J Geod (2012) 86:393–408
DOI 10.1007/s00190-011-0526-3

ORIGINAL ARTICLE

On the combination of high-resolution and satellite-only global gravity models

Mirko Reguzzoni · Fernando Sansò

Received: 9 September 2010 / Accepted: 2 November 2011 / Published online: 24 December 2011

Abstract The issue of combining high-resolution gravity models, based on observations taken on the Earth surface, with those derived from satellite-only observations is of increasing importance, due to the new data provided by gravity satellite missions, CHAMP, GRACE and GOCE. The paper addresses this issue with a twofold purpose. On the one hand, it is an attempt to discuss and assess general concepts, well known in literature, such as achievable resolution, regularization in the least-squares sense or in an infinite dimensional setup, combination criteria, symmetry and block diagonal structures. In particular, as for the symmetry question, a well-defined result, generalizing known facts, is derived. On the other hand, the outcomes of the general discussion are specifically applied to the combination of a high-resolution model (e.g. EGM08) with a GOCE gravity model estimated by the so-called space-wise approach. Small numerical examples are developed to clarify the property of the proposed solution.

Keywords Data combination · GOCE mission · Collocation · Space-wise approach · Block diagonal structures

M. Reguzzoni (✉)
Politecnico di Milano, DIIAR, P.za Leonardo da Vinci 32, 20133 Milano, Italy
e-mail: mirko@geomatica.como.polimi.it

F. Sansò
Politecnico di Milano, Polo Territoriale di Como, DIIAR, Via Valleggio 11, 22100 Como, Italy
e-mail: fernando.sanso@polimi.it

1 Introduction and definition of the problem

One of the main achievements of the past few years in the process of improving the knowledge of the gravity field of the Earth is constituted by the establishment of the global gravity model, EGM08, see Pavlis et al. (2008), complete up to the maximum degree 2,159 with some coefficients estimated up to degree 2,190, roughly implying a ground resolution of $5' \times 5'$, and an overall accuracy expressed by a commission error of about 9 cm.

Yet, the approach used to combine the huge and inhomogeneous "ground" material (primarily gravity, digital terrain models and mean radar altimetry) goes through a number of approximation steps culminating in the final estimation by the so-called block diagonal least-squares techniques (Pavlis et al. 2008). The success of this global model could not be so important without the joint use of the information derived from satellite observations, namely the spherical harmonic coefficients of the GRACE model ITG-GRACE03S (Mayer-Gürr 2006). In the near future, the problem of producing a new global model that will incorporate the information derived from the GOCE mission (ESA 1999) in the form of a GOCE-only model (Pail et al. 2011) will have to be solved.

The problem is indeed not new and it has been discussed many times in literature (see for instance the pioneering work by Kaula 1966a) as well as already applied in one form or the other in the computation of different families of global models (for instance the EGM series: see Lemoine et al. 1998, the EIGEN series: see Reigber et al. 2005, the OSU series: see Rapp 1984, or the TEG series: see Tapley et al. 1997, Ch. 10, just to mention a few).

The approach applied to produce such a combined model has been basically a least-squares approach, leading to a direct sum of normal matrices, usually requiring the solution of the delicate problem of determining their relative weighting,

due to the imperfect knowledge of the error covariance structure of the estimated coefficients.

On the other hand, there are approaches to the production of satellite-only global models that do not follow a least-squares criterion and therefore for them there is no normal matrix available. For instance, for the GOCE mission one of the ways in which a global model is produced is the so-called space-wise approach (SpW) in which no normal equations are implemented and so no normal matrix is available (Migliaccio et al. 2004; Reguzzoni and Tselfes 2009). Nevertheless, an error covariance matrix of the spherical harmonic coefficients is computed for the SpW solution. This in fact, following a line of thought introduced in geodesy by Gundlich et al. (2003) to work out the covariance of the EGM96 model (Lemoine et al. 1998), can be obtained via Monte Carlo methods (Alkhatib and Schuh 2007; Migliaccio et al. 2009).

This paper tries to propose a solution to the problem of combining two global models under the above conditions. In order to properly define such a solution we found it important to answer a number of questions which we list below:

- before merging two existing solutions, should we resort to a pure finite dimensional deterministic theory, namely a least-squares method after reduction of the solution space to a finite number of parameters, or should we use infinite dimensional methods like in collocation? Moreover, to what extent are the solutions different from one another? See also Colombo (1981); Moritz (1989); Bosch (1993); Krarup (2006);
- is there a point where the two approaches, least-squares and collocation, diverge significantly, either algorithmically or numerically? See also Rapp (1975);
- in order to implement the merging, in particular for the case at hand, is it necessary to resort to the observations or can we just use the estimated coefficients and their covariance matrices? See also Koch (1999);
- since the reduction of the numerical effort to the solution of a block diagonal (or almost block diagonal) "normal" system is a key issue for its feasibility, what is the hypothesis that we implicitly impose on the data by using such a technique? In particular, is a block diagonal structure of the normal system of collocation theory necessarily related to the (certainly unrealistic) hypothesis of isotropy of the underlying field? See also Colombo (1981); Sansò and Tscherning (2003); Boxhammer and Schuh (2006);
- what is the best way of applying a block diagonal approximation? In particular, should we average out the factors that produce non-zero entries out of the block diagonal pattern or should we only compute a full normal matrix and then disregard the non-block diagonal terms?
- what is the effect of using a block diagonal approximation to the combination of models like EGM08 and GOCE-only? In particular, is this approximation restricting the influence of the newly achieved GOCE model on the combined solution coefficients to the maximum degree and order of the satellite-only model?

To all these questions we try to give an answer. In particular the first four questions are already well discussed in geodetic literature, but we recall them trying to provide some improvement and always comparing the least-squares with the collocation approach. To the last two questions we try to answer by means of small numerical examples, attempting to capture the essential features of the numerically much heavier real case; similar studies on the impact of the block diagonal approximation can be found for instance in Gruber (2001).

Closing this section, we introduce some elements of the notation that we shall employ, in order to keep formulas as concise as possible. In particular, we shall assume that the gravity anomalous potential $T(\boldsymbol{x})$ on a reference sphere, with radius $r = R$, can be represented by a possibly infinite dimensional series of spherical harmonics

$$T(\boldsymbol{x}) = \sum_{n=2}^{+\infty} \sum_{m=-n}^{n} T_{nm} S_{nm}(\boldsymbol{x}), \tag{1}$$

where $\boldsymbol{x} = (R, \vartheta, \lambda)$,

$$S_{nm}(\boldsymbol{x}) = \overline{P}_{n,|m|}(\vartheta) \begin{cases} \cos m\lambda & m \geq 0 \\ \sin |m|\lambda & m < 0 \end{cases} \tag{2}$$

and $\overline{P}_{n,|m|}$ are the fully normalized Legendre functions (Heiskanen and Moritz 1967).

To shorten the formula (1) one can introduce the vector of the harmonic coefficients $\boldsymbol{T} \equiv \{T_{nm}\}$, originally ordered degree by degree, i.e.

$$\boldsymbol{T}^{\mathrm{T}} \equiv [\ldots; T_{n,-n}, \ldots T_{n,0} \ldots T_{n,n}; \ldots] \tag{3}$$
$$n = 2, 3, \ldots$$

The upper limit of n in (3), depending on the context, can be either ∞ or just a very high number like $\overline{N} = 2{,}159$, i.e. the maximum degree of the high-resolution prior model.

Correspondingly, and keeping the same ordering scheme, one can define the vector $\boldsymbol{S}(\boldsymbol{x}) = \{S_{nm}(\boldsymbol{x})\}$, so that (1) simply writes

$$T(\boldsymbol{x}) = \boldsymbol{T}^{\mathrm{T}} \boldsymbol{S}(\boldsymbol{x}). \tag{4}$$

When useful the vector $\boldsymbol{T}$ will be split into two parts. One, $\boldsymbol{T}_{(N)}$, is the same as (3) but with n running from 2 to N, while the other $\boldsymbol{T}^{(N)}$ is the same as (3) but with n running from $N+1$ up to the upper limit, $\overline{N}$ or $+\infty$. So we can always put

$$\boldsymbol{T} = \left| \begin{array}{c} \boldsymbol{T}_{(N)} \\ \boldsymbol{T}^{(N)} \end{array} \right|. \tag{5}$$

The same splitting can be applied to the vector $\boldsymbol{S}(\boldsymbol{x})$ or even to any matrix $\boldsymbol{A}$ multiplying $\boldsymbol{T}$, to the extent that we can write

$$\boldsymbol{A} = [\boldsymbol{A}_{(N)}, \boldsymbol{A}^{(N)}] \tag{6}$$

and

$$\boldsymbol{AT} = \boldsymbol{A}_{(N)}\boldsymbol{T}_{(N)} + \boldsymbol{A}^{(N)}\boldsymbol{T}^{(N)}. \tag{7}$$

Since the case will be often present that N is just the maximum order of the new satellite-only global model (e.g. $N = 250$), we shall adopt the notation $\boldsymbol{T}_{\mathrm{S}}$ for the corresponding vector of coefficients. When on the contrary the maximum degree is $\overline{N}$, the degree of the prior global model, we shall call the vector $\boldsymbol{T}_{\mathrm{G}}$.

In this particular case we shall write

$$\boldsymbol{T}_{\mathrm{G}} = \left| \begin{matrix} \boldsymbol{T}_{\mathrm{S}} \\ \boldsymbol{T}_{\mathrm{r}} \end{matrix} \right| \tag{8}$$

to indicate that $\boldsymbol{T}_{\mathrm{r}}$ contains the remnant of $\boldsymbol{T}_{\mathrm{S}}$, e.g. all the degrees between 251 and 2,159.

A similar notation will then be adopted for vectors or matrices that multiply $\boldsymbol{T}_{\mathrm{G}}$. Other reorderings of $\boldsymbol{T}$ will be needed in the text, e.g. grouping coefficients with the same order, but they will be clarified from time to time.

2 Finite versus infinite dimensional solutions

It has been mentioned in the previous section that there are alternative approaches to the estimation of the coefficients of a global model; either the reduction to finite dimensional spaces followed by a least-squares application, or the use of collocation concepts, which work directly into an infinite dimensional Hilbert space (Sansò 1986).

The purpose of this section is to recall the relation between the two solutions and to analyse under what conditions they are equivalent and in particular under what conditions a truncated solution $\widehat{\boldsymbol{T}}_{(N)}$ can be considered to provide a good approximation to the sought global model.

There are two different interpretations of the collocation approach to treat the linear problem

$$\boldsymbol{Y} = \boldsymbol{AT} + \boldsymbol{v}, \tag{9}$$

with $\boldsymbol{T} \in \mathcal{R}^{\infty}$. In one of them, $\boldsymbol{T}$ is treated as an $\mathcal{R}^{\infty}$ random vector with known covariance matrix $\boldsymbol{C}_T$ that for the sake of simplicity we shall assume to be diagonal. In this case $\widehat{\boldsymbol{T}}$ is determined as the linear predictor that minimizes the mean square prediction error. In the other approach, the predictor $\widehat{T}(\boldsymbol{x}) = \widehat{\boldsymbol{T}}^{\mathrm{T}}\boldsymbol{S}(\boldsymbol{x})$ is restricted to some Hilbert space H_K, with reproducing kernel $K(\boldsymbol{x}, \boldsymbol{y})$ and determined on the basis of the hybrid norm or regularized least-squares collocation principle

$$\underset{\boldsymbol{T}}{\mathrm{Min}} \left\{ (\boldsymbol{Y} - \boldsymbol{AT})^{\mathrm{T}}\boldsymbol{C}_v^{-1}(\boldsymbol{Y} - \boldsymbol{AT}) + \| \boldsymbol{T}^{\mathrm{T}}\boldsymbol{S}(\boldsymbol{x}) \|^2_{H_K} \right\}. \tag{10}$$

When $K(\boldsymbol{x}, \boldsymbol{y})$ is isotropic, i.e.

$$K(\boldsymbol{x}, \boldsymbol{y}) = \sum_n k_n S_{nm}(\boldsymbol{x}) S_{nm}(\boldsymbol{y}), \tag{11}$$

the principle (10) can be re-formulated as

$$\underset{\boldsymbol{T}}{\mathrm{Min}} \left\{ (\boldsymbol{Y} - \boldsymbol{AT})^{\mathrm{T}}\boldsymbol{C}_v^{-1}(\boldsymbol{Y} - \boldsymbol{AT}) + \boldsymbol{T}^{\mathrm{T}}\boldsymbol{K}^{-1}\boldsymbol{T} \right\} \tag{12}$$

where $\boldsymbol{K}^{-1}$ is an infinite dimensional diagonal matrix with diagonal elements k_n^{-1}. It is known since long that the two approaches described above are equivalent on condition that the two matrices $\boldsymbol{C}_T$ and $\boldsymbol{K}$ are equal, i.e. when we interpret T_{nm} as uncorrelated random variables with zero mean and variances equal to k_n (Sansò 1986; Sansò et al. 2000). So we shall not dwell on this proof of equivalence, but we shall investigate the relation between this solution and a least-squares solution when the full available information is exploited.

In fact if we add to the observation equations

$$\boldsymbol{Y} = \boldsymbol{A}_{(N)}\boldsymbol{T}_{(N)} + \boldsymbol{v} \tag{13}$$

the further pseudo-observation equations

$$\boldsymbol{T}_{(N)} = \boldsymbol{\eta}, \tag{14}$$

with the assumption that the average $\mathrm{E}\{\boldsymbol{\eta}\} = 0$ and given the covariances $\boldsymbol{C}_v$, $\boldsymbol{C}_\eta$ [diagonal, with variances $\sigma_n^2(T)$] and $\boldsymbol{C}_{v\eta} = 0$, we can write the full least-squares principle as

$$\mathrm{Min} \left\{ (\boldsymbol{Y} - \boldsymbol{A}_{(N)}\boldsymbol{T}_{(N)})^{\mathrm{T}}\boldsymbol{C}_v^{-1}(\boldsymbol{Y} - \boldsymbol{A}_{(N)}\boldsymbol{T}_{(N)}) + \boldsymbol{T}_{(N)}^{\mathrm{T}}\boldsymbol{C}_\eta^{-1}\mathbf{T}_{(N)} \right\} \tag{15}$$

leading to the normal system

$$(\boldsymbol{D}_{(N)} + \boldsymbol{C}_\eta^{-1})\widehat{\boldsymbol{T}}_{(N)} = \boldsymbol{A}_{(N)}^{\mathrm{T}}\boldsymbol{C}_v^{-1}\boldsymbol{Y} \tag{16}$$

$$(\boldsymbol{D}_{(N)} = \boldsymbol{A}_{(N)}^{\mathrm{T}}\boldsymbol{C}_v^{-1}\boldsymbol{A}_{(N)}). \tag{17}$$

Note that (15) is nothing but a Tykhonov regularized principle (Phillips 1962; Tikhonov 1963) with $\boldsymbol{C}_\eta^{-1}$ as a regularizer; see also Kirsch (1996), Sect. 2.2.

If we compare (12) and (15) we see that the two principles are formally identical, with the only difference that in (12) $\widehat{\boldsymbol{T}}$ is infinite dimensional, while in (16) the unknown $\widehat{\boldsymbol{T}}_{(N)}$ is finite dimensional, in fact with dimension $(N+1)^2 - 4$. As it is obvious, we can state an equivalence between the two only on condition that the approximation holds

$$\boldsymbol{A}\widehat{\boldsymbol{T}} = \boldsymbol{A}_{(N)}\widehat{\boldsymbol{T}}_{(N)} + \boldsymbol{A}^{(N)}\widehat{\boldsymbol{T}}^{(N)} \cong \boldsymbol{A}_{(N)}\widehat{\boldsymbol{T}}_{(N)}, \tag{18}$$

i.e. that the addendum $\boldsymbol{A}^{(N)}\widehat{\boldsymbol{T}}^{(N)}$ can be neglected in (18) with respect to the noise term $\boldsymbol{v}$. This requirement could be understood either roughly in terms of orders of magnitude, or made more precise, as shortly developed in the Appendix.

In any way the relation (18) becomes strictly true if $\boldsymbol{A}$ has from the beginning the form

$$\boldsymbol{A} = |\boldsymbol{A}_{(N)}\boldsymbol{0}|. \tag{19}$$

In this case in fact one writes (12) in the form

$$\begin{aligned}\text{Min}\Big\{&(\boldsymbol{Y}-\boldsymbol{A}_{(N)}\boldsymbol{T}_{(N)})^{\mathrm{T}}\boldsymbol{C}_{\nu}^{-1}(\boldsymbol{Y}-\boldsymbol{A}_{(N)}\boldsymbol{T}_{(N)})\\&+\boldsymbol{T}_{(N)}^{\mathrm{T}}\boldsymbol{K}_{(N)}^{-1}\boldsymbol{T}_{(N)}+\boldsymbol{T}^{(N)\,\mathrm{T}}(\boldsymbol{K}^{(N)})^{-1}\boldsymbol{T}^{(N)}\Big\};\end{aligned} \tag{20}$$

since $\boldsymbol{T}^{(N)}$ appears in (20) only in the last quadratic term, we see that the minimum condition imposes

$$\widehat{\boldsymbol{T}}^{(N)} \equiv \boldsymbol{0} \tag{21}$$

while $\widehat{\boldsymbol{T}}_{(N)}$ is given by the same normal equation (16) on condition that

$$\boldsymbol{C}_{\eta} = \boldsymbol{K}_{(N)} \tag{22}$$

or

$$\sigma_n^2(T) = k_n. \tag{23}$$

One interesting observation here is that the equivalence between the least-squares collocation principle and the simple regularized least-squares lies on the fact that $\boldsymbol{A}^{(N)}\widehat{\boldsymbol{T}}^{(N)}$ is negligible with respect to the noise $\boldsymbol{\nu}$, which traces back to the sensitivity criterion more than to the commission/omission error criterion, as discussed in the Appendix too.

Another interesting remark is that if one does not want to use the sharp truncation hypothesis (19), one is led to the infinite dimensional solution

$$\widehat{\boldsymbol{T}} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}_{\nu}^{-1}\boldsymbol{A}+\boldsymbol{K}^{-1})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}_{\nu}^{-1}\boldsymbol{Y} \tag{24}$$

which might seem not easily computable. This, however, is not the case on account of the identity (see for instance Colombo 1981, p. 50)

$$(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}_{\nu}^{-1}\boldsymbol{A}+\boldsymbol{K}^{-1})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}_{\nu}^{-1} \equiv \boldsymbol{K}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{C}_{\nu}+\boldsymbol{A}\boldsymbol{K}\boldsymbol{A}^{\mathrm{T}})^{-1}, \tag{25}$$

showing that only finite dimensional inversions are required. In addition the form (25) of the predictor is exactly the same as that of the standard collocation theory.

The conclusions of this section are twofold. The first is that finite or infinite dimensional solutions are admissible approximations of one another, if observation equations are suitably truncated, disregarding terms that can be considered irrelevant, in the sense worked out in the Appendix. In practice, this means that considering $\boldsymbol{T}$ as infinite dimensional or finite dimensional, with any large dimensions, will be for us one and the same thing, at least for the problem at hand.

The second conclusion, issuing from the first, is that basically using a least-squares collocation approach or just a least-squares approach augmented with pseudo-observation equations (14) leads to the same solution. So, no particular choice has to be made up to here, between the two.

3 Least-squares combination, sufficient statistics and innovation solutions

In this section we address the question of combining different data sets, that, to come closer to our real problem, we will tag with indexes G or S thinking of them as data pertaining to "global" prior observations or to new "satellite" observations that we want to merge with the former. Let us remark in any event that although the matters discussed in the paper are oriented to the solution of a specific problem, justifying the use of the indexes G and S, yet the same reasonings do apply to a more general situation in which global models with different spectral resolutions have to be combined. So the idea is that $\boldsymbol{Y}_{\mathrm{G}}$ is a global data set with model

$$\boldsymbol{Y}_{\mathrm{G}} = \boldsymbol{A}_{\mathrm{G}}\boldsymbol{T}_{\mathrm{G}} + \boldsymbol{\nu}_{\mathrm{G}}, \quad \mathrm{E}\{\boldsymbol{\nu}_{\mathrm{G}}\boldsymbol{\nu}_{\mathrm{G}}^{\mathrm{T}}\} = \boldsymbol{C}_{\mathrm{G}} \tag{26}$$

leading to a global model, e.g. EGM08,

$$\begin{aligned}&\widehat{\boldsymbol{T}}_{\mathrm{G}} = \boldsymbol{D}_{\mathrm{G}}^{-1}\boldsymbol{A}_{\mathrm{G}}^{\mathrm{T}}\boldsymbol{C}_{\mathrm{G}}^{-1}\boldsymbol{Y}_{\mathrm{G}}\\&(\boldsymbol{D}_{\mathrm{G}} = \boldsymbol{A}_{\mathrm{G}}^{\mathrm{T}}\boldsymbol{C}_{\mathrm{G}}^{-1}\boldsymbol{A}_{\mathrm{G}})\end{aligned} \tag{27}$$

up to some high degree, e.g. 2,159, while $\boldsymbol{Y}_{\mathrm{S}}$ is a "new" satellite data set with model

$$\boldsymbol{Y}_{\mathrm{S}} = \boldsymbol{A}_{\mathrm{S}}\boldsymbol{T}_{\mathrm{S}} + \boldsymbol{\nu}_{\mathrm{S}}, \quad \mathrm{E}\{\boldsymbol{\nu}_{\mathrm{S}}\boldsymbol{\nu}_{\mathrm{S}}^{\mathrm{T}}\} = \boldsymbol{C}_{\mathrm{S}} \tag{28}$$

and with solution

$$\begin{aligned}&\widehat{\boldsymbol{T}}_{\mathrm{S}} = \boldsymbol{D}_{\mathrm{S}}^{-1}\boldsymbol{A}_{\mathrm{S}}^{\mathrm{T}}\boldsymbol{C}_{\mathrm{S}}^{-1}\boldsymbol{Y}_{\mathrm{S}}\\&(\boldsymbol{D}_{\mathrm{S}} = \boldsymbol{A}_{\mathrm{S}}^{\mathrm{T}}\boldsymbol{C}_{\mathrm{S}}^{-1}\boldsymbol{A}_{\mathrm{S}})\end{aligned} \tag{29}$$

to fix the ideas up to some degree N_{S} between 200 and 250. Moreover, we can assume that

$$\mathrm{E}\{\boldsymbol{\nu}_{\mathrm{G}}\boldsymbol{\nu}_{\mathrm{S}}^{\mathrm{T}}\} = \boldsymbol{C}_{\mathrm{GS}} \equiv 0. \tag{30}$$

Our problem is how to combine all the available information to get the best estimate of the coefficients. A first idea would be to form with (26) and (28) a unique least-squares system. However, in order to have the same unknown parameter vector, recalling the notation (8) with $\dim \boldsymbol{T}_{\mathrm{G}} \gg \dim \boldsymbol{T}_{\mathrm{S}}$, we can introduce the projector

$$\boldsymbol{\Pi} = [\boldsymbol{I}_{\mathrm{S}}, \boldsymbol{0}] \tag{31}$$

such that

$$\boldsymbol{T}_{\mathrm{S}} = \boldsymbol{\Pi}\boldsymbol{T} \tag{32}$$

and rewrite (28) as

$$\boldsymbol{Y}_{\mathrm{S}} = \boldsymbol{A}_{\mathrm{S}}\boldsymbol{\Pi}\boldsymbol{T} + \boldsymbol{\nu}_{\mathrm{S}}. \tag{33}$$

Standard least-squares theory now (see Koch 1999), taking (30) into account, gives the normal equations system

$$(\boldsymbol{D}_{\mathrm{G}} + \boldsymbol{\Pi}^{\mathrm{T}}\boldsymbol{D}_{\mathrm{S}}\boldsymbol{\Pi})\widehat{\boldsymbol{T}} = \boldsymbol{A}_{\mathrm{G}}^{\mathrm{T}}\boldsymbol{C}_{\mathrm{G}}^{-1}\boldsymbol{Y}_{\mathrm{G}} + \boldsymbol{\Pi}^{\mathrm{T}}\boldsymbol{A}_{\mathrm{S}}^{\mathrm{T}}\boldsymbol{C}_{\mathrm{S}}^{-1}\boldsymbol{Y}_{\mathrm{S}}. \tag{34}$$

Another idea would be to use the already estimated coefficients, $\widehat{\boldsymbol{T}}_{\mathrm{G}}$, with its error covariance $\boldsymbol{D}_{\mathrm{G}}^{-1}$, and $\widehat{\boldsymbol{T}}_{\mathrm{S}}$, with its

error covariance matrix $\boldsymbol{D}_{\mathrm{S}}^{-1}$, and write as "observation equations"

$$\begin{cases} \widehat{\boldsymbol{T}}_{\mathrm{G}} = \boldsymbol{T} + \boldsymbol{e}_{\mathrm{G}}, & (\boldsymbol{C}_{e_{\mathrm{G}}} = \boldsymbol{D}_{\mathrm{G}}^{-1}) \\ \widehat{\boldsymbol{T}}_{\mathrm{S}} = \boldsymbol{\Pi}\boldsymbol{T} + \boldsymbol{e}_{\mathrm{S}}, & (\boldsymbol{C}_{e_{\mathrm{S}}} = \boldsymbol{D}_{\mathrm{S}}^{-1}) \end{cases} \tag{35}$$

where indeed $\boldsymbol{e}_{\mathrm{G}}$ and $\boldsymbol{e}_{\mathrm{S}}$ are uncorrelated. The application of the least-squares principle to (35) then gives

$$\begin{aligned} (\boldsymbol{D}_{\mathrm{G}} + \boldsymbol{\Pi}^{\mathrm{T}}\boldsymbol{D}_{\mathrm{S}}\boldsymbol{\Pi})\widehat{\boldsymbol{T}} &= \boldsymbol{D}_{\mathrm{G}}\widehat{\boldsymbol{T}}_{\mathrm{G}} + \boldsymbol{\Pi}^{\mathrm{T}}\boldsymbol{D}_{\mathrm{S}}\widehat{\boldsymbol{T}}_{\mathrm{S}} \\ &= \boldsymbol{A}_{\mathrm{G}}^{\mathrm{T}}\boldsymbol{C}_{\mathrm{G}}^{-1}\boldsymbol{Y}_{\mathrm{G}} + \boldsymbol{\Pi}^{\mathrm{T}}\boldsymbol{A}_{\mathrm{S}}^{\mathrm{T}}\boldsymbol{C}_{\mathrm{S}}^{-1}\boldsymbol{Y}_{\mathrm{S}}. \end{aligned} \tag{36}$$

Notice that, by assuming that we know the full covariances of the observations and therefore of $\boldsymbol{e}_{\mathrm{G}}$ and $\boldsymbol{e}_{\mathrm{S}}$ too, see (35), we do not need to introduce relative weights of the normal matrices in (36). This is also the reason why we do not use the notation $\widehat{\boldsymbol{D}}_{\mathrm{G}}^{-1}$ and $\widehat{\boldsymbol{D}}_{\mathrm{S}}^{-1}$ to indicate estimated covariance matrices. This is true at least in principle. In the real situation the relative weighting is often solved in an empirical way (see for instance Koch and Kusche 2002).

As we see (36) is the same as (34), implying that also the two estimates, $\widehat{\boldsymbol{T}}$, are the same. This is not surprising and reflects the well-known fact that $\widehat{\boldsymbol{T}}_{\mathrm{G}}$ and $\widehat{\boldsymbol{T}}_{\mathrm{S}}$ are sufficient statistics for $\boldsymbol{T}$ in the models (26) and (33), respectively (see Cox and Hinkley 1974, Sect. 2.2).

So the problem now is not which of the two ways we want to choose, but rather the numerical solvability of the normal system (36). In this respect it is particularly useful to write the solution of (36) by exploiting the identity

$$\begin{aligned} &(\boldsymbol{D}_{\mathrm{G}} + \boldsymbol{\Pi}^{\mathrm{T}}\boldsymbol{D}_{\mathrm{S}}\boldsymbol{\Pi})^{-1} \\ &\quad = \boldsymbol{D}_{\mathrm{G}}^{-1} - \boldsymbol{D}_{\mathrm{G}}^{-1}\boldsymbol{\Pi}^{\mathrm{T}}(\boldsymbol{D}_{\mathrm{S}}^{-1} + \boldsymbol{\Pi}\boldsymbol{D}_{\mathrm{G}}^{-1}\boldsymbol{\Pi}^{\mathrm{T}})^{-1}\boldsymbol{\Pi}\boldsymbol{D}_{\mathrm{G}}^{-1} \end{aligned} \tag{37}$$

to arrive at the solution in innovation form

$$\widehat{\boldsymbol{T}} - \widehat{\boldsymbol{T}}_{\mathrm{G}} = \boldsymbol{D}_{\mathrm{G}}^{-1}\boldsymbol{\Pi}^{\mathrm{T}}(\boldsymbol{D}_{\mathrm{S}}^{-1} + \boldsymbol{\Pi}\boldsymbol{D}_{\mathrm{G}}^{-1}\boldsymbol{\Pi}^{\mathrm{T}})^{-1}(\widehat{\boldsymbol{T}}_{\mathrm{S}} - \boldsymbol{\Pi}\widehat{\boldsymbol{T}}_{\mathrm{G}}). \tag{38}$$

All that is not new as it is a standard trick applied for instance in Kalman filtering (Kalman and Bucy 1961), but it is the interpretation and the form (38) that makes it useful to our case.

We do not dwell on the problem of computing $\boldsymbol{D}_{\mathrm{G}}^{-1}$, although we observe that this problem has already been solved in some approximate way, because we have already been able to compute $\widehat{\boldsymbol{T}}_{\mathrm{G}}$.

On the other hand, we note that in (38) the known term is just the variation of the estimate $\widehat{\boldsymbol{T}}_{\mathrm{S}}$ with respect to the same coefficients contained in the previous model $\widehat{\boldsymbol{T}}_{\mathrm{G}}$ that we want to update. We note also that, apart from $\boldsymbol{D}_{\mathrm{G}}^{-1}$, the central matrix-inversion in (38) has the same dimension of $\widehat{\boldsymbol{T}}_{\mathrm{S}}$ and not of $\widehat{\boldsymbol{T}}_{\mathrm{G}}$. Moreover, in (38) we never meet $\boldsymbol{D}_{\mathrm{S}}$, but only $\boldsymbol{D}_{\mathrm{S}}^{-1}$. This is of utmost importance because in analysing the satellite data often the operative analysis chain does not lead directly to normal equations. An example is the so-called space-wise approach to gravity mission analysis (Migliaccio et al. 2004; Reguzzoni and Tselfes 2009). However, the solution algorithm allows to give an estimate of the error covariance matrix of $\widehat{\boldsymbol{T}}_{\mathrm{S}}$, namely of $\boldsymbol{D}_{\mathrm{S}}^{-1}$, by means of a Monte Carlo method (Migliaccio et al. 2009). This approximation, though reasonably good for individual coefficients, is not uniform in the sense of a matrix norm (Pertusini et al. 2010). Furthermore, even when one can use more samples than unknowns (40,000 samples for a coefficient vector up to degree 200), still it has generally a very unstable inverse, as it has been proved numerically. Therefore, it could never be used to derive $\boldsymbol{D}_{\mathrm{S}}$ from $\boldsymbol{D}_{\mathrm{S}}^{-1}$.

We note as well that the same reasoning holds for $\boldsymbol{D}_{\mathrm{G}}^{-1}$ too. Therefore also for the "global" observations $\boldsymbol{Y}_{\mathrm{G}}$ and the corresponding global model, $\widehat{\boldsymbol{T}}_{\mathrm{G}}$, we see that we do not need the original normal matrix but only its inverse.

For the above reasons, the solution (38) is feasible, while the solution of (36) might not be.

A final remark is that, in case we believe we have enough information to write equations like (14), i.e. to build the matrix $\boldsymbol{K}_{(N)}$ according to (22) and (23) up to the maximum degree N, the two basic formulas (36) and (38) have to be modified simply by changing $\boldsymbol{D}_{\mathrm{G}}$ into $\boldsymbol{D}_{\mathrm{G}} + \boldsymbol{K}_{(N)}^{-1}$, i.e. by adding $k_n^{-1} = \sigma_n^{-2}(T)$ at the proper places to the diagonal terms of $\boldsymbol{D}_{\mathrm{G}}$. In particular this has no impact on the implementation of the solution (38), whether the shape of the normal matrix therein is block diagonal or not.

Three conclusions can be derived from the discussion of this section. The first is that one can directly use the estimated coefficients from two different data sets "G" and "S" to produce a combined solution, without resorting to the original observations, if we believe that covariances have been correctly propagated. The second is that, by writing the combination solution in the "updating" form (38), the numerical complexity of its computation is of the same order as that of deriving the coefficients $\widehat{\boldsymbol{T}}_{\mathrm{S}}$ from the data $\boldsymbol{Y}_{\mathrm{S}}$ only. The third one is that the computation of formula (38) does not require the knowledge of the normal matrices $\boldsymbol{D}_{\mathrm{S}}$ and $\boldsymbol{D}_{\mathrm{G}}$, but only of the covariance matrices $\boldsymbol{D}_{\mathrm{S}}^{-1}$ and $\boldsymbol{D}_{\mathrm{G}}^{-1}$. These in fact might be available, at least in approximated form, without $\boldsymbol{D}_{\mathrm{S}}$ and $\boldsymbol{D}_{\mathrm{G}}$ being known.

The search for a solution of the combination problem by inverting $(\boldsymbol{D}_{\mathrm{S}}^{-1} + \boldsymbol{\Pi}\boldsymbol{D}_{\mathrm{G}}^{-1}\boldsymbol{\Pi}^{\mathrm{T}})$ is extensively discussed in Sect. 5.

4 Symmetry and block diagonal structures

The idea that normal systems in both least-squares and collocation concepts, when data sets are axi-symmetrical, take the form of a discrete convolution in longitude, λ, is present in the geodetic literature since a long time, see for instance Colombo (1981), Sneeuw (2000), Sansò and Tscherning (2003), just to mention a few. This leads to a block diagonal

structure of normal matrices, where blocks have a maximum dimension equal to $\overline{N}$, decreasing to dimension 1 for the coefficients $T_{\overline{N},\pm\overline{N}}$. This makes at once feasible the solution of normal systems and even the explicit inversion and storage of the normal matrix by inverting the individual blocks. For this reason, it is very important to investigate clearly this item in the context of global models combination.

One remark is basic to understand this section, namely when we say that a data set is axi-symmetric we mean that the points where data are given form a grid along parallels (i.e. circles around the Z axis) at constant $\Delta\lambda$ intervals, though the radius of each parallel and the distance between any two parallels can vary. For instance a regular geographic grid on the ellipsoid is axi-symmetrical. Note also that we could even have different observation functionals from parallel to parallel, though on each parallel the functional has to be the same.

We could split the section in two parts: block diagonal least-squares systems and block diagonal matrices for random fields. Since the first part is well described in literature, we send the reader to the mentioned works, while we will be more analytical on the second case because to our knowledge the result is new in this general form.

Assume that a grid with nodes

$$\boldsymbol{t} \equiv (\vartheta_i, k\Delta),\ (i \text{ parallel index},\ k \text{ meridian index})$$
$$1 \le k \le \overline{M} = \frac{2\pi}{\Delta}$$

is given and at each node we have an observation equation

$$Y_{\boldsymbol{t}} = \sum_{n,m} \gamma_{nm} T_{nm} Y_{nm}(\boldsymbol{x}_{\boldsymbol{t}}) W_{\boldsymbol{t},n} + \nu_{\boldsymbol{t}} \tag{39}$$

where

$$W_{\boldsymbol{t},n} = \left(\frac{R}{r_i}\right)^{n+1}. \tag{40}$$

The form (39) of the observation equations is particular in the sense that we assume that the observational functionals are diagonalised by the solid spherical harmonics, their effect being expressed by the numerical coefficients γ_{nm}. This is the case for most of the useful functionals, at least in spherical approximation.

Assume further $T(\boldsymbol{x}) \equiv \boldsymbol{T}^{\mathrm{T}}\boldsymbol{S}(\boldsymbol{x})$ to be a random field, i.e. $\boldsymbol{T}$ to be a random vector in $\mathcal{R}^{\infty}$. What we shall prove in a moment is that under conditions of axi-symmetry, plus one slightly more restrictive condition, we have, using the notation (39), that the covariance of $T(\boldsymbol{x})$ can be written as (see Lemma 1)

$$\begin{aligned} C(\vartheta, \vartheta', \lambda - \lambda') &= \sum_{m=0}^{+\infty} \sum_{\ell,n=m}^{+\infty} C_{\ell n}^{m} \overline{P}_{\ell m}(\vartheta) \overline{P}_{nm}(\vartheta') W_{\ell\vartheta} W_{n\vartheta'} \\ &\quad \cdot \cos m(\lambda - \lambda') \\ &\equiv \sum_{m=0}^{+\infty} C^{m}(\vartheta, \vartheta') \cos m(\lambda - \lambda') \end{aligned} \tag{41}$$

where

$$W_{\ell\vartheta} = \left(\frac{R}{r_\vartheta}\right)^{\ell+1}, \tag{42}$$

$$C^{m}(\vartheta, \vartheta') = C^{m}(\vartheta', \vartheta), \quad C_{n\ell}^{m} = C_{\ell n}^{m}. \tag{43}$$

Therefore, the normal matrix of collocation, with functionals of the form (39), will read [with $\boldsymbol{t} = (\vartheta_i, k\Delta)$, $\boldsymbol{t}' = (\vartheta_r, s\Delta)$]

$$\begin{aligned} C(L_{\boldsymbol{t}}, L_{\boldsymbol{t}'}) &= \sum_{m=0}^{+\infty} \left(\sum_{\ell,n=m}^{+\infty} \gamma_{\ell m}\gamma_{nm} C_{\ell n}^{m} \overline{P}_{\ell m}(\vartheta_i) \overline{P}_{nm}(\vartheta_r) \right) \\ &\quad \cdot \cos m(k - s)\Delta \\ &\equiv \sum_{m=0}^{+\infty} G_{ir}^{m} \cos m(k - s)\Delta \\ &\equiv \sum_{m=0}^{\overline{M}-1} \widetilde{G}_{ir}^{m} \cos m(k - s)\Delta. \end{aligned} \tag{44}$$

In the last step of (44) aliasing is taken into account. In fact, noting that the meridian indexes k, s run from 0 to $\overline{M} - 1$ and $\Delta = \frac{2\pi}{\overline{M}}$, so that

$$\cos(m + h\overline{M})(k - s)\Delta = \cos m(k - s)\Delta, \quad \forall h \ge 0 \tag{45}$$

one can put

$$\widetilde{G}_{ir}^{m} = \sum_{h=0}^{+\infty} G_{ir}^{m+h\overline{M}}, \quad 0 \le m \le \overline{M} - 1. \tag{46}$$

Now it is just a matter of simple algebra to prove that the inverse of the matrix

$$C_{\boldsymbol{t}\boldsymbol{t}'} \equiv C(L_{\boldsymbol{t}}, L_{\boldsymbol{t}'}) \tag{47}$$

has the explicit shape [with $\boldsymbol{t}'' = (\vartheta_p, q\Delta)$]

$$C_{\boldsymbol{t}'\boldsymbol{t}''}^{(-1)} = \frac{1}{\overline{M}^2} \sum_{m'=0}^{\overline{M}-1} H_{rp}^{m'} \frac{\cos m'(s - p)\Delta}{1 + \delta_{m'0}}. \tag{48}$$

In this endeavour the identity (Papoulis 1984, Sect. 3.4)

$$\begin{aligned} &\sum_{s=0}^{\overline{M}-1} \cos m(k - s)\Delta \cos m'(s - q)\Delta \\ &\quad = \overline{M}\delta_{mm'}(1 + \delta_{m0}) \cos m(k - q)\Delta \end{aligned} \tag{49}$$

plays a central role. Moreover, the matrices $\{H^m_{rp}\}$ are defined to be the inverse of the matrices $\{\widetilde{G}^m_{ir}\}$, namely

$$\sum_r \widetilde{G}^m_{ir} H^m_{rp} \equiv \delta_{ip}. \tag{50}$$

It has to be pointed out that the indexes i, r, p are referred to parallels and therefore they vary from 1 up to the number of parallels present in the grid. We notice as well that even if the noise has a more general structure, for instance it is the sum of an uncorrelated component, possibly variable from point to point, and another component with the same covariance structure as in (41), then formulas (44) and (48) still remain true.

Formulas (48), (49) prove that the solution of the normal system can then be obtained by inverting only blocks with the same dimension as the number of parallels. All that agrees with the findings of Sansò and Tscherning (2003). However, in the context of the mentioned paper also the noise was supposed to be constant along parallels, a very heavy and often not met hypothesis. This requirement was necessary there to apply an FFT algorithm which here is not used and therefore there is no need that the noise variance be constant along the main diagonal. So now the problem left is to prove (41) as a consequence of axi-symmetry. We note that this is new since we do not assume that $T(\boldsymbol{x})$ is isotropic, i.e. that it depends on the spherical distance ψ only

$$C(\boldsymbol{x}, \boldsymbol{x}') \equiv C(\psi_{\boldsymbol{x}\boldsymbol{x}'}), \tag{51}$$

from which (41) descends easily, but we only suppose that

$$C(\boldsymbol{x}, \boldsymbol{x}') = C(\vartheta, \vartheta', \lambda - \lambda'), \tag{52}$$

which is a consequence of axi-symmetry, together with the more restrictive condition that

$$C(\vartheta, \vartheta', \lambda - \lambda') = C(\vartheta, \vartheta', \lambda' - \lambda). \tag{53}$$

Physically this means that, by using the symbols of Fig. 1, the following equalities hold

$$C(\boldsymbol{x}, \boldsymbol{x}') \equiv C(\boldsymbol{x}'', \boldsymbol{x}''') \equiv C(\boldsymbol{x}, {}'\boldsymbol{x}). \tag{54}$$

In other words, $C(\vartheta, \vartheta', \lambda - \lambda')$ is an even function of $\lambda - \lambda'$. We are ready now for the proof of (41).

Lemma 1 *Assume that the random field $T(\mathbf{x})$ enjoys the symmetries implied by (54), so that its covariance function satisfies (52) and (53), then $C(\vartheta, \vartheta', \lambda - \lambda')$ must take the form (41), with the coefficients $C^m_{\ell n}$ satisfying the symmetry relation (43).*

Proof To prove the Lemma we need preliminarily a proposition, which is easily found on any book on Sturm–Liouville theory, for instance Nikiforov and Uvarov (1988) (Ch. II, Sect. 8–10).

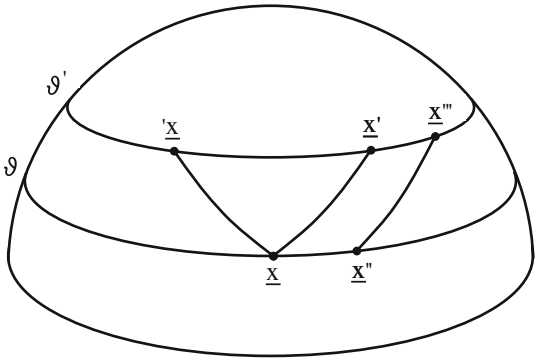

Fig. 1 The symmetries of the random field $T(\boldsymbol{x})$

Proposition 1 *For any fixed m, the sequence*

$$\left\{ \widetilde{P}_{nm}(\vartheta) = \frac{\sqrt{1+\delta_{m0}}}{2} \overline{P}_{nm}(\vartheta),\ n = m, m+1, \ldots \right\} \tag{55}$$

is a complete orthonormal sequence in $L^2(0,\pi)_{\sin\vartheta}$ with weight $(\sin\vartheta)$. This means that $\forall f(\vartheta) \in L^2(0,\pi)_{\sin\vartheta}$ one has the convergent representation

$$\begin{cases} f(\vartheta) = \displaystyle\sum_{n=m}^{+\infty} \widetilde{P}_{nm} \left(f, \widetilde{P}_{nm}\right) \\ \left(f, \widetilde{P}_{nm}\right) = \displaystyle\int_0^{\pi} f(\vartheta) \widetilde{P}_{nm}(\vartheta) \sin\vartheta \mathrm{d}\vartheta. \end{cases} \tag{56}$$

Here we only remark that it is easy to derive (56), when we already know that $\{Y_{nm}(\vartheta, \lambda)\}$ is a complete orthonormal system in L^2 over the unit sphere with scalar product

$$(f, g)_\sigma = \frac{1}{4\pi} \int_\sigma f(\vartheta, \lambda) g(\vartheta, \lambda) \sin\vartheta \, \mathrm{d}\vartheta \, \mathrm{d}\lambda. \tag{57}$$

As a matter of fact, such a property is well known among geodesists (see Heiskanen and Moritz 1967, Sect. 1.3).

Now assume that the variance of $T(\boldsymbol{x})$, namely $C(\boldsymbol{x}, \boldsymbol{x})$ is bounded everywhere so that $C(\vartheta, \vartheta', \lambda - \lambda')$ is certainly square integrable in each of its variables. Therefore, we can use the Fourier development, with cosine elements only,

$$C(\vartheta, \vartheta', \lambda - \lambda') = \sum_{m=0}^{+\infty} C^m(\vartheta, \vartheta') \cos m(\lambda - \lambda'), \tag{58}$$

because C is even in $\lambda - \lambda'$.

We note as well that we always have

$$C(\boldsymbol{x}, \boldsymbol{x}') = C(\vartheta, \vartheta', \lambda - \lambda') = C(\vartheta', \vartheta, \lambda' - \lambda) = C(\boldsymbol{x}', \boldsymbol{x}) \tag{59}$$

because $C(\boldsymbol{x}, \boldsymbol{x}')$ is a covariance function but at the same time

$$C(\vartheta', \vartheta, \lambda - \lambda') = C(\vartheta', \vartheta, \lambda' - \lambda) \tag{60}$$

because of (53). This clearly implies

$$C^m(\vartheta, \vartheta') = C^m(\vartheta', \vartheta). \tag{61}$$

Now, using Proposition 1, we can claim that there is a sequence $h_n^m(\vartheta')$ such that

$$C^m(\vartheta, \vartheta') = \sum_{n=m}^{+\infty} \widetilde{P}_{nm}(\vartheta) h_n^m(\vartheta'). \tag{62}$$

From (62) and (56) we have

$$\begin{aligned} h_n^m(\vartheta') &= \left(C^m(\vartheta, \vartheta'), \widetilde{P}_{nm}(\vartheta)\right) = \left(C^m(\vartheta', \vartheta), \widetilde{P}_{nm}(\vartheta)\right) \\ &= \sum_{\ell=m}^{+\infty} \widetilde{P}_{\ell m}(\vartheta') \left(h_\ell^m(\vartheta), \widetilde{P}_{nm}(\vartheta)\right). \end{aligned} \tag{63}$$

On the other hand, again using Proposition 1 and (56), it is

$$h_n^m(\vartheta') = \sum_{\ell m}^{+\infty} \left(h_n^m(\vartheta), \widetilde{P}_{\ell m}(\vartheta)\right) \widetilde{P}_{\ell m}(\vartheta'). \tag{64}$$

Comparing (63) and (64) we see that

$$\left(h_\ell^m(\vartheta), \widetilde{P}_{nm}(\vartheta)\right) = \left(h_n^m(\vartheta), \widetilde{P}_{\ell m}(\vartheta)\right) = \overline{C}_{\ell n}^m = \overline{C}_{n\ell}^m. \tag{65}$$

So substituting back into (63) and (62) we get

$$\begin{aligned} C^m(\vartheta, \vartheta') &= \sum_{n=m}^{+\infty} \sum_{\ell=m}^{+\infty} \overline{C}_{n\ell}^m \widetilde{P}_{nm}(\vartheta) \widetilde{P}_{\ell m}(\vartheta') \\ &\equiv \sum_{n,\ell=m}^{+\infty} C_{n\ell}^m \overline{P}_{nm}(\vartheta) \overline{P}_{\ell m}(\vartheta') \end{aligned} \tag{66}$$

where we have put

$$C_{n\ell}^m = \overline{C}_{n\ell}^m \cdot \frac{1 + \delta_{m0}}{4}. \tag{67}$$

Because of (65), $C_{n\ell}^m$ satisfies the symmetry relation (43) and the Lemma is proved. □

Concluding the section we can claim that, under the conditions of axi-symmetry described at the beginning, both approaches, least-squares and collocation, lead to block diagonal normal systems solvable by inverting smaller block matrices, which is a feasible numerical task.

We call the attention to the fact that based on the above discussion, axi-symmetry and block diagonal structure correspond to one another. So even the use of a block diagonal approximation is not presupposing the isotropy of the underlying signal, but rather a form of axi-symmetry of its stochastic characteristics. In particular it has to be underlined that if $T_{\ell m}$ are independent, although not with equal variances, yet (41) can still be satisfied; in fact the symmetry requirement (60) implies only that $T_{\ell m}$ and $T_{\ell,-m}$ have the same variance.

As for the noise, if it has a correlated component, this must have the same axi-symmetric behaviour of the signal, while an uncorrelated component can be arbitrary.

A final remark is in order here on the comparison of a collocation solution with a pure least-squares solution. In some sense the two are complementary as for the requirements we have to put on signal and noise. For the sake of simplicity assume the noise to be just uncorrelated. Then in forming the ordinary least-squares normals we get a block diagonal structure only if the noise is axi-symmetric, i.e. it has constant variance on parallels. On the other hand, we do not need to put any requirement on the signal, which is just represented by deterministic parameters. On the contrary, a collocation solution will have a block diagonal "normal" matrix only if the signal is stochastically axi-symmetric, but the noise is not required to satisfy any particular constraint.

5 Small examples on the way of a numerical solution

It is time now to look at the problem of how to implement numerically the solution, for instance in the form (38), which we rewrite as

$$\begin{aligned} &\delta\widehat{\boldsymbol{T}} = \boldsymbol{C}_{\mathrm{G}}\boldsymbol{\Pi}^{\mathrm{T}}(\boldsymbol{C}_{\mathrm{S}} + \boldsymbol{\Pi}\boldsymbol{C}_{\mathrm{G}}\boldsymbol{\Pi}^{\mathrm{T}})^{-1}\delta\widehat{\boldsymbol{T}}_{\mathrm{S}}, \\ &\delta\widehat{\boldsymbol{T}} = \widehat{\boldsymbol{T}} - \widehat{\boldsymbol{T}}_{\mathrm{G}}, \quad \delta\widehat{\boldsymbol{T}}_{\mathrm{S}} = \widehat{\boldsymbol{T}}_{\mathrm{S}} - \boldsymbol{\Pi}\widehat{\boldsymbol{T}}_{\mathrm{G}}, \\ &\boldsymbol{\Pi} = [\boldsymbol{I}_{\mathrm{S}}, 0], \quad \boldsymbol{C}_{\mathrm{S}} = \boldsymbol{D}_{\mathrm{S}}^{-1}, \quad \boldsymbol{C}_{\mathrm{G}} = \boldsymbol{D}_{\mathrm{G}}^{-1}. \end{aligned} \tag{68}$$

We recall that $\boldsymbol{C}_{\mathrm{S}}$ is the error covariance matrix of the satellite solution $\widehat{\boldsymbol{T}}_{\mathrm{S}}$, that is in fact available, and $\boldsymbol{C}_{\mathrm{G}}$ the error covariance matrix of $\widehat{\boldsymbol{T}}_{\mathrm{G}}$, which we assume to be available too.

As we see, there are two main numerical problems in the implementation of (68):

(1) the computation of $\boldsymbol{C}_{\mathrm{G}}\boldsymbol{\Pi}^{\mathrm{T}}$;
(2) the computation of $(\boldsymbol{C}_{\mathrm{S}} + \boldsymbol{\Pi}\boldsymbol{C}_{\mathrm{G}}\boldsymbol{\Pi}^{\mathrm{T}})^{-1}\delta\widehat{\boldsymbol{T}}_{\mathrm{S}}$.

We tackle separately the two problems.

(1) here we must take advantage of the discussion of Sect. 4. In building a high-resolution global model, for instance EGM08, the "ground" data are ultimately downward continued to a regular grid at the ellipsoid level (Rapp 1994; Pavlis 1997). The form of the observation equations at this point is such as to support an application of the discussion of Sect. 4, namely such that we can expect a block diagonal normal matrix $\boldsymbol{D}_{\mathrm{G}}$.

This, however, requires to disregard the inhomogeneity of the accuracy and spatial distribution of the data from which mean anomalies are estimated. Moreover, the effect of the downward continuation is strongly dependent on the topography and therefore it displays an important geographic signature. All in all, the predicted residual anomalies are affected by errors that are certainly not constant in longitude, particularly when crossing high mountain chains like the Himalayas or the Andes.

The possible combination with previously existing satellite models does not modify the above statement since these are intrinsically almost block diagonal. So, the conditions for cylindrical symmetry are violated and we cannot expect an exactly block diagonal $\boldsymbol{D}_{\mathrm{G}}$. Nevertheless, the success of the solution obtained by using only the block diagonal part of $\boldsymbol{D}_{\mathrm{G}}$ (Pavlis et al. 2008) strongly supports the hypothesis that such an inhomogeneity is only a minor problem.

The following simulated example goes along with this conclusion.

Example 1 This example is meant to clarify how large can be the degradation of a simple block diagonal solution of harmonic coefficients estimation from ground data, despite the presence of a noise with significantly non axi-symmetric variances.

The idea is basically that once we write (26), even if the best estimator is the one given by (27), according to the Gauss Markov theorem (Koch 1999), still an unbiased estimator is

$$\widehat{\boldsymbol{T}}_{\mathrm{G}} = (\boldsymbol{A}_{\mathrm{G}}^{\mathrm{T}}\boldsymbol{A}_{\mathrm{G}})^{-1}\boldsymbol{A}_{\mathrm{G}}^{\mathrm{T}}\boldsymbol{Y}_{\mathrm{G}}. \tag{69}$$

The point is that if $\boldsymbol{Y}_{\mathrm{G}}$ is a gridded data set, e.g. of gravity anomalies on the ellipsoid, then, according to the discussion of Sect. 4, the normal matrix $\boldsymbol{A}_{\mathrm{G}}^{\mathrm{T}}\boldsymbol{A}_{G}$ is block diagonal and the solution of (69) is computable even for very high degrees.

The question is how much we lose in this procedure. In order to gain insight into the loss, we have performed a small simulation, in which data are δg values on a $3^\circ \times 3^\circ$ grid on the ellipsoid. The anomalies are computed from EGM96 (Lemoine et al. 1998) up to degree 59, since in this way no aliasing is present in the data. Then a random independent error has been added to the data, sampling with the following rules:

Land (general)	$\sigma_v = 2$ mGal
Ocean	$\sigma_v = 5$ mGal
Himalayas–Andes	$\sigma_v = 12$ mGal

The data have then been inverted with three different methods and error degree variances have been computed:

(a) exact least-squares solution, i.e. with the correct use of the covariance matrix of the observations;
(b) the block diagonal solution (69);
(c) the solution by discretisation of integration formulas (Colombo 1981).

The results are presented in Fig. 2 where we can see that the solutions (b) and (c) have practically the same performance and that their differences with respect to the solution (a) are perceivable but small.

On the other hand, estimated formal errors derived from the approximated block diagonal normal matrix could be of poor quality and this is one of the reasons why the formal errors of the resulting solution sometimes need to be empirically calibrated, as performed, e.g. in EGM08 or in the EIGEN models.

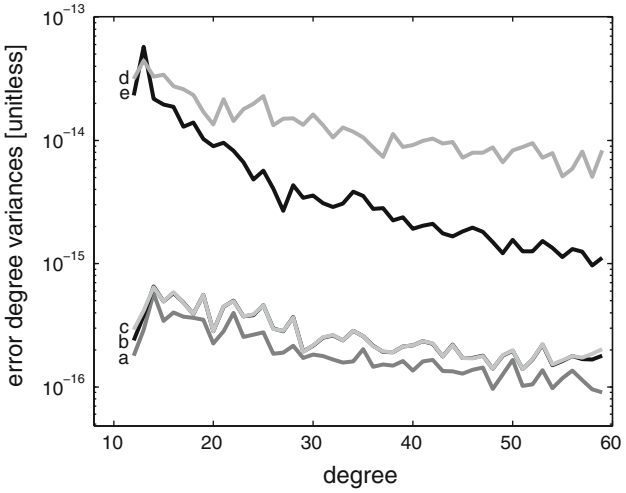

Fig. 2 Error degree variances of the exact least-squares solution (*a*), of the block diagonal solution assuming a unique observation error variance (*b*), of the numerical integration solution (*c*), of the block diagonal solution neglecting off-diagonal blocks (*d*); EGM96 signal degree variances (*e*)

An interesting remark can be drawn if we try to compute a block diagonal solution by calculating first the correct normal system and then neglecting the off-diagonal blocks. This is displayed as (d) in Fig. 2 and, as we can see, this solution is totally wrong with error degree variances larger than those of the signal.

A further important remark here is that for the collocation approach there is not the analogue of an unbiased but non-optional solution, leading to a block diagonal form like (69). As a matter of fact, while in Markov's theory the class of linear estimators is restricted by the condition of unbiasedness before applying the optimum principle, in collocation the class of linear estimators is, by assumption, always unbiased because the average of each coefficient T_{nm} is zero; in other words, in collocation the condition of unbiasedness is not adding any information (Rao et al. 2008).

Now we can approach the problem of computing $\boldsymbol{C}_{\mathrm{G}}\boldsymbol{\Pi}^{\mathrm{T}}$ assuming that $\boldsymbol{D}_{\mathrm{G}}$ can be reordered into a block diagonal form, namely the same block diagonal matrix used in the computation of the coefficients of EGM08. At this point it is important to take into account the ordering of the vector $\boldsymbol{T}$. In fact we know that the normal matrix is authentically block diagonal if and only if we organize the vector $\boldsymbol{T}$ by orders, e.g. $m = 0, m = 1, m = -1, \ldots, m = N_{\mathrm{G}}, m = -N_{\mathrm{G}}$. But the projector $\boldsymbol{\Pi}$ has the particular form $\boldsymbol{\Pi} = [\boldsymbol{I}_{\mathrm{S}}, 0]$, only if we put first all orders and degrees up to N_{S} and then the others, as illustrated in Fig. 3.

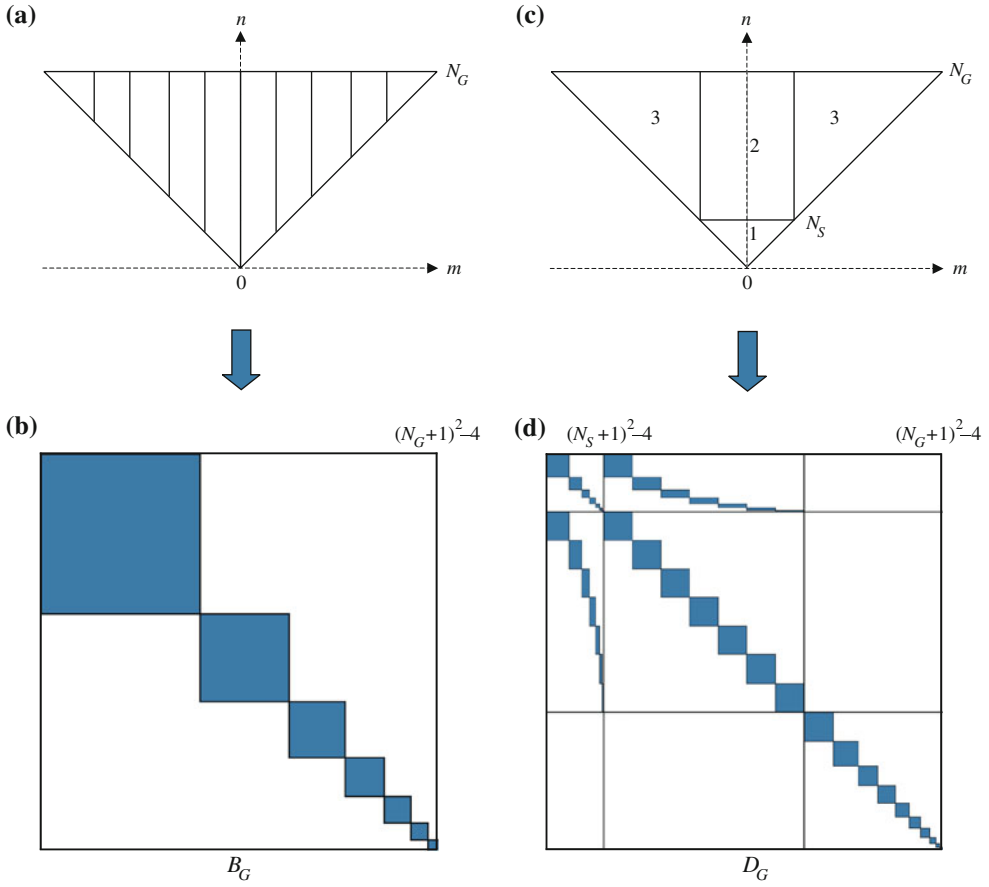

Fig. 3 **a** Organization of $\boldsymbol{T}$ by orders ($n \leq m \leq N_G$); **b** corresponding block diagonal normal matrix; **c** re-organization of $\boldsymbol{T}$ by orders up to degree N_S (area 1), then same orders but degrees $N_S < n \leq N_G$ (area 2), then all orders $m > N_S$ (area 3); **d** corresponding normal matrix; **b**, **d** non-zero elements only in drawn blocks

We call $\boldsymbol{B}_G$ the normal matrix with the first ordering and $\boldsymbol{D}_G$ that with the second one and we note that $\boldsymbol{D}_G$ can be obtained from $\boldsymbol{B}_G$ by reordering rows and columns, i.e. we have by some permutation matrix $\boldsymbol{P}$

$$\boldsymbol{D}_G = \boldsymbol{P}\boldsymbol{B}_G\boldsymbol{P}^T. \tag{70}$$

Here $\boldsymbol{P}$ is a unitary matrix, preserving the Cartesian modulus of vectors, so that

$$\boldsymbol{P}^{-1} = \boldsymbol{P}^T \tag{71}$$

and then

$$\boldsymbol{C}_G = \boldsymbol{D}_G^{-1} = \boldsymbol{P}\boldsymbol{B}_G^{-1}\boldsymbol{P}^T. \tag{72}$$

On the other hand $\boldsymbol{B}_G^{-1}$ is a block diagonal matrix with the same profile as $\boldsymbol{B}_G$; therefore (72) says that also $\boldsymbol{D}_G^{-1}$ has the same shape, i.e. the same blocks of non-zero elements, as $\boldsymbol{D}_G$.

We shall put

$$\boldsymbol{C}_G = \boldsymbol{D}_G^{-1} = \begin{vmatrix} \boldsymbol{A}_S & \boldsymbol{A}_{Sr} \\ \boldsymbol{A}_{rS} & \boldsymbol{A}_r \end{vmatrix} \tag{73}$$

and notice that $\boldsymbol{A}_S$ is block diagonal with block dimension determined by the choice of the triangle 1 in Fig. 3c, while in $\boldsymbol{A}_{rS}$ the non-zero elements are designed in such a way that the only non-zero outputs (rows of $\boldsymbol{A}_{rS}$) refer to the components of $\boldsymbol{T}$ placed in the area 2 in Fig. 3c, i.e. with $m \leq N_S$ and $N_S < n \leq N_G$. This gives to the normal matrix the peculiar shape displayed in Fig. 3d, known in literature as kite shape (Boxhammer and Schuh 2006).

We must remark here that a reverse ordering, giving rise to a falling kite figure of the normal matrix, has some computational advantages over the presented ordering of Fig. 3d, as discussed in Lemoine et al. (1998), Sect. 8.2.4.

Now recalling the definition of $\boldsymbol{\Pi}$, see (68), one finds immediately that

$$\boldsymbol{C}_G\boldsymbol{\Pi}^T = \begin{vmatrix} \boldsymbol{A}_S \\ \boldsymbol{A}_{rS} \end{vmatrix}. \tag{74}$$

This therefore closes the first problem.

(2) Due to (74) we find that

$$\boldsymbol{\Pi} \boldsymbol{C}_{\mathrm{G}} \boldsymbol{\Pi}^{\mathrm{T}} = \boldsymbol{A}_{\mathrm{S}} \tag{75}$$

is a block diagonal matrix of dimension $(N_{\mathrm{S}}+1)^2-4$ exactly as $\boldsymbol{C}_{\mathrm{S}}$. Now we have to solve a system of the form

$$(\boldsymbol{C}_{\mathrm{S}} + \boldsymbol{A}_{\mathrm{S}})\boldsymbol{\lambda} = \delta \widehat{\boldsymbol{T}}_{\mathrm{S}}, \tag{76}$$

so as to arrive, using (68) and (74), at

$$\delta \widehat{\boldsymbol{T}} = \begin{vmatrix} \boldsymbol{A}_{\mathrm{S}} & \boldsymbol{\lambda} \\ \boldsymbol{A}_{\mathrm{rS}} & \boldsymbol{\lambda} \end{vmatrix}. \tag{77}$$

It is interesting to observe that, whatever is $\boldsymbol{\lambda}$, the corrections $\delta \widehat{\boldsymbol{T}}$ different from zero will be only those for which $0 \le m \le N_{\mathrm{S}}$, i.e. for the unknowns of the areas 1 and 2 of Fig. 3c. The unknowns relative to the area 3 in the same figure will not change under any circumstance.

We are now reduced to find the solution of (76). Naturally if $\boldsymbol{C}_{\mathrm{S}}$ would be block diagonal, as $\boldsymbol{A}_{\mathrm{S}}$ is, then the solution of (76) would be trivial. The situation, however, is that $\boldsymbol{C}_{\mathrm{S}}$ is not exactly block diagonal, although it is close to such a condition. The reason is basically that while the satellite is taking observations regularly at 1 Hz along the orbit, the observation points after a sufficiently long time tend to distribute evenly in parallel bands (Bouman 2000). Since this is the parameter that mostly affects the prediction error on a grid at satellite level, one understands that observations on such a grid would tend to produce an almost block diagonal normal matrix and then an almost block diagonal covariance matrix $\boldsymbol{C}_{\mathrm{S}}$.

Such a hypothesis has been fully confirmed by Monte Carlo simulations (see for instance Pertusini et al. 2010). So we shall assume that one can write

$$\boldsymbol{C}_{\mathrm{S}} = \boldsymbol{C}_{0\mathrm{S}} + \boldsymbol{R} \tag{78}$$

where $\boldsymbol{C}_{0\mathrm{S}}$ is the prevailing block diagonal part of $\boldsymbol{C}_{\mathrm{S}}$. Then (76) can be rewritten as

$$(\boldsymbol{G}_0 + \boldsymbol{R})\boldsymbol{\lambda} = \delta \widetilde{\boldsymbol{T}}_{\mathrm{S}} \tag{79}$$

where

$$\boldsymbol{G}_0 = \boldsymbol{C}_{0\mathrm{S}} + \boldsymbol{A}_{\mathrm{S}} \tag{80}$$

and $\boldsymbol{R}$ is much smaller, hopefully also in norm, than $\boldsymbol{G}_0$.

Moreover, $\boldsymbol{G}_0$ is block diagonal and since we expect $\boldsymbol{A}_{\mathrm{S}}$ to be well conditioned, we do have the same for $\boldsymbol{G}_0$. So, even if some of the blocks of $\boldsymbol{C}_{0\mathrm{S}}$ would not be well conditioned or even not invertible, the addition of the corresponding blocks of $\boldsymbol{A}_{\mathrm{S}}$ should have a regularizing effect.

At this point the easiest and feasible numerical solution of (79) is obtained from solving sequentially the system

$$\boldsymbol{G}_0 \boldsymbol{\lambda}_{n+1} = \delta \widehat{\boldsymbol{T}}_{\mathrm{S}} - \boldsymbol{R}\boldsymbol{\lambda}_n. \tag{81}$$

It is immediate to verify that the recursion (81) is equivalent to the use of the known Neumann series

$$(\boldsymbol{G}_0 + \boldsymbol{R})^{-1} = \boldsymbol{G}_0^{-1} - \boldsymbol{G}_0^{-1}\boldsymbol{R}\boldsymbol{G}_0^{-1} + \boldsymbol{G}_0^{-1}\boldsymbol{R}\boldsymbol{G}_0^{-1}\boldsymbol{R}\boldsymbol{G}_0^{-1} - \cdots \tag{82}$$

In turn (82) is convergent when the norm bound holds

$$\| \boldsymbol{G}_0^{-1/2}\boldsymbol{R}\boldsymbol{G}_0^{-1/2} \| \le q < 1. \tag{83}$$

Indeed, (82) is not the best numerical solver of our system and sometimes it is not even applicable because condition (83) might not be satisfied. Nevertheless, we mention it here because in the subsequent example we will compare the result of $(\boldsymbol{G}_0+\boldsymbol{R})^{-1}$ with the first term, $\boldsymbol{G}_0^{-1}$, to derive some interesting suggestions. A better numerical solution can be obtained by the pre-conditioned conjugate gradient (Hestenes and Stiefel 1952; Kirsch 1996) or similar methods.

Finally let us report, without proof, the formula for the covariance of the new estimate $\widehat{\boldsymbol{T}}$, where $\widehat{\boldsymbol{T}}_{\mathrm{G}}$ and $\widehat{\boldsymbol{T}}_{\mathrm{S}}$ are merged. Actually, we have

$$\boldsymbol{C}_{\widehat{\boldsymbol{T}}} = \boldsymbol{C}_{\mathrm{G}} - \boldsymbol{C}_{\mathrm{G}}\boldsymbol{\Pi}^{\mathrm{T}}(\boldsymbol{C}_{\mathrm{S}} + \boldsymbol{A}_{\mathrm{S}})^{-1}\boldsymbol{\Pi}\boldsymbol{C}_{\mathrm{G}}. \tag{84}$$

This is a computationally tractable formula, at least for the diagonal of $\boldsymbol{C}_{\widehat{T}}$; yet, the point here is that from (84) we read that $\boldsymbol{C}_{\widehat{T}} < \boldsymbol{C}_{\mathrm{G}}$, namely merging the information of $\widehat{\boldsymbol{T}}_{\mathrm{S}}$ reduces, as it has to be, the covariance of the estimation error.

Now that all the main practical problems have been illustrated, we show how it works by a numerical example, much smaller in dimension but correctly reflecting pros and cons of the method.

Example 2 In this example we have tried to reproduce a somewhat more realistic situation, though restricting the computations to a simple and manageable case. The reason is that we wanted a case where the exact solution could be computed without a large effort.

The simulation starts with the "true" spherical harmonic coefficients taken from EGM96, from degree 10 up to degree 179. A regular grid with size $1^\circ \times 1^\circ$ of δg values has been computed on a spherical boundary at ground level and a 1 mGal white noise has been added. The least-squares solution $\widehat{\boldsymbol{T}}_{\mathrm{G}}$, up to degree and order 179, and its covariance matrix $\boldsymbol{C}_{\mathrm{G}}$ have been computed. It is then no surprise that $\boldsymbol{C}_{\mathrm{G}}$ has an exact block diagonal structure. At the same time a simulation has been performed on a 2-month GOCE-like orbit for satellite-to-satellite (SST) observations with a realistic noise (Catastini et al. 2003). As known, these observations ultimately provide T along the orbit when using the energy integral approach (Jekeli 1999; Visser et al. 2003). By applying a collocation gridding at satellite level (which is part of the so-called space-wise approach), a grid with size $0.5^\circ \times 0.5^\circ$ of predicted values for the evaluation functional of T has been produced. Spherical harmonic coefficients from degree 10 up to degree 60 are finally computed by discretisation of integration formulas. By a classical application of the Monte Carlo method, we could derive from 10,000 samples, along

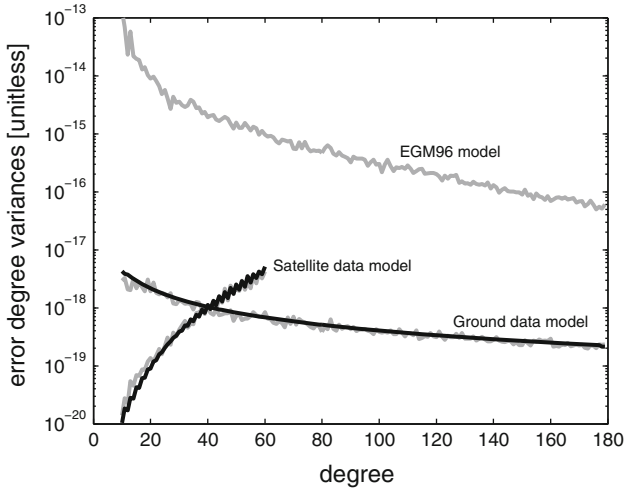

Fig. 4 Empirical (*grey*) and predicted (*black*) error degree variances of the estimated model either from satellite or from ground data. EGM96 signal degree variances are plotted as reference

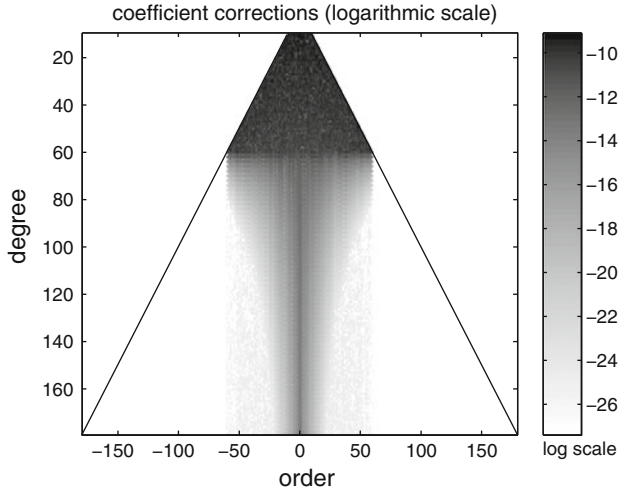

Fig. 5 Corrections to the ground-data model from the incorporation of the satellite-data model. Plotted quantity: $\log_{10} \delta\widehat{\boldsymbol{T}}$

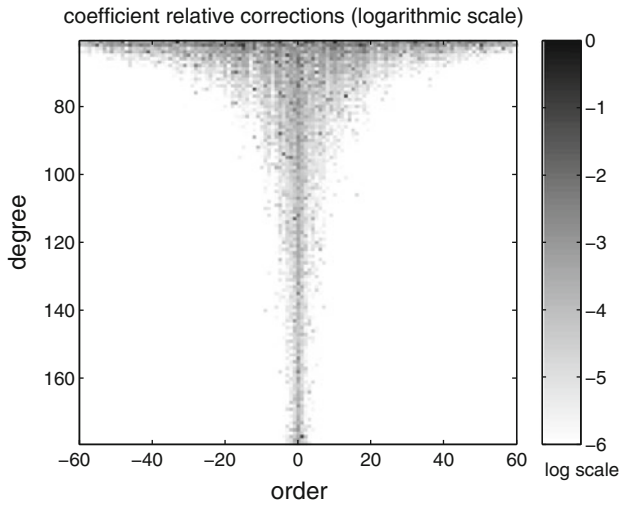

Fig. 6 Relative corrections with respect to the true coefficients in the area 2 ($n \geq 61$, $m \leq 60$), see Fig. 3c. Plotted quantity: $\log_{10}[\delta\widehat{\boldsymbol{T}}/\boldsymbol{T}]$

the above line, an estimate of a realistic, in the framework of this example, covariance matrix $\boldsymbol{C}_S$. Although, due to the nature of the observational functional, $\boldsymbol{C}_S$ is dominated by long wavelength errors not exactly cylindrically symmetric, yet the almost block diagonal outlook of $\boldsymbol{C}_S$, apart from a few orders close to 0, is quite pronounced. In other words, the matrix $\boldsymbol{R}$ of equation (78) is here non zero, but small. A comparison between the error degree variances of the two sets of estimated coefficients is shown in Fig. 4.

Since the number of unknowns of the example is not large, namely 3,621 unknowns, the solution of the system (79) and the final solution (77) can be computed in an exact manner, providing the corrections $\delta\widehat{\boldsymbol{T}}$ to the input coefficients $\widehat{\boldsymbol{T}}_G$, displayed in Fig. 5. As one can see, the most important corrections are in the area 1 of Fig. 3c, but also many coefficients in the area 2 undergo a significant variation. All in all, the signal correction contained in the area 2 has a mean square power of the order of 1 mm, which is certainly not very important. Nevertheless, the change of individual coefficients with respect to their own size ranges up to some per cent (cf. Fig. 6) and therefore it is worthwhile to be computed.

Two further comments could be made on this small simulated experiment. The first is that the solution $\boldsymbol{\lambda}$ of (79) computed by completely disregarding $\boldsymbol{R}$, i.e. by exploiting only the block diagonal part of the normal matrix, is extremely close to the exact solution. Nevertheless, the use of (77) still produces a visible effect on the coefficients of area 2 and should therefore be applied.

A second remark is that in this example the full series solution (82) cannot be directly applied since the matrix $\boldsymbol{G}_0^{-1/2}\boldsymbol{R}\boldsymbol{G}_0^{-1/2}$ has a single eigenvalue much larger than 1. A suitable strategy based on more refined numerical methods should be therefore applied (see for instance Schuh 1996; Klees et al. 2004).

Such conclusions are indeed not of general nature, because the authors do not know whether they are related to the peculiar features of this example. Further conclusions of the section are postponed to the next section.

6 Conclusions

First of all, we would like to underline that this study, although motivated by the need of understanding how to approach the problem of combining two global models like the high-resolution EGM08 and the SpW satellite-only model from the GOCE data analysis, is in fact more general and could be applied to other cases, e.g. the combination of two satellite models, at least as far as approximate values of the inverse normal matrices are available.

The specific findings for the problem that motivated the study though are mainly two:

(1) by applying the estimation/prediction formula (38), issuing either from a least-squares or from a collocation approach, we can combine two global models with a numerical complexity (measured by the dimension of the normal system to be solved) which is as large as the number of parameters (determinist or stochastic, according to the interpretation) of the smaller of the two models;

(2) the application of the combination formula (38) does not require the direct knowledge of the least-squares normal matrices relative to the two original data sets but only of their inverse, namely of the covariance matrices of the already estimated parameters.

Neither of two findings is in itself particularly new, only it is their consideration for the problem at hand which makes them appealing to us. In fact, the processes of estimating separately the two global models, in the paper represented by the two vectors $\widehat{\boldsymbol{T}}_{\mathrm{G}}$ and $\widehat{\boldsymbol{T}}_{\mathrm{S}}$, are so complicated and use so many steps of approximation that one is not entitled to say that the true normal matrices are available. In addition, in our case, $\widehat{\boldsymbol{T}}_{\mathrm{S}}$ is not coming from a straight application of least-squares principles, but rather from a stepwise collocation approach (the SpW approach), so that the corresponding normal matrix is not even defined. All in all, in such complicated procedures it seems to us more reliable and realistic to derive direct estimates of the covariances by means of Monte Carlo sampling methods than by pure numerical covariance propagation.

More specifically to answer the questions we have put in the introduction we can draw the following conclusions:

- finite versus infinite dimensional solutions, least-squares versus collocation: if the full information (including prior information) is treated either by least-squares or by collocation, it provides exactly the same solution; a difference occurs only whether we use infinite dimensional covariance functions.
  This conclusion was already known in literature. Nevertheless, one point has been elucidated: namely in reducing the problem to finite dimensions, one cannot use the simple criterion of equating the commission to the omission error. Rather a sensitivity analysis has to be conducted and this generally requires global models with a higher degree to be resolved. Only afterwards, part of the estimated coefficients, i.e. those above the degree of equal omission and commission error, can be neglected. In other words, to avoid biases larger than the noise, the criterion of the best maximum degree to which perform the estimation is not the same as that of an acceptable maximum degree;
- merging our global model by resorting to the original observation equations versus merging the coefficients: the answer to the question is clear and very well known in estimation theory (both least-squares and collocation), and it is traced back to the concept of sufficient statistics or to projections on nested subspaces. The two approaches are equivalent when covariances are correctly propagated. So the point is only whether the combination is numerically feasible. This has been discussed at the point (1), (2) above;
- symmetry and block diagonal structures: since block diagonal (or almost block diagonal) structures simplify the numerical work, it is important to understand what are the implicit conditions (or approximations) imposed on the problem to which we apply them. The result that they correspond to a cylindrical symmetry is not new. However, in the paper the collocation case is worked out in detail showing that such a symmetry, with an additional hypothesis, is sufficient to guarantee the block diagonal shape of the "normal" matrix. The detailed proof was not known to the authors.
  An important remark has been made on this point, showing that there is a different behaviour of least-squares with respect to collocation when analysing gridded data; the first method gives a block diagonal normal only if the noise is constant along parallel. The second method gives rise to a block diagonal normal if the signal is axisymmetric, but the noise can be whatever, as far as it is uncorrelated;
- on how to perform a block diagonal approximation: Example 1, though very small, shows that a normal system in least-squares theory corresponding to non-optimal estimation, i.e. without an exact inverse noise matrix, can still provide a good approximation to the exact solution, while cutting the block diagonal form from a complete exact normal matrix can give a very bad solution. In this respect least-squares theory seems to depart significantly from collocation, where there are no analogous concepts;
- the coefficients updated by the new satellite data: the question is whether only coefficients contained in $\boldsymbol{T}_{\mathrm{S}}$ will be updated by the introduction of the $\boldsymbol{Y}_{\mathrm{S}}$ data. Example 2 clearly shows that, although quantitatively small, yet the effect of combining $\widehat{\boldsymbol{T}}_{\mathrm{S}}$ with $\widehat{\boldsymbol{T}}_{\mathrm{G}}$ has an impact on all the coefficients with order $|m| \leq N_{\mathrm{S}}$, even for degrees $n > N_{\mathrm{S}}$.

Naturally, although some conclusions have been drawn in the direction of a strategy for the combination in particular of EGM08 and GOCE global models, a final word still requires much more research especially on the numerical side, which in the present paper is certainly of smaller importance.

Acknowledgments This work has been performed in the framework of the Italian project GOCE-ITALY, supported by the Italian Space Agency (ASI). Moreover, the authors would like to thank the reviewers for their important contribution from which the paper has benefited a lot.

Appendix

In this appendix we will make some quantitative considerations on model errors in the observation equations used to estimate a global model. On the same issue the reader could consult also Rummel (1997) and Koch (2005).

As we did in Sect. 4, we assume to have observation equations

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{T} + \boldsymbol{v} = \boldsymbol{A}_{(N)}\boldsymbol{T}_{(N)} + \boldsymbol{A}^{(N)}\boldsymbol{T}^{(N)} + \boldsymbol{v} \tag{85}$$

and we would like to reason on the truncation condition consisting in deleting the term $\boldsymbol{A}^{(N)}\boldsymbol{T}^{(N)}$. Such a term then will become a bias $\boldsymbol{b} = \boldsymbol{A}^{(N)}\boldsymbol{T}^{(N)}$ in our model.

The ordinary criterion would be to require that the commission error at degree $N+1$ would become larger than the omission error (see for instance Rummel 1997).

But this is more a criterion derived for reasons of optimal estimation rather than a choice based on the requirement that $\boldsymbol{b} = 0$ will not significantly change the model (85). This rather asks for the statement that $\boldsymbol{b}$ is "small" with respect to $\boldsymbol{v}$; in this case in fact any criterion tending to reduce $\boldsymbol{v}$ will give almost equivalent answers.

The target can be reached either by comparing $\boldsymbol{b}$ with $\boldsymbol{v}$ as vectors or component by component. In the first case we have to use the metric $\boldsymbol{C}_v^{-1}$ which is natural in the observation space according to a least-squares principle. So one should write

$$\boldsymbol{b}^{\mathrm{T}}\boldsymbol{C}_v^{-1}\boldsymbol{b} = \boldsymbol{T}^{(N)\mathrm{T}}\boldsymbol{A}^{(N)\mathrm{T}}\boldsymbol{C}_v^{-1}\boldsymbol{A}^{(N)}\boldsymbol{T}^{(N)} \ll \boldsymbol{v}^{\mathrm{T}}\boldsymbol{C}_v^{-1}\boldsymbol{v}. \tag{86}$$

If we assume to have prior information on $\boldsymbol{T}^{(N)}$ in the form of its (diagonal) covariance matrix $\boldsymbol{K}^{(N)}$ (Kaula 1966b; Tscherning and Rapp 1974), (86) can be duly averaged with respect to $\boldsymbol{T}^{(N)}$ and to $\boldsymbol{v}$, to give

$$\mathrm{Tr}\boldsymbol{K}^{(N)}(\boldsymbol{A}^{(N)\mathrm{T}}\boldsymbol{C}_v^{-1}\boldsymbol{A}^{(N)}) \ll \mathrm{Tr}\boldsymbol{C}_v^{-1}\boldsymbol{C}_v = M, \tag{87}$$

where M is the number of measurements, i.e. the dimension of $\boldsymbol{v}$.

Any further elaboration along this line depends on the shape of the matrices therein; however, it is useful to notice that if we have the direct case in which $\boldsymbol{A}^{(N)\mathrm{T}}\boldsymbol{C}_v^{-1}\boldsymbol{A}^{(N)} = \sigma_v^{-2}\boldsymbol{I}$, then (87) gives

$$\mathrm{Tr}\boldsymbol{K}^{(N)} \ll M\sigma_v^2. \tag{88}$$

In this formula we recognize that $\mathrm{Tr}\boldsymbol{K}^{(N)}$ is just the mean square omission error, while $M\sigma_v^2$ is the mean square commission error. We can observe though that (88) says that the omission error should be much smaller than the commission error, forcing us to retain a higher number of unknowns than those required by the ordinary equality criterion.

Coming to a comparison of individual components b_i, v_i, we are pushed to use more restrictive inequalities implying also more severe conditions. In fact, assuming that in analogy to the hypothesis used in Sect. 4 the observation equations can be written in the form

$$\begin{aligned} Y_i &= \sum_{n=2}^{N}\sum_{m=-n}^{n}\gamma_{nm}T_{nm}S_{nm}(\boldsymbol{x}_i) \\ &\quad + \sum_{n=N+1}^{+\infty}\sum_{m=-n}^{n}\gamma_{nm}T_{nm}S_{nm}(\boldsymbol{x}_i) + v_i \\ &= (\boldsymbol{A}_{(N)}\boldsymbol{T}_{(N)})_i + (\boldsymbol{A}^{(N)}\boldsymbol{T}^{(N)})_i + v_i, \end{aligned} \tag{89}$$

the following condition issues

$$\left|\sum_{n=N+1}^{+\infty}\sum_{m=-n}^{n}\gamma_{nm}T_{nm}S_{nm}(\boldsymbol{x}_i)\right| \ll \sigma_{v_i}. \tag{90}$$

In order to elaborate on (90), one is forced to establish an a-fortiori condition. So we can write

$$\begin{aligned} |b_i| &= \left|\sum_{n=N+1}^{+\infty}\sum_{m=-n}^{n}\gamma_{nm}T_{nm}S_{nm}(\boldsymbol{x}_i)\right| \\ &\le \sum_{n=N+1}^{+\infty}\sum_{m=-n}^{n}|\gamma_{nm}||T_{nm}||S_{nm}(\boldsymbol{x}_i)|. \end{aligned} \tag{91}$$

On the other hand, from the known relation

$$\sum_{m=-n}^{n}S_{nm}(\boldsymbol{x})^2 = W_n^2(2n+1), \tag{92}$$

where

$$W_n = \left(\frac{R}{r}\right)^{n+1}, \tag{93}$$

we see that

$$|S_{nm}(\boldsymbol{x})| \le W_n\sqrt{2n+1}. \tag{94}$$

Using (94) in (91) and applying the Cauchy–Schwarz inequality, we get

$$|b_i| \le \sum_{n=N+1}^{+\infty}W_n\sqrt{2n+1}\sqrt{\sum_{m=-n}^{n}\gamma_{nm}^2}\sqrt{\sum_{m=-n}^{n}T_{nm}^2}. \tag{95}$$

Now we recognize that, according to a standard definition,

$$\sqrt{\sum_{m=-n}^{n}T_{nm}^2} = k_n, \tag{96}$$

and we put

$$\gamma_n = \sqrt{\frac{1}{2n+1}\sum_{m=-n}^{n}\gamma_{nm}^2}. \tag{97}$$

With these positions (95) becomes

$$|b_i| \le \sum_{n=N+1}^{+\infty}W_nk_n(2n+1)\gamma_n; \tag{98}$$

indeed (90) is a-fortiori satisfied if

$$\sum_{n=N+1}^{+\infty} W_n k_n (2n+1)\gamma_n \ll \sigma_{v_i}. \tag{99}$$

An identical relation could be derived by squaring (90) and averaging over $\{T_{nm}\}$. We underline that (99) is generally a much stronger condition than (87).

The conclusion of this appendix, as for the matter of estimating global gravity model, is that despite the criterion of keeping the maximum degree N of a global model at the value at which omission and commission errors balance each other, it is always better to include into the estimation model a higher maximum degree N' and then disregard the estimated coefficients from degree $N+1$ to N' than sharply putting $N' = N$. In any way, it is wise to perform a direct sensitivity analysis to ascertain by a direct check that biases introduced by truncation are in fact negligible with respect to noise.

References

Alkhatib H, Schuh W-D (2007) Integration of the Monte Carlo covariance estimation strategy into tailored solution procedures for large-scale least squares problems. J Geod 81(1):53–66

Bosch W (1993) A rigorous least squares combination of low and high degree spherical harmonics. Presented at XXI IAG General Assembly, 8–15 August 1993, Beijing, China

Bouman J (2000) Quality assessment of satellite-based global gravity fields models. In: Publications on Geodesy, vol 48, The Netherlands Geodetic Commission, Delft

Boxhammer Ch, Schuh W-D (2006) GOCE gravity field modeling: computational aspects—free kite numbering scheme. In: Flury J, Rummel R, Reigber Ch, Rothacher M, Boedecker G, Schreiber U (eds) Observation of the Earth system from space. Springer, Berlin, pp 209–224

Catastini G, Cesare S, De Sanctis S, Dumontel M, Parisch M, Sechi G (2007) Predictions of the GOCE in-flight performances with the End-to-End System Simulator. In: Proceedings of the 3rd International GOCE User Workshop, 6–8 November 2006, Frascati, Italy, pp 9–16

Colombo OL (1981) Numerical methods for harmonic analysis on the sphere. Report No. 310, Department of Geodetic Science and Surveying, The Ohio State University, Columbus

Cox DR, Hinkley DV (1974) Theoretical statistics. Chapman and Hall, London

ESA (1999) Gravity field and steady-state ocean circulation mission. ESA SP-1233(1). ESA Publication Division, c/o ESTEC, Noordwijk

Gruber T (2001) High-resolution gravity field modeling with full variance-covariance matrices. J Geod 75(9–10):505–514

Gundlich N, Koch K-R, Kusche J (2003) Gibbs sampler for computing and propagating large covariance matrices. J Geod 77(9):514–528

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

Hestenes MR, Stiefel E (1952) Methods of conjugate gradients for solving linear systems. J Res Natl Bureau Stand 49(6):409–436

Jekeli C (1999) The determination of gravitational potential differences from satellite-to-satellite tracking. Celest Mech Dyn Astron 75(2):85–101

Kalman RE, Bucy RS (1961) New results in linear filtering and prediction theory. Trans ASME Ser D J Basic Eng 83:95–107

Kaula WM (1966) Tests and combination of satellite determinations of the gravity field with gravimetry. J Geophys Res 71(22):5303–5314

Kaula WM (1966) Theory of satellite geodesy. Blaisdell Publishing Company, Waltham

Kirsch A (1996) An introduction to the matematical theory of inverse problems. In: Applied Mathematical Sciences, vol 120. Springer, Berlin

Klees R, Ditmar P, Kusche J (2004) Numerical techniques for large least-squares problems with applications to GOCE. In: Sansò F (ed) International Association of Geodesy symposia, "V Hotine-Marussi Symposium on Mathematical Geodesy", vol 127. Springer, Berlin, pp 12–21

Koch K-R (1999) Parameter estimation and hypothesis testing in linear models, 2nd edn. Springer, Berlin

Koch K-R, Kusche J (2002) Regularization of geopotential determination from satellite data by variance components. J Geod 76(5):259–268

Koch K-R (2005) Determining the maximum degree of harmonic coefficients in geopotential models by Monte Carlo methods. Studia Geophysica et Geodaetica 49(3):256–275

Krarup T (2006) A convergence problem in collocation theory. In: Borre K (ed) Mathematical foundation of geodesy. Springer, Berlin

Lemoine FG, Kenyon SC, Factor JK, Trimmer RG, Pavlis NK, Chinn DS, Cox CM, Klosko SM, Luthcke SB, Torrence MH, Wang YM, Williamson RG, Pavlis EC, Rapp RH, Olson TR (1998) The development of the joint NASA GSFC and NIMA geopotential model EGM96. NASA/TP-1998-206861, Goddard Space Flight Center, Greenbelt, Maryland

Mayer-Gürr T (2006) Gravitationsfeldbestimmung aus der Analyse kurzer Bahnbögen am Beispiel der Satellitenmissionen CHAMP und GRACE. PhD thesis, University of Bonn, Germany

Migliaccio F, Reguzzoni M, Sansò F (2004) Space-wise approach to satellite gravity field determination in the presence of coloured noise. J Geod 78(4–5):304–313

Migliaccio F, Reguzzoni M, Sansò F, Tselfes N (2009) An error model for the GOCE space-wise solution by Montecarlo methods. In: Sideris MG (ed) International Association of Geodesy symposia, "Observing our Changing Earth", vol 133. Springer, Berlin, pp 337–344

Moritz H (1989) Advanced physical geodesy, 2nd edn. Wichmann Verlag, Karlsruhe

Nikiforov AF, Uvarov VB (1988) Special functions of mathematical physics. Birkhäuser, Basel-Boston

Pail R, Bruinsma S, Migliaccio F, Förste C, Goiginger H, Schuh W-D, Hoeck E, Reguzzoni M, Brockmann JM, Abrikosov O, Veicherts M, Fecher T, Mayrhofer R, Krasbutter I, Sansò F, Tscherning CC (2011) First GOCE gravity field models derived by three different approaches. J Geod 85(11):819–843

Papoulis A (1984) Signal analysis. McGraw Hill, New York

Pavlis NK (1997) Development and applications of geopotential models. In: Lecture notes of the international school for the determination and use of the geoid, International Geoid Service, DIIAR—Politecnico di Milano, pp 101–147

Pavlis NK, Holmes SA, Kenyon SC, Factor JK (2008) An Earth gravitational model to degree 2160: EGM2008. Presented at the 2008 General Assembly of the European Geophysical Union (EGU), pp 13–18 April 2008, Vienna

Pertusini L, Reguzzoni M, Sansò F (2010) Analysis of the covariance structure of the GOCE space-wise solution with possible applications. In: Mertikas SP (ed) International Association of Geodesy symposia, "Gravity, Geoid and Earth Observation", vol 135. Springer-Verlag, Berlin, pp 95–202

Phillips DL (1962) A technique for the numerical solution of certain integral equations of the first kind. J Assoc Comput Mach 9:84–97

Rao CR, Toutenburg H, Shalabh, Heumann C (2008) Linear models and generalizations, 3rd edn. Springer, Berlin

Rapp RH (1975) Comparison of least squares and collocation estimated from potential coefficients. In: Methoden und Verfahren der Mathematischen Physik, Band 14, Bibliographisches Institut, Zürich, pp 133–148

Rapp RH (1984) The determination of high degree potential coefficient expansions from the combination of satellite and terrestrial gravity information. Report No. 361, Department of Geodetic Science and Surveying, The Ohio State University, Columbus

Rapp RH (1994) The use of potential coefficient models in computing geoid undulations. In: Lecture notes of the international school for the determination and use of the geoid, International Geoid Service, DIIAR—Politecnico di Milano, pp 71–99

Reguzzoni M, Tselfes N (2009) Optimal multi-step collocation: application to the space-wise approach for GOCE data analysis. J Geod 83(1):13–29

Reigber Ch, Schmidt R, Flechtner F, König R, Meyer U, Neumayer KH, Schwintzer P, Zhu SY (2005) An Earth gravity field model complete to degree and order 150 from GRACE: EIGEN-GRACE02S. J Geodyn 39(1): 1–10

Rummel R (1997) Spherical spectral properties of the Earth's gravitational potential and its first and second derivatives. In: Sansò F, Rummel R (eds) Lecture notes in Earth sciences, "Geodetic Boundary Value Problems in View of the One Centimeter Geoid", vol 65. Springer, Berlin, pp 359–404

Sansò F (1986) Statistical methods in physical geodesy. In: Sünkel H (ed) Lecture notes in Earth sciences, "Mathematical and Numerical Techniques in Physical Geodesy", vol 7. Springer, Berlin, pp 49–155

Sansò F, Venuti G, Tscherning CC (2000) A theorem of insensitivity of the collocation solution to variations of the metric of the interpolation space. In: Schwarz KP (ed) International Association of Geodesy symposia, vol 121, Springer, Berlin, pp 233–240

Sansò F, Tscherning CC (2003) Fast spherical collocation: theory and examples. J Geod 77(1–2):101–112

Schuh W-D (1996) Tailored numerical solution strategies for the global determination of the Earth's gravity field. Mitteilungen der Geodätischen Institute der TU Graz, Folge 81, Graz

Sneeuw N (2000) A semi-analytical approach to gravity field analysis from satellite observations. Deutsche Geodätische Kommission, Reihe C, No. 527, Verlag der Bayerischen Akademie der Wissenschaften, München

Tapley BD, Shum CK, Ries JC, Poole SR, Abusali PAM, Bettadpur SV, Eanes RJ, Kim MC, Rim HJ, Shutz BE (1997) The TEG-3 Geopotential Model. In: Segawa J, Fujimoto H, Okubo S (eds) International Association of Geodesy symposium, "Gravity, Geoid, and Marine Geodesy", vol 117. Springer, Berlin, pp 453–460

Tikhonov AN (1963) Regularization of incorrectly posed problems. Sov Math Doklady 4(6):1624–1627

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

Visser PNAM, Sneeuw N, Gerlach C (2003) Energy integral method for gravity field determination from satellite orbit coordinates. J Geod 77(3–4):207–216