Journal of Geodesy (2019) 93:1739–1757
https://doi.org/10.1007/s00190-019-01279-1

ORIGINAL ARTICLE

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

Martin Willberg[1] · **Philipp Zingerle**[1] · **Roland Pail**[1]

Received: 24 August 2018 / Accepted: 30 June 2019 / Published online: 16 July 2019

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.

Keywords Least-squares collocation · Regional geoid modeling · Covariance function · Remove–compute–restore · XGM2016 · High-resolution GGM

1 Introduction

In this paper, we present a method for improving regional geoid modeling by including full covariance information from a high-resolution global geopotential model (GGM) in least-squares collocation (LSC). Since its foundation by Krarup (1969), and the key publication by Moritz (1980) LSC is considered as one of the most important methods for local and regional geoid modeling. Although the main concept of LSC has never changed, a few adaptations have been introduced recently. Nowadays, frequently a satellite-only model is used in LSC as background for the long wavelength part of the Earth's gravity field. LSC thereby benefits from the good model quality in the long wavelength part that is mainly provided by the Gravity Recovery And Climate Experiment (GRACE; Tapley et al. 2004) mission and the Gravity field and steady-state Ocean Circulation Explorer (GOCE; Drinkwater et al. 2003) mission.

Compared to satellite-only GGMs, high-resolution GGMs (we use this term for models with a maximum degree of 719 or higher) have included a wider spectral range of the Earth's gravity signal and therefore exhibit a higher commission error. For many of these high-resolution GGMs, the associated variance–covariance information is not fully available. As an example, in the case of EGM2008 only grids of geographic error variances of gravity anomalies and geoid undulations are provided (Pavlis et al. 2012). Additionally, EGM2008 and other high-resolution GGMs like EIGEN-6c (Förste et al. 2014) assume constant accuracy for their input ground data. However, with GOCO05c (Fecher et al. 2017) and its successor XGM2016 (Pail et al. 2018) we have two high-resolution GGMs that apply regional vary-

✉ Martin Willberg
martin.willberg@tum.de

[1] Institute of Astronomical and Physical Geodesy, Technical University of Munich, Arcisstrasse 21, 80333 Munich, Germany

ing weighting which results in an improved, more realistic and location-dependent accuracy estimation that enables us to use covariance matrices from these models for LSC. In this paper, we present the first inclusion of high-resolution GGMs with their full covariance matrices in LSC, develop the corresponding methodology and demonstrate its improvement to standard LSC in a numerical closed-loop simulation. The benefit of consistently including covariance matrices from a satellite-only GGM has already been demonstrated in various publications. This was first performed with a full covariance matrix by Haagmans and van Gelderen (1991). Pail et al. (2010) then demonstrated the rigorous inclusion of a GGM and its full accuracy information to LSC in a remove–compute–restore (RCR) approach. However, unlike other methods, our approach is specifically adjusted to work with high-resolution GGMs which leads us to an extended formalism of the LSC problem. Gerlach and Fecher (2012) showed that for GOCE covariance information, the very high computation effort of calculating full covariance matrices can be significantly reduced using approximations such as sparse or block-diagonal covariance matrices without losing much benefit. However, this simplification is not valid in our case since GGMs with regional weighting cause strong correlations among the gravity field parameters (Fecher et al. 2017).

The LSC approach is in practice still a frequently applied method for regional gravity geoid modeling. According to Moritz (1980) and Sansò (1986) the disturbing gravity field T of the Earth can be described as a random field. Also, it is possible to derive a statistical and homogeneous description of T that determines a global average part of the gravity field (Moritz 1980; Tscherning 1999). However, this does not coincide with reality because the correlation length and the covariance change with location. As a result, the use of homogeneous and isotropic covariance functions (e.g., in Moritz 1980) does not always give an optimal result in LSC (Tscherning 1999). Therefore, we adapt the LSC notation from Moritz (1980) in several, closely connected ways: (1) we include consistent treatment of covariance information in an RCR approach, (2) we change the definition of covariance matrices by replacing the total average operator with the mathematical expectation value, and (3) we use only residuals as LSC input which changes the nature of the covariance matrices as they describe only uncertainties instead of signal content. Here, our approach differs from previous studies by Moritz (1980), Pail et al. (2010), Forsberg and Tscherning (1981), Tscherning and Rapp (1974), or Sansò (1986). Only through these adaptations, we are able to use the location-dependent covariance matrices that are derived directly from the normal equation system of a high-resolution GGM. Furthermore, our notation offers the advantage that every input quantity in LSC is directly described by a covariance matrix which is in contrast with, for example, the definition by Haagmans and van Gelderen (1991). We see these adaptations as a necessary step towards using modern high-quality models effectively in LSC. At the present stage, we use the XGM2016 (Pail et al. 2018) and the topographic gravity model dV_ELL_Earth2014 (Rexer et al. 2016) for the RCR approach.

This paper is structured as follows. In Sect. 2 we derive in detail the adaptations to LSC by Moritz (1980). Next, in Sect. 3 we describe the specifications of a synthetic test case scenario in South America as well as the data sets that we use to show the benefit of our method. The results of three numerical test cases are visualized and outlined in Sect. 4, and finally Sect. 5 draws conclusions, describes ways to benefit from the demonstrated approach, and provides an outlook.

2 Theoretical background

2.1 LSC according to Moritz (1980)

This section adopts the content and notation of Moritz (1980). For our purposes we rewrite only those parts that are essential for the next sections. The formulas of the least-squares prediction for random observations $\mathbf{l}$ and a random output signal $\mathbf{s}$ are given in Moritz (1980, Chapter 9) with

$$\mathbf{s} = \mathbf{C}_{\mathrm{sl}}\,\mathbf{C}_{\mathrm{ll}}^{-1}\mathbf{l}, \tag{1}$$

$$\mathbf{E}_{\mathrm{ss}} = \mathbf{C}_{\mathrm{ss}} - \mathbf{C}_{\mathrm{sl}}\mathbf{C}_{\mathrm{ll}}^{-1}\mathbf{C}_{\mathrm{ls}}, \tag{2}$$

where $\mathbf{E}_{\mathrm{ss}}$ is the error covariance of the output $\mathbf{s}$, and $\mathbf{C}$ are covariance matrices with the subscripts giving the positions and functionals of the related points. This is valid for all $\mathbf{l}$ and $\mathbf{s}$ with

$$\mathrm{E}\{\mathbf{l}\} = \mathbf{0}, \quad \mathrm{E}\{\mathbf{s}\} = \mathbf{0}, \tag{3}$$

where E describes the expectation operator in form of a mathematical probability distribution. Afterwards, in the section 'Collocation with random errors' (Moritz 1980, Chapter 14) the observations $\mathbf{l}$ as a functional of the gravitational potential T are redefined as a combination of the true input signal $\mathfrak{t}$ and a random (stochastic) noise $\mathbf{n}$. So that, in symbolic notation we have

$$\mathbf{l} = \mathfrak{t} + \mathbf{n}. \tag{4}$$

In analogy to the definition of $\mathfrak{t}$, $\mathfrak{s}$ is defined as the true output signal. The gravity field functionals $\mathfrak{t}$ and $\mathfrak{s}$ are not random in a mere statistical manner since every evaluation point has a value without uncertainty. However, as it is done by Moritz (1980, Chapter 14), we treat $\mathfrak{t}$ and $\mathfrak{s}$ as statistical values in a

formal sense. This justifies the usage of the expectation operator E which, however, leaves the signals $\mathbf{t}$ and $\mathfrak{s}$ unaffected

$$\begin{aligned}
&\mathrm{E}\{\mathfrak{s}\} = \mathfrak{s},\\
&\mathrm{E}\{\mathbf{t}\} = \mathbf{t},\\
&\mathrm{E}\{\mathbf{n}\} = \mathbf{0},\\
&\mathrm{E}\{\mathbf{l}\} = \mathrm{E}\{\mathbf{t}\} + \mathrm{E}\{\mathbf{n}\} = \mathbf{t}.
\end{aligned} \tag{5}$$

In contrast to the expectation in a probabilistic sense, the operator M describes a homogeneous, isotropic average over the sphere which gives the mean global behavior of the gravitational field. $\mathbf{t}$, $\mathfrak{s}$ and $\mathbf{l}$ are all functionals of the disturbing potential. Therefore, applying the operator M to $\mathbf{t}$, $\mathfrak{s}$ and $\mathbf{l}$ gets zero. Thus, we write for the spatial average M

$$\begin{aligned}
&\mathrm{M}\{\mathfrak{s}\} = 0,\\
&\mathrm{M}\{\mathbf{t}\} = 0,\\
&\mathrm{M}\{\mathbf{l}\} = 0,\\
&\mathrm{M}\{\mathbf{n}\} = n.
\end{aligned} \tag{6}$$

In Eq. 5, we can see that in general the observations $\mathbf{l}$ are not centered, but in a global view with the definition of a total average operator $\bar{\mathrm{E}}$

$$\bar{\mathrm{E}} = \mathrm{EM}, \tag{7}$$

which is an average over the probability distribution and the global distribution (sphere), they can be considered as centered quantities

$$\begin{aligned}
&\bar{\mathrm{E}}\{\mathfrak{s}\} = \mathrm{EM}\{\mathfrak{s}\} = 0,\\
&\bar{\mathrm{E}}\{\mathbf{t}\} = \mathrm{EM}\{\mathbf{t}\} = 0,\\
&\bar{\mathrm{E}}\{\mathbf{l}\} = \mathrm{EM}\{\mathbf{t}\} + \mathrm{EM}\{\mathbf{n}\} = 0.
\end{aligned} \tag{8}$$

To be able to distinguish between different forms of covariance matrices in the following sections, we write covariance matrices that are derived from the total average $\bar{\mathrm{E}}$ with $\bar{\mathbf{C}}$ instead of $\mathbf{C}$. Consequently, the covariance matrices of the centered quantities defined by the total average $\bar{\mathrm{E}}$ are

$$\begin{aligned}
&\bar{\mathbf{C}}_{\mathfrak{s}\mathfrak{s}} = \bar{\mathrm{E}}\{\mathfrak{s}\mathfrak{s}^{\mathrm{T}}\} = \mathrm{EM}\{\mathfrak{s}\mathfrak{s}^{\mathrm{T}}\} = \mathrm{M}\{\mathfrak{s}\mathfrak{s}^{\mathrm{T}}\},\\
&\bar{\mathbf{C}}_{\mathrm{tt}} = \bar{\mathrm{E}}\{\mathbf{t}\mathbf{t}^{\mathrm{T}}\} = \mathrm{EM}\{\mathbf{t}\mathbf{t}^{\mathrm{T}}\} = \mathrm{M}\{\mathbf{t}\mathbf{t}^{\mathrm{T}}\},\\
&\bar{\mathbf{C}}_{\mathrm{nn}} = \bar{\mathrm{E}}\{\mathbf{n}\mathbf{n}^{\mathrm{T}}\} = \mathrm{EM}\{\mathbf{n}\mathbf{n}^{\mathrm{T}}\} = \mathrm{E}\{\mathbf{n}\mathbf{n}^{\mathrm{T}}\} = \mathbf{C}_{\mathrm{nn}}.
\end{aligned} \tag{9}$$

By considering the fact that $\mathbf{t}$ and $\mathfrak{s}$ are uncorrelated to $\mathbf{n}$, we write by using Eq. 4

$$\begin{aligned}
\bar{\mathbf{C}}_{\mathrm{ll}} &= \bar{\mathrm{E}}\{\mathbf{l}\mathbf{l}^{\mathrm{T}}\}\\
&= \bar{\mathrm{E}}\{(\mathbf{t}+\mathbf{n})(\mathbf{t}+\mathbf{n})^{\mathrm{T}}\}\\
&= \bar{\mathrm{E}}\{(\mathbf{t}\mathbf{t}^{\mathrm{T}})\} + \bar{\mathrm{E}}\{(\mathbf{t}\mathbf{n}^{\mathrm{T}})\} + \bar{\mathrm{E}}\{(\mathbf{n}\mathbf{t}^{\mathrm{T}})\} + \bar{\mathrm{E}}\{(\mathbf{n}\mathbf{n}^{\mathrm{T}})\}\\
&= \bar{\mathbf{C}}_{\mathrm{tt}} + \underbrace{\bar{\mathbf{C}}_{\mathrm{tn}}}_{=\mathbf{0}} + \underbrace{\bar{\mathbf{C}}_{\mathrm{nt}}}_{=\mathbf{0}} + \mathbf{C}_{\mathrm{nn}}\\
&= \bar{\mathbf{C}}_{\mathrm{tt}} + \mathbf{C}_{\mathrm{nn}},
\end{aligned} \tag{10}$$

$$\begin{aligned}
\bar{\mathbf{C}}_{\mathfrak{s}\mathrm{l}} &= \bar{\mathrm{E}}\{\mathfrak{s}\mathbf{l}^{\mathrm{T}}\}\\
&= \bar{\mathrm{E}}\{(\mathfrak{s})(\mathbf{t}+\mathbf{n})^{\mathrm{T}}\}\\
&= \bar{\mathrm{E}}\{(\mathfrak{s}\mathbf{t}^{\mathrm{T}})\} + \bar{\mathrm{E}}\{(\mathfrak{s}\mathbf{n}^{\mathrm{T}})\}\\
&= \bar{\mathbf{C}}_{\mathfrak{s}\mathrm{t}} + \underbrace{\bar{\mathbf{C}}_{\mathfrak{s}\mathrm{n}}}_{=\mathbf{0}}.
\end{aligned} \tag{11}$$

Applying these expressions, the fundamental formula for least-squares collocation with noise is obtained from Eq. 1

$$\bar{\mathbf{s}} = \bar{\mathbf{C}}_{\mathfrak{s}\mathrm{t}}\,(\bar{\mathbf{C}}_{\mathrm{tt}} + \mathbf{C}_{\mathrm{nn}})^{-1}\mathbf{l}. \tag{12}$$

For consistency with the following sections and in contrast to Moritz (1980) we retain the bar over all quantities with inclusion of the global average operator $\bar{\mathrm{E}}$, and write for the corresponding error covariance matrix $\bar{\mathbf{E}}_{\bar{\mathfrak{s}}\bar{\mathfrak{s}}}$ according to Eq. 2

$$\bar{\mathbf{E}}_{\bar{\mathfrak{s}}\bar{\mathfrak{s}}} = \bar{\mathbf{C}}_{\mathfrak{s}\mathfrak{s}} - \bar{\mathbf{C}}_{\mathfrak{s}\mathrm{t}}(\bar{\mathbf{C}}_{\mathrm{tt}} + \mathbf{C}_{\mathrm{nn}})^{-1}\bar{\mathbf{C}}_{\mathfrak{s}\mathrm{t}}^{\mathrm{T}}. \tag{13}$$

This description of least-squares collocation by Moritz (1980, Chapter 14) has been established as standard procedure in the literature (see Tscherning 2015; Arabelos and Tscherning 2009; Hofmann-Wellenhof and Moritz 2006; Rieser 2015) and we follow this notation because in our opinion it is the most consistent and detailed description. However, it should be kept in mind that these definitions are only valid for centered observations $\mathbf{l}$ and a centered output $\mathbf{s}$, because this was set as a requirement (Eq. 3) and is used in the definitions of the covariance matrices (Eqs. 9–11). Furthermore, the covariance matrices as defined in this section assume a normal distribution of gravitational functionals and describe an average part of the Earth's surface. Accordingly, these covariance matrices are mainly independent of the location on the Earth (homogeneous) as well as of the direction (isotropic), and as a consequence may not be optimal for local gravity field collocation (Tscherning 1999). Note, that most of the covariance matrices in Moritz (1980) are calculated from signal degree variances, but the discussion in this paragraph holds also for analytical covariance matrices which are calculated by empirical covariance fitting, e.g., from a Tscherning–Rapp model (Tscherning and Rapp 1974). While the latter do not consider a global view, they use the same assumptions, namely centered and normally distributed observations, for a target regional area, therefore resulting in homogeneous and isotropic covariance functions as well.

2.2 The remove–compute–restore (RCR) approach

For many applications, LSC is combined with a remove–compute–restore (RCR) approach (Forsberg and Tscherning 1981; Forsberg 1984). According to the name, RCR means that a part $\hat{\mathbf{l}}$ of the signal is removed from the observations $\mathbf{l}$ before the computation

$$\Delta\mathbf{l} = \mathbf{l} - \hat{\mathbf{l}}. \tag{14}$$

Accordingly, the collocation Γ is performed only with the residuals of the input signal $\Delta\mathbf{l}$

$$\Delta\mathbf{s} = \Gamma\,\Delta\mathbf{l}, \tag{15}$$

and afterwards, the removed part $\hat{\mathbf{s}}$ is restored again

$$\mathbf{s} = \Delta\mathbf{s} + \hat{\mathbf{s}}, \tag{16}$$

to yield the output $\mathbf{s}$ of the collocation. The main reason for using RCR in collocation is that LSC with residuals is more accurate than it would be with the full signal content. For more detailed background, we refer to Hofmann-Wellenhof and Moritz (2006, Chapter 11) or Rieser (2015, Chapter 4). Usually, the remove and restore steps describe different functionals of the gravity field, and the RCR thereby implicitly includes a field transformation. Moreover, because a field transformation moves signal energy between different degrees in the frequency domain an error in the remove step may not be consistently restored after a field transformation. This might even be the case when remove and restore steps are calculated consistently. Therefore, we see the need to model the accuracy of the remove step consistently in LSC. Only some of the LSC approaches take the accuracy of the reduction model into account (e.g., Haagmans and van Gelderen 1991; Pail et al. 2010; Sansò 2013) while other approaches do not (e.g., Forsberg and Tscherning 1981; Hofmann-Wellenhof and Moritz 2006; Rieser 2015).

In general, the corresponding covariance matrices in all these approaches are calculated either from signal degree variances (Heiskanen and Moritz 1967, Chapter 7.3; Moritz 1980, Chapter 10) or from empirical covariance fitting (Tscherning and Rapp 1974). Those approaches that include the accuracy of a reduction model directly in the calculation, additionally introduce the covariance matrix of a satellite-only model to the method (e.g., Haagmans and van Gelderen (1991)). Since all of these LSC approaches are formulated differently and use various notations, we do not go into further detail but point out only the differences of our approach to the existing ones in the following sections.

2.3 Residual least-squares collocation

Although observations $\mathbf{l}$ and output $\mathbf{s}$ of a LSC as functionals of the disturbing gravity field T of the Earth are centered globally (Heiskanen and Moritz 1967, Chapter 2.19), it is not possible to automatically assume the same for regional gravity field modeling. However, the definition of LSC according to Sect. 2.1 is only valid if the observations and the output are centered over the target area (Eq. 3). We propose rather to use covariance matrices from a purely stochastic point of view and thereby follow the standard definition of a covariance with an expectation value in a mathematical sense E (Eqs. 17–19). As a result, we are no longer limited to the requirements of Eq. 3. The gravity functionals $\mathbf{t}$ and $\mathfrak{s}$ are furthermore regarded as statistical quantities in a formal sense. As they are defined as true signal content of the gravity field, repeated error-free measurements at one point always give the same result (see Eq. 5). Considering this, we write for the covariance matrices $\mathbf{C}_{\mathfrak{s}\mathfrak{s}}$, $\mathbf{C}_{\mathrm{tt}}$ and $\mathbf{C}_{\mathrm{ll}}$

$$\mathbf{C}_{\mathfrak{s}\mathfrak{s}} = \mathrm{E}\{(\mathfrak{s} - \underbrace{\mathrm{E}\{\mathfrak{s}\}}_{=\mathfrak{s}})(\mathfrak{s} - \underbrace{\mathrm{E}\{\mathfrak{s}\}}_{=\mathfrak{s}})^{\mathrm{T}}\} = \{\mathbf{0}\}, \tag{17}$$

$$\mathbf{C}_{\mathrm{tt}} = \mathrm{E}\{(\mathbf{t} - \underbrace{\mathrm{E}\{\mathbf{t}\}}_{=\mathbf{t}})(\mathbf{t} - \underbrace{\mathrm{E}\{\mathbf{t}\}}_{=\mathbf{t}})^{\mathrm{T}}\} = \{\mathbf{0}\}, \tag{18}$$

$$\mathbf{C}_{\mathrm{ll}} = \mathrm{E}\{(\mathbf{l} - \underbrace{\mathrm{E}\{\mathbf{l}\}}_{=\mathbf{t}})(\mathbf{l} - \underbrace{\mathrm{E}\{\mathbf{l}\}}_{=\mathbf{t}})^{\mathrm{T}}\} = \mathrm{E}\{\mathbf{n}\mathbf{n}^{\mathrm{T}}\} = \mathbf{C}_{\mathrm{nn}}, \tag{19}$$

which differs from the equivalent formulation in Moritz (1980) (Eqs. 9, 10). The covariance of the uncorrelated observations that is named $\mathbf{C}_{\mathrm{nn}}$ in Moritz (1980) is hereinafter called $\mathbf{C}_{\mathrm{ll}}$. We consider this notation as more consistent, because it describes the accuracy of the observations. Note at this point, that $\mathbf{t}$ and $\mathbf{l}$ are not centered (Eq. 5). Now, we assume that we can calculate an unbiased GGM without systematic errors that is able to describe the full signal content of the Earth's gravity field. From this model we derive different gravity field functionals $\mathbf{t}$ at the input points. Since this variable is not error-free, we denote the quantity $\hat{\mathbf{l}}$ and write for its expectation value E

$$\mathrm{E}\{\hat{\mathbf{l}}\} = \mathbf{t}. \tag{20}$$

We use the hat operator for quantities that are derived from an introduced model in order to distinguish them from measurements and error-free quantities. Since we assumed that $\hat{\mathbf{l}}$ does not contain systematic errors, the difference $\Delta\mathbf{l}$

$$\Delta\mathbf{l} = \mathbf{l} - \hat{\mathbf{l}}, \tag{21}$$

is centered and describes a remove step (see Sect. 2.2). By analogy, we write $\hat{\mathbf{s}}$ as the result of the same model as an approximation of the true gravity signal $\mathfrak{s}$ and obtain

$$\Delta\mathbf{s} = \mathfrak{s} - \hat{\mathbf{s}}. \tag{22}$$

The resulting covariance matrices $\mathbf{C}_{\hat{\mathbf{l}}\hat{\mathbf{l}}}$, $\mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{l}}}$ and $\mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}$ of our estimated values $\hat{\mathbf{l}}$ and $\hat{\mathbf{s}}$ are defined by

$$\begin{aligned}
\mathbf{C}_{\hat{\mathbf{l}}\hat{\mathbf{l}}} &= \mathrm{E}\{(\hat{\mathbf{l}} - \mathrm{E}\{\hat{\mathbf{l}}\})(\hat{\mathbf{l}} - \mathrm{E}\{\hat{\mathbf{l}}\})^{\mathrm{T}}\} \\
&= \mathrm{E}\{(\hat{\mathbf{l}} - \mathbf{t})(\hat{\mathbf{l}} - \mathbf{t})^{\mathrm{T}}\}, && (23)\\
\mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{l}}} &= \mathrm{E}\{(\hat{\mathbf{s}} - \mathrm{E}\{\hat{\mathbf{s}}\})(\hat{\mathbf{l}} - \mathrm{E}\{\hat{\mathbf{l}}\})^{\mathrm{T}}\} \\
&= \mathrm{E}\{(\hat{\mathbf{s}} - \mathfrak{s})(\hat{\mathbf{l}} - \mathbf{t})^{\mathrm{T}}\}, && (24)\\
\mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{s}}} &= \mathrm{E}\{(\hat{\mathbf{s}} - \mathrm{E}\{\hat{\mathbf{s}}\})(\hat{\mathbf{s}} - \mathrm{E}\{\hat{\mathbf{s}}\})^{\mathrm{T}}\} \\
&= \mathrm{E}\{(\hat{\mathbf{s}} - \mathfrak{s})(\hat{\mathbf{s}} - \mathfrak{s})^{\mathrm{T}}\}. && (25)
\end{aligned}$$

They are used in the following to form the covariance matrices $\mathbf{C}_{\Delta\mathbf{l}\Delta\mathbf{l}}$, $\mathbf{C}_{\Delta\mathbf{s}\Delta\mathbf{l}}$ and $\mathbf{C}_{\Delta\mathbf{s}\Delta\mathbf{s}}$. For the transcription of $\mathbf{C}_{\Delta\mathbf{l}\Delta\mathbf{l}}$ we use Eqs. 4, 19, 21 and 23 and the fact that the random noise $\mathbf{n}$ is uncorrelated to a signal.

$$\begin{aligned}
\mathbf{C}_{\Delta\mathbf{l}\Delta\mathbf{l}} &= \mathrm{E}\{(\Delta\mathbf{l} - \underbrace{\mathrm{E}\{\Delta\mathbf{l}\}}_{=\mathbf{0}})(\Delta\mathbf{l} - \underbrace{\mathrm{E}\{\Delta\mathbf{l}\}}_{=\mathbf{0}})^{\mathrm{T}}\} \\
&= \mathrm{E}\{\Delta\mathbf{l}\Delta\mathbf{l}^{\mathrm{T}}\} \\
&= \mathrm{E}\{((\mathbf{t}+\mathbf{n}) - \hat{\mathbf{l}})((\mathbf{t}+\mathbf{n}) - \hat{\mathbf{l}})^{\mathrm{T}}\} \\
&= \mathrm{E}\{((\mathbf{t} - \hat{\mathbf{l}}) + \mathbf{n})((\mathbf{t} - \hat{\mathbf{l}}) + \mathbf{n})^{\mathrm{T}}\} \\
&= \mathrm{E}\{(\mathbf{t} - \hat{\mathbf{l}})(\mathbf{t} - \hat{\mathbf{l}})^{\mathrm{T}}\} + \underbrace{\mathrm{E}\{(\mathbf{t} - \hat{\mathbf{l}})\mathbf{n}^{\mathrm{T}}\}}_{=\mathbf{0}} \\
&\quad + \underbrace{\mathrm{E}\{\mathbf{n}(\mathbf{t} - \hat{\mathbf{l}})^{\mathrm{T}}\}}_{=\mathbf{0}} + \mathrm{E}\{\mathbf{n}\mathbf{n}^{\mathrm{T}}\} \\
&= \mathbf{C}_{\hat{\mathbf{l}}\hat{\mathbf{l}}} + \mathbf{C}_{\mathrm{ll}} && (26)
\end{aligned}$$

$\Delta\mathbf{l}$ contains the uncertainty of the observations $\mathbf{l}$ as well as the uncertainty of observations $\hat{\mathbf{l}}$ synthesized from a GGM. Correspondingly, we can divide the covariance matrix $\mathbf{C}_{\Delta\mathbf{l}\Delta\mathbf{l}}$ into the covariance of the model accuracy $\mathbf{C}_{\hat{\mathbf{l}}\hat{\mathbf{l}}}$, and the covariance $\mathbf{C}_{\mathrm{ll}}$ that describes the observation noise according to Eq. 19. Again applying the fact that the random noise $\mathbf{n}$ is uncorrelated to $\mathfrak{s}$ and $\hat{\mathbf{s}}$, we write analogously for $\mathbf{C}_{\Delta\mathbf{s}\Delta\mathbf{l}}$ and $\mathbf{C}_{\Delta\mathbf{s}\Delta\mathbf{s}}$

$$\begin{aligned}
\mathbf{C}_{\Delta\mathbf{s}\Delta\mathbf{l}} &= \mathrm{E}\{(\Delta\mathbf{s} - \underbrace{\mathrm{E}\{\Delta\mathbf{s}\}}_{=\mathbf{0}})(\Delta\mathbf{l} - \underbrace{\mathrm{E}\{\Delta\mathbf{l}\}}_{=\mathbf{0}})^{\mathrm{T}}\} \\
&= \mathrm{E}\{\Delta\mathbf{s}\Delta\mathbf{l}^{\mathrm{T}}\} \\
&= \mathrm{E}\{(\mathfrak{s} - \hat{\mathbf{s}})((\mathbf{t}+\mathbf{n}) - \hat{\mathbf{l}})^{\mathrm{T}}\} \\
&= \mathrm{E}\{(\mathfrak{s} - \hat{\mathbf{s}})((\mathbf{t} - \hat{\mathbf{l}}) + \mathbf{n})^{\mathrm{T}}\} \\
&= \mathrm{E}\{(\mathfrak{s} - \hat{\mathbf{s}})(\mathbf{t} - \hat{\mathbf{l}})^{\mathrm{T}}\} + \underbrace{\mathrm{E}\{(\mathfrak{s} - \hat{\mathbf{s}})\mathbf{n}^{\mathrm{T}}\}}_{=\mathbf{0}} \\
&= \mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{l}}}, && (27)\\
\mathbf{C}_{\Delta\mathbf{s}\Delta\mathbf{s}} &= \mathrm{E}\{(\Delta\mathbf{s} - \underbrace{\mathrm{E}\{\Delta\mathbf{s}\}}_{=\mathbf{0}})(\Delta\mathbf{s} - \underbrace{\mathrm{E}\{\Delta\mathbf{s}\}}_{=\mathbf{0}})^{\mathrm{T}}\} \\
&= \mathrm{E}\{\Delta\mathbf{s}\Delta\mathbf{s}^{\mathrm{T}}\} \\
&= \mathrm{E}\{(\mathfrak{s} - \hat{\mathbf{s}})(\mathfrak{s} - \hat{\mathbf{s}})^{\mathrm{T}}\} \\
&= \mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}, && (28)
\end{aligned}$$

with $\mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{l}}}$ being the covariance of the introduced model that describes the uncertainties and correlations between the positions and functionals of the input to those of the output. Analogously, the covariance $\mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}$ describes the uncertainties and the correlations of $\hat{\mathbf{s}}$ among different output positions (and functionals).

Thus, we can rewrite the definition of LSC from Eq. 12 with $\Delta\mathbf{l}$ and $\Delta\mathbf{s}$ instead of $\mathbf{l}$ and $\mathbf{s}$ and use the findings from Eqs. 21, 26 and 27, resulting in a notation of LSC that uses only residuals as input

$$\begin{aligned}
\Delta\mathbf{s} &= \mathbf{C}_{\Delta\mathbf{s}\Delta\mathbf{l}}\,(\mathbf{C}_{\Delta\mathbf{l}\Delta\mathbf{l}})^{-1}\Delta\mathbf{l} \\
&= \mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{l}}}\,(\mathbf{C}_{\mathrm{ll}} + \mathbf{C}_{\hat{\mathbf{l}}\hat{\mathbf{l}}})^{-1}(\mathbf{l} - \hat{\mathbf{l}}). && (29)
\end{aligned}$$

After restoring the subtracted signal part $\hat{\mathbf{s}}$ we define the residual least-squares collocation (RLSC)

$$\mathbf{s} = \mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{l}}}\,(\mathbf{C}_{\mathrm{ll}} + \mathbf{C}_{\hat{\mathbf{l}}\hat{\mathbf{l}}})^{-1}(\mathbf{l} - \hat{\mathbf{l}}) + \hat{\mathbf{s}}. \tag{30}$$

Instead of centered observations and a centered output which are used as requirements in Sect. 2.1, we introduce the assumption that it is possible to describe the observations $\mathbf{l}$ and the output $\mathbf{s}$ by means of an unbiased model ($\hat{\mathbf{l}}$ and $\hat{\mathbf{s}}$). In this way, the input $(\mathbf{l} - \hat{\mathbf{l}})$ of RLSC stays centered. We have two key factors of RLSC in comparison with other LSC methods: (1) the input consists only of residuals, and (2) the covariance matrices $\mathbf{C}_{\mathrm{ll}}$ and $\mathbf{C}_{\hat{\mathbf{l}}\hat{\mathbf{l}}}$ describe the accuracy of the input $\mathbf{l}$ and $\hat{\mathbf{l}}$ directly. With these considerations, we differ from previous LSC approaches (e.g., Moritz 1980; Haagmans and van Gelderen 1991; Pail et al. 2010; Sansò 2013; Rieser 2015), which always include signal covariance matrices in LSC, while in our approach the covariance matrices contain only the uncertainties of input and output. For example, Haagmans and van Gelderen (1991) also include a reduction step with full covariance information in LSC, but at the same time they use a third covariance matrix inside the brackets of Eq. 30 that describes the covariance of the gravity signal itself (equal to $\bar{\mathbf{C}}_{\mathrm{tt}}$ from Sect. 2.1). The reason for the difference at this point is that we assume the existence of a high-quality GGM, whose uncertainties can be fully described by a variance–covariance matrix (no systematic errors), which other approaches do not.

The interpretation of Eq. 30 is quite different from the definitions in Sect. 2.1 where the covariance matrices in the collocation describe the signal content instead of the error characteristics. Nevertheless, RLSC is consistent with the theory of Moritz (1980), which can be demonstrated by the following thought experiment. If the model $\hat{\mathbf{l}}$ in Eq. 30 becomes worse, then the elements of the covariance matrix $\mathbf{C}_{\hat{\mathbf{l}}\hat{\mathbf{l}}}$ that describe its uncertainties will become larger. This is also valid for an extreme case where we do not subtract a model at all, so that $\hat{\mathbf{l}}$ becomes zero. In this case, the covariance $\mathbf{C}_{\hat{\mathbf{l}}\hat{\mathbf{l}}}$ describes the full signal content of the observations $\mathbf{l}$ which is basically the same idea as in Moritz (1980). However, the covariance $\mathbf{C}_{\hat{\mathbf{l}}\hat{\mathbf{l}}}$ is still regarded as the uncertainty of the bad (or missing) model instead of the covariance of the signal content (as it is in Sect. 2.1). Even with $\hat{\mathbf{l}} = \mathbf{0}$ there is a difference in the approaches, since in general the expectation value E is not equal to the total average $\bar{\mathrm{E}}$ (Eq. 7). However, it is possible to use the total average $\bar{\mathrm{E}}$ as an approximation of the expectation value E, which means that for $\hat{\mathbf{l}} = \mathbf{0}$ we can approximate

$$\begin{aligned} \mathbf{C}_{\hat{\mathbf{l}}\hat{\mathbf{l}}} &\approx \bar{\mathbf{C}}_{\mathfrak{t}\mathfrak{t}}, \\ \mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{l}}} &\approx \bar{\mathbf{C}}_{\mathfrak{s}\mathfrak{t}}. \end{aligned} \tag{31}$$

With this approximation our formulation of least-squares collocation with errors (Eq. 30) becomes identical to standard LSC from Moritz (1980) (Eq. 12) including the RCR approach (Eqs. 14, 16). Accordingly, we write for the error covariance matrix $\mathbf{E}_{\mathrm{ss}}$ of the output $\mathbf{s}$ in Eq. 30 and refer to Moritz (1980, Chapter 14) for an analogous and detailed derivation

$$\mathbf{E}_{\mathrm{ss}} = \mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{s}}} - \mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{l}}} \left(\mathbf{C}_{\mathrm{ll}} + \mathbf{C}_{\hat{\mathbf{l}}\hat{\mathbf{l}}}\right)^{-1} \mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{l}}}^{\mathrm{T}}. \tag{32}$$

Note that, by definition there is no difference between $\mathbf{E}$ and $\mathbf{C}$ in our approach since both describe error covariance matrices. Nevertheless, we retain this notation because it clarifies that $\mathbf{E}_{\mathrm{ss}}$ refers to the covariance of the output $\mathbf{s}$ instead of the true gravity signal $\mathfrak{s}$ or the model $\hat{\mathbf{s}}$.

2.4 Inclusion of a GGM and a topographic gravity model into RLSC

To calculate RLSC from Eq. 30 we require estimates of the input $\hat{\mathbf{l}}$ and the output $\hat{\mathbf{s}}$ as well as the related covariance matrices for the input $\mathbf{C}_{\hat{\mathbf{l}}\hat{\mathbf{l}}}$, the output $\mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}$ and the combination $\mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{l}}}$. All of these can be derived from the normal equation system related to a GGM

$$\mathbf{N}_{\mathrm{ff}}^{\mathrm{m}} \hat{\mathbf{x}}_{\mathrm{f}}^{\mathrm{m}} = \mathbf{q}_{\mathrm{f}}^{\mathrm{m}}. \tag{33}$$

Here, $\mathbf{N}_{\mathrm{ff}}^{\mathrm{m}}$ is the normal equation matrix, $\mathbf{q}_{\mathrm{f}}^{\mathrm{m}}$ is the right-hand side, and $\hat{\mathbf{x}}_{\mathrm{f}}^{\mathrm{m}}$ are the estimated Spherical Harmonic (SH) coefficients. The superscript 'm' refers to quantities that are derived from a GGM. In the following we continue to use superscripts to clarify the origin of covariance matrices and vectors. In contrast, the subscripts are continuously applied to describe the corresponding functionals and their positions with the subscript 'f' standing for the frequency domain of the SH coefficients. For more details about the normal equation system describing the SH coefficients we refer to Fecher et al. (2015). Next, the normal equation matrix $\mathrm{N}_{\mathrm{ff}}^{\mathrm{m}}$ is inverted

$$\mathbf{C}_{\mathrm{ff}}^{\mathrm{m}} = \left(\mathbf{N}_{\mathrm{ff}}^{\mathrm{m}}\right)^{-1}, \tag{34}$$

to obtain the covariance matrix $\mathbf{C}_{\mathrm{ff}}^{\mathrm{m}}$ of the SH coefficients, which is used to solve the normal equation system and estimate the SH coefficients $\hat{\mathbf{x}}_{\mathrm{f}}^{\mathrm{m}}$

$$\hat{\mathbf{x}}_{\mathrm{f}}^{\mathrm{m}} = \mathbf{C}_{\mathrm{ff}}^{\mathrm{m}}\, \mathbf{q}_{\mathrm{f}}^{\mathrm{m}}. \tag{35}$$

Afterwards, we write the transformations of the SH coefficients to different functionals and point positions in the space domain with the design matrices $\mathbf{A}_{\mathrm{to}}^{\mathrm{from}}$.

$$\begin{aligned} \hat{\mathbf{l}}^{\mathrm{m}} &= \mathbf{A}_{\mathrm{l}}^{\mathrm{f}}\, \hat{\mathbf{x}}_{\mathrm{f}}^{\mathrm{m}}, \\ \hat{\mathbf{s}}^{\mathrm{m}} &= \mathbf{A}_{s}^{\mathrm{f}}\, \hat{\mathbf{x}}_{\mathrm{f}}^{\mathrm{m}}, \end{aligned} \tag{36}$$

and for the calculation of the covariance matrices

$$\begin{aligned} \mathbf{C}_{\hat{\mathbf{l}}\hat{\mathbf{l}}}^{\mathrm{m}} &= \mathbf{A}_{\mathrm{l}}^{\mathrm{f}}\, \mathbf{C}_{\mathrm{ff}}^{\mathrm{m}}\, \mathbf{A}_{\mathrm{l}}^{\mathrm{f\,T}}, \\ \mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{l}}}^{\mathrm{m}} &= \mathbf{A}_{\mathrm{s}}^{\mathrm{f}}\, \mathbf{C}_{\mathrm{ff}}^{\mathrm{m}}\, \mathbf{A}_{\mathrm{l}}^{\mathrm{f\,T}}, \\ \mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}^{\mathrm{m}} &= \mathbf{A}_{\mathrm{s}}^{\mathrm{f}}\, \mathbf{C}_{\mathrm{ff}}^{\mathrm{m}}\, \mathbf{A}_{\mathrm{s}}^{\mathrm{f\,T}}. \end{aligned} \tag{37}$$

Here, the covariance $\mathbf{C}_{\hat{\mathbf{l}}\hat{\mathbf{l}}}^{\mathrm{m}}$ describes the uncertainties and correlations of the GGM at the input points, $\mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}^{\mathrm{m}}$ the covariance at the output points, and $\mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{l}}}^{\mathrm{m}}$ their cross correlations. The square root of the main diagonal of $\mathbf{C}_{\hat{\mathbf{l}}\hat{\mathbf{l}}}^{\mathrm{m}}$ gives the standard deviations of the GGM parameters at the input points in terms of the respective functional. Note that as the result of a GGM, $\hat{\mathbf{l}}^{\mathrm{m}}$ is a representative of the gravity field up to a certain spherical harmonic degree N, but does not consider the full frequency spectrum of $\mathbf{l}$. However, we can use $\hat{\mathbf{l}}^{\mathrm{m}}$ from a GGM, because it centers the difference $\Delta\mathbf{l}$ (in the degrees up to N)

$$\Delta\mathbf{l} = \mathbf{l} - \hat{\mathbf{l}}^{\mathrm{m}}, \tag{38}$$

as introduced in Eq. 14. When a GGM is considered, there are two possibilities: either a GGM from a satellite-only solution that describes the gravity field up to degrees around 200–280 (e.g., GOCO05s; Mayer-Gürr et al. 2015) is used, or a GGM with much higher spatial resolution that already contains terrestrial information like EGM2008 (Pavlis et al. 2012), EIGEN-6c4 (Förste et al. 2014) or XGM2016 (Pail et al. 2018). In both cases, the approximation of the true gravity

field $\mathbf{t}$ by $\hat{\mathbf{l}}$ can be improved by additionally considering the information in the frequencies above the maximum degree of the GGM. In general, large parts of the gravity signal beyond a GGM's resolution are related to the topography (Forsberg and Tscherning 1981; Hirt et al. 2013; Rexer et al. 2016), which is why we use an additional model for the topographic gravity effect above the maximum degree N of the GGM and calculate its effect $\hat{\mathbf{l}}^{\mathrm{t}}$ (with 't' standing for topography) at the input points

$$\hat{\mathbf{l}} = \hat{\mathbf{l}}^{\mathrm{m}} + \hat{\mathbf{l}}^{\mathrm{t}}, \tag{39}$$

and the output points

$$\hat{\mathbf{s}} = \hat{\mathbf{s}}^{\mathrm{m}} + \hat{\mathbf{s}}^{\mathrm{t}}. \tag{40}$$

The same requirements that apply for the definition of $\hat{\mathbf{l}}^{\mathrm{m}}$ also apply for $\hat{\mathbf{l}}^{\mathrm{t}}$, so that the resulting $\Delta\mathbf{l}$ is normally distributed and centered (up to the maximum degree of the topographic gravity model)

$$\Delta\mathbf{l} = \mathbf{l} - \hat{\mathbf{l}}^{\mathrm{m}} - \hat{\mathbf{l}}^{\mathrm{t}}. \tag{41}$$

From a practical point of view it is also important that $\hat{\mathbf{l}}^{\mathrm{m}}$ and $\hat{\mathbf{l}}^{\mathrm{t}}$ can be considered as independent of each other so that the variance propagation from Eq. 39 yields

$$\mathbf{C}_{\hat{\mathbf{l}}\hat{\mathbf{l}}} = \mathbf{C}^{\mathrm{m}}_{\hat{\mathbf{l}}\hat{\mathbf{l}}} + \mathbf{C}^{\mathrm{t}}_{\hat{\mathbf{l}}\hat{\mathbf{l}}}, \tag{42}$$

with $\mathbf{C}^{\mathrm{t}}_{\hat{\mathbf{l}}\hat{\mathbf{l}}}$ being the covariance of the topographic gravity model that describes the models uncertainties and correlations. This is analogous to the character of $\mathbf{C}^{\mathrm{m}}_{\hat{\mathbf{l}}\hat{\mathbf{l}}}$ for the GGM, and the same is valid for the covariance matrices $\mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{l}}}$ and $\mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}$

$$\begin{aligned}\mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{l}}} &= \mathbf{C}^{\mathrm{m}}_{\hat{\mathbf{s}}\hat{\mathbf{l}}} + \mathbf{C}^{\mathrm{t}}_{\hat{\mathbf{s}}\hat{\mathbf{l}}},\\ \mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{s}}} &= \mathbf{C}^{\mathrm{m}}_{\hat{\mathbf{s}}\hat{\mathbf{s}}} + \mathbf{C}^{\mathrm{t}}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}.\end{aligned} \tag{43}$$

As we use the topographic gravity model only in the degrees above the maximum degree N of the GGM, we regard the two models as uncorrelated and write for their degree n

$$\begin{aligned}n^{\mathrm{m}} &\in \{2, N\},\\ n^{\mathrm{t}} &\in \{N+1, N_{\max}\},\end{aligned} \tag{44}$$

where $N_{\max}$ is the maximum degree of $\hat{\mathbf{l}}^{\mathrm{t}}$. Usually, there is neither a normal equation system available for the degrees above a GGM nor another source for a covariance matrix that describes the accuracy of the topographic gravity model. Therefore, without the possibility of deriving direct accuracy or covariance information for the topographic gravity model, we must use the total average $\bar{\mathrm{E}}$ for the derivation of a covariance matrix. The resulting covariance is an approximation for the accuracy and the correlations of the topographic gravity model and is derived under the assumptions mentioned in Sect. 2.1. This covariance can for example be calculated from the residuals $\Delta\mathbf{l}$, because the topography is usually associated with the largest source of uncertainties in $\Delta\mathbf{l}$. The approach of empirical covariance fitting is quite common in regional geoid modeling and is described in Tscherning and Rapp (1974). Note that the resulting covariance matrices are designed to describe a finite dimensional space and therefore disregard the gravity signal above degree $N_{\max}$. Adopting this approach our notation changes as follows, again marking the covariance matrices derived from the total average $\bar{\mathrm{E}}$ with a bar, as in Sect. 2.1.

$$\begin{aligned}\mathbf{C}^{\mathrm{t}}_{\hat{\mathbf{l}}\hat{\mathbf{l}}} &\approx \bar{\mathbf{C}}^{\mathrm{t}}_{\hat{\mathbf{l}}\hat{\mathbf{l}}}\\ \mathbf{C}^{\mathrm{t}}_{\hat{\mathbf{s}}\hat{\mathbf{l}}} &\approx \bar{\mathbf{C}}^{\mathrm{t}}_{\hat{\mathbf{s}}\hat{\mathbf{l}}}\\ \mathbf{C}^{\mathrm{t}}_{\hat{\mathbf{s}}\hat{\mathbf{s}}} &\approx \bar{\mathbf{C}}^{\mathrm{t}}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}\end{aligned} \tag{45}$$

By inserting $\mathbf{C}_{\hat{\mathbf{s}}\hat{\mathbf{l}}}$, $\mathbf{C}_{\hat{\mathbf{l}}\hat{\mathbf{l}}}$, $\hat{\mathbf{l}}$ and $\hat{\mathbf{s}}$ from the previous equations into Eq. 30, we obtain the final formulation of RLSC including the GGM and the topographic gravity model

$$\mathbf{s} = \underbrace{(\mathbf{C}^{\mathrm{m}}_{\hat{\mathbf{s}}\hat{\mathbf{l}}} + \bar{\mathbf{C}}^{\mathrm{t}}_{\hat{\mathbf{s}}\hat{\mathbf{l}}})}_{\text{Part 1}}\ \underbrace{(\mathbf{C}_{\mathrm{ll}} + \mathbf{C}^{\mathrm{m}}_{\hat{\mathbf{l}}\hat{\mathbf{l}}} + \bar{\mathbf{C}}^{\mathrm{t}}_{\hat{\mathbf{l}}\hat{\mathbf{l}}})^{-1}}_{\text{Part 2}}\ \underbrace{(\mathbf{l} - \hat{\mathbf{l}}^{\mathrm{m}} - \hat{\mathbf{l}}^{\mathrm{t}})}_{\text{Part 3}} + \underbrace{\hat{\mathbf{s}}^{\mathrm{m}} + \hat{\mathbf{s}}^{\mathrm{t}}}_{\text{Part 4}}. \tag{46}$$

Since Eq. 46 is very important in the following sections, we look at its various parts in more detail. Part 3 contains the residual input to RLSC subject to the condition that it is centered. Here, the remove step is performed by reducing $\hat{\mathbf{l}}^{\mathrm{m}}$ and $\hat{\mathbf{l}}^{\mathrm{t}}$ from the observations $\mathbf{l}$. Part 2 describes the uncertainties of part 3 accordingly. Every quantity from the input $(\mathbf{l}, \hat{\mathbf{l}}^{\mathrm{m}}, \hat{\mathbf{l}}^{\mathrm{t}})$ has its own covariance matrix $(\mathbf{C}_{\mathrm{ll}}, \mathbf{C}^{\mathrm{m}}_{\hat{\mathbf{l}}\hat{\mathbf{l}}}, \bar{\mathbf{C}}^{\mathrm{t}}_{\hat{\mathbf{l}}\hat{\mathbf{l}}})$. Similarly, we have the covariance matrices between input and output uncertainties in part 1, except that the covariance between the accuracy of the observations and the output $\mathbf{s}$ is missing because the two are assumed to be uncorrelated (Eq. 27). Finally, part 4 describes the restore step in the output functional $(\hat{\mathbf{s}}^{\mathrm{m}}, \hat{\mathbf{s}}^{\mathrm{t}})$.

The approach taken from Eq. 46 uses the full covariance information of a GGM up to a certain degree N, and topographic information above, from degree $N+1$ up to degree $N_{\max}$. The approach therefore delivers several advantages compared to the approach of Moritz (1980):

1. The observations $\mathbf{l}$ themselves do not have to be centered. Instead, we use the condition that GGM and topographic gravity model have the same offset as the observations $\mathbf{l}$ so that $\mathbf{l} - \hat{\mathbf{l}}^{\mathrm{m}} - \hat{\mathbf{l}}^{\mathrm{t}}$ is centered.
2. The covariance matrices $\mathbf{C}^{\mathrm{m}}_{\hat{\mathbf{s}}\hat{\mathbf{l}}}$ and $\mathbf{C}^{\mathrm{m}}_{\hat{\mathbf{l}}\hat{\mathbf{l}}}$ are neither homogeneous nor isotropic but fit perfectly to a target area and can benefit from the continuously improving quality of (high-resolution) GGMs.

3. This approach uses only residuals as input for LSC, because it can be shown that this yields the best results (see Sect. 2.2).
4. RLSC can give a realistic formal error estimate, because with the inclusion of the GGM uncertainties all the stochastic information is included in the resulting error covariance matrix $\mathbf{E}_{ss}$, which is calculated (analogously to Eq. 46) by inserting $\mathbf{C}^{m}_{\hat{l}\hat{l}}$, $\mathbf{C}^{m}_{\hat{s}\hat{l}}$ and $\mathbf{C}^{m}_{\hat{s}\hat{s}}$ into Eq. 32

$$\begin{aligned}\mathbf{E}_{ss} &= \mathbf{C}_{\hat{s}\hat{s}} - \mathbf{C}_{\hat{s}\hat{l}}\,(\mathbf{C}_{ll} + \mathbf{C}_{\hat{l}\hat{l}})^{-1}\mathbf{C}^{T}_{\hat{s}\hat{l}} \\ &= (\mathbf{C}^{m}_{\hat{s}\hat{s}} + \bar{\mathbf{C}}^{t}_{\hat{s}\hat{s}}) - (\mathbf{C}^{m}_{\hat{s}\hat{l}} + \bar{\mathbf{C}}^{t}_{\hat{s}\hat{l}}) \\ &\quad \times (\mathbf{C}_{ll} + \mathbf{C}^{m}_{\hat{l}\hat{l}} + \bar{\mathbf{C}}^{t}_{\hat{l}\hat{l}})^{-1}(\mathbf{C}^{m}_{\hat{s}\hat{l}} + \bar{\mathbf{C}}^{t}_{\hat{s}\hat{l}})^{T}. \end{aligned} \tag{47}$$

In comparison with the existing literature that has already included accuracy information from a satellite-only model (e.g., Haagmans and van Gelderen 1991; Pail et al. 2010; Sansò 2013) our approach still maintains the advantages 1 and 2. Point 4 is only advantageous in comparison with approaches that do not include all stochastic parts in LSC (e.g., Forsberg and Tscherning 1981; Hofmann-Wellenhof and Moritz 2006; Rieser 2015).

3 Data and simulation concept

We show the benefit of including full covariance matrices from a high-resolution GGM in collocation by a comparison between RLSC (Eq. 46) and a comparable approach without GGM covariance matrices, which we refer to as standard LSC. We derive standard LSC

$$\bar{\mathbf{s}} = \bar{\mathbf{C}}^{t}_{\hat{s}\hat{l}}\,(\mathbf{C}_{ll} + \bar{\mathbf{C}}^{t}_{\hat{l}\hat{l}})^{-1}(\mathbf{l} - \hat{\mathbf{l}}^{m} - \hat{\mathbf{l}}^{t}) + \hat{\mathbf{s}}^{m} + \hat{\mathbf{s}}^{t}, \tag{48}$$

from the approach by Moritz (1980) (Eq. 12), but include an RCR approach for the GGM and the topographic gravity model that is analogous to Eq. 46. Standard LSC is thereby equivalent to RLSC (Eq. 46) except that the two covariance matrices $\mathbf{C}^{m}_{\hat{l}\hat{l}}$ and $\mathbf{C}^{m}_{\hat{s}\hat{l}}$ are missing, because the GGM is either assumed to be error-free or its noise component is implicitly included in the model covariance function, which is used to fit the empirical covariance function (Tscherning and Rapp 1974). In case of an error-free GGM the corresponding error covariance matrix $\bar{\mathbf{E}}_{\bar{s}\bar{s}}$ results in accuracy estimates of $\bar{\mathbf{s}}$ that are overly optimistic

$$\bar{\mathbf{E}}_{\bar{s}\bar{s}} = \bar{\mathbf{C}}^{t}_{\hat{s}\hat{s}} - \bar{\mathbf{C}}^{t}_{\hat{s}\hat{l}}\,(\mathbf{C}_{ll} + \bar{\mathbf{C}}^{t}_{\hat{l}\hat{l}})^{-1}\bar{\mathbf{C}}^{t\,T}_{\hat{s}\hat{l}}. \tag{49}$$

The comparison of the two methods works best within a synthetic test case scenario that allows us to compute the residuals of the two methods by knowing the pre-defined truth. Moreover, it makes it possible to evaluate and assess the formal error estimate of the stochastic part of the collocation (compare RLSC: Eq. 47 and standard LSC: Eq. 49). To obtain realistic results from a synthetic test case, we add noise to the observations and to the GGM (as shown below).

The synthetic test case is calculated for one of the most common challenges of regional geoid modeling, i.e., the calculation of the geoid from gravity anomalies given at selected input points. As the calculation should be an evaluation test for a real geoid computation, we use the input positions of actual gravity observations for our synthetic test case. Output is a regular grid of geoid heights in the target area, the South American Andes. This is a useful test area for the study since the Andes are one of the most demanding regions worldwide in terms of gravity signal variations, heterogeneous data distribution, and topographic effects. Also, we have access to the actual terrestrial database for this area (Hosse et al. 2014) and can therefore realistically simulate the data distribution on land.

The terrestrial gravity observation points are inhomogeneously distributed in an area between longitude [− 72° to − 66°] and latitude [− 29° to − 18°] that is the northern part of Chile (CHL), north-western Argentina (ARG) and south-western Bolivia (BOL, Fig. 1a). Furthermore, we add altimetry observations to cover the ocean region of the test area. The altimetry observations are given as a regular grid with 5′ spacing in the Pacific Ocean (PAC) bordered by the − 74° longitude meridian and the − 31° parallel. Altimetric gravity data on a regular grid are available from various ocean products, e.g., DTU13 (Andersen et al. 2015) or Sandwell and Smith (2009) and can be found for download at the corresponding websites. The distribution of the input points in Fig. 1a is displayed together with the terrain height. The image shows the Pacific Ocean to the West and the main ridge of the Andes from north to south. The test case includes areas with quite dense terrestrial observations, while others show large data gaps. Overall, there are 14,613 gravity anomaly input points with 7814 from terrestrial observations and 6799 altimetry grid points. As output, geoid heights $\mathbf{N}$ are estimated for the whole study area as a regular 5′ grid, which results in 21,901 output points (Fig. 1b).

As a GGM we use the XGM2016 (Pail et al. 2018), because we have its full normal equation system available that allows the calculation of the covariance matrices $\mathbf{C}^{m}_{\hat{l}\hat{l}}$, $\mathbf{C}^{m}_{\hat{s}\hat{l}}$ and $\mathbf{C}^{m}_{\hat{s}\hat{s}}$ (Eq. 37), as well as input and output functionals ($\hat{\mathbf{l}}^{m}$, $\hat{\mathbf{s}}^{m}$) in Eq. 36. The XGM2016 is a combined gravity field model up to degree 719 which uses relative regional weighting for the combination of terrestrial and satellite information. Compared to the GOCO05c model (Fecher et al. 2017), the XGM2016 includes an improved terrestrial data set provided by the National Geospatial-Intelligence Agency (NGA). In our opinion the XGM2016 is one of the most consistently calculated high-resolution GGMs, and its good

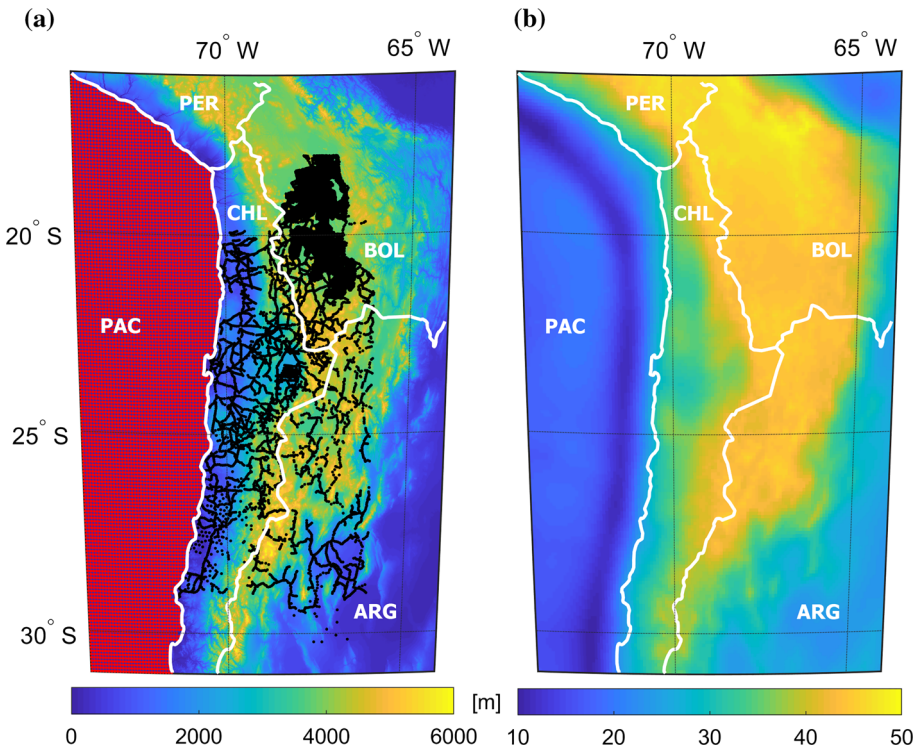

Fig. 1 LSC input points and pre-defined truth of the test scenario with borders as white lines. **a** Terrain heights with the distribution of the input points with altimetry observations (red dots) and terrestrial measurements (black dots). **b** Regular 5′ resolution grid of geoid heights that are the pre-defined truth $\mathbf{s}$ for the synthetic test case

performance in comparison with other high-resolution gravity field models such as EGM2008 (Pavlis et al. 2012), EIGEN-6c4 (Förste et al. 2014) and GOCO05c (Fecher et al. 2017) is demonstrated in Pail et al. (2018). In a post-analysis of the variance–covariance matrix of XGM2016, however, we found out that the accuracy estimates in the higher degrees (beyond the resolution of the satellite data) underestimate the true noise behavior. The regionally varying weights of XGM2016 have been computed empirically from comparison of satellite and low-pass filtered ground data (Pail et al. 2018). This procedure ensures an adequate relative weighting of satellite and ground data, but implicitly disregards signal content in the high degrees. In a new version of XGM2016, which shall be released soon, we will take this omission error into account by means of scaling the error estimates by their high-frequency signal content. From this analysis, we consider a factor of 3 as a reasonable value to calibrate the error estimates in the area of our synthetic test case. Therefore, we scale the XGM2016 accuracy estimates in Sects. 4.1 and 4.2 by a factor of 3 and hence must increase the elements of the covariance matrices $\mathbf{C}^{\mathrm{m}}_{\hat{\mathbf{l}}\hat{\mathbf{l}}}$, $\mathbf{C}^{\mathrm{m}}_{\hat{\mathbf{s}}\hat{\mathbf{l}}}$ and $\mathbf{C}^{\mathrm{m}}_{\hat{\mathbf{s}}\hat{\mathbf{s}}}$ by a factor of 9. However, in Sect. 4.3 we present results with the original XGM2016 accuracy and thus investigate the impact of the accuracy of the GGM.

For the definition of the true gravity signal, we choose a XGM2016 and EIGEN-6c4 combination model ['GOCE-OGMOC', Gruber and Willberg (2019)] up to degree 2190 and calculate $\mathbf{s}$ and $\mathbf{t}$ from it. The main reason for using this model is that we can be sure it does not contain systematic errors relative to the XGM2016 model because the long wavelength parts are identical. The geoid heights $\mathbf{s}$ that we use as pre-defined truth, and to which we subsequently compare our results, are presented in Fig. 1b. In the target area, the geoid heights vary from 10 m in the Atacama Trench up to almost 50 m in the plateau of Bolivia. The true gravity anomalies $\mathbf{t}$ are used to calculate the simulated gravity observations $\mathbf{l}$ by adding white noise $\mathbf{n}$ which is calculated from a random noise vector $\mathbf{e}^{\mathrm{l}}$ and the accuracy of the observations σ^{l}

$$\mathbf{l} = \mathbf{t} + \mathbf{n} = \mathbf{t} + \sigma^{\mathrm{l}}\mathbf{e}^{\mathrm{l}}. \tag{50}$$

σ^{l} is also used in the covariance matrix $\mathbf{C}_{\mathrm{ll}}$ of the observations where we assume the accuracies of different observations $\mathbf{l}$ to be uncorrelated and therefore obtain the diagonal matrix $\mathbf{C}_{\mathrm{ll}}$

$$\mathbf{C}_{\mathrm{ll}} = \sigma^{\mathrm{l}}\mathbf{I}\,. \tag{51}$$

We must consider this as a simplification because we assume that we know the accuracy σ^{l} of the input observations which is not guaranteed in reality.

For $\hat{\mathbf{l}}^{\mathrm{t}}$ and $\hat{\mathbf{s}}^{\mathrm{t}}$ we use a spherical harmonic synthesis from the topographic gravity model dV_ELL_Earth2014 (Rexer et al. 2016) from degree $N+1$ to $N_{\max}$. Note that it is not necessary for $\hat{\mathbf{l}}^{\mathrm{t}}$ (or $\hat{\mathbf{s}}^{\mathrm{t}}$) to contain all remaining gravity signals above the degree N of the GGM provided it is without systematic error and independent of $\hat{\mathbf{l}}^{\mathrm{m}}$ (conditions in the

Table 1 Overview of all quantities, their definition and source in our synthetic test case scenario

Quantity	Description	Source	Equations
$\mathbf{t}$	Pre-defined truth synthesized in input functional	GOCE-OGMOC	–
$\mathbf{s}$	Pre-defined truth synthesized in output functional	GOCE-OGMOC	–
$\mathbf{C}^{\mathrm{m}}_{\hat{\mathrm{l}}\hat{\mathrm{l}}}$	GGM: error covariance function for input points	XGM2016	37
$\mathbf{C}^{\mathrm{m}}_{\hat{\mathrm{s}}\hat{\mathrm{l}}}$	GGM: error covariance function for input and output	XGM2016	37
$\mathbf{C}^{\mathrm{m}}_{\hat{\mathrm{s}}\hat{\mathrm{s}}}$	GGM: error covariance function for output points	XGM2016	37
$\hat{\mathbf{l}}^{\mathrm{m}}$	GGM: synthesized in input functional (remove step)	XGM2016	36
$\hat{\mathbf{s}}^{\mathrm{m}}$	GGM: synthesized in output functional (restore step)	XGM2016	36
N	Maximum degree of the GGM	XGM2016	44
$\mathbf{l}^{\mathrm{n}}$	Colored noise of GGM at input points (synthetic case only)	White noise and XGM2016	56
$\mathbf{s}^{\mathrm{n}}$	Colored noise of GGM at output points (synthetic case only)	White noise and XGM2016	56
$\boldsymbol{\kappa}^{\mathrm{n}}$	Combined colored noise of GGM (synthetic case only)	White noise and XGM2016	54
$\bar{\mathbf{C}}^{\mathrm{t}}_{\hat{\mathrm{l}}\hat{\mathrm{l}}}$	Topo. gravity: error covariance function for input points	Fit to empirical covariance	45
$\bar{\mathbf{C}}^{\mathrm{t}}_{\hat{\mathrm{s}}\hat{\mathrm{l}}}$	Topo. gravity: error covariance function for input and output	Fit to empirical covariance	45
$\bar{\mathbf{C}}^{\mathrm{t}}_{\hat{\mathrm{s}}\hat{\mathrm{s}}}$	Topo. gravity: error covariance function for output points	Fit to empirical covariance	45
$\hat{\mathbf{l}}^{\mathrm{t}}$	Topo. gravity: synthesized in input functional (remove step)	dV_ELL_Earth2014	39
$\hat{\mathbf{s}}^{\mathrm{t}}$	Topo. gravity: synthesized in output functional (restore step)	dV_ELL_Earth2014	40
$N_{\max}$	Maximum degree of the topographic gravity model	dV_ELL_Earth2014	44
σ^{l}	Assumed accuracy of input observations	Accuracy of observations	50
$\mathbf{C}_{\mathrm{ll}}$	Error covariance function of observations, diagonal matrix	Accuracy of observations	51
$\mathbf{e}^{\mathrm{l}}$	White noise vector (synthetic case only)	White noise	50
$\mathbf{e}^{\mathrm{m}}$	White noise vector (synthetic case only)	White noise	54

definition of Sect. 2.4). An overview of all quantities of the synthetic test case can be found in Table 1.

Finally, we add noise to the GGM. Since the XGM2016 model uses regional varying weighting, colored noise that actually describes the regionally varying accuracy of the model is added in terms of a (random) realization based on the variance–covariance information. For the calculation of this covariance, we combine the observations $\mathbf{l}$ and the output $\mathbf{s}$ in one vector $\boldsymbol{\kappa}$

$$\boldsymbol{\kappa} = \begin{pmatrix} \mathbf{l} \\ \mathbf{s} \end{pmatrix}, \tag{52}$$

and propagate the variance–covariance matrix of the GGM $\mathbf{C}^{\mathrm{m}}_{\mathrm{ff}}$ to the combined covariance matrix $\mathbf{C}^{\mathrm{m}}_{\hat{\kappa}\hat{\kappa}}$ of input and output by attaching the three formulas of Eq. 37 to give

$$\mathbf{C}^{\mathrm{m}}_{\hat{\kappa}\hat{\kappa}} = \mathbf{A}^{\mathrm{f}}_{\kappa}\, \mathbf{C}^{\mathrm{m}}_{\mathrm{ff}} \mathbf{A}^{\mathrm{f\,T}}_{\kappa}, \tag{53}$$

where $\mathbf{A}^{\mathrm{f}}_{\kappa}$ is the design matrix for the transformation from SH coefficients to the input and output points $\boldsymbol{\kappa}$. We obtain the colored noise $\boldsymbol{\kappa}^{\mathrm{n}}$ of input and output by multiplying a random vector $\mathbf{e}^{\mathrm{m}}$ with white noise characteristics and the Cholesky decomposition of the combined covariance matrix $\mathbf{C}^{\mathrm{m}}_{\hat{\kappa}\hat{\kappa}}$

$$\boldsymbol{\kappa}^{\mathrm{n}} = \mathrm{chol}(\mathbf{C}^{\mathrm{m}}_{\hat{\kappa}\hat{\kappa}})\, \mathbf{e}^{\mathrm{m}} = \mathrm{chol}\begin{pmatrix} \mathbf{C}^{\mathrm{m}}_{\hat{\mathrm{l}}\hat{\mathrm{l}}} & \mathbf{C}^{\mathrm{m\,T}}_{\hat{\mathrm{s}}\hat{\mathrm{l}}} \\ \mathbf{C}^{\mathrm{m}}_{\hat{\mathrm{s}}\hat{\mathrm{l}}} & \mathbf{C}^{\mathrm{m}}_{\hat{\mathrm{s}}\hat{\mathrm{s}}} \end{pmatrix} \mathbf{e}^{\mathrm{m}}, \tag{54}$$

which is defined by

$$\mathbf{X} = \mathrm{chol}(\mathbf{X})\, \mathrm{chol}(\mathbf{X})^{\mathrm{T}}, \tag{55}$$

where $\mathbf{X}$ is a positive definite matrix (e.g., covariance matrix). Consequently, we obtain the noise of the GGM for the input $\mathbf{l}^{\mathrm{n}}$ and the output $\mathbf{s}^{\mathrm{n}}$ analogously to the definition in Eq. 52 by

$$\boldsymbol{\kappa}^{\mathrm{n}} = \begin{pmatrix} \mathbf{l}^{\mathrm{n}} \\ \mathbf{s}^{\mathrm{n}} \end{pmatrix}, \tag{56}$$

and add them correspondingly to $\hat{\mathbf{l}}^{\mathrm{m}}$ and $\hat{\mathbf{s}}^{\mathrm{m}}$. Thus, for a synthetic test case we adjust the formula of RLSC (Eq. 46) with the inclusion of the observation noise $\mathbf{n}$ (Eq. 50) and the noise of the XGM2016 model $\mathbf{l}^{\mathrm{n}}$ and $\mathbf{s}^{\mathrm{n}}$ (Eqs. 54, 56)

$$\begin{aligned} \mathbf{s} = {} & (\mathbf{C}^{\mathrm{m}}_{\hat{\mathrm{s}}\hat{\mathrm{l}}} + \bar{\mathbf{C}}^{\mathrm{t}}_{\hat{\mathrm{s}}\hat{\mathrm{l}}})\, (\mathbf{C}_{\mathrm{ll}} + \mathbf{C}^{\mathrm{m}}_{\hat{\mathrm{l}}\hat{\mathrm{l}}} + \bar{\mathbf{C}}^{\mathrm{t}}_{\hat{\mathrm{l}}\hat{\mathrm{l}}})^{-1} ((\mathbf{t} + \mathbf{n}) - (\hat{\mathbf{l}}^{\mathrm{m}} + \mathbf{l}^{\mathrm{n}}) - \hat{\mathbf{l}}^{\mathrm{t}}) \\ & + (\hat{\mathbf{s}}^{\mathrm{m}} + \mathbf{s}^{\mathrm{n}}) + \hat{\mathbf{s}}^{\mathrm{t}}, \end{aligned} \tag{57}$$

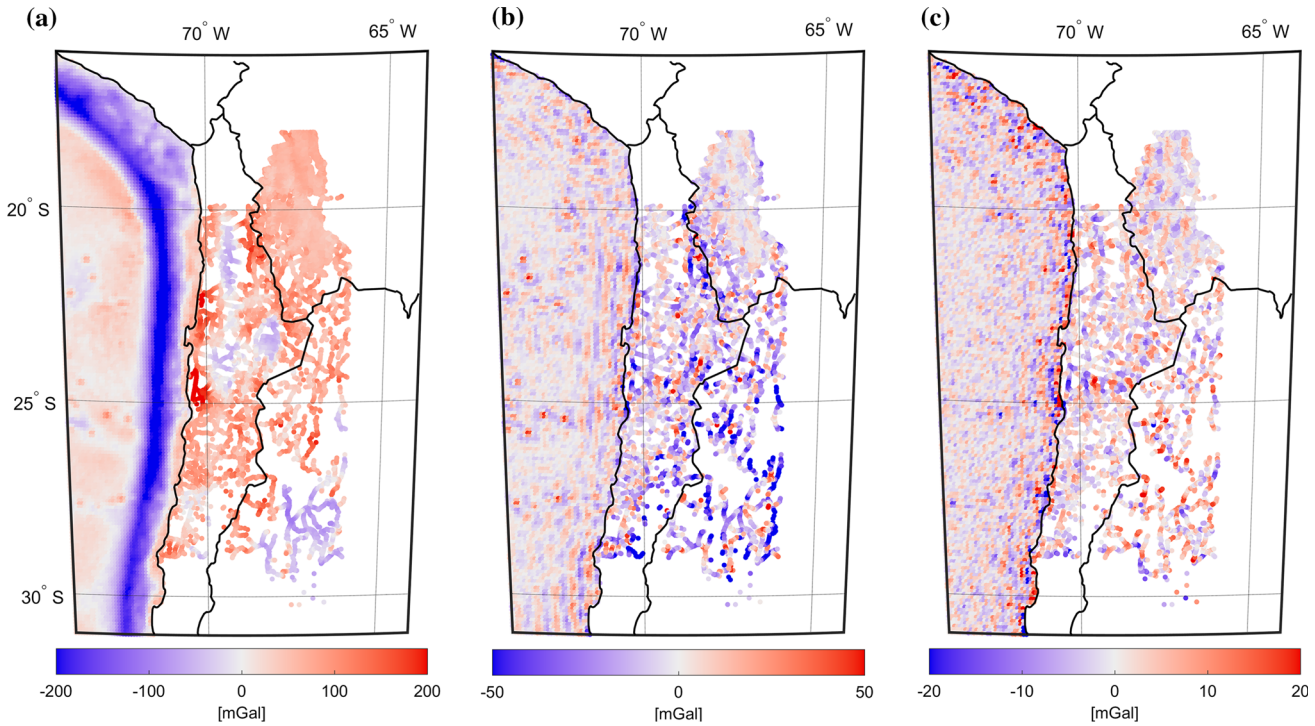

Fig. 2 LSC remove step for the gravity anomalies: **a** terrestrial observations with $\sigma_n = 1$ mGal, **b** observations after reduction of XGM2016 model, **c** LSC input $\mathbf{\Delta l}$ after reducing XGM2016 and topographic gravity

and analogously, for standard LSC from Eq. 48

$$\bar{\mathbf{s}} = \bar{\mathbf{C}}^{\mathrm{t}}_{\hat{\mathrm{s}}\hat{\mathrm{l}}} \left(\mathbf{C}_{\mathrm{ll}} + \bar{\mathbf{C}}^{\mathrm{t}}_{\hat{\mathrm{l}}\hat{\mathrm{l}}}\right)^{-1} ((\mathbf{t}+\mathbf{n}) - (\hat{\mathbf{l}}^{\mathrm{m}} + \mathbf{l}^{\mathrm{n}}) - \hat{\mathbf{l}}^{\mathrm{t}}) + (\hat{\mathbf{s}}^{\mathrm{m}} + \mathbf{s}^{\mathrm{n}}) + \hat{\mathbf{s}}^{\mathrm{t}}. \tag{58}$$

In both cases, the stochastic part of the collocation does not change, so that the calculation of the error covariance matrices $\mathbf{E}_{\mathrm{ss}}$ (Eq. 47) and $\bar{\mathbf{E}}_{\bar{\mathrm{s}}\bar{\mathrm{s}}}$ (Eq. 49) remain unchanged in the synthetic test case.

The effect of the remove step in Eqs. 57 and 58 is presented in Fig. 2, where the observations always include a white noise of $\sigma^1 = 1$ mGal (see Eq. 50). The XGM2016 model up to degree $N = 719$ contains the colored noise that is calculated from the Cholesky decomposition (Eq. 54), and the topographic gravity reduction is used from $N + 1$ to 2190. Figure 2a shows the original, noisy observations $\mathbf{l}$. We reduce the observations $\mathbf{l}$ first by the high-resolution XGM2016 model $(\hat{\mathbf{l}}^{\mathrm{m}} + \mathbf{l}^{\mathrm{n}})$ (Fig. 2b) and afterwards also by the gravity signal that is related to the topography, resulting in the residual input vector $\mathbf{l} - (\hat{\mathbf{l}}^{\mathrm{m}} + \mathbf{l}^{\mathrm{n}}) - \hat{\mathbf{l}}^{\mathrm{t}}$ (Fig. 2c). Note that the color scale decreases significantly between Fig. 2a and Fig. 2c. This is also apparent in Table 2, which shows the mean value and the standard deviation of the corresponding data sets in Fig. 2. Regarding its standard deviation (SD), the signal is reduced by about 80% by subtracting the XGM2016 model and more than 90% in combination with the topography. With a mean value of only 0.2 mGal the condition of a centered LSC input is almost fulfilled in this case. In our numerical simulation, the final LSC input contains the noise of the observations and the GGM as well as the inaccuracy of the topographic gravity model. These three effects are modeled in the related error covariance matrices of our approach (Eq. 57). The inaccuracy of the topographic gravity model thereby also includes density anomalies in the spectral range from $N + 1$ to $N_{\max}$, which are contained in the residual observations. These anomalies are not adequately represented in the topographic gravity model, because it assumes a constant topographic density (Rexer 2017, Chapter 3.2).

Currently we do not have accuracy information for the topographic gravity model (see Sect. 2.4) and this kind of information is not available for any of the topographic gravity models (Rexer et al. 2016; Grombein et al. 2016). However, because we assume the largest part of the LSC input results from differences between the topographic gravity model and the observations in the high frequencies (as they contain also the effect of density anomalies), we calculate a Model Covariance Function (MCF) that fits the LSC input. This is considered to be a standard approach in LSC which is used for the calculation with signal degree variances (Moritz 1980, Chapter 12) as well as for empirical covariance fitting with the most common approach by Tscherning and Rapp (1974). The result is an isotropic and homogeneous covariance matrix for the spectral range above degree 719 that fits

Table 2 Consistent numerical analysis of the LSC remove step in Fig. 2

Signal description	Notation	Mean value (mGal)	SD (mGal)	Figures
Observations	$\mathbf{l} = \mathbf{t} + \mathbf{n}$	10.2	78.7	2a
Reduced by XGM2016 only	$\mathbf{l} - (\hat{\mathbf{l}}^{\mathrm{m}} + \mathbf{l}^{\mathrm{n}})$	− 3.3	15.2	2b
Reduced by XGM2016 + topographic gravity	$\mathbf{l} - (\hat{\mathbf{l}}^{\mathrm{m}} + \mathbf{l}^{\mathrm{n}}) - \hat{\mathbf{l}}^{\mathrm{t}}$	0.2	5.5	2c

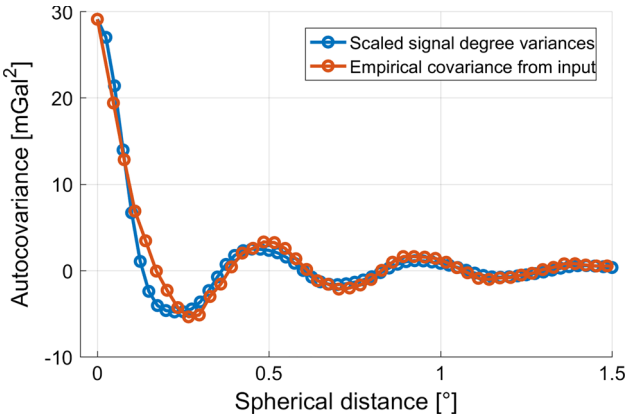

Fig. 3 Empirical covariance function that is calculated from the LSC residuals in the target area (red) and the scaled covariance from signal degree variances (blue) as it is used for LSC. Both covariance functions are shown dependent on the spherical distance between two points

to our input residuals. In general, we calculate the covariance matrices $\bar{\mathbf{C}}^{\mathrm{t}}_{\hat{\mathbf{l}}\hat{\mathbf{l}}}$ and $\bar{\mathbf{C}}^{\mathrm{t}}_{\hat{\mathbf{s}}\hat{\mathbf{l}}}$ from signal degree variances and refer to Moritz (1980) for more details. However, global signal degree variance models describe an average Earth, but the South American Andes are not an average area in terms of gravity signal (due to large topographic masses). Thus, we have to use a scale factor to fit model degree variances (blue, Fig. 3) to the Empirical Covariance Function (ECF) of the LSC input residuals (red). In Fig. 3, we calculate signal degree variances, according to our data simulation, from the difference between the GOCE-OGMOC and the dV_ELL_Earth2014 model from degree 720 to 2190 and scale them by a factor of 4.25. Figure 3 then shows the ECF and MCF in dependence of the spherical distance between two points. In this case the correlation length amounts to only 0.12°.

4 Results of the synthetic test scenario

Section 3 describes the method, the parameters and the sources that are used to evaluate RLSC (Eq. 57) in comparison with standard LSC (Eq. 58). Both methods use the RCR concept where a GGM and the topographic gravity are removed before the collocation, and their effects are restored (to the output functional) afterwards. For the calculation we set the degree N of the GGM first and use the topographic gravity model always from degree $N + 1$ to $N_{\max} = 2190$. The results of $\mathbf{s}$ (or $\bar{\mathbf{s}}$) are compared with the assumed truth $\mathfrak{s}$, and the difference indicates the accuracy of the collocation result under these pre-defined conditions. In terms of numerical classification we define the Root Mean Square (RMS) over a target area with n^{points} points as a quality criterion for the LSC result $\mathbf{s}$ (or $\bar{\mathbf{s}}$)

$$\mathrm{RMS} = \sqrt{\frac{\sum_{\mathrm{points}} (\mathbf{s} - \mathfrak{s})^2}{n^{\mathrm{points}}}}. \tag{59}$$

We use an equivalent formula to characterize the Mean estimated standard deviation (MSD), which describes the stochastic accuracies and is derived from $\mathbf{E}_{\mathrm{ss}}$ (or $\bar{\mathbf{E}}_{\bar{\mathrm{s}}\bar{\mathrm{s}}}$) with $\mathrm{diag}(\mathbf{X})$ giving the main diagonal of a square matrix $\mathbf{X}$

$$\mathrm{MSD} = \sqrt{\frac{\sum_{\mathrm{points}} \mathrm{diag}(\mathbf{E}_{\mathrm{ss}})}{n^{\mathrm{points}}}}. \tag{60}$$

4.1 Simulation with satellite-only model resolution

At first we compare LSC for calculations with the XGM2016 model but use the model only up to degree 200 (case A). Consequently, RLSC resembles other approaches in which covariance functions from satellite-only global gravity fields are included to LSC (e.g., Pail et al. 2010; Sansò 2013). However, the actual computation of the covariance matrices in our case is still different (Eq. 37). Standard LSC in case A is similar to LSC approaches that do not account for the accuracy of the GGM (e.g., Rieser 2015; Hofmann-Wellenhof and Moritz 2006). The main reason why it is so common to use a GGM up to degree of 180–250 in LSC is the high accuracy of GGMs in this frequency range (Gruber et al. 2011), which is mainly due to the GOCE mission (Drinkwater et al. 2003). Thus, in the first case we use the XGM2016 model with $n^{\mathrm{m}} \in \{2, 200\}$ for all quantities with the superscript 'm', and $n^{\mathrm{t}} \in \{201, 2190\}$ accordingly. The accuracy of the input observations σ_n is assumed to be 1 mGal. For the calculations in Sects. 4.1 and 4.2, the accuracy of the XGM2016 model is multiplied by a factor of 3 (see Sect. 3). It is added

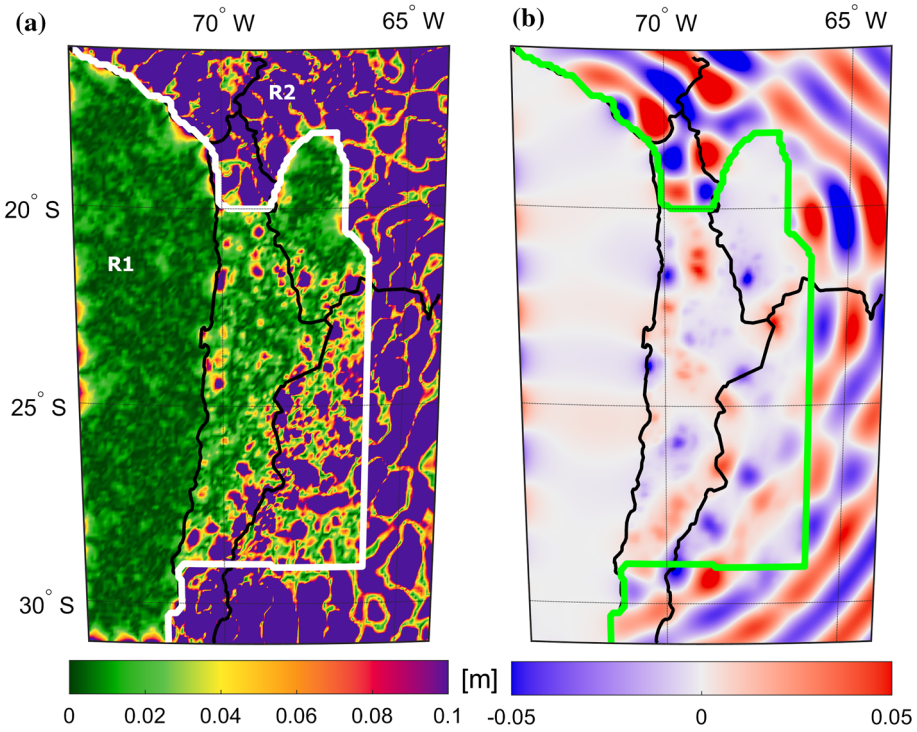

Fig. 4 Case A: Geoid height residuals with the XGM2016 model up to degree 200. Presented is **a** the absolute difference between the pre-defined truth $\mathbf{s}$ and the RLSC result $\mathbf{s}$, **b** the difference between the result of standard LSC $\bar{\mathbf{s}}$ and RLSC $\mathbf{s}$

as described in Sect. 3 but limited to degree 200 in case A. An overview of all quantities is given in Table 1.

The collocation result $\mathbf{s}$ of RLSC from Eq. 57 in case A is presented as an absolute geoid height difference (residuals) to the pre-defined truth $\mathbf{s}$ in Fig. 4a. In Fig. 4 and the following images, we separate a region with terrestrial data coverage (R1, left side) from an area without ground data (R2, right side) by a white line, because we see large differences between these two regions. For the interpretation we focus mainly on the region R1. In Fig. 4a, we see a high correlation between the LSC residuals and the positions of the input observations (see Fig. 1a): in the Pacific Ocean and the bulge in the north we have dense observations and therefore only small residuals mainly below 1 cm. In region R2, the residuals amount to more than 10 cm which is also the case for large parts of R1 in the Argentinian area (southeast). We conclude that the number of terrestrial observations in this area is not dense enough to describe the geoid height signals above degree 200. The RMS (Eq. 59) of the region R1 accounts for 6.0 cm, and is obviously significantly larger in R2.

In case A, the RLSC results $\mathbf{s}$ (Fig. 4a) and standard LSC $\bar{\mathbf{s}}$ (Eq. 58) show almost the same RMS (Table 3) and only small differences among each other. The difference $\bar{\mathbf{s}}$ minus $\mathbf{s}$ in Fig. 4b shows maximum values of about ± 5.0 cm, but these values occur to a large extent in areas without observations. In areas with dense gravity observations (e.g., Pacific Ocean) the differences between the two methods are of the order of just a few millimeters. The reason for the differences being so small is the fact that the extracted variances from $\mathbf{C}^{\mathrm{m}}_{\hat{l}\hat{l}}$ and $\mathbf{C}^{\mathrm{m}}_{\hat{s}\hat{l}}$ of the GGM are insignificantly small compared to the corresponding variances of the topographic gravity in $\bar{\mathbf{C}}^{\mathrm{t}}_{\hat{l}\hat{l}}$ and $\bar{\mathbf{C}}^{\mathrm{t}}_{\hat{s}\hat{l}}$. Therefore, $\bar{\mathbf{C}}^{\mathrm{m}}_{\mathrm{ll}}$ and $\bar{\mathbf{C}}^{\mathrm{m}}_{\mathrm{sl}}$ from RLSC do not have an essential effect on the summation in Eq. 57.

$$\begin{aligned} \mathrm{diag}(\mathbf{C}^{\mathrm{m}}_{\hat{l}\hat{l}}) &<< \mathrm{diag}(\bar{\mathbf{C}}^{\mathrm{t}}_{\hat{l}\hat{l}}) \\ \mathrm{diag}(\mathbf{C}^{\mathrm{m}}_{\hat{s}\hat{l}}) &<< \mathrm{diag}(\bar{\mathbf{C}}^{\mathrm{t}}_{\hat{s}\hat{l}}) \end{aligned} \tag{61}$$

This is demonstrated in Fig. 5 which shows the variances that are extracted from $\bar{\mathbf{C}}^{\mathrm{t}}_{\hat{l}\hat{l}}$ (Fig. 5a) and $\mathbf{C}^{\mathrm{m}}_{\hat{l}\hat{l}}$ (Fig. 5b). The variances from the topographic gravity in $\bar{\mathbf{C}}^{\mathrm{t}}_{\hat{l}\hat{l}}$ are in the order of 1000 times larger than the variances of the GGM in $\mathbf{C}^{\mathrm{m}}_{\hat{l}\hat{l}}$. We can conclude that in case A the additional benefit of adding the covariance matrices $\mathbf{C}^{\mathrm{m}}_{\hat{l}\hat{l}}$ and $\mathbf{C}^{\mathrm{m}}_{\hat{s}\hat{l}}$ to LSC is small. This is also the explanation why LSC approaches often ignore the accuracy of a satellite-only resolution GGM completely (e.g., Hofmann-Wellenhof and Moritz 2006; Rieser 2015).

4.2 Simulation with full GGM resolution

However, this changes in case B where we use the full degree $N = 719$ of the XGM2016 model for reduction ($n^{\mathrm{m}} \in \{2, 719\}$) and thus apply the topographic gravity for $n^{\mathrm{t}} \in \{720, 2190\}$. In principle, the procedure remains the same, but we would like to emphasize that this set-up increases all matrix elements in $\mathbf{C}^{\mathrm{m}}_{\hat{l}\hat{l}}$, $\mathbf{C}^{\mathrm{m}}_{\hat{s}\hat{l}}$ and $\mathbf{C}^{\mathrm{m}}_{\hat{s}\hat{s}}$, and therefore also the noise of the XGM2016 model (see Eq. 54).

Table 3 Overview of the three synthetic test cases and the numerical results for region R1

Case	Sections	N	Signal	Method	RLSC (cm)	Standard LSC (cm)	Figures
A	4.1	200	Residuals	RMS	6.0	6.1	4
		200	Formal error	MSD	5.2	5.2	–
B	4.2	719	Residuals	RMS	3.9	6.9	6
		719	Formal error	MSD	3.6	2.2	7
C	4.3	719	Residuals	RMS	2.9	3.3	8
		719	Formal error	MSD	2.6	2.2	–

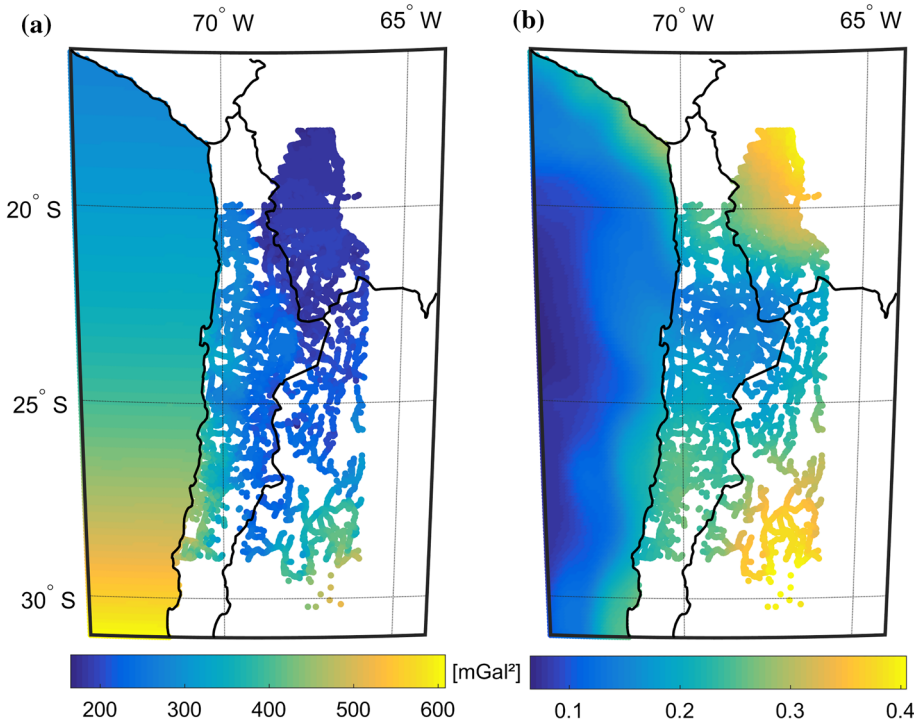

Fig. 5 Case A: Comparison of the extracted variances that are used in RLSC. **a** Variances extracted from the covariance matrix $\bar{\mathbf{C}}^{\mathrm{t}}_{\tilde{\imath}\tilde{\imath}}$ of the topographic gravity model which depends only on latitude and height. **b** Regional varying variance elements from the covariance matrix $\mathbf{C}^{\mathrm{m}}_{\tilde{\imath}\tilde{\imath}}$ of the XGM2016 model (to degree 200)

The results of a collocation with XGM2016 to full degree (case B) are visualized in Fig. 6, where we show again the absolute differences to the pre-defined truth $\mathfrak{s}$. From this point, we retain the form in which we present RLSC with GGM covariance (Eq. 57) on the left-hand side (Fig. 6a) and standard LSC (Eq. 58) on the right-hand side (Fig. 6b). It is obvious that RLSC performs much better in case B. In Fig. 6a, we again see a high correlation with the distribution of the input points. In areas with dense terrestrial observation, e.g., the Pacific Ocean or the bulge in the north, the residuals are mainly below 1 cm. The residuals over the well-covered Chilean area are much smaller than on the Argentinian side, which also results from the corresponding point distributions. In the region R2 the residuals are largest and often above 10 cm. In Fig. 6b, we have in general much larger residuals, so that even areas with dense gravity observations show residuals of at least 2 cm. The higher residuals primarily result from the noise of the GGM since standard LSC regards the noisy XGM2016 model as error-free and therefore fully reflects its error in the result. The RMS of standard LSC in region R1 accounts for 6.9 cm, while the one of RLSC is only 3.9 cm (Table 3).

In summary, RLSC and standard LSC perform very similarly with a satellite-only resolution GGM in case A, but when using a high-resolution GGM as in case B, there are significant improvements when including the GGM covariance (Table 3, Fig. 6). Furthermore, the benefit of including a high-resolution GGM itself is demonstrated when we compare the RLSC result in case B (Fig. 6a) with the result in case A (Fig. 4a). Even with the much higher error from the GGM in case B, the RMS for RLSC in region R1 is reduced from 6.0 to 3.9 cm. This improvement results at least partly from the area with sparse terrestrial measurements in Argentina that has clearly higher residuals in case A (Fig. 4a). However, even areas with small data gaps which are common in mountainous areas and the edges of observed areas benefit highly from the inclusion of a high-resolution GGM in RLSC. The latter is for example visible at the southern end of the Pacific Ocean where case A clearly shows edge effects while case B does not. Standard LSC, on the

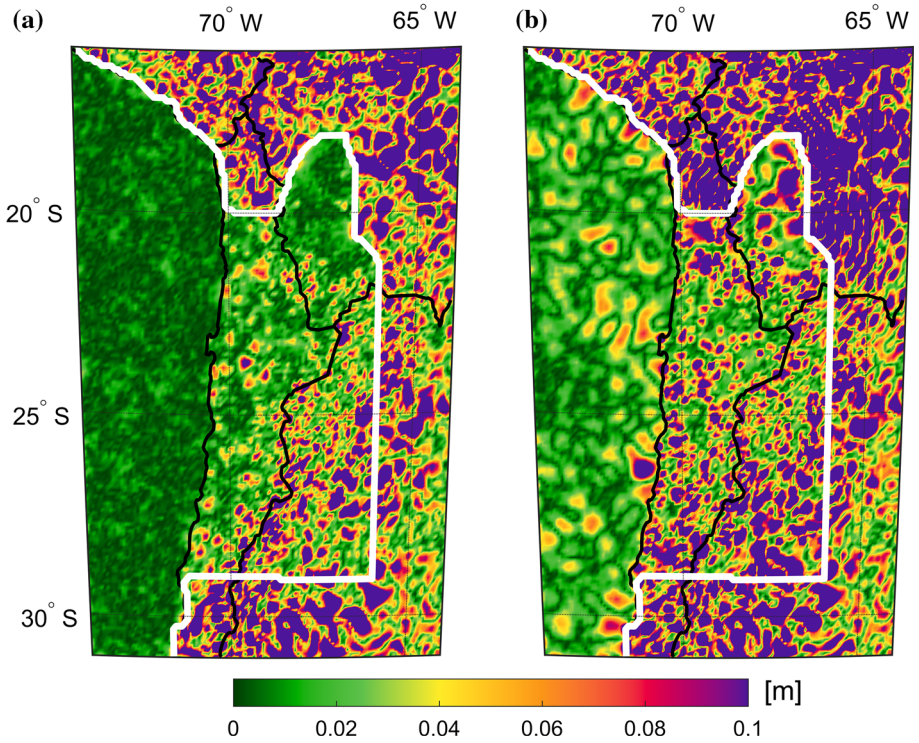

Fig. 6 Case B: Geoid height residuals with the XGM2016 model that is the absolute difference between the pre-defined truth $\mathbf{s}$ and **a** the RLSC result **s**, **b** the standard LSC result $\bar{\mathbf{s}}$

other hand, performs worse in case B, and its RMS increases due to the noise $\mathbf{l}^{\mathrm{n}}$ of the GGM from 6.1 cm in case A to 6.9 cm in case B. We conclude that, in contrast to standard LSC, RLSC is able to handle the noise in the GGM and gives good results in areas with a sufficient number of observations.

Figure 7 shows the formal error of case B, which is derived from the error covariance matrix $\mathbf{E}_{\mathrm{ss}}$ or respectively $\bar{\mathbf{E}}_{\bar{\mathrm{s}}\bar{\mathrm{s}}}$ as square root of the main diagonal elements (Eq. 60). The difference between Fig. 7a and Fig. 7b results from the neglected covariance matrices $\mathbf{C}^{\mathrm{m}}_{\hat{\mathrm{l}}\hat{\mathrm{l}}}$ and $\mathbf{C}^{\mathrm{m}}_{\hat{\mathrm{s}}\hat{\mathrm{l}}}$ in Eq. 58, and since covariance matrices are positive definite, the estimated error for standard LSC (Fig. 7b) is always smaller than the one from RLSC (Fig. 7a). The corresponding MSD (Eq. 60) in region R1 is 3.6 cm for RLSC and 2.2 cm for standard LSC. Thus, we see that the residuals and the formal errors agree much better (see Table 3) for RLSC (Figs. 6a + 7a) than for standard LSC (Figs. 6b + 7b). We again point out the Pacific Ocean and the bulge in Bolivia where in both cases the dark green values of Fig. 7b do not fit to the corresponding error in Fig. 6b. In contrast, we see that for RLSC essentially all peaks in the residuals (Fig. 6a) are indicated by higher values in Fig. 7a as well, which demonstrates that the formal error of RLSC fits much better to its residuals. We conclude that standard LSC cannot realistically represent the formal error since the accuracy of the GGM is not included. Therefore, the result of $\bar{\mathbf{E}}_{\bar{\mathrm{s}}\bar{\mathrm{s}}}$ is always too optimistic. The consistent calculation of $\mathbf{E}_{\mathrm{ss}}$ is considered as a main advantage of RLSC.

4.3 Simulation with a different XGM2016 accuracy

To analyze the behavior of RLSC in dependence of the accuracy of the GGM, we repeat the computations with the original XGM2016 accuracy, i.e., without multiplying it by a factor of 3. Therefore, we recalculate case B with the original XGM2016 covariance matrices $\mathbf{C}^{\mathrm{m}}_{\hat{\mathrm{l}}\hat{\mathrm{l}}}$, $\mathbf{C}^{\mathrm{m}}_{\hat{\mathrm{s}}\hat{\mathrm{l}}}$ and $\mathbf{C}^{\mathrm{m}}_{\hat{\mathrm{s}}\hat{\mathrm{s}}}$ and name it case C. Note, that this will generally improve the results since the downscaling of the covariance matrices also decreases the noise $\mathbf{l}^{\mathrm{n}}$ and $\mathbf{s}^{\mathrm{n}}$ of the GGM which is used in the RCR step.

Figure 8 shows again the resulting absolute geoid height differences between the results of RLSC and standard LSC to the pre-defined truth $\mathbf{s}$. At first glance the two images in Fig. 8 look similar. As in case B, in general we have small errors in the Pacific Ocean, medium errors in the land areas with observations, and the largest errors in region R2. But nevertheless, we see that the result of RLSC (Fig. 8a) is better than standard LSC (Fig. 8b). The RMS from region R1 amounts to 2.9 cm (Fig. 8a) and 3.3 cm (Fig. 8b), respectively. Especially, the dark green area in the Pacific which indicates residuals below 1 cm is more uniform in Fig. 8a. Also, for RLSC the bulge in the north shows mainly residuals below 1 cm, while for standard LSC it shows much higher residuals.

Table 3 summarizes the RMS and MSD values of the three synthetic test cases in region R1, from which the following conclusions can be drawn. Generally, in all three test cases RLSC performs better than standard LSC. However,

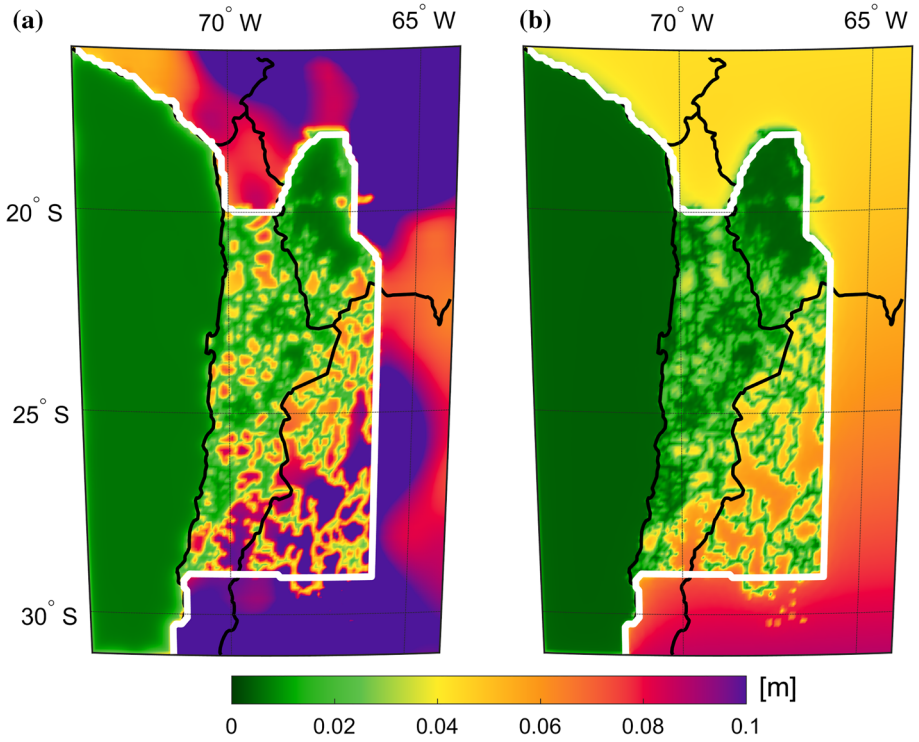

Fig. 7 Case B: Formal error of the geoid height calculation. This derives **a** from $\mathbf{E}_{ss}$ for RLSC, **b** from $\bar{\mathbf{E}}_{\bar{s}\bar{s}}$ for standard LSC, as the square root of the main diagonal

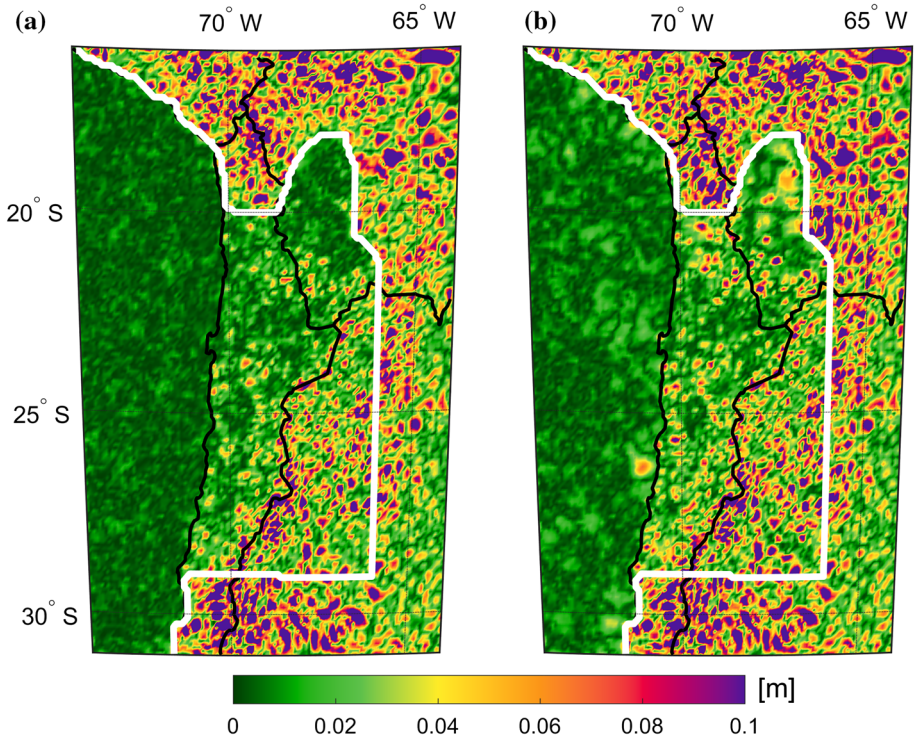

Fig. 8 Case C: Geoid height residuals with the original XGM2016 accuracy. Presented is the absolute difference between the pre-defined truth $\mathbf{s}$ and **a** the RLSC result $\mathbf{s}$, **b** the standard LSC result $\bar{\mathbf{s}}$

the benefit in case A is negligibly small. With the given point distribution in the target area we see a benefit from including a high-resolution GGM in RLSC which decreases the RMS from 6.0 cm (case A) to 3.9 cm (case B). This benefit is particularly apparent in areas with sparse or medium-dense data distributions and at the edges of the terrestrial observations. The inclusion of a high-resolution GGM without using RLSC yields with 6.9 cm the worst result among the three cases. Figure 6b demonstrates this even more clearly than the RMS comparison. As expected, the inclusion of a higher quality GGM (case C) gives better results in both of the LSC methods, but still favors RLSC. In contrast to standard LSC,

in all three cases the accuracy of RLSC is well approximated by the formal error estimates. In case B and C the MSD of the formal error differs by only 0.3 cm from the RMS of the residuals (Table 3).

5 Conclusion and outlook

In this paper, we derive and evaluate an approach named residual least-squares collocation that includes several adaptations to previous LSC methods. Principally, we use only residuals as input to RLSC and the stochastic properties of all inputs in the RCR step are separately described by a corresponding covariance matrix. As a result, this adapted formulation only uses error covariance matrices instead of covariance matrices that describe the gravity signal as used in Moritz (1980), Pail et al. (2010), Sansò (2013) or Haagmans and van Gelderen (1991).

We derive the formulation of RLSC from the basics of Moritz (1980) but use the mere stochastic expectation operator E instead of the total average operator $\bar{\mathrm{E}}$ for the definition of the covariance matrices. On this basis, for the first time we include a full covariance matrix of a high-resolution GGM in regional geoid modeling. The method also allows us to use covariance matrices that are tailored to a target area. It was already formulated by Tscherning (1999) that the use of isotropic covariance functions does not yield the optimal result in estimating the non-isotropic functionals of the Earth's disturbing gravity field T. Therefore, we derive covariance functions directly from the normal equation system of the high-resolution GGM (here: XGM2016) and use them for the collocation. In general, these location-dependent and non-isotropic covariance matrices fit better to a target area than covariance matrices derived purely from signal degree variances (Moritz 1980; Heiskanen and Moritz 1967) or empirical covariance fitting (Tscherning and Rapp 1974), which both describe only an average part of the Earth (or respectively an average part of a target area). Furthermore, similar to the LSC methods in Pail et al. (2010) or Sansò (2013) RLSC offers the advantage that all stochastic effects are included in a consistent way directly in LSC, which should lead to a realistic accuracy estimation of the output quantities. The downside of the presented approach is that the computation of the covariance matrices for the input and the output points ($\mathbf{C}^{\mathrm{m}}_{\hat{l}\hat{l}}$, $\mathbf{C}^{\mathrm{m}}_{\hat{s}\hat{l}}$ and $\mathbf{C}^{\mathrm{m}}_{\hat{s}\hat{s}}$) is very CPU-intensive. For the calculation we exploited the LRZ supercomputing system SuperMUC phase 2 where we used 80 Haswell nodes (Xeon E5-2697 v3). These nodes have 28 cores each with a peak performance of 41.6 GFlops/s, which results in a computation time less than an hour. Additionally, it is necessary to have access to the full normal equation system of the GGM which is frequently not available to the public.

For an optimum evaluation of the benefit of the RLSC method we formulate a synthetic test case with a pre-defined truth. To obtain realistic results from it we add noise to the GGM and to the terrestrial gravity observations. However, a number of assumptions must be included in the scenario. In particular, we assume that the variance–covariance information of the XGM2016 model is correct, since we use it to derive the noise of the GGM. Furthermore, we band-limit the gravity signal to degree 2190. In a case where we use real gravity data, we would have to extend the topographic gravity reduction to higher frequencies. Nevertheless, the simulation of three different test cases yields meaningful results and allows us to evaluate RLSC compared to a standard LSC. We see that regional gravity modeling can benefit from the inclusion of a high-resolution GGM with RLSC. In our test case, the numerical advantage of this is an RMS reduction from 6.0 cm (case A) to 3.9 cm (case B) for RLSC where the RMS values are largely affected by the areas with sparse ground data distribution. In areas with a good data coverage RLSC gives residuals mainly below 1 cm in all of the three test cases. In general, by including a high-resolution GGM in RLSC, we obtain better results especially in areas with sparse observations or close to data gaps. In our opinion, this constitutes a big advantage, because terrestrial gravity measurements are often inhomogeneously distributed and data gaps are almost inevitable especially in mountainous regions. Furthermore, it is shown that the stochastic error estimates of the RLSC approach provide realistic uncertainty estimates, which becomes very important when real gravity data is used.

An additional advantage of the inclusion of a high-resolution GGM in LSC and thus working with rather high-frequency residuals is, that it can significantly reduce the amount of terrestrial data that is needed for the calculation. As shown in Sect. 3, the correlation length when using XGM2016 and a topographic gravity model in the remove step is only 0.12°. Therefore, it is possible to reduce the amount of data that is included to RLSC locally and for example use only those observations for the calculation that are within a specified distance from the output. Considering this, we regard the calculation of height reference stations for the International Height Reference System (IHRS; more information in Ihde et al. 2017 or Sánchez and Sideris 2017) as a predestined application for RLSC. The high-resolution GGM could be used as the basis of a worldwide height unification, and terrestrial observations would be used for the local refinement around a height reference station. The unique advantage of RLSC in this context is that terrestrial data coverage would not be needed in an area as large as about 200 km around an IHRF station, which is the current default for a satellite-only GGM. Furthermore, IHRS stations will be distributed globally, so that numerous IHRF stations can either benefit from the in general good quality of high-resolution GGMs over oceans that results from satellite altimetry, or are set to areas

with little terrestrial data coverage where RLSC appears to perform best. Finally, a correct stochastic accuracy estimation is very important for the calculation of potential values at the IHRF stations.

The next step will be to validate the performance of the presented approach within the IAG Joint Working Group (JWG, 2.2.2: the 1 cm geoid experiment) with real measurements. JWG 2.2.2 aims to assess the calculation of gravity potential values at IHRS stations from different calculation methods, and therefore enables another comparison of RLSC to other regional gravity field determination methods. In any case, with continuously improving high-resolution GGMs and especially with the announcement of the Earth Gravitational Model 2020 (EGM2020) by the NGA, we see the impact of high-resolution GGMs for regional geoid modeling even increasing in the future.

Acknowledgements A large part of the investigations presented in this paper was conducted in the framework of the project 'Optimally combined regional geoid models for the realization of height systems in developing countries' funded by the German Research Foundation (DFG Project No. PA 1543/14-1). We also acknowledge the provision of computer resources by the Leibniz Supercomputing Centre (LRZ; Address: Boltzmannstrasse 1, 85748 Garching, Germany).

References

Andersen O, Knudsen P, Stenseng L (2015) The DTU13 MSS (mean sea surface) and MDT (mean dynamic topography) from 20 years of satellite altimetry. In: Jin S, Barzaghi R (eds) IGFS 2014. International association of geodesy symposia, vol 144. Springer, Cham, pp 111–120. https://doi.org/10.1007/1345_2015_182

Arabelos DN, Tscherning CC (2009) Error-covariances of the estimates of spherical harmonic coefficients computed by LSC, using second-order radial derivative functionals associated with realistic GOCE orbits. J Geod 83(5):419–430. https://doi.org/10.1007/s00190-008-0250-9

Drinkwater MR, Floberghagen R, Haagmans R, Muzi D, Popescu A (2003) GOCE: ESA's first earth explorer core mission. In: Beutler GB et al (eds) Earth gravity field from space—from sensors to earth sciences, space sciences series of ISSI, vol 17, pp 419–432. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-1333-7_36

Fecher T, Pail R, Gruber T (2015) Global gravity field modeling based on GOCE and complementary gravity data. Int J Appl Earth Obs Geoinf 35A:120–127. https://doi.org/10.1016/j.jag.2013.10.005 (ISSN 0303-2434)

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(3):571–590. https://doi.org/10.1007/s10712-016-9406-y

Forsberg R (1984) A study of terrain reductions, density anomalies and geophysical inversion methods in gravity field modelling. Reports of the Department of Geodetic Science and Surveying, No. 355, Ohio State University, Columbus

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

Förste C, Bruinsma SL, Abrikosov O, Lemoine JM, Marty JC, Flechtner F, Balmino G, Barthelmes F, Biancale R (2014): EIGEN-6C4 the latest combined global gravity field model including GOCE data up to degree and order 2190 of GFZ Potsdam and GRGS Toulouse. GFZ Data Services. https://doi.org/10.5880/icgem.2015.1

Gerlach C, Fecher T (2012) Approximations of the GOCE error variance–covariance matrix for least-squares estimation of height datum offsets. J Geod Sci 2(4):247–256. https://doi.org/10.2478/v10156-011-0049-0

Grombein T, Seitz K, Heck B (2016) The rock–water–ice topographic gravity field model RWI_TOPO_2015 and its comparison to a conventional rock-equivalent version. Surv Geophys 37(5):937–976. https://doi.org/10.1007/s10712-016-9376-0

Gruber T, Willberg M (2019) Signal and error assessment of GOCE-based high resolution gravity field models. In: International symposium gravity, geoid and height systems 2, J Geod Sci (**accepted**)

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. https://doi.org/10.1007/s00190-011-0486-7

Haagmans RHN, van Gelderen M (1991) Error variances–covariances of GEM-TI: their characteristics and implications in geoid computation. J Geophys Res 96(B12):20011–20022. https://doi.org/10.1029/91JB01971

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

Hirt C, Claessens SJ, Fecher T, Kuhn M, Pail R, Rexer M (2013) New ultra-high resolution picture of Earth's gravity field. Geophys Res Lett 40(16):4279–4283. https://doi.org/10.1002/grl.50838

Hofmann-Wellenhof B, Moritz H (2006) Physical geodesy. Springer, Wien ISBN 10 2-211-33544-7

Hosse M, Pail R, Horwath M, Holzrichter N, Gutknecht BD (2014) Combined regional gravity model of the andean convergent subduction zone and its application to crustal density modelling in active plate margins. Surv Geophys 35(6):1393–1415. https://doi.org/10.1007/s10712-014-9307-x

Ihde J, Sánchez L, Barzaghi R, Drewes H, Förste 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

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. https://doi.org/10.1007/3-540-33767-9

Mayer-Gürr T, and the GOCO consortium (2015) The new combined satellite only model GOCO05s. EGU General Assembly, Vienna. https://doi.org/10.13140/RG.2.1.4688.6807

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

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. https://doi.org/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. 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 117:B04406. https://doi.org/10.1029/2011JB008916

Rexer M (2017) Spectral solutions to the topographic potential in the context of high-resolution global gravity field modelling. Dissertation, Technical University of Munich

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(6):1035–1074. https://doi.org/10.1007/s10712-016-9382-2

Rieser D (2015) GOCE gravity gradients for geoid and Moho determination applying the Least Squares Collocation approach. Dissertation, Graz University of Technology

Sánchez L, Sideris MG (2017) Vertical datum unification for the international height reference system (IHRS). Geophys J Int 209(2):570–586. https://doi.org/10.1093/gji/ggx025

Sandwell DT, Smith WHF (2009) Global marine gravity from retracked Geosat and ERS-1 altimetry: ridge segmentation versus spreading rate. J Geophys Res 114:B01411. https://doi.org/10.1029/2008JB006008

Sansò F (1986) Statistical methods in physics geodesy. In: Suenkel H (ed) Mathematical and numerical techniques in physical geodesy. Lecture notes in earth sciences, vol 7. Springer, Berlin, pp 49–155. https://doi.org/10.1007/BFb0010132

Sansò F (2013) The local modelling of the gravity field by collocation. In: Sansò F, Sideris MG (eds) Geoid determination: theory and methods. Springer, Heidelberg. https://doi.org/10.1007/978-3-540-74700-0

Tapley BD, Bettadpur S, Watkins M, Reigber C (2004) The gravity recovery and climate experiment: mission overview and early results. Geophys Res Lett 31(9):L09607. https://doi.org/10.1029/2004GL019920

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 (2015) Least-squares collocation. In: Grafarend E (ed) Encyclopedia of geodesy. Springer, Cham. https://doi.org/10.1007/978-3-319-02370-0_51-1

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