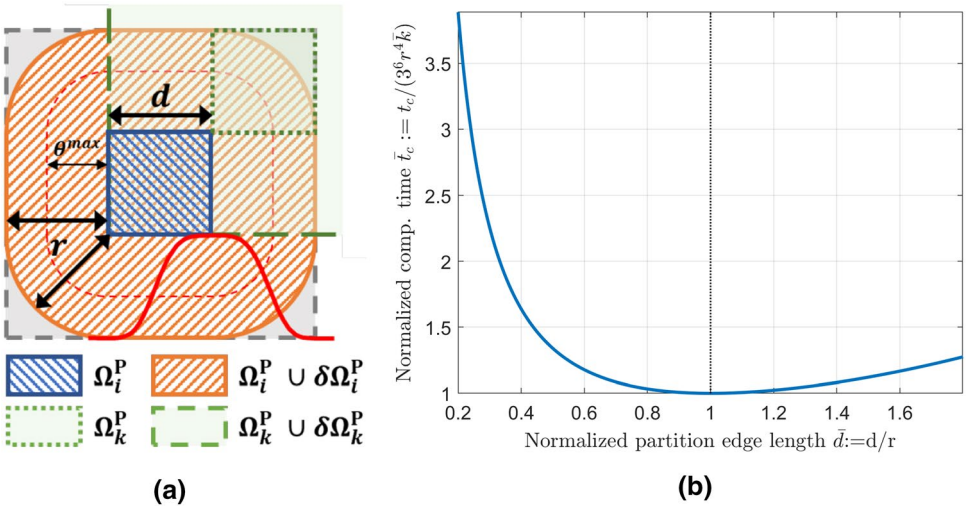


Fig. 5 **a** Scheme of the partitioning items. Blue hatched: central area of a partition with edge length d. Orange hatched: extended area with buffer distance r. Light green: central and buffer area of an adjacent partition. Solid red curve: 1D visualization of the attenuation function (Eq. 16) for an evaluation point located on the right edge of the partition. Dashed red line: visualization of the parameter $\theta^{\max}$ (1σ, see Fig. 4) for the attenuation function. **b** Normalized computation time as function of the normalized partition edge length (see Eq. 19)

$$m(d) = m_0 \left(\frac{d_0}{d} \right)^2 \tag{18}$$

As mentioned above, the overall size of the partition must include its surrounding $\delta\Omega_i^P$. This shall be taken into account by a buffer distance r (presumably depending on $\theta^{\max}$), yielding the edge length of the extended partition $d_e = 2r + d$. As a rule of thumb, choosing $r = 5/3\theta^{\max}$ seems to yield a good trade-off between omission error and buffer distance when applying Eq. 16 as weighting function (see Fig. 4). For the sake of simplicity, $\delta\Omega_i^P$ shall also be a square (omitting the roundings of the buffer, cf. Fig. 5a); hence, the area of the extended partition is d_e^2, and with that the number of (observation) points per partition $n = \rho d_e^2$. By inserting n and m into Eq. 17 one finds:

$$t_c(d) = \overbrace{k m_0 d_0^2 \rho^3}^{\bar{k}} \frac{d_e^6}{d^2} = \bar{k} \frac{(2r+d)^6}{d^2} \tag{19}$$

The minimum of this equation is obtained by setting the derivative $\frac{\partial t_c}{\partial d}$ to zero:

$$\frac{\partial t_c}{\partial d} = 2\bar{k} \frac{(2r+d)^5}{d^2} \left(3 - \frac{2r+d}{d} \right) = 0 \;\rightarrow d = r \tag{20}$$

Hence, the optimal edge length is obviously the chosen buffer distance itself. Naturally, this is only valid under the assumed simplified conditions, where particularly the special treatment of fringe partitions was neglected. Nevertheless, for larger regions, this assessment should be sufficient in most cases, and for smaller regions, reaching the perfect optimum should not be important anyway. Be aware that in this derivation, n is primarily related to the number of observation points $\{\underline{L}_i^P \in \Omega_i^P, \delta\underline{L}_i^P \in \delta\Omega_i^P\}$ and not to the number of estimation points $\underline{S}_i^P \in \Omega_i^P$. If the number of the latter is significantly larger than the number of the former (i.e., $\geq n^2$), the optimum can be missed by this estimate, generally favoring smaller partitions (hence making the matrix multiplication less expensive).

From Eq. 19, one can further discern the most expensive parts within the LSC, which are represented by (1) the buffer distance r (resp. $\theta^{\max}$) that increases the cost by the power of six, and (2) the point density that increases the cost by the power of three. Consequently, if computation time is still problematic after partitioning, one should primarily try to further reduce the correlation length (e.g., by improved reductions) and/or to decrease the observation point density (e.g., by thinning the data). Additionally, even if the optimality criterion as stated above is met for a sub-region, in practice one might be forced to further partition it, as the system memory might not be sufficient (e.g., if the local point density is high). This is critical, as the overall computation time (Eq. 19) increases significantly for $d < r/2$ (see Fig. 3). Therefore, having enough system memory available might be crucial in such cases. In order to be able to dynamically adapt the partition size when the density is too high, it is proposed to use a divide and conquer approach. In the reference implementation the divide and conquer algorithm works recursively in the 3D Cartesian space, always dividing a partition along the dimension with the largest extent until all criteria are met (i.e., max. extents or number of points, e.g., see Fig. 9). Obviously, this algorithm tends to create partitions to be smaller than optimal regarding the chosen maximum extents. Hence, choosing these maximum extents to be slightly larger than the buffer distance r might help to better reach the optimum (i.e., r) on average. As shown in Fig. 3, choosing larger extents is less problematic than choosing them too small in terms of computation time penalty.

4 Validation

Most of the modifications described in Sect. 3 are simple to validate. For instance, the validation of the accelerated covariance calculation (Sect. 3.1) is performed by cross-comparison to the rigorous calculation method. It has been found that one can approximate the rigorous result to any desired accuracy level by increasing the grid sampling or using more elaborate interpolation algorithms. The correct estimation of the degree variances (Sect. 3.2) is verified by inspecting the fit to the empirical covariance function where the residuals are small and do not show any systematics (Fig. 1c, d). The improvement by using the modified coordinates (Sect. 3.3) is also directly discernible by inspecting the a priori variances (Fig. 3).

The validation becomes more difficult when studying the modification of the LSC estimator (Sect. 3.4) as this alteration directly influences the outcome of the LSC especially in combination with the partitioning approach (Sect. 3.5). Hence, an in-depth validation of this topic is performed focusing on the behavior of the LSC estimator (Sect. 4.1) and the impact of the localization regarding the result and the partitioning (Sect. 4.2). Note that all investigations within this section are performed using gravity observations in selected regions of Antarctica, which already served as input to the 2016 AntGG grid (Scheinert et al. 2016). In particular, we are making use of gravity disturbances resulting from airborne gravimetry in West Antarctica (Bell et al. 1999; Studinger et al. 2002) and over a transect from the Transantarctic Mts. to the South Pole (Studinger et al. 2006). The data is reduced beforehand by a high-resolution gravity model combining satellite and topographic information (see Zingerle et al. 2019 for the model, Sect. 2.2 for the RCR-LSC theory). The obtained observation residuals fit to the empirical covariance function and degree variances as shown in Figs. 1 and 2. For the localization of the LSC estimator and the partitioning, the parameters $r = d = 2.5°$ and $\theta^{\max} = 1.5°$ are chosen if not stated otherwise (cf. Sects. 3.4, 3.5, Figs. 4, 5, 6, 7, 8). Gravity disturbances on or near the ice surface form also the output functional in the estimation. Insights found by the following evaluations do not necessarily hold for other setups (i.e., other combinations of input/output functionals, estimates with extensive upward or downward continuation or different covariance functions), as the shape of the LSC estimator might strongly vary between different scenarios. Nevertheless, it can be stated that the following conclusions shall be valid in all cases where the pattern of the LSC estimator is locally dominated (cf. Sect. 4.1, Fig. 6).

4.1 Behavior of the localized LSC estimator

Without explicitly calculating the LSC estimator (Eq. 5), its behavior is generally hard to predict as it involves matrix inversion and thus may produce highly variable results even when just slightly altering the input configuration. Nevertheless, when inspecting the LSC estimator (i.e., the kernel) for different locations (see Fig. 6), at least one general rule of thumb could be derived, which is valid in most cases (when the covariance function is decreasing with increasing distance). The better the observation coverage (i.e., density and quality) surrounding an estimation point, the smaller the effective influence radius of the kernel (cf. Fig. 6a, point A). Vice versa, when the nearby observation coverage is poor this influence radius can significantly increase in size (cf. Fig. 6d, point B). As a direct consequence, the kernel of estimates within good coverage (e.g., point A) can be considered to be already localized, while the kernel of estimates within poor coverage (e.g., point B) does not show this property. Subsequently, localizing the kernel (cf. Sect. 3.4) for points within good coverage has barely any influence (cf. Fig. 6b, c) while localizing those within poor coverage significantly alters the kernel (cf. Fig. 6e, f). To have an objective measure of how strongly the kernel is modified by the localization, a kernel disparity measure Δ_i^{ker} is introduced for an estimate i as:

$$\Delta_i^{\text{ker}} = \sqrt{\frac{\sum_k \left(\tilde{a}_{ik}\left(1 - w_{ik}^G\right)\right)^2}{\sum_k \tilde{a}_{ik}^2}} = \sqrt{\frac{\sum_k \left(\tilde{a}_{ik} - \tilde{a}_{ik}^{loc}\right)^2}{\sum_k \tilde{a}_{ik}^2}} \tag{21}$$

where $\tilde{a}_{ik}$ denotes the elements of the original kernel matrix $\tilde{A}_S^L$, and $\tilde{a}_{ik}^{loc}$ the elements of the localized kernel. Since the relation $0 \leq w_{ik}^G \leq 1$ holds for all weighting coefficients (cf. Eq. 16, Fig. 4), Δ_i^{ker} ranges between zero and one. A value of 0 means that the localized kernel is identical to the original kernel, and a value of 1 indicates a completely different kernel (cf. Fig. 7b). Therefore, as the weighting function always attenuates kernel elements toward zero with increasing distance, the appropriate estimate will also be attenuated toward zero in comparison with the estimate derived from the unmodified kernel (see also Sect. 4.2, Fig. 8). Ultimately, the weighting function ensures that the further away the closest observation is located from the estimation point, the more the corresponding estimate is attenuated toward zero. In the context of RCR-LSC (Sect. 2.2.b), this behavior can be considered as desirable, as it is unlikely that the LSC is able to significantly improve such estimates in comparison with the a priori reductions (e.g., see Fig. 7). In fact, the unmodified LSC kernel tends to unnaturally overshoot (the estimate) in such locations, especially when the observations contain long-wavelength errors. This becomes obvious when examining the incisive long-range patterns of the original

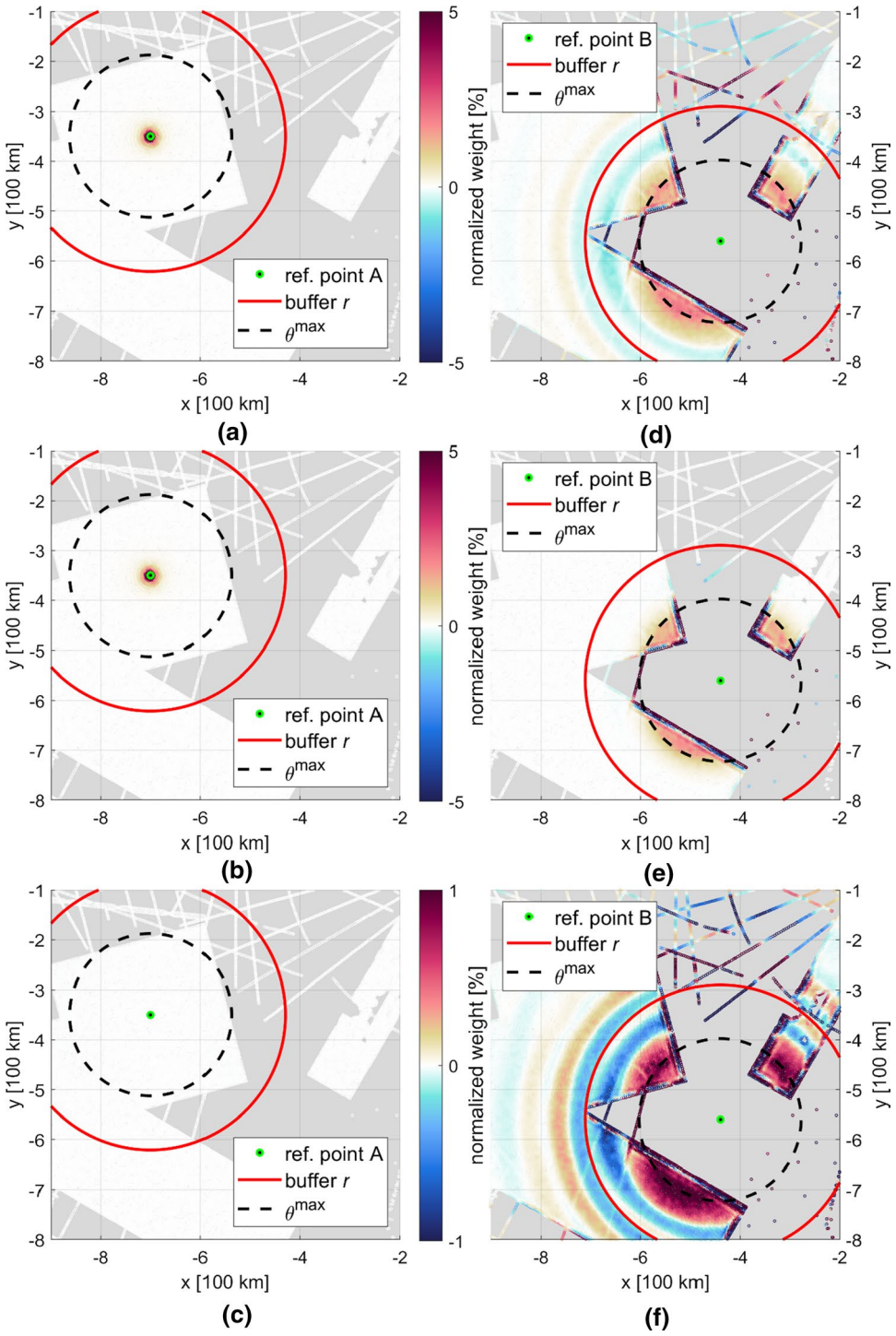


Fig. 6 Influence of the localization of the LSC estimator $\tilde{A}_S^L$ (Eq. 5) for two different evaluation sites in Antarctica: site A is located within a densely measured area (**a-c**), while site B is located in an area where no observations are present (**d-f**). **a**, **d** Original LSC estimator. **b**, **e** Localized LSC estimator (see Sect. 3.4, Fig. 4). **c**, **f** Difference between original and localized estimator. For visualization, the individual weights of the LSC estimator are normalized regarding the appropriate largest weight. As a reference, the applied weighting parameter $\theta^{\max}$ (1.5° black dashed circle) and appropriate buffer distance r (2.5°, red circle) are shown. For $\tilde{A}_S^L$, the covariance function chosen is according to Fig. 1 using degree variances as of Fig. 2. Reference points are located on an ellipsoidal height of 5 km (slightly above the ice layer). The figure's background is shaded in gray (where no observations are present). Grid coordinates refer to a polar stereographic projection where the positive y-axis coincides with the Greenwich meridian. The disparity Δ_i^{ker}(see Eq. 21, Fig. 7b) between original and localized kernel is 0.16% for point A and 58.49% for point B

kernel function in such locations (cf. Fig. 6b): signals that are in the same wavelength range as the extents of the areas with positive or negative weights are able to strongly amplify the estimate due to possible resonances with the kernel.

In summary, one can conclude that the localization of the LSC estimator is probably beneficial in any situation (see Fig. 7): in areas with good coverage the original LSC estimate is widely preserved, while in areas with poor coverage the estimate is attenuated and thus prevents the previously mentioned overshooting. A further strong indication that the localization does not negatively impact the estimate is given when inspecting the increase in the formal error (Eq. 6) when applying the localized kernel: obviously, this increase (Fig. 7d) is negligible compared to the formal error amplitude (Fig. 7c).

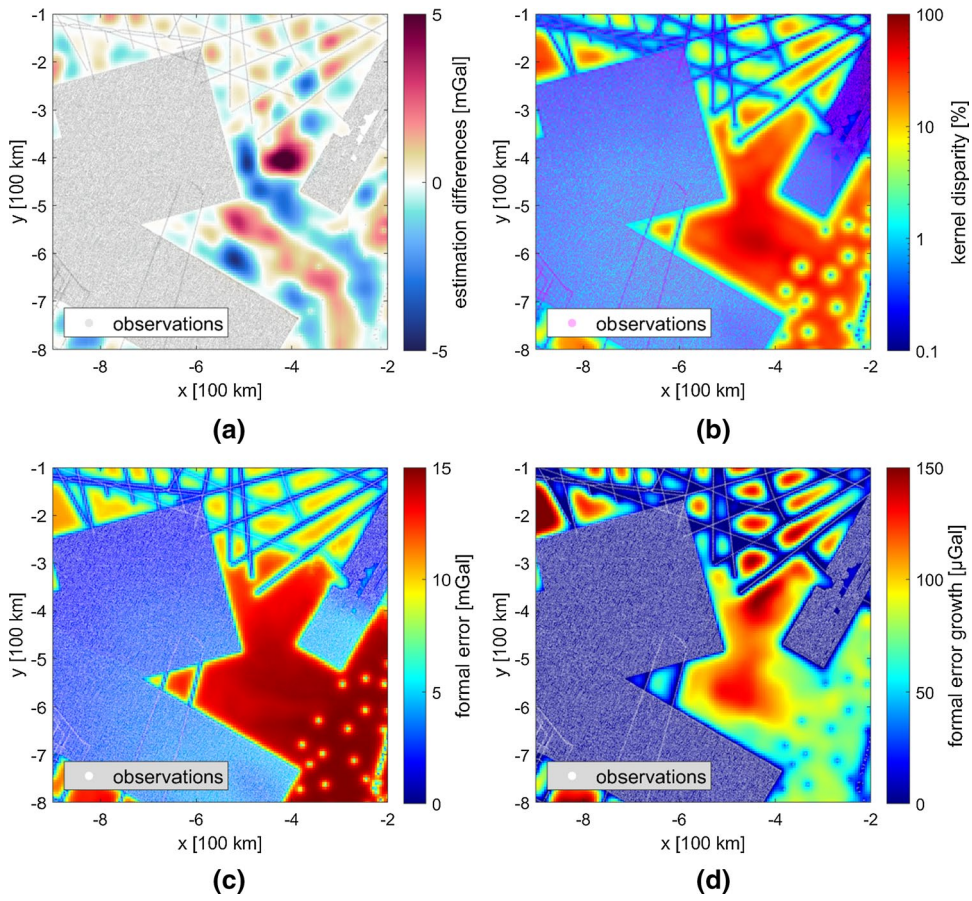


Fig. 7 Impact of the localized kernel on the estimation result (i.e., a 5 km regular polar stereographic grid of reduced gravity disturbances on the surface). The localization is identical to Figs. 4, 6 (i.e., $\theta^{\max} = 1.5°$). **a** Estimation difference between original and localized kernel. **b** Kernel disparity Δ_i^{ker} (see Eq. 21) on the estimation grid (logarithmic scale). **c** Formal error of the estimation with the original kernel (see Eq. 6). **d** Formal error increase when localizing the kernel (regarding the original result, see (**c**))

4.2 Influence of the localization to the partitioning

Even though the localization of the LSC estimator already benefits from the classical LSC method (see Sect. 4.1), the primary reason for introducing the localization is to enable the partitioning of the collocation domain. As explained in Sect. 4.1, this localization limits the influence of the LSC estimator (i.e., kernel function) according to the chosen localization criterion (e.g., $\theta^{\max}$), and with this also to the buffer distance r (cf. Fig. 6). Obviously, applying a partitioning with buffer distance r to the collocation domain (cf. Fig. 5a) prevents the LSC estimator from including observations outside this region (e.g., see Fig. 6d, elements outside the red circle). Hence, it may be concluded that the partitioning itself directly impacts the collocation result in case that the kernel relies on observations beyond the buffer distance r. Albeit those alterations of the result mostly preserve the overall quality of the outcome, they introduce disparities between adjacent partitions, which manifest as fringe effects. In agreement with the findings in Sect. 4.1, these effects can be large especially in regions with sparse observation coverage where the LSC estimator tends to have a wide pattern (see Figs. 6d, 8a). There, the discontinuities on transitions between partitions might be problematic particularly when evaluating numerical derivatives of the result (cf. Fig. 8c) or predicting higher-order gravity functionals. As explained in Sect. 3.4, an adequate localization of the LSC estimator can prevent most of these fringe effects since correlations of the kernel beyond the buffer distance are removed in the first place (see Fig. 6e). In practical applications, residual fringe effects become virtually untraceable when using an appropriate kernel localization (cf. Fig. 8b, d).

As already mentioned in Sect. 3.4, despite the obvious large reduction in fringe effects due to the localized kernel, there is no mathematical guarantee of continuity between estimates of adjacent partitions (due to matrix inversion). Nevertheless, one can ask if there is at least convergence of the result when increasing the buffer distance r. In theory, when the buffer includes the whole collocation domain (i.e., an infinite buffer), fringe effects are again impossible (since the matrix to be inverted would always be the same in this case). Convergence is also desirable, as it indicates 'how stable' an obtained solution is when adding additional far-distance observations to the estimator. It is finally expected, the better the convergence, the better the fit to a result obtained by an infinite buffer and hence to a rigorous solution without partitioning. For evaluating the convergence behavior, LSC estimates are calculated

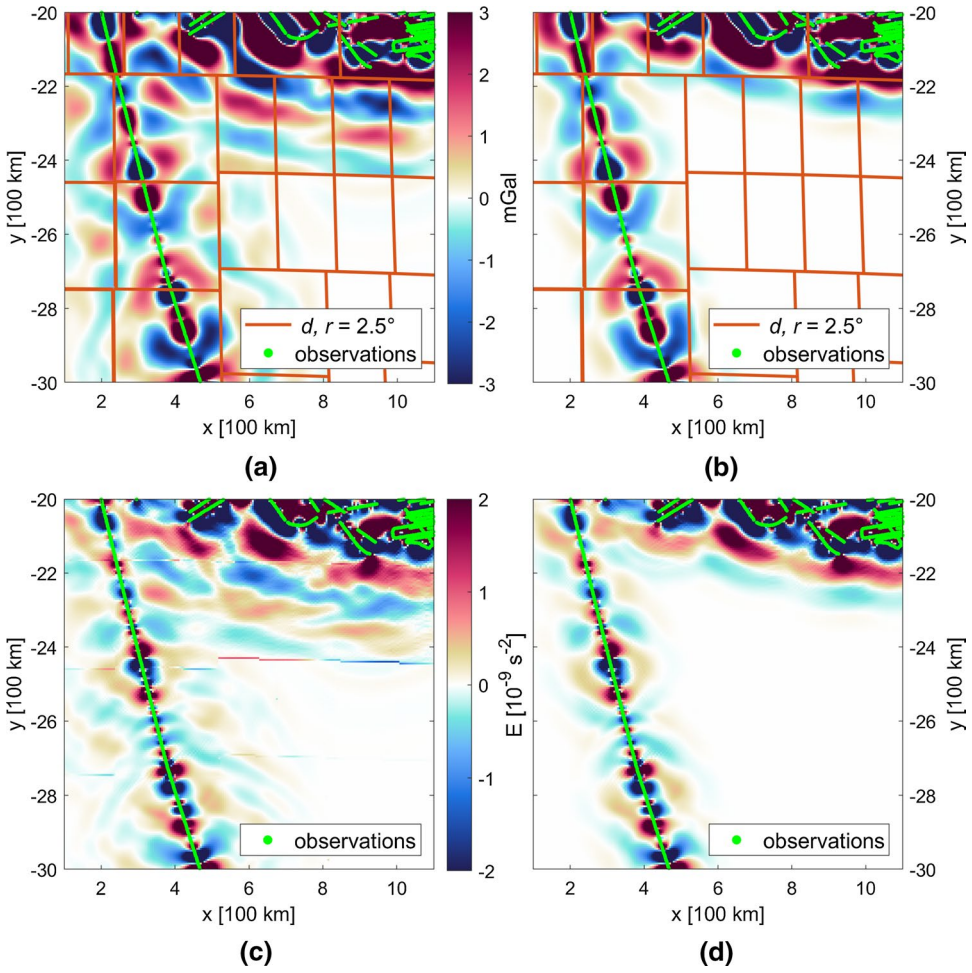


Fig. 8 Fringe effects on partition boundaries with and without kernel localization in a peripheral area of Antarctica (same setup as in Fig. 7). Partition boundaries are highlighted in red using $d = r = 2.5°$ as partitioning parameters (see Fig. 5a). **a** Collocated residuals obtained by the original kernel (gravity disturbances). **b** Collocated residuals obtained by the localized kernel. **c** Numerical derivative of **a** (with original kernel) in y-direction. **d** Numerical derivative of **b** (with localized kernel) in y-direction

for two different buffer distances r (and hence partition edge lengths d) while the localization parameter $\theta^{\max}$ is preserved. The difference between these two results can then be interpreted as a convergence resp. stability measure (see Fig. 9). As expected, the convergence is strongly improved when using the localized kernel (Fig. 9b) instead of the original one (Fig. 9a).

4.3 Statistical evaluation

As last validation, the behavior of the localization is inspected from a statistical perspective over the whole collocation area (see Fig. 3) by examining frequency distributions regarding the non-localized LSC solution (cf. Fig. 10). Comparing the convergence when increasing the

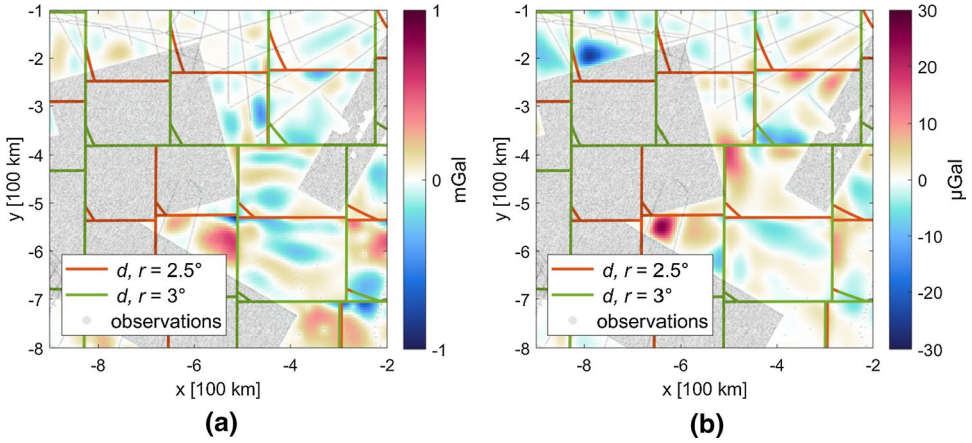


Fig. 9 Difference between LSC estimates when increasing the buffer distance r as well as the edge length d from 2.5° (red lines) to 3.0° (green lines). The rest of the setup is identical to Fig. 7. **a** Differences when using the original kernel. **b** Differences when using the localized kernel (with $\theta^{\max} = \text{const.} = 1.5°$)

buffer distance (see Fig. 10a) as in Sect. 4b, the localized kernel shows a much higher consistency with a 99% percentile of just about ±10 µGal (in contrast to ±320 µGal when using the original kernel). This improved consistency by a factor of about 30 is also in good agreement with the amplitudes of the deviations shown in Fig. 9. The findings from Sect. 4.1 are statistically substantiated in Fig. 10b where estimates in proximity to observations (<10 km) show just small differences to the original kernel with a 99% percentile of about ±0.3 mGal (cf. also Fig. 7a). These differences are getting gradually larger with the increasing distance to the nearest observation and become as large as about ±3 mGal (99% quantile) for far off estimates (>20km). Together with the evaluations in Sects. 4.1 and 4.2, this is the final evidence that the LSC kernel localization works as expected and produces stable results, where residual fringe effects are strongly reduced and can be further minimized by increasing the buffer distance r.

5 Discussion

In this last section, we briefly discuss the impact of PE-LSC to actual application scenarios regarding the feasibility and the computational effort, and we compare it to other LSA techniques commonly used in this field of study (Sect. 5.1). Subsequently, we conclude the paper by summarizing the benefits of PE-LSC and by discussing remaining disadvantages. Finally, we present some ideas of how to possibly overcome those drawbacks which might be subject of future investigations (Sect. 5.2).

5.1 Consequences for the computational effort

In geodesy, large-scale gravity field modeling using scattered observations is usually performed either applying spectral methods like Fourier analysis, spherical harmonic analysis or statistical approaches such as LSC. In the scattered case, the spectral methods need to be solved using a general least squares adjustment (LSA) approach, since in this case no orthogonality assumptions can be used to accelerate these methods. Comparing classical LSC and LSA it is discernible that both methods are in the same order of magnitude of computational cost (in terms of effort and memory requirements) as both require matrix multiplication and inversion. As mentioned in Sect. 1, the computational effort of those operations is in the cubic order, and memory requirements are in the quadratic order (regarding the number of observation). Solving them becomes numerically expensive with an increasing number of observations (see Fig. 11). In practice, the available memory usually poses a hard limit on the computing feasibility because conventional linear algebra routines cannot exceed the RAM limit of the system (or just with a large penalty in terms of computation time). Therefore, even current-generation high-performance computing

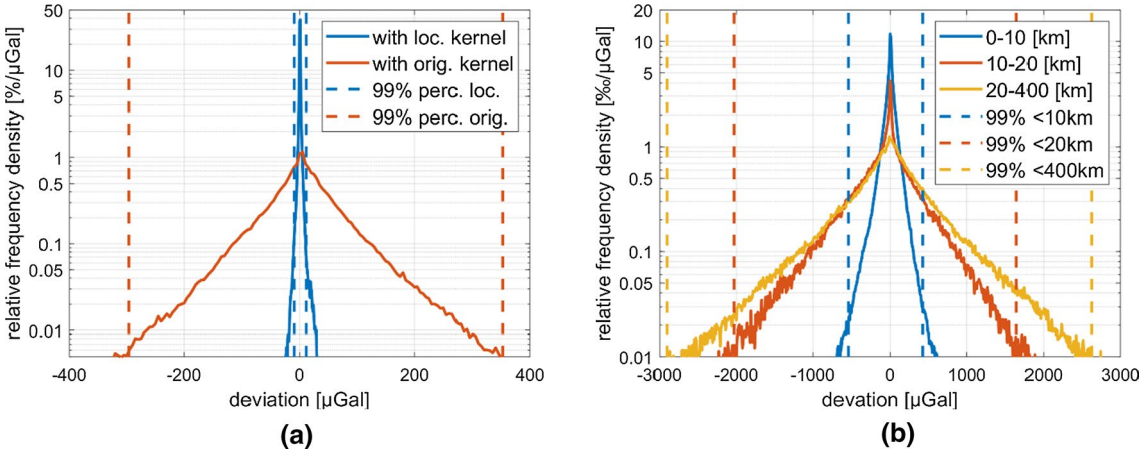


Fig. 10 Statistical comparison between the original and the localized kernel over the entire collocation area (see Fig. 3), covering the entire Antarctica from 60° to 90°S (for estimates on the same grid as in Fig. 7). The frequency distribution shown are generated by computing histograms from the deviations between different solutions using a bin size on a µGal level. A logarithmic scale was chosen to enable the visualization of frequency distributions of a broader range. **a** Frequency distribution of the deviation between the solution with $r = d = 3.0°$ and the solution with $r = d = 2.5°$ ($\theta^{\max} = \text{const.} = 1.5°$, cf. Fig. 9) for the localized kernel (solid blue line) and the original kernel (solid red line). **b** Frequency distribution of the deviation between the solution using the localized kernel ($\theta^{\max} = 1.5°$) and the solution using the original kernel ($r = d = 3.0°$ for both solutions, cf. Fig. 7). Solid blue line: frequency distribution for the set of estimation points where the nearest observation is less than 10 km away. Solid red line: frequency distribution for the set where nearest observation is 10 to 20 km away. Solid yellow line: frequency distribution for the set where the nearest observation is 20 to 400 km away. The dashed lines in figure (**a**) and (**b**) depict the appropriate 99% percentiles (two sided, > 0.5% and > 99.5%)

(HPC) systems, having, e.g., about 256 TB of memory, are limited to approximately 5 million observations. Their computation would take an infeasible time of ~400 years (single core, cf. Fig. 11, dashed blue and dotted green line).

On the other hand, PE-LSC is typically not limited by the system memory when using the partitioning approach and would accomplish the computation of the same 5 million observations in about 2 months (single core, cf. Fig. 11, solid blue line). Having a multi-core workstation CPU (e.g., 64 cores, assuming an ideal linear scale-up) available this task can be performed within a reasonable timeframe of 1–2 days. In the appropriate partitioning scenario, an exemplary but realistic point density of a 4×4 km grid is assumed, where 5 million observations would correspond to a covered area of $8 \cdot 10^7$ km^2 which is roughly the area of Asia and Africa together. Extending this scenario further to a global scale ($\sim 5 \cdot 10^8$ km^2), it would include about 6 times as many points (~30 mio.) and hence, would just need about 6 times as much time to finish for the PE-LSC method, which still yields a reasonable timeframe of 1–2 weeks (on a 64 core workstation CPU). Currently, the same calculation would not be feasible with the classical approach (~100,000 years of single-core computation time

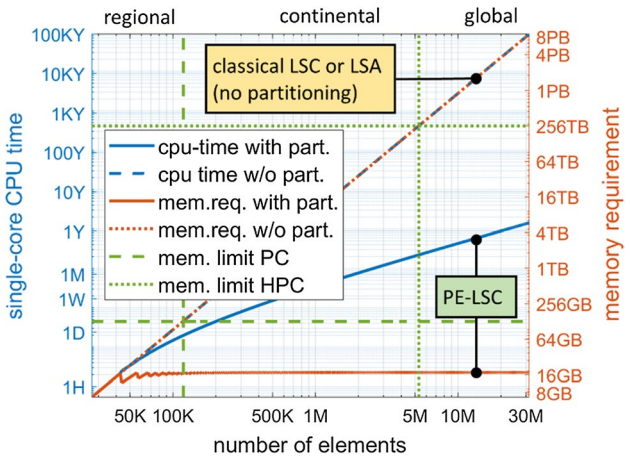


Fig. 11 Computational requirements regarding the number of observations to process (extrapolated) of the classical LSC or LSA approach (unpartitioned) in comparison with the PE-LSC method (partitioned). Single-core CPU times are scaled to the Intel Skylake-SP processor family with AVX512 enabled (using double precision). For the partitioning, a point density of $1/16$ obs./km^2 is assumed (i.e., a regular 4×4 km grid) with the partition parameters $r = d = 2.5° \approx 278$ km. It is assumed that the number of estimates is in the same order as the number of observations. Solid blue: single-core CPU times for the partitioned approach (PE-LSC). Dashed blue: single-core CPU times for the unpartitioned approaches (LSC, LSA). Solid red: memory requirements (RAM) for the partitioned approach (PE-LSC). Dotted red: memory requirements (RAM) for the unpartitioned approaches (LSC, LSA). Dashed green: typical memory limit (RAM) of current-gen. workstation PCs (128 GB). Dotted green: typical memory limit (RAM) of current-gen. high-performance computers (256 TB)

and ~8PB of memory required). This leads to the conclusion, that for general high-resolution and large-scale gravity field modeling only a partitioned LSC approach is suited to perform this task (for scattered observations). With PE-LSC, such an approach is given which not only enables the optimal partitioning, but also tries to minimize fringe effect (cf. Fig. 8) caused by the partitioning. Using PE-LSC in the actual scenario of the AntGG project where more than 800,000 observations must be processed, a solution with approx. 1.4 Mio. estimates can be obtained within a few hours (on a 56 core CPU).

5.2 Conclusion and outlook

We elaborated a partition-enhanced LSC method (PE-LSC) to improve the stability and computational efficiency. The discussed modifications, i.e., the accelerated calculation of covariance matrices and the partitioning strategy together with the kernel localization make it feasible to conduct collocations on large scales with relatively low computational effort (see Sect. 5.1 and Fig. 11) while minimizing fringe effects. In fact, with the shown advancements the effort only increases linearly with the size of the study area, which makes it even possible to think of applying LSC to scattered data on a global scale (cf. Fig. 11).

Although practical limits are not left for PE-LSC (at least not in the shown scenario), there is still one major inherent shortcoming of the (PE-)LSC method to consider, which is the modeling of the covariance function. By now, for PE-LSC there is no practical alternative to homogeneous-isotropic covariance functions, since (1) the information of the actual ('true') covariance function is not available, and (2) there are usually not enough observations present to empirically derive general covariance functions comprehensively. (3) Even if such a general covariance function might be derived empirically, without a spectral or closed-form representation, deriving cross-covariances among different functionals would also be problematic. Using homogeneous-isotropic covariance functions means to disregard the local and directional variations of the investigated physical quantity. In gravity field modeling, for example, even if it is common knowledge that in mountainous regions the expected variation in the gravity field signal is larger, by now this fact is not considered within PE-LSC. This deficiency becomes even more prominent with increasing size of the collocation area as it can be expected that the covered terrain also becomes more heterogeneous (mountains, plains, sea, ice, etc.).

Therefore, for future work in the field of LSC, overcoming this issue is considered as the most important task to make (PE-)LSC a feasible stand-alone alternative even for global gravity field modeling. Darbeheshti (2009) provides a comprehensive overview of studies already performed on

this subject. Unfortunately, none of these seems fully appropriate for PE-LSC as every method has its own limitations. A promising approach for the application in PE-LSC might be the classification of the collocation area into classes of similar behavior (e.g., mountains, hills, plains, ocean, ice, etc.) and a separate estimation of 'local' degree variances (see Sect. 3.2) for each class. A non-homogeneous and non-isotropic covariance function might then be derived by some sort of interpolation between the covariance functions derived from the different degree variances and based on the actual point locations (regarding the previous classification, similar as proposed by Tscherning 1999).

Until such a technique becomes available, the shortcoming of the suboptimal covariance modeling might be overcome by combining (PE-)LSC with the aforementioned spectral methods (as discussed, e.g., by Reguzzoni and Tselfes 2009). While those methods are rigorous in the way they describe the gravity field, they require the observations usually to be given on regular grids in order to be applicable efficiently (not as in the example above, e.g., see Sneeuw 1994). At this point, (PE-)LSC can help by pre-calculating the observations beforehand on the required grid and, thus, producing a best guess of what to expect on those locations. Subsequently, this best guess can be used by the spectral methods to derive the final spectral representation of the gravity field (e.g., cf. Zingerle et al. 2020). If needed, the obtained gravity field model can be further improved iteratively by recalculating reductions for the RCR-LSC from the newly derived model and by restarting the whole procedure anew using the improved (narrowed) reductions.

Appendix

Partial derivatives of homogeneous-isotropic covariances regarding a local ENU-frame

Evaluation of the generalized chain rule (see Sect. 3.1) requires the partial derivatives of the covariance c (cf. Eq. 9) with respect to the chosen local (reference) frame coordinates. In case of gravity field modeling, which usually assumes the chosen local frame to be an ENU-frame (east, north, up) tangential to the sphere, the partial derivatives needed are:

$$c_{ik}^{n,m} := \frac{\partial^{(m+n)} c_{ik}}{\partial t^m \partial u^n} = \sum_{l=0}^{l_{\max}} c_l \frac{(l+1)!}{(l-n+1)!} u_{ik}^{(l-n+1)} P_{lm}\left(t_{ik}\right) \quad (22)$$

in terms of intermediate frame coordinates,

$$\frac{\partial^m \overline{u}_i}{\partial x_{Ui}^m} = \frac{(-1)^m m! R}{r_i^{m+1}}, \quad \frac{\partial^{(m+n)} u_{ik}}{\partial x_{Ui}^m \partial x_{Uk}^n} = \frac{\partial^m \overline{u}_i}{\partial x_{Ui}^m} \frac{\partial^n \overline{u}_k}{\partial x_{Uk}^n} \quad (23)$$

for the up component (x_U) and

$$\frac{\partial^{(m+n)} \boldsymbol{e}_i}{\partial x_{Ei}^m \partial x_{Ni}^n} = \begin{pmatrix} \frac{\partial^m \cos\lambda_i}{\partial \lambda_i^m} \frac{\partial^n \cos\varphi_i}{\partial \varphi_i^n} \\ \frac{\partial^m \sin\lambda_i}{\partial \lambda_i^m} \frac{\partial^n \cos\varphi_i}{\partial \varphi_i^n} \\ \frac{\partial^m (1)}{\partial \lambda_i^m} \frac{\partial^n \sin\varphi_i}{\partial \varphi_i^n} \end{pmatrix} \frac{1}{r_i^n \left(r_i \cos\varphi_i\right)^m}, \quad (24)$$

$$\frac{\partial^{(m+n+o+q)} t_{ik}}{\partial x_{Ei}^m \partial x_{Ni}^n \partial x_{Ek}^o \partial x_{Nk}^q} = \left\langle \frac{\partial^{(m+n)} \boldsymbol{e}_i}{\partial x_{Ei}^m \partial x_{Ni}^n}, \frac{\partial^{(o+q)} \boldsymbol{e}_k}{\partial x_{Ek}^o \partial x_{Nk}^q} \right\rangle$$

for the east component and north component (x_E and x_N). Here, $P_{lm}\left(t_{ik}\right)$ denotes the mth derivate of $P_l\left(t_{ik}\right)$ regarding t_{ik} (not to confuse with the associated Legendre function which is defined slightly different), $\left(\lambda_i, \varphi_i\right)$ are the appropriate spherical coordinates (longitude and latitude) of v_i. In Eq. 22, in opposition to Eq. 9, the letter n for the spherical-harmonic degree was exchanged with the letter l to reuse n for the order of derivative of u_{ik}. Again, it is noteworthy that also partial derivatives are expressible as outer (Cartesian) products of $\underline{v} \times \underline{v}$ which allows for efficient evaluation of all pairs over indices i, k. For the sake of performance and accuracy, it is advisable to pre-calculate all the necessary derivatives $\underline{c}_G^{n,m}$ of the intermediate grid $\underline{c}_G$ (see Eq. 9, in the same fashion as $\underline{c}_G$ is calculated). With some additional programming effort, also linear out-of-scheme quantities (i.e., quantities that are not directly expressible as derivative such as, e.g., gravity anomalies) can be implemented.

Acknowledgements The investigations were supported by funding of the German Research Foundation (DFG), Projects PA 1543/18-1 and SCHE 1426/24-1, within the DFG Priority Program "Antarctic Research" (SPP 1158). Computations were carried out using resources of the Leibniz Supercomputing Centre (LRZ).

Author contributions P.Z. performed research and wrote the paper; R.P., M.W. and M.S. contributed to the final manuscript with suggestions and corrections.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability Data associated with this manuscript will be made available within a future publication which is designated for presenting the results of the corresponding DFG projects.

References

Bell RE, Childers VA, Arko RA, Blankenship DD, Brozena JM (1999) Airborne gravity and precise positioning for geologic applications. J Geophys Res 104(B7):15281–15292. https://doi.org/10.1029/1999JB900122

Darbeheshti N (2009) Modification of the least-squares collocation method for non-stationary gravity field modelling [Doctoral dissertation, Curtin University]. Curtin Theses. http://hdl.handle.net/20.500.11937/2228

de Boor C (1978) A practical guide to spline. Applied Mathematical Sciences, vol 27. Springer, New York

Forsberg R, Tscherning CC (1981) The use of height data in gravity field approximation by collocation. J Geophys Res 86(B9):7843–7854

Hardy M (2006) Combinatorics of Partial Derivatives. Electron J Comb 13(1):R1

Krarup T (1969) A contribution to the mathematical foundation of physical geodesy. Danish Geodetic Institute, Report no. 44, Copenhagen

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

Reguzzoni M, Tselfes N (2009) Optimal multi-step collocation: application to the space-wise approach for GOCE data analysis. J Geod 83:13–29. https://doi.org/10.1007/s00190-008-0225-x

Scheinert M, Ferraccioli F, Schwabe J, Bell R, Studinger M, Damaske D, Jokat W, Aleshkova N, Jordan T, Leitchenkov G, Blankenship DD, Damiani TM, Young D, Cochran JR, Richter TD (2016) New Antarctic gravity anomaly grid for enhanced geodetic and geophysical studies in Antarctica. Geophys Res Lett 43:600–610

Sneeuw N (1994) Global spherical harmonic analysis by least-squares and numerical quadrature methods in historical perspective. J Geophys Res 118:707–716

Studinger M, Bell RE, Finn C, Blankenship D (2002) Mesozoic and Cenozoic extensional tec-tonics of the West Antarctic Rift System from high-resolution airborne geophysical mapping. Roy Soc New Zealand Bull 35:563–569

Studinger M, Bell RE, Fitzgerald PG, Buck WR (2006) Crustal architecture of the Transantarctic Mountains between the Scott and Reedy Glacier region and South Pole from aero-geophysical data. Earth Plan Sci Lett 250(1–2):182–199. https://doi.org/10.1016/j.epsl.2006.07.035

Tscherning CC (1999) Construction of anisotropic covariance functions using Riesz representers. J Geod 73(6):332–336. https://doi.org/10.1007/s001900050250

Tscherning CC, Rapp RH (1974) Closed covariance expressions for gravity anomalies, geoid undulations, and deflections of the vertical implied by anomaly degree variance models. Ohio State University, Department of Geodetic Science and Surveying, OSU Report No. 208, Columbus

Willberg M (2020) Enhanced methodologies of least-squares collocation for the realization of height systems. [Doctoral dissertation, Technical University of Munich]. mediaTUM. http://mediatum.ub.tum.de/node?id=1551033

Willberg M, Zingerle P, Pail R (2019) Residual least-squares collocation: use of covariance matrices from high-resolution global geopotential models. J Geod 93:1739–1757. https://doi.org/10.1007/s00190-019-01279-1

Zingerle P, Pail R, Scheinert M, Schaller T (2019) Evaluation of terrestrial and airborne gravity data over Antarcticaa generic approach. J Geod Sci 9:29–40. https://doi.org/10.1515/jogs-2019-0004

Zingerle P, Pail R, Gruber T, Oikonomidou X (2020) The combined global gravity field model XGM2019e. J Geod 94:66. https://doi.org/10.1007/s00190-020-01398-0

Publication P-3: The combined global gravity field model XGM2019e

Reference
Zingerle P., Pail R., Gruber T. and Oikonomidou X. (2020) The combined global gravity field model XGM2019e. J Geod 94, 66. doi 10.1007/s00190-020-01398-0

Abstract
XGM2019e is a combined global gravity field model represented by spheroidal harmonics up to degree and order (d/o) 5399, corresponding to a spatial resolution of 2' ($\sim$4 km). As data sources, it includes the satellite model GOCO06s in the longer wavelength range up to d/o 300 combined with a ground gravity grid which also covers the shorter wavelengths. The ground data consist over land and ocean of gravity anomalies provided by courtesy of NGA (15' resolution, identical to XGM2016) augmented with topographically derived gravity information over land (EARTH2014). Over the oceans, gravity anomalies derived from satellite altimetry are used (DTU13 with a resolution of 1'). The combination of the satellite data with the ground gravity observations is performed by using full normal equations up to d/o 719 (15'). Beyond d/o 719, a block-diagonal least squares solution is calculated for the high-resolution ground gravity data (from topography and altimetry). All calculations are performed in the spheroidal harmonic domain. In the spectral band up to d/o 719, the new model shows a slightly improved behaviour in the magnitude of a few mm RMS over land as compared to preceding models such as XGM2016, EIGEN6c4 or EGM2008 when validated with independent geoid information derived from GNSS/levelling. Over land and in the spectral range above d/o 719, the accuracy of XGM2019e marginally suffers from the sole use of topographic forward modelling, and geoid differences at GNSS/levelling stations are increased in the order of several mm RMS in well-surveyed areas, such as the US and Europe, compared to models containing real gravity data over their entire spectrum, e.g. EIGEN6c4 or EGM2008. However, GNSS/levelling validation also indicates that the performance of XGM2019e can be considered as globally more consistent and independent of existing high-resolution global models. Over the oceans, the model exhibits an enhanced performance (equal or better than preceding models), which is confirmed by comparison of the MDT's computed from CNES/CLS 2015 mean sea surface and the high-resolution geoid models. The MDT based on XGM2019e shows fewer artefacts, particularly in the coastal regions, and fits globally better to DTU17MDT which is considered as an independent reference MDT.

Declaration of own contribution
(PZ: Philipp Zingerle; RP: Roland Pail; TG: Thomas Gruber; XO: Xanthi Oikonomidou)
RP and TG initiated the study. PZ created the study design and developed the ideas with support of RP and TG. PZ implemented the needed methods and carried out the calculations. The validation of XGM2019e over the ocean was performed by PZ with the help of XO who derived geostrophic currents. The GNSS/levelling validation of XGM2019e over land was calculated by TG with suggestion of PZ. PZ and TG created the interpretation of the results. Figures and text were compiled by PZ with contribution of TG, RP and XO. **The overall own contribution of PZ for *P-3* is estimated at 84 %**, which is the average value of the percentage values estimated for the five criteria listed in the table below (Tab. P.3).

Criteria	**Estimated own contribution**
Computation and results	90 %
Ideas and study design	90 %
Analysis and interpretation	70 %
Text	90 %
Figures and tables	80 %
Total	**84 %**

Tab. P.3 – Criteria and estimated contribution share of Philipp Zingerle for *P-3*

Confirmation by the authors

We hereby confirm the correctness of the declaration of own contribution (of 84 % for Philipp Zingerle) for the publication

Zingerle P., Pail R., Gruber T. and Oikonomidou X. (2020) The combined global gravity field model XGM2019e. J Geod 94, 66. doi 10.1007/s00190-020-01398-0

Roland Pail

Institute of Astronomical and Physical Geodesy, Technical University of Munich, Germany

Signature: ..... Date: 23.09.2021

Thomas Gruber

Institute of Astronomical and Physical Geodesy, Technical University of Munich, Germany

Signature: ..... Date: 7.9.2021

Xanthi Oikonomidou

Institute of Astronomical and Physical Geodesy, Technical University of Munich, Germany

Signature: ..... Date: 30.9.2021

Journal of Geodesy (2020) 94:66
https://doi.org/10.1007/s00190-020-01398-0

ORIGINAL ARTICLE

The combined global gravity field model XGM2019e

P. Zingerle[1] · R. Pail[1] · T. Gruber[1] · X. Oikonomidou[1]

Received: 17 October 2019 / Accepted: 3 July 2020 / Published online: 20 July 2020

Abstract
XGM2019e is a combined global gravity field model represented by spheroidal harmonics up to degree and order (d/o) 5399, corresponding to a spatial resolution of 2′ (~4 km). As data sources, it includes the satellite model GOCO06s in the longer wavelength range up to d/o 300 combined with a ground gravity grid which also covers the shorter wavelengths. The ground data consist over land and ocean of gravity anomalies provided by courtesy of NGA (15′ resolution, identical to XGM2016) augmented with topographically derived gravity information over land (EARTH2014). Over the oceans, gravity anomalies derived from satellite altimetry are used (DTU13 with a resolution of 1′). The combination of the satellite data with the ground gravity observations is performed by using full normal equations up to d/o 719 (15′). Beyond d/o 719, a block-diagonal least squares solution is calculated for the high-resolution ground gravity data (from topography and altimetry). All calculations are performed in the spheroidal harmonic domain. In the spectral band up to d/o 719, the new model shows a slightly improved behaviour in the magnitude of a few mm RMS over land as compared to preceding models such as XGM2016, EIGEN6c4 or EGM2008 when validated with independent geoid information derived from GNSS/levelling. Over land and in the spectral range above d/o 719, the accuracy of XGM2019e marginally suffers from the sole use of topographic forward modelling, and geoid differences at GNSS/levelling stations are increased in the order of several mm RMS in well-surveyed areas, such as the US and Europe, compared to models containing real gravity data over their entire spectrum, e.g. EIGEN6c4 or EGM2008. However, GNSS/levelling validation also indicates that the performance of XGM2019e can be considered as globally more consistent and independent of existing high-resolution global models. Over the oceans, the model exhibits an enhanced performance (equal or better than preceding models), which is confirmed by comparison of the MDT's computed from CNES/CLS 2015 mean sea surface and the high-resolution geoid models. The MDT based on XGM2019e shows fewer artefacts, particularly in the coastal regions, and fits globally better to DTU17MDT which is considered as an independent reference MDT.

Keywords Gravity · Combined gravity field model · Spherical harmonics · Spheroidal harmonics · Full normal equation systems · High-performance computing

1 Introduction

The precise knowledge of Earth's gravity field is crucial for a multitude of geosciences as it can be used to deploy vertical reference frames and to give insights into the distribution of masses in the system Earth. As an example, a precise high-resolution gravity model is fundamental for a global height unification (e.g. Gruber et al. 2012; Ihde et al. 2017) or a consistent sea level analysis (e.g. Woodworth et al. 2012). In oceanographic applications, for instance, the gravity field can be used as a physical reference surface to derive the mean dynamic topography (MDT, e.g. Siegismund 2013). In geophysics, gravity field information is used for lithospheric modelling where it serves as a constraining boundary value (e.g. McKenzie et al. 2014).

Since the emergence of satellite gravity field missions, especially the Gravity Recovery And Climate Experiment (GRACE, Tapley et al. 2004) and the Gravity field and steady-state Ocean Circulation Explorer (GOCE, Drinkwater et al. 2003), the quality of global gravity field models has significantly improved. For instance, GRACE data augmented with a comprehensive collection of ground gravity observations contributed to the widely used high-resolution

✉ P. Zingerle
zingerle@tum.de

[1] Institute of Astronomical and Physical Geodesy, Technical University of Munich, Munich, Germany

combined global gravity field model EGM2008 (Pavlis et al. 2012). Other subsequent high-resolution models like EIGEN6-C4 (Förste et al. 2014) or GECO (Gilardoni et al. 2016) extended EGM2008 later on by additionally including GOCE data, which resulted in a further improvement in the longer wavelengths.

With the introduction of the release 6 series of GOCE gravity field models, new improved combined satellite-only models such as GOCO06s (Kvas et al. 2019b) have emerged. Consequently, this allows the calculation of new high-resolution combined gravity field models, with the intention to optimally merge satellite-only models with terrestrial, altimetric and topographic gravity information, such that each data source keeps its high-quality information without using information from pre-existing high-resolution models. This triggered the development of the high-resolution combined gravity field model XGM2019e.

At the time of writing, only three models next to EGM2008 with maximum d/o larger than 2000 were published at the International Centre for Global Earth Models (ICGEM, Ince et al. 2019). All these models are using EGM2008 as data source above the resolution of satellite models and consequently are highly dependent on that model. With XGM2019e, we present a new model that is independent of EGM2008: as ground gravity data, it solely includes a primary 15′ gravity anomaly grid augmented with topographic information over land and gravity anomalies derived from altimetry over sea (cf. Sect. 2). The inclusion of these augmentation datasets allows the calculation of XGM2019e up to the remarkable resolution of 2′ resp. d/o 5400 (cf. Sect. 3).

In comparison with the precursor model XGM2016 (Pail et al. 2018), some new techniques regarding the combination strategy are introduced with XGM2019e. These include the new and improved empirical weighting approach for the ground gravity observations and a newly developed spectral filter technique in the spatial domain (cf. Sect. 3). Other changes are the application of spheroidal harmonics instead of spherical harmonics and the complete reimplementation of the associated Legendre function routines, which eventually enable the calculation of spherical/spheroidal harmonics up to very high d/o (remark: the term spheroid is used to denote ellipsoids of revolution and to differentiate from arbitrary 3-axis ellipsoids).

2 Data sources

XGM2019e is composed of three main data sources: the combined satellite-only model GOCO06s, the 15′ ground gravity anomaly dataset provided by NGA, and the 1′ min augmentation dataset consisting of gravity anomalies derived from altimetry over the oceans and topography over the continents. All three datasets are briefly presented in the following.

2.1 The GOCO06s satellite-only gravity field model

GOCO06s is a combined satellite-only model consisting mainly of data from the GOCE (TIM6, Brockmann et al. 2019) and GRACE (ITSG-Grace2018s, Kvas et al. 2019a) missions. For the lower-degree coefficients also, satellite laser ranging (SLR) and satellite-to-satellite tracking (SST) observations (using the Global Positioning Satellite (GPS) system) of several other satellite missions were included. For the combination in XGM2019e, the unconstrained normal equation (NEQ) system of GOCO06s up to d/o 300 is used. Since in the original GOCO06s also temporal gravity variations are parameterized, all non-static parameters have been pre-eliminated. The correctness of this procedure is confirmed by comparing the original GOCO06s static coefficients with the recomputed solution by adding back the Kaula regularization to the NEQ system. As expected, the resulting coefficient differences are in the magnitude of numerical precision.

2.2 The 15′ ground gravity dataset

As the primary source for the ground gravity observations (land and oceans), a pre-compiled 15′ (∼ 30 km at the equator) global geographic grid provided by the US National Geospatial-Intelligence Agency (NGA) is used. The dataset has already been spectrally limited beforehand to d/o 719 (in the spheroidal harmonic domain) and reduced to the spheroid's surface. This ground gravity dataset is identical to the dataset used in XGM2016 and is mainly based on observed gravity anomalies over land and DTU13 (Andersen et al. 2013) gravity anomalies derived from altimetry over sea (Fig. 1).

2.3 The 1′ augmentation dataset

To further extend the spatial resolution, a 1′ (∼ 2 km at the equator) global geographic grid of gravity anomalies is compiled in a preliminary step, containing forward-modelled gravity anomalies from topography over land, and gravity anomalies derived from altimetry over sea.

For the topography-derived gravity, the EARTH2014 (Rexer et al. 2017) spherical harmonic model is used up to d/o 5480, and gravity anomalies are synthesized on the 1′ target grid. An in-depth evaluation of the performance of EARTH2014 can be found in Rexer et al. 2016 and Hirt et al. 2017. To reduce the largest uncertainties in the longer wavelengths which are induced by the hypothetical density assumptions in the course of the numerical forward modelling, EARTH2014 is combined with GOCO06s in the

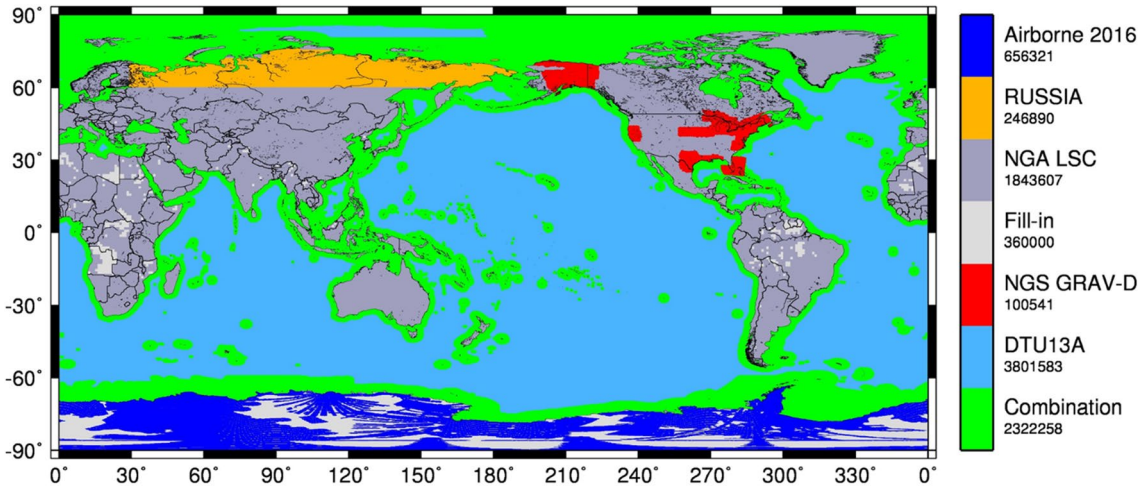


Fig. 1 Data composition of NGA's primary 15′ dataset (see Pail et al. 2018). Dark blue: airborne gravity datasets over Antarctica (Scheinert et al. 2016). Orange: contributed pre-compiled dataset covering parts of Siberia (Pavlis et al. 2012). Dark grey: areas processed (collocated) by the NGA (Pail et al. 2018). Light grey: areas with sparse/inaccurate data coverage where topographic information is filled in (EARTH2014, Rexer et al. 2017). Red: GRAV-D airborne data over North America from the US National Geodetic Survey (e.g. Li et al. 2016). Light blue: altimetric gravity anomalies derived from DTU13 (Andersen et al. 2013). Green: Combination of DTU13 and other datasets (Pail et al. 2018)

spectral domain beforehand. For more details regarding the applied combination strategy used for the combined model (SATOP1), the reader is referred to Zingerle et al. 2019b. The combination with the satellite model is necessary to minimize the (long wavelength) gravity anomaly differences to the altimetry dataset in the coastal areas because they would hamper the subsequent merging of both datasets.

To be consistent with the altimetric data of the primary 15′ dataset, the same DTU13 gravity anomalies are used within the augmentation dataset, but now using the full resolution (1′). Both the altimetric and topographic datasets are combined using a land–ocean mask derived from the GSHHS database (Wessel and Smith 1996). In order to avoid discontinuities at the coastline (i.e. to smoothen the land–ocean transition), a distance-dependent tapering function is applied, which linearly decreases until 30 km into the ocean. The distance of 30 km is chosen because it roughly corresponds to the resolution of the primary 15′ dataset. The tapering is necessary not only to avoid high-frequency artefacts induced by such discontinuities but also to account for the decreasing quality of altimetric observations towards the coasts.

3 Combination strategy and results

The calculation of the XGM2019e model can conceptually be split into two steps: Firstly, the combination of the GOCO06s with the primary 15′ ground gravity dataset leading to a d/o 719 model is referred to as XGM2019, as it logically succeeds the XGM series of gravity field models. Secondly, and independently of XGM2019, the higher-resolution part of the model in the spectral band between d/o 720 and 5480 can be derived through the augmentation dataset.

3.1 Calculation of XGM2019 (up to d/o 719)

The calculation of the XGM2019 spheroidal harmonic model coefficients up to d/o 719 consists of a weighted least squares adjustment of GOCO06s with the primary 15′ NGA ground gravity dataset. Introducing individual point weights into the least squares adjustment (LSA) approach leads to the loss of the orthogonality of spheroidal harmonics and therefore results in a dense normal equation system with more than 500.000 unknowns (cf. Pail et al. 2018).

One of the challenges of this combination method is the realistic choice of ground gravity observation error variances $\mathrm{evar}_{\mathrm{gr}}$ relative to the satellite component. In the previous model GOCO05c (Fecher et al. 2017), variance component estimation was used for a relative weighting of different regional data sets of the world. In XGM2016, the regionally varying relative weights were empirically derived from differences of the ground gravity dataset with the global satellite-only model GOCO05s up to d/o 200. The main drawback of latter method is that only their "commission error" was considered, but the gravity signal and corresponding errors beyond this cut-off degree were not taken into account. Therefore, for XGM2019 the computation strategy has been improved in order to include also errors of the higher-frequency signals: the comparison to the 15′ ground observations $\delta g_{\mathrm{ground}}$ is now performed using a preliminary solution of XGM2019 up to d/o 719, called XGM19a in

the following. XGM19a is obtained by using the same LSA approach as for XGM2019 by combining GOCO06s and δg_{ground}, but assigning an equal error of 2 mGal to all observations of the ground gravity grid. Based on gravity anomalies δg_{XGM19a}, that are synthesized from XGM19a, local error variances $\text{evar}_0(\theta, \lambda)$ are estimated applying a Gaussian kernel function W_G with a filter strength of 50 km half-width at half-maximum (HWHM, empirically derived):

$$\text{evar}_0(\theta, \lambda) \approx \left(W_G * \left(\delta g_{\text{XGM19a}} - \delta g_{\text{ground}} \right)^2 \right)(\theta, \lambda) \tag{1}$$

Since the solution of XGM19a converges strongly towards δg_{ground} above the resolution of the satellite model, δg_{XGM19a} differs from δg_{ground} mainly in the spectral band of the satellite model (up to about d/o 200). Correspondingly, also the thereby derived error variances evar_0 only differ in this spectral band (similar to XGM2016). Therefore, they are not fully representative for the ground observations which contain the full spectral content up to d/o 719. To account for this and to restore the missing spectral content within the local error variances, an omission/commission (signal) variance ratio $f_{\text{o/c}}$ is additionally estimated and applied to the uncorrected error variances evar_0, such that:

$$\text{evar}_{\text{gr}}(\theta, \lambda) = f_{\text{o/c}}(\theta, \lambda) \text{evar}_0(\theta, \lambda) \tag{2}$$

The location-dependent factor $f_{\text{o/c}}(\theta, \lambda)$ is determined by the ratio of signal variances:

$$f_{\text{o/c}}(\theta, \lambda) = \frac{\text{var up to model res.}}{\text{var in satellite band}} = \\ = \frac{\text{var}_{\text{XGM19a}}(\theta, \lambda)}{\text{var}_{\text{sat}}(\theta, \lambda)} \approx \frac{\left(W_G * \delta g^2_{\text{XGM19a}} \right)(\theta, \lambda)}{\left(W_G * \delta g^2_{\text{rsat}} \right)(\theta, \lambda)}. \tag{3}$$

Like the local error variances evar_0, also the local signal variances $\text{var}_{\text{XGM19a}}$ and var_{sat} are derived using the same Gaussian filter (cf. Eqs. 1, 3). For the estimation of var_{sat}, the satellite-influenced spectral part of the XGM19a solution is needed. In the scope of LSA theory, this part is described by the redundancy component of the satellite system (GOCO06s) within XGM19a. Applying this satellite redundancy component to the solution of XGM19a (through multiplication) leads to the gravity anomalies δg_{rsat} which are assumed to contain the same spectral content as the error variances evar_0. The basic assumption of this scaling approach is that the error variances of the ground observations have no spectral correlations and therefore scale identically to the signal variances in the spatial domain. Figure 2 shows the resulting ground gravity error estimates $\sigma_{\text{gr}} = \sqrt{\text{evar}_{\text{gr}}}$.

Finally, we note that the LSA for XGM2019 is performed in the spheroidal harmonic domain, implying that the whole GOCO06s normal equation system has been transformed to the spheroidal harmonic domain beforehand (Jekeli 1988). After the LSA, the entire variance–covariance matrix is transformed back to spherical harmonics to provide the final formal errors.

3.2 Calculation of the coefficients above d/o 719

The higher d/o coefficients for XGM2019e are calculated solely from the augmentation dataset, using the rigorous block-diagonal technique (Sneeuw 1994). As this high-resolution dataset is primarily created in the spatial domain (cf. Sect. 2.3), it cannot be assumed that it is spectrally limited to a certain d/o. Thus, trying to perform a spheroidal harmonic analysis will inevitably introduce aliasing as well as

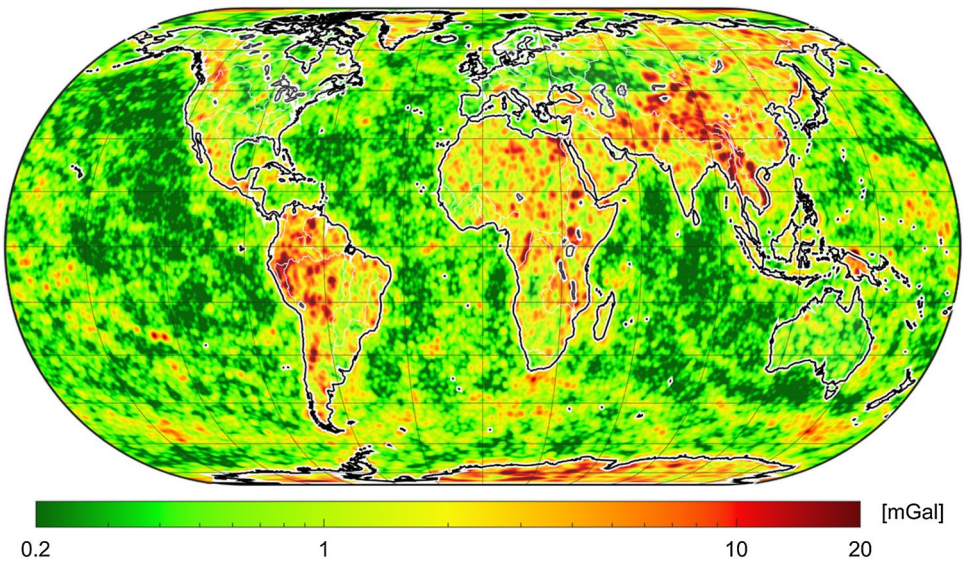


Fig. 2 Standard deviations assumed for the NGA ground gravity dataset in the LSA. As the dataset itself, the deviations are provided in terms of gravity anomalies. Colours are scaled logarithmically

spectral leakage effects (even when using the block-diagonal technique, as the 2:1 relation of observations to estimated coefficients is inevitable on geographic grids).

To avoid this, we introduce a three-step filter strategy, called the SLASH approach (**S**patial **L**ow pass—**A**nalysis—**S**pectral **H**igh pass): in the first step, a Gaussian filter is applied to the combined grid with the aim to remove most of the grid's content beyond the Nyquist wavelength ($< 1'$ in this case). In the second step, the resulting low-pass filtered $1'$ grid can be safely analysed up to d/o 10700 (cf. Fig. 3, green line). Due to the existence of an analytical correspondence of the Gaussian filter in the spatial and spectral domain (Jekeli 1981), the influence of the filter on the analysed signal can be reverted in the spectral domain (step three). This is done by multiplying the appropriate spectral Gaussian filter factors (one per degree, cf. Fig. 3, orange line) to the analysed coefficients, forming the restored signal (cf. Fig. 3, blue line). This signal represents an optimally low-pass filtered result in the sense that the obtained coefficients (1) exactly match the unfiltered coefficients in the case when the grid is spectrally limited beforehand and (2) are minimally influenced by aliasing and spectral leakage effects from signals beyond the Nyquist frequency otherwise. Figure 3 visualizes this spectral limitation procedure for the $1'$ augmentation dataset in the spheroidal harmonic domain (the findings of Jekeli (1981) are also valid for spheroidal harmonics). Please note the increase in the signal degree variances of the result above d/o ~ 3000 is attributed to the increasing

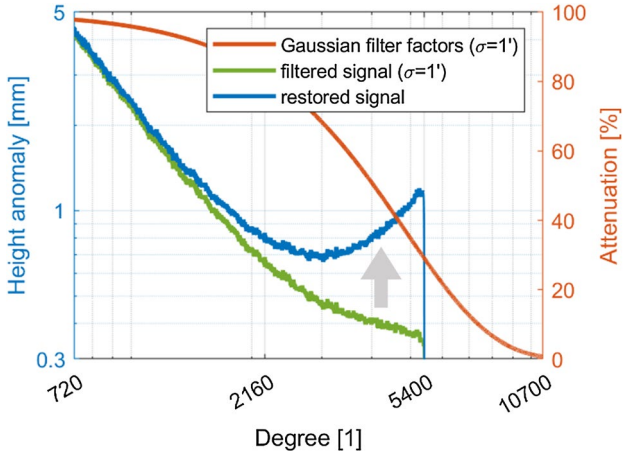


Fig. 3 Illustration of the applied spectral limitation procedure in the spheroidal harmonic domain. Orange: the degree-dependent factor of a Gaussian filter with $\sigma = 1'$. Green: signal degree variances in terms of height anomalies after the analysis (up to d/o 10700) of the spatially filtered grid ($\sigma = 1'$). As the attenuation of the filtered signal at d/o 10700 is very high, aliasing and spectral leakage effects are minimized. Blue: final signal degree variances after rescaling using the inverse Gaussian filter factors. In the case of the augmentation dataset, degree variances above d/o 5400 are very small (< 0.01 mm) for filtered and restored signal degree variances

divergence of the spheroidal harmonic series expansion (cf. Sect. 3.3) and not related to the SLASH procedure.

3.3 Compilation of XGM2019e

Since the topographic data (EARTH2014) are limited to d/o 5480, the high-degree coefficients (that were calculated to d/o 10700) are also only used up to this d/o, resulting in the target resolution of XGM2019e. Theoretically, a model up to d/o 10700 would be feasible, but due to the limited resolution of EARTH2014 and the fact that DTU13 gravity anomalies do not show any geophysical signal content beyond d/o 5400 (as DTU13 is filtered to 6.5 km HWHM, see Andersen et al. 2013), it has been decided to cut the final model at d/o 5480.

In the last step, XGM2019e is created by merging the coefficients of XGM2019 up to d/o 719 and the high-resolution coefficients starting at d/o 720. No tapering function is applied. This is possible due to the use of spheroidal harmonics, since the reference surface of the expansion (spheroid) coincides with the surface where the ground data are located, and thus, cut-off effects as experienced with spherical harmonics are avoided. Additionally, to avoid further artefacts, the $1'$ augmentation dataset was compiled with the requirement to be most compatible with the primary NGA dataset. Therefore, the same altimetric gravity anomalies DTU13GRA were used, a wide tapering function towards the land was chosen, and a similar replacement of the longest wavelengths took place, cf. Sect. 2.3. Finally, the whole spectrum of XGM2019e is transformed into the spherical harmonic domain (Jekeli 1988).

The error degree variances of XGM2019e beyond d/o 719 are derived by comparing the high-resolution coefficients $c_{\mathrm{lm}}^{\mathrm{AUG}}$ (from the augmented dataset, cf. Sect. 3.2) to the XGM2019(e) coefficients $c_{\mathrm{lm}}^{\mathrm{XGM2019}}$ in the lower band (up to d/o 719, cf. Sect. 3.1) and applying an extrapolation function. The extrapolation is performed through a first-order polynomial fit (of parameters a_0, a_1) in the double logarithmic domain:

$$\mathrm{evar}_l^{\mathrm{AUG}} \sim \sum_{m=-l}^{l} \frac{\left(c_{\mathrm{lm}}^{\mathrm{AUG}} - c_{\mathrm{lm}}^{\mathrm{XGM2019}}\right)^2}{2l+1}, \quad \mathrm{var}_l^{\mathrm{AUG}} := \sum_{m=-l}^{l} \frac{\left(c_{\mathrm{lm}}^{\mathrm{AUG}}\right)^2}{2l+1}$$

$$\mathrm{Fit}\left(a_0, a_1\right) \text{with } l \leq 719 \text{ in } : \log\left(\frac{\mathrm{evar}_l^{\mathrm{AUG}}}{\mathrm{var}_l^{\mathrm{AUG}}}\right) = a_0 + a_1 \log\left(l\right)$$

$$\text{Extrapolate } l > 719 : \mathrm{evar}_l^{\mathrm{XGM2019e}} = \log^{-1}\left(a_0 + a_1 \log\left(l\right)\right) \mathrm{var}_l^{\mathrm{AUG}} \tag{4}$$

where c_{lm} denotes a spheroidal harmonic coefficient of degree l and order m. The estimated error degree variances evar_l are then transformed to the spherical domain and

merged with the XGM2019 error degree variances, forming the final error degree variances of XGM2019e. The results in terms of degree signals and errors in the spherical harmonic domain are depicted in Fig. 4. Important to note here is:

(1) The good agreement of the coefficient differences with the formal errors of GOCO06s in the spectral band of the transition (above d/o 100), which proves the correctness of the LSA approach.
(2) The convergence of XGM2019 solution towards a (block-diagonal) ground-only solution with increasing d/o (especially above d/o 500).
(3) The jump in the formal error at d/o 720 that is inevitable due to the change of the data source.
(4) The slowed convergence (commencing divergence) of the harmonic series in the highest degrees as the Earth's surface exceeds the Brillouin sphere (in the spherical case) resp. spheroid (in the spheroidal case) due to high elevations in the topography. The manifestation of this effect is stronger in terms of spheroidal harmonics as the reference spheroid is on average further below Earth's surface than the reference sphere (since its radius is per definition the semi-major axis). This is discernible by comparing the blue line of Fig. 3 (spheroidal case, already diverging) with the blue line of Fig. 4 (spherical case, still converging).
(5) The increase in the formal error in the highest degrees which is an effect of the spherical harmonic transformation, as it is not visible in the spheroidal harmonic domain. Eventually, this is also related to the spheroidal harmonic series divergence.

4 Validation and discussion

The fact that XGM2019e only contains forward-modelled topographic gravity anomalies over land beyond d/o 719 likely shifts the primary use of the model more towards oceanographic applications. With its 2′ resolution, it covers all the signal present in the DTU13 gravity anomaly dataset (cf. Sect. 3.3). Together with the weighted LSA combination strategy (cf. Sect. 3), this leads to an improved oceanic geoid which allows the derivation of a higher-quality MDT (Fig. 5b). The MDT itself is a central parameter of the maritime system and inherently connected to dynamic processes like ocean currents (Fig. 5d–f), heat transport and sea level rise. Hence, it is of major importance for Earth system sciences like oceanography or climatology and may ultimately lead to a deeper comprehension of the Earth system as a whole, thus also allowing for better predictions in the future.

4.1 MDT comparisons

To demonstrate the model's performance over the ocean, geostrophic currents are derived by comparing the XGM2019e geoid to the independent CNES/CLS 2015 mean sea surface (Schaeffer et al. 2016, see Fig. 5) and thereby generating an MDT: it is clearly discernible that the unfiltered MDT derived from XGM2019e (Fig. 5b) shows fewer artefacts and delivers a smoother result than the MDT derived from EGM2008 (Fig. 5a, the most comparable model in terms of performance, all other high-resolution models show even larger artefacts). Consequently, the

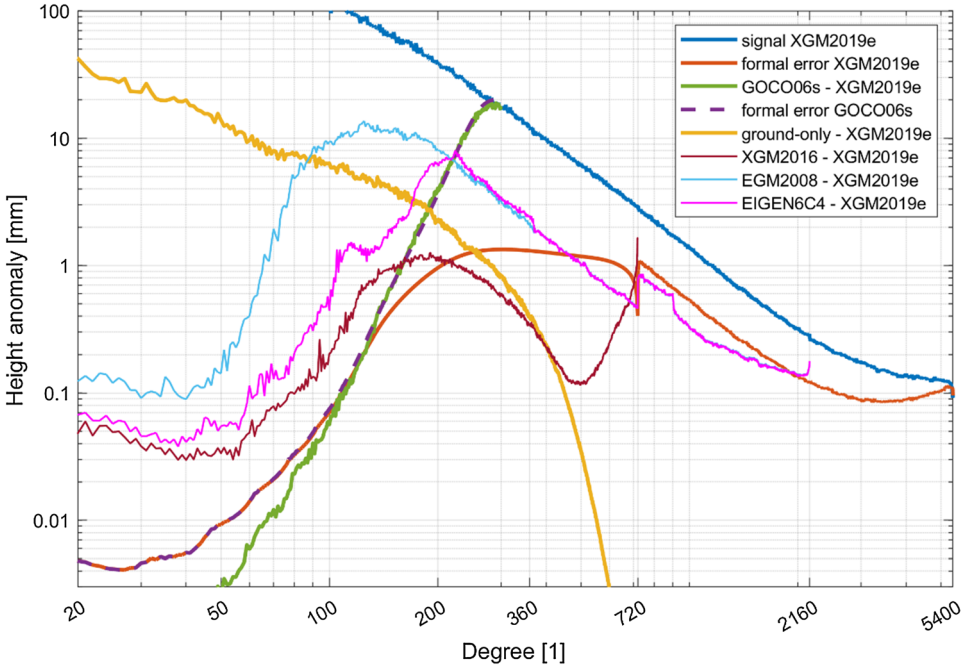


Fig. 4 Spherical harmonic degree signals and errors in terms of height anomalies. Dark blue: signal XGM2019e. Light red: formal error XGM2019e. Green: signal difference GOCO06s-XGM2019e. Dashed violet: formal error GOCO06s. Yellow: signal difference of a ground-only solution versus XGM2019e. The ground-only solution is obtained using the unweighted block-diagonal analysis technique. Dark red: signal difference XGM2016-XGM2019e. Light blue: signal difference EGM2008-XGM2019e. Magenta: signal difference EIGEN6C4-XGM2019e

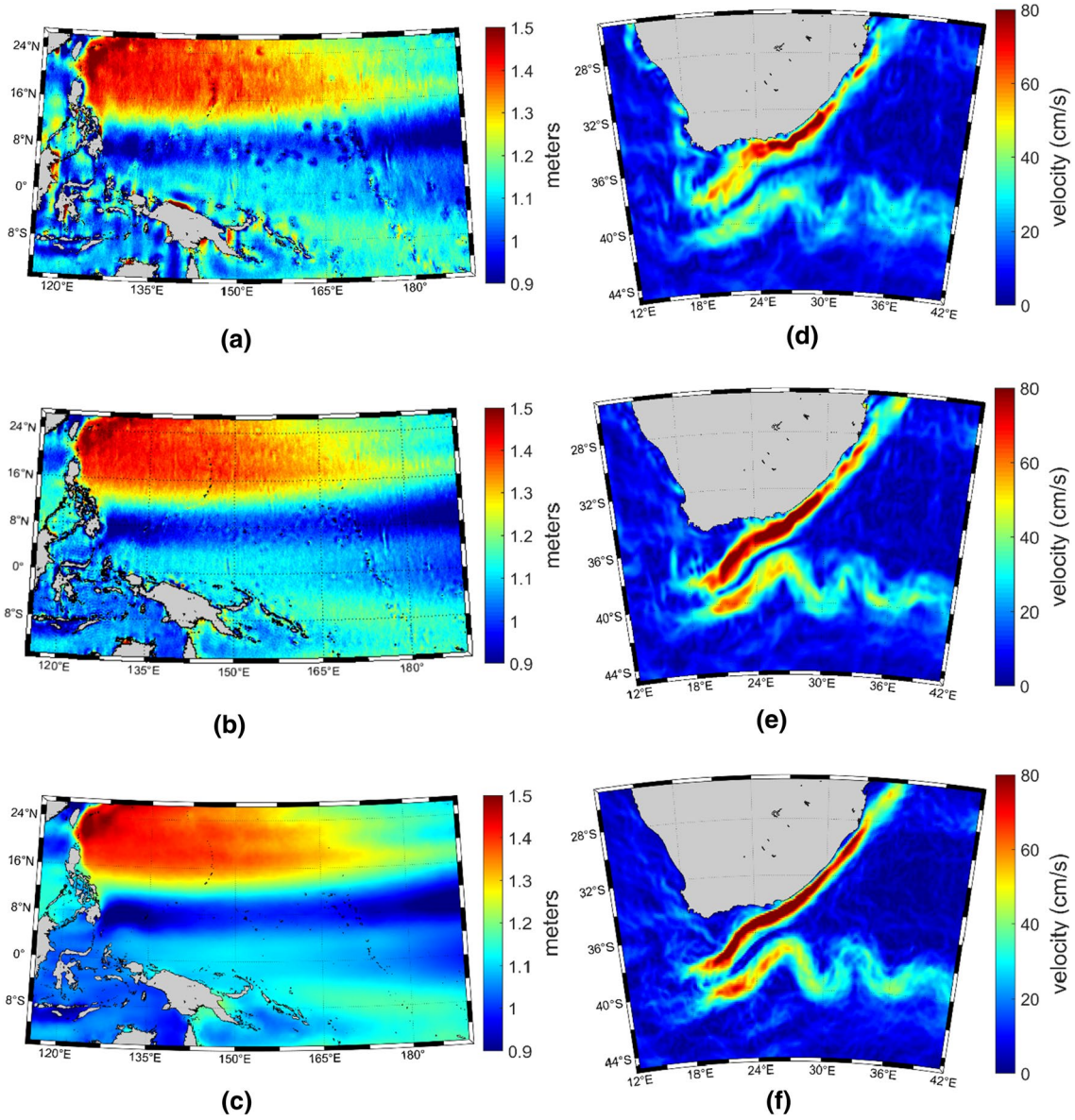
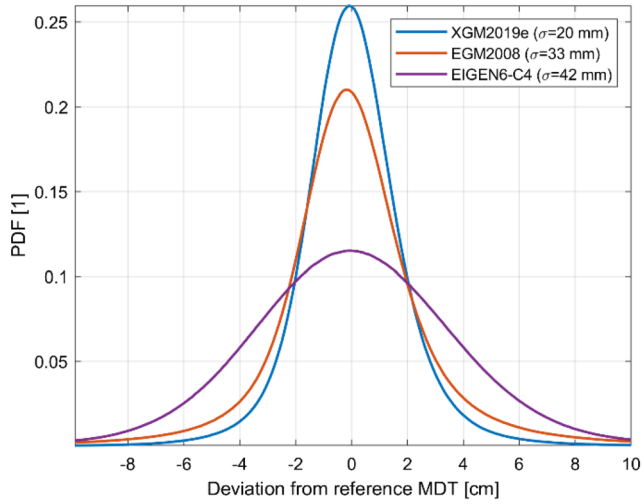


Fig. 5 Unfiltered geodetic MDT solutions from the CNES/CLS 2015 means sea surface and the EGM2008 (**a**) resp. XGM2019e (**b**) ocean geoid (shown for the Western Pacific). **c** reference MDT obtained from the drifter-optimized DTU17MDT (Knudsen et al. 2018). Geostrophic current velocities for the Agulhas current derived from (identically) Gaussian-filtered EGM2008 MDT (**d**) resp. XGM2019e MDT (**e**). **f** reference geostrophic current velocities from DTU17MDT

improved MDT allows one to derive improved geostrophic currents, as is shown in the comparison of the Agulhas current (Fig. 5d, e). This visual evidence is proved by comparing both MDTs to the drifter-optimized DTU17MDT (Fig. 5c, Knudsen et al. 2018). Within the open ocean (up to 60° northern/southern latitude and 30 km away from coasts), the XGM2019e-derived MDT shows a global standard deviation to DTU17MDT of 2.02 cm, while the MDTs derived from EGM2008 and EIGEN6-C4 have standard deviations of 3.34 cm and 4.25 cm, respectively. This means that the performance in the ocean has improved by ~ 40% in comparison with EGM2008 (~ 52% to EIGEN6-C4) when validating against DTU17MDT. A statistically complete evaluation of the comparisons to DTU17MDT is found in Fig. 6a in terms of empirically derived probability density functions (PDFs).

Please be aware that due to the use of the altimetric gravity anomalies (DTU13A) in the ground observations (cf. Fig. 1), the XGM2019e model may be biased to a certain degree to the a priori MDT used within DTU13GRA. In the course of the processing of DTU13A, an a priori MDT is removed from the mean sea surface up to d/o 100 (cf.



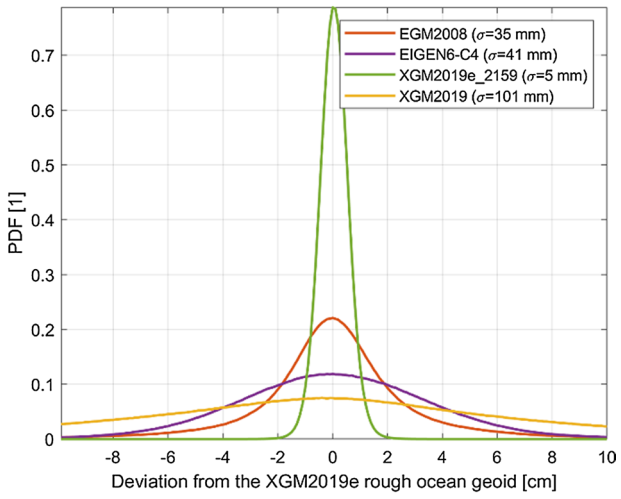


Fig. 6 **a** Empirically derived probability density functions (PDFs) for deviations of unfiltered geoid derived MDTs (using CNES/CLS 2015 MSS) from the reference DTU17MDT. Mean values of MDT deviations were eliminated beforehand. Blue: deviation of the XGM2019e derived MDT. Red: deviation of the EGM2008 derived MDT. Violet: deviation of the EIGEN6-C4 derived MDT. **b** Empirically derived PDFs for deviations from the XGM2019e geoid calculated from the roughest 10% of the ocean's geoid surface. Red: deviation of the EGM2008 derived geoid. Violet: deviation of the EIGEN6-C4 derived geoid. Green: deviation of the XGM2019e (up to d/o 2159) derived geoid. Yellow: deviation of the XGM2019 (up to d/o 719) derived geoid

Andersen et al. 2013). Since up to d/o 100 XGM2019e is dominated by the satellite model (cf. Fig. 4), the bias towards this a priori MDT is completely removed. Nevertheless, there is still a bias to expect, as removing an MDT up to d/o 100 implies to neglect all MDT signal present above d/o 100. Through the inclusion of GOCO06s data in the XGM2019e model, it is possible to restore the MDT signal up to the satellite resolution of d/o ~ 200 (cf. Figs. 4, 5b), meaning that all actual MDT signal above this resolution remains as bias within XGM2019e. As the actual MDT has in general a long-wavelength character (cf. Fig. 5c), the magnitude of this bias can be considered as small (cf. differences to DTU17MDT), but one may expect that an MDT derived from XGM2019e is still somewhat too smooth compared to the actual MDT.

Even though the ocean's geoid can be considered smooth compared to the land geoid (since the signal of the seabed gets attenuated due to upward continuation of the gravity field onto the ocean's surface), there is still significant signal left above d/o 2159. This can be observed in XGM2019e especially over rough seabed structures (e.g. oceanic trenches or ridges) where deviations in the ocean's geoid can reach up to about 2 cm when neglecting gravity field signal above d/o 2159 (cf. Fig. 7). In the proximity of such rough structures, one can expect a global standard deviation of about 5 mm induced by the residual signal above d/o 2159 (cf. Fig. 6b, green line). This standard deviation increases strongly to about 10 cm when further reducing the maximal spectral resolution of XGM2019e to d/o 719 (cf. Fig. 6b, yellow line).

4.2 GNSS/levelling performance

GNSS/levelling-derived geoid comparisons pose a good way to evaluate the performance of global models over land. For this, we apply the procedure described in Gruber and Willberg 2019 and compare four models with some regional GNSS/levelling geoid heights. It is noted that the XGM2019e model has only been used up to a spectral resolution of d/o 2190 (see also below) in order to be comparable with the resolution of EGM2008 and EIGEN6-C4. From Fig. 8, which shows the RMS differences between a regional GNSS/levelling-derived geoid data set and geoid heights computed from the models, the following observations can be made:

(1) In case of XGM2019e, the RMS of geoid height differences to the GNSS/levelling values is always constant for all model truncation degrees because this model was also used for estimating the omission error above this degree and order. In case the RMS of geoid differences for other gravity models is below this line, a model performs better than XGM2019e and vice versa.
(2) In areas where one can assume high-quality ground observations (Fig. 8a–d), the XGM2019e model up to

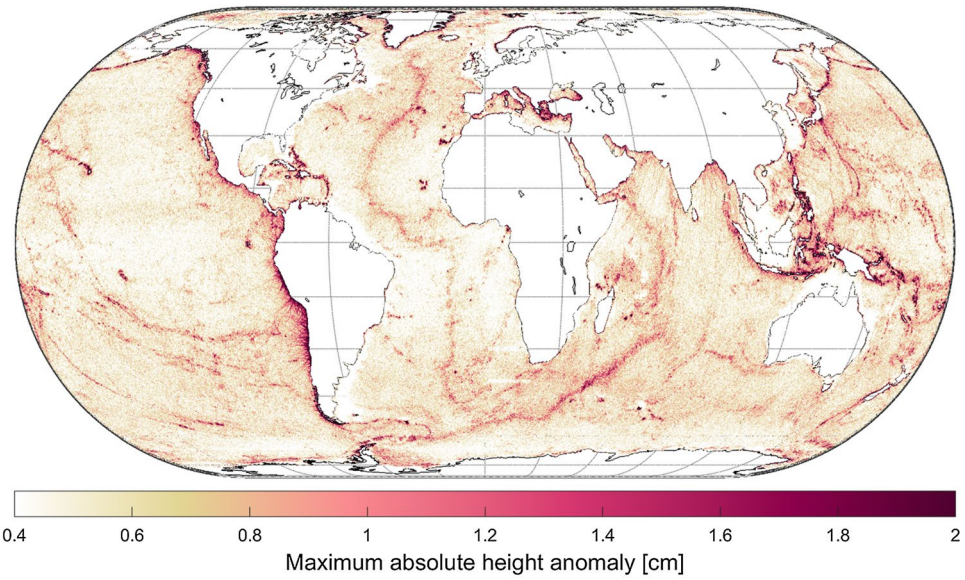


Fig. 7 Residual signal of XGM2019e above d/o 2159 over the ocean in terms of maximum absolute height anomalies (i.e. geoid heights) on a 5′ raster

d/o 719 performs better than all other tested models in most cases. This is rather obvious for the EGM2008 model, for which the RMS of geoid height differences is significantly higher than for XGM2019e in all areas. As this is observed already for low truncation degrees (e.g. between d/o 50 and 200), one can assume that the largest impact stems from the inclusion of GOCE data, which was not yet available at the time of EGM2008. In these areas, XGM2019e also performs slightly better than XGM2016 indicating some improved modelling as for both models used an identical land data set. For EIGEN6-C4, one can identify that for 3 areas XGM2019e outperforms this model as well, which is mainly due to the use of the latest GOCE solution (release 6 instead of release 5) and probably also due to an improved modelling approach. Only for Australia (Fig. 8d) EIGEN6-C4 provides slightly better results.

(3) When looking to the RMS of geoid differences in well-observed areas for degrees above 719, it becomes obvious that EGM2008 and EIGEN6-C4 outperform XGM2019e. This is visible in Fig. 8a–d as the reduced RMS for both models between degree 720 and their full resolution. This clearly shows the impact of using observed instead of topography-derived gravity anomalies in the range between d/o 719 and d/o 2190 in EGM2008 and EIGEN6-C4.

(4) In less well-surveyed areas one can identify from the geoid differences RMS that the XGM2019e model outperforms all other models (Fig. 8e, f). This shows on the one hand again the impact of the GOCE satellite data (up to d/o 200) and on the other hand that topography-derived gravity anomalies can provide better information than less accurate ground gravity data. For example, in Brazil the RMS reduction from the EGM2008 model to the XGM2019e model is at a remarkable level of 7 cm. The major part of this (about 80% of the total reduction) can be attributed to the GOCE data, which is nicely shown by the EIGEN6-C4 performance in the lower degrees and the remaining reduction results from improved gravity data in the area including the topography-derived gravity anomalies for the very high frequencies.

In summary, one can conclude from the GNSS/levelling comparisons that despite the use of topographic information, XGM2019e shows a solid performance even over land. For the longer wavelengths up to d/o 719, XGM2019e exhibits a slightly better performance as compared to previous models. Above d/o 719 and in well-surveyed areas, XGM2019e cannot fully compete with EGM2008 and EIGEN6-C4 due to the lack of gravity measurements with a higher resolution than 15′. In areas with poor data coverage/quality, the performance of XGM2019e might be considered as identical, or even better than that of other models. Thus, it can be stated that XGM2019e performs globally more consistently than other models (i.e. accuracies within XGM2019e vary less than accuracies within EGM2008 or EIGEN6-C4, cf. Fig. 8). This seems reasonable since the topographic information used within XGM2019e is available globally with nearly constant quality while the availability (and quality) of direct gravity field measurements is strongly location dependent. This better consistency and the fact that gravity field information provided within XGM2019e is globally available up to ~ 4 km can be important, e.g. for a more consistent gravity field reduction in the frame of a compute–remove–restore process of regional gravity field modelling (especially in areas where the terrestrial data quality is low or data access is restricted). In the scope of

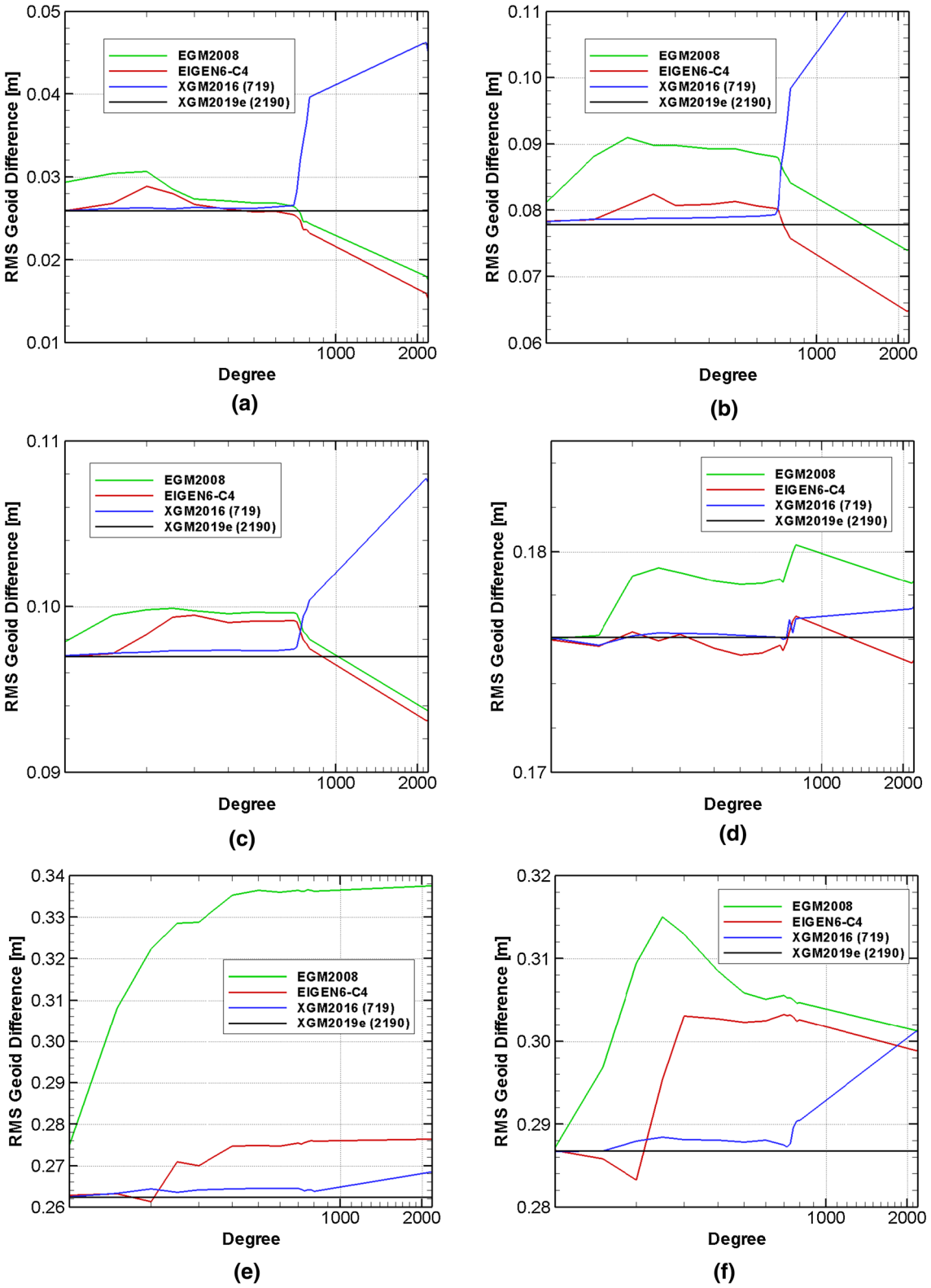


Fig. 8 GNSS-levelling performance of XGM2019e, XGM2016, EGM2008 and EIGEN6-C4 for different spectral resolutions up to d/o 2190 (lower RMS values of geoid differences mean better performance, for details see Gruber and Willberg 2019). RMS of geoid differences **a** in Germany, **b** in Japan, **c** in the US, **d** in Australia, **e** in Brazil, **f** in Mexico. GNSS/levelling comparisons in other regions are also performed. As they support the drawn conclusions and thus give no further insights, they are not shown within this work

regional gravity field modelling, it is also noteworthy that for XGM2019e full variance–covariance information is available up to d/o 719 which can be used for more realistic error propagations into the spatial domain to improve regional modelling approaches (see e.g. Willberg et al. 2019).

4.3 Provision of XGM2019e

To comply with the existing standard, XGM2019e is published as spherical harmonic model. As a matter of fact, it is problematic to truncate the spherical harmonic series expansion at a certain degree when evaluating functionals on a spheroid since artefacts are introduced (cf. Pavlis et al. 2012). In order to avoid these problems, spectral truncations must be performed in the spheroidal harmonic domain. The resulting procedure is considered problematic for most end-users, as they may not be familiar with the theory of spheroidal harmonics and/or the spectral transformation formulas involved. It is therefore decided to provide the model pre-calculated at three different spectral resolutions: d/o 5540, 2190 and 760. All three resolutions are available on ICGEM (Zingerle et al. 2019a).

5 Outlook

It is planned to continue the series of high-resolution XGM models in the future. On the processing branch, improvements can still be made by increasing the maximum d/o of the densely modelled part (soon up to d/o 2159, as new supercomputing resources at the Leibniz Supercomputing Center are available). One point to focus within the preparation of future models will be the task of data acquisition resp. compilation. As of now (January 2020), access to gravity field information is still restricted in many regions of the world, so the outcome of this endeavour is open. The situation over land might improve after the public release of EGM2020. In the oceanic regions, we are confident to achieve further enhancements in the future by switching to updated altimetric products and changing the processing strategy. As an example, it is planned to directly use mean sea surface heights and individual MDT products to derive the gravity field instead of using pre-calculated gravity anomalies.

Acknowledgements Open Access funding provided by Projekt DEAL. The work presented in this paper was performed in the framework of the project "GOCE High-Level Processing Facility" (GOCE-HPF), funded by the European Space Agency (ESA main Contract No. 18308/04/NL/MM). Ground observation dataset was provided by courtesy of the US National Geospatial-Intelligence Agency (NGA). Computations were carried out using resources of the Leibniz Supercomputing Centre (LRZ).

References

Andersen O, Knudsen P, Kenyon S, Factor J, Holmes S (2013) The DTU13 global marine gravity field—first evaluation. Ocean Surf Topogr Sci Team Meet, Boulder, Colorado

Brockmann JM, Schubert T, Mayer-Gürr T, Schuh W-D (2019) The Earth's gravity field as seen by the GOCE satellite: an improved sixth release derived with the time-wise approach (GO_CONS_GCF_2_TIM_R6). GFZ Data Serv. https://doi.org/10.5880/ICGEM.2019.003

Drinkwater MR, Floberghagen R, Haagmans R, Muzi D, Popescu A (2003) GOCE: ESA's first earth explorer core mission. In: Beutler G et al (eds) Earth gravity field from space—from sensors to earth science, space sciences series of ISSI, vol 18. Kluwer Academic Publishers, Dordrecht, pp 419–432. ISBN 1-4020-1408-2

Fecher T, Pail R, Gruber T, the GOCO Consortium (2017) GOCO05c: a new combined gravity field model based on full normal equations and regionally varying weighting. Surv Geophys 38:571–590. https://doi.org/10.1007/s10712-016-9406-y

Förste C, Bruinsma S, Abrikosov O, Flechtner F, Marty J-C, Lemoine J-M, Dahle C, Neumayer H, Barthelmes F, König R et al (2014) EIGEN-6C4-The latest combined global gravity field model including GOCE data up to degree and order 1949 of GFZ Potsdam and GRGS Toulouse. EGU Gen Assembly Conf Abstr 16:3707

Gilardoni M, Reguzzoni M, Sampietro D (2016) GECO: a global gravity model by locally combining GOCE data and EGM2008. Stud Geophys Geod 60:228–247. https://doi.org/10.1007/s11200-015-1114-4

Gruber T, Willberg M (2019) Signal and error assessment of GOCE-based high resolution gravity field models. J Geod Sci 9(1):71–86. https://doi.org/10.1515/jogs-2019-0008

Gruber T, Gerlach C, Haagmans R (2012) Intercontinental height datum connection with GOCE and GPS-levelling data. J Geod Sci. https://doi.org/10.2478/v10156-012-0001-y

Hirt C, Rexer M, Claessens S, Rummel R (2017) The relation between degree-2160 spectral models of Earth's gravitational and topographic potential: a guide on global correlation measures and their dependency on approximation effects. J Geodesy 91:1179–1205. https://doi.org/10.1007/s00190-017-1016-z

Ihde J, Sánchez L, Barzaghi R, Drewes H, Foerste C, Gruber T, Liebsch G, Marti U, Pail R, Sideris M (2017) Definition and proposed realization of the International Height Reference System (IHRS). Surv Geophys 38(3):549–570. https://doi.org/10.1007/s10712-017-9409-3

Ince ES, Barthelmes F, Reißland S, Elger K, Förste C, Flechtner F, Schuh H (2019) ICGEM: 15 years of successful collection and distribution of global gravitational models, associated services and future plans. Earth Syst Sci Data 11:647–674. https://doi.org/10.5194/essd-11-647-2019

Jekeli C (1981) Alternative methods to smooth the Earth's gravity field. NASA, Grant No. NGR 36-008-161, OSURF Proj. No. 783210, 48 pp, Dec 1981, N82-22821/4

Jekeli C (1988) The exact transformation between ellipsoidal and spherical harmonic expansions. Manuscr Geod 1988(13):106–113

Knudsen P, Andersen O, Fecher T, Gruber T, Maximenko N (2018) A new OGMOC mean dynamic topography model: DTU17MDT. In: 25 years of progress in radar altimetry symposium, Portugal, 24/09/2018–29/09/2018, pp 213–214

Kvas A, Behzadpour S, Ellmer M, Klinger B, Strasser S, Zehentner N, Mayer-Gürr T (2019a) ITSG-Grace2018: overview and evaluation of a new GRACE-only gravity field time series. J Geophys Res Solid Earth 124:9332–9344. https://doi.org/10.1029/2019JB017415

Kvas A, Mayer-Gürr T, Krauss S, Brockmann JM, Schubert T, Schuh W-D, Pail R, Gruber T, Jäggi A, Meyer U (2019b) The satellite-only gravity field model GOCO06s. GFZ Data Serv. https://doi.org/10.5880/ICGEM.2019.002

Li X, Crowley JW, Holmes SA, Wang YM (2016) The contribution of the GRAV-D airborne gravity to geoid determination in the Great Lakes region. Geophys Res Lett 43:4358–4365. https://doi.org/10.1002/2016GL068374

McKenzie D, Yi W, Rummel R (2014) Estimates of Te from GOCE data. Earth Planet Sci Lett 399:116–127. https://doi.org/10.1016/j.epsl.2014.05.003

Pail R, Fecher T, Barnes D, Factor JF, Holmes SA, Gruber T, Zingerle P (2018) Short note: the experimental geopotential model XGM2016. J Geodesy 92:443. https://doi.org/10.1007/s00190-017-1070-6

Pavlis NK, Holmes SA, Kenyon SC, Factor JK (2012) The development and evaluation of the Earth Gravitational Model 2008 (EGM2008). J Geophys Res Solid Earth 1978–2012:117. https://doi.org/10.1029/2011JB008916

Rexer M, Hirt C, Claessens S, Tenzer R (2016) Layer-based modelling of the earth's gravitational potential up to 10-km scale in spherical harmonics in spherical and ellipsoidal approximation. Surv Geophys 37:1035–1074. https://doi.org/10.1007/s10712-016-9382-2

Rexer M, Hirt C, Pail R (2017) High-resolution global forward modelling: a degree-5480 global ellipsoidal topographic potential model. In: EGU general assembly conference abstracts, 19, p 7725. https://ui.adsabs.harvard.edu/abs/2017EGUGA..19.7725R

Schaeffer P, Pujol MI, Faugere Y, Picot N, Guillot A (2016) New mean sea surface CNES_CLS 2015 focusing on the use of geodetic missions of CryoSat-2 and Jason-1. ESA Living Planet Symposium, 2016

Scheinert M, Ferraccioli F, Schwabe J, Bell R, Studinger M, Damaske D, Jokat W, Aleshkova N, Jordan T, Leitchenkov G, Blankenship D, Damiani T, Young D, Cochran J, Richter T (2016) New Antarctic gravity anomaly grid for enhanced geodetic and geophysical studies in Antarctica. Geophys Res Lett 43(2):600–610. https://doi.org/10.1002/2015GL067439

Siegismund F (2013) Assessment of optimally filtered recent geodetic mean dynamic topographies. J Geophys Res Oceans 118(1):108–117. https://doi.org/10.1029/2012JC008149

Sneeuw N (1994) Global spherical harmonic analysis by least-squares and numerical quadrature methods in historical perspective. Geophys J Int 118:707–716. https://doi.org/10.1111/j.1365-246X.1994.tb03995.x

Tapley BD, Bettadpur S, Watkins M, Reigber C (2004) The gravity recovery and climate experiment: mission overview and early results. Geophys Res Lett 31:L09607. https://doi.org/10.1029/2004GL019920

Wessel P, Smith WHF (1996) A global, self-consistent, hierarchical, high-resolution shoreline database. J Geophys Res 101(B4):8741–8743. https://doi.org/10.1029/96jb00104

Willberg M, Zingerle P, Pail R (2019) Residual least-squares collocation: use of covariance matrices from high-resolution global geopotential models. J Geod 93(9):1739–1757. https://doi.org/10.1007/s00190-019-01279-1

Woodworth PL, Hughes CW, Bingham RJ, Gruber T (2012) Towards worldwide height system unification using ocean information. J Geod Sci. https://doi.org/10.2478/v10156-012-0004-8

Zingerle P, Pail R, Gruber T, Oikonomidou X (2019a) The experimental gravity field model XGM2019e. GFZ Data Serv. https://doi.org/10.5880/ICGEM.2019.007

Zingerle P, Pail R, Scheinert M, Schaller T (2019b) Evaluation of terrestrial and airborne gravity data over Antarctica: a generic approach. J Geod Sci 9:29–40. https://doi.org/10.1515/jogs-2019-0004

Publication P-4: Residual least-squares collocation: use of covariance matrices from high-resolution global geopotential models

Reference
Willberg M., Zingerle P. and Pail R. (2019) Residual least-squares collocation: use of covariance matrices from high-resolution global geopotential models. J Geod 93, 1739–1757 (2019). doi 10.1007/s00190-019-01279-1

Abstract
The paper presents a modified formulation of Least-Squares Collocation. This Residual Least-Squares Collocation (RLSC) includes a remove-compute-restore procedure with a high-resolution Global Geopotential Model (GGM) and a topographic gravitational potential model. In contrast to previous approaches, in RLSC, the remaining input residuals are modeled with error covariance matrices instead of signal covariance matrices. Therefore, we include the full variance-covariance information of a high-resolution GGM, namely the XGM2016, to the procedure. The included covariance matrices are anisotropic and location-dependent and enable a realistic error modeling of a target area. This fact represents an advantage over covariance matrices derived from signal degree variances or empirical covariance fitting. Additionally, due to the stochastic modeling of all involved components, RLSC provides realistic accuracy estimates. In a synthetic closed-loop test case with a realistic data distribution in the Andes we demonstrate the advantages of RLSC for regional geoid modeling and quantify the benefit which results mainly from a rigorously handled high-resolution GGM. In terms of root mean square deviations from the true reference solution, RLSC delivers an improvement of about 30% compared to a standard LSC approach, where the benefit is particularly pronounced in areas with a sparse data distribution. This improved performance, together with the fact that the resulting stochastic error estimates better reflect the true errors, might be an important aspect for the application of RLSC to derive gravity potential values and their uncertainties at reference stations of the International Height Reference System.

Declaration of own contribution
(MW: Martin Willberg; PZ: Philipp Zingerle; RP: Roland Pail)
MW and PZ derived the mathematical formulation of RLSC and the study design together, which are mostly the results of joint discussions (between MW and PZ). MW designed the test case in South America and the comparison between RLSC and LSC. PZ produced covariance matrices for XGM2016. Apart from that, MW performed the majority of all computations and created the results. MW was responsible for most of the analysis and interpretations, where PZ and RP contributed with discussions, corrections and support. MW wrote the text with improvements from RP and PZ. MW created figures and tables for the paper.

The overall own contribution of PZ for *P-4* is estimated at 15 %, which is the average value of the percentage values estimated for the five criteria listed in the table below (Tab. P.4).

Criteria	**Estimated own contribution**
Computation and results	10 %
Ideas and study design	40 %
Analysis and interpretation	20 %
Text	5 %
Figures and tables	0 %
Total	**15 %**

Tab. P.4 – Criteria and estimated contribution share of Philipp Zingerle for *P-4*

Confirmation by the authors

We hereby confirm the correctness of the declaration of own contribution (of 15 % for Philipp Zingerle) for the publication

Willberg M., Zingerle P. and Pail, R. (2019) Residual least-squares collocation: use of covariance matrices from high-resolution global geopotential models. J Geod 93, 1739–1757 (2019). doi 10.1007/s00190-019-01279-1

Martin Willberg

Institute of Astronomical and Physical Geodesy, Technical University of Munich, Germany

Signature: Martin Willberg Date: 27.09.2021

Roland Pail

Institute of Astronomical and Physical Geodesy, Technical University of Munich, Germany

Signature: Roland Pail Date: 23.09.2021

Publication P-5: Integration of airborne gravimetry data filtering into residual least-squares collocation - example from the 1 cm geoid experiment

Reference
Willberg M., Zingerle P. and Pail R. (2020) Integration of airborne gravimetry data filtering into residual least-squares collocation: example from the 1 cm geoid experiment. J Geod 94, 75. doi 10.1007/s00190-020-01396-2

Abstract
Low-pass filters are commonly used for the processing of airborne gravity observations. In this paper for the first time, we include the resulting correlations consistently in the functional and stochastic model of residual least-squares collocation (RLSC). We demonstrate the necessity of removing high-frequency noise from airborne gravity observations, and derive corresponding parameters for a Gaussian low-pass filter. Thereby, we intend an optimal combination of terrestrial and airborne gravity observations in the mountainous area of Colorado. We validate the combination in the frame of our participation in 'the 1 cm geoid experiment'. This regional geoid modeling inter-comparison exercise allows the calculation of a reference solution, which is defined as the mean value of 13 independent height anomaly results in this area. Our result performs among the best and with 7.5 mm shows the lowest standard deviation to the reference. From internal validation we furthermore conclude that the input from airborne and terrestrial gravity observations is consistent in large parts of the target area, but not necessarily in the highly mountainous areas. Therefore, the relative weighting between these two data sets turns out to be a main driver for the final result, and is an important factor in explaining the remaining differences between various height anomaly results in this experiment.

Declaration of own contribution
(MW: Martin Willberg; PZ: Philipp Zingerle; RP: Roland Pail)
MW had the idea to contribute to the 1 cm geoid experiment, prepared the gravity observations and performed initial results, where high-frequency noise was detected in the airborne observations. Accordingly, MW and PZ formulated the methodology section which handles high-frequency noise consistently in the RLSC approach. MW performed most of the calculations and created the results in the paper. The coauthors supported MW in analysis and interpretation, and provided detailed comments and corrections to the manuscript from MW. MW created figures and tables for the paper, whereby PZ had the idea for the figures 2 and 3.

The overall own contribution of PZ for *P-5* is estimated at 12 %, which is the average value of the percentage values estimated for the six criteria listed in the table below (Tab. P.5).

Criteria	**Estimated own contribution**
Computation and results	5 %
Ideas and study design	25 %
Analysis and interpretation	15 %
Text	5 %
Figures and tables	10 %
Total	**12 %**

Tab. P.5 – Criteria and estimated contribution share of Philipp Zingerle for *P-5*

Confirmation by the authors

We hereby confirm the correctness of the declaration of own contribution (of 12 % for Philipp Zingerle) for the publication

Willberg M., Zingerle P. and Pail R. (2020) Integration of airborne gravimetry data filtering into residual least-squares collocation: example from the 1~cm geoid experiment. J Geod 94, 75. doi 10.1007/s00190-020-01396-2

Martin Willberg

Institute of Astronomical and Physical Geodesy, Technical University of Munich, Germany

Signature: Martin Willberg Date: 27.09.2021

Roland Pail

Institute of Astronomical and Physical Geodesy, Technical University of Munich, Germany

Signature: Roland Pail Date: 23.09.2021