

# How to handle colored observation noise in large least-squares problems

R. Klees<sup>1</sup>, P. Ditmar<sup>1</sup>, P. Broersen<sup>2</sup>

<sup>1</sup> Department of Geodesy, Physical, Geometric and Space Geodesy, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Thijsseweg 11, 2629 JA Delft, The Netherlands  
e-mails: r.klees@citg.tudelft.nl, ditmar@citg.tudelft.nl; Tel.: +31-15-2785100; Fax: +31-15-2783711

<sup>2</sup> Department of Applied Physics, Delft University of Technology, P.O. Box 5046, 2600 GA Delft, The Netherlands  
e-mail: broersen@tn.tudelft.nl

Received: 10 January 2002 / Accepted: 1 August 2002

**Abstract.** An approach to handling colored observation noise in large least-squares (LS) problems is presented. The handling of colored noise is reduced to the problem of solving a Toeplitz system of linear equations. The colored noise is represented as an auto regressive moving-average (ARMA) process. Stability and invertability of the ARMA model allows the solution of the Toeplitz system to be reduced to two successive filtering operations using the inverse transfer function of the ARMA model. The numerical complexity of the algorithm scales proportionally to the order of the ARMA model times the number of observations. This makes the algorithm particularly suited for LS problems with millions of observations. It can be used in combination with direct and iterative algorithms for the solution of the normal equations. The performance of the algorithm is demonstrated for the computation of a model of the Earth's gravity field from simulated satellite-derived gravity gradients up to spherical harmonic degree 300.

**Keywords:** Large least-squares problems – auto regressive moving-average process – Noise power spectral density function – satellite gravity gradiometry

problem challenging is the combination of colored (i.e. frequency-dependent) observation noise with a huge number of observations and unknown gravity field parameters. The question answered in this paper is: How should colored observation noise be properly taken into account without significantly increasing the numerical effort?

It will be shown that when the normal equations are solved iteratively, the problem of colored noise reduces to the solution of a Toeplitz system of linear equations per iteration. This system is solved by exploiting an auto regressive moving-average (ARMA) model of the colored noise, which has a numerical complexity of the order of the number of observations. This makes it particularly suited for the solution of large least-squares (LS) problems with millions of observations and thousands of unknowns. The approach may also be used in combination with a direct method for the solution of the normal equations. However, the numerical effort is then increased as each column of the design matrix has to be filtered.

The material presented is organized as follows. In Sect. 2, large LS problems are addressed and the benefit of iterative methods for the solution of the normal equations is discussed. In Sect. 3, it is shown that, when iterative methods are used, the proper handling of stationary colored noise reduces to the solution of a Toeplitz system of equations per iteration. In Sect. 4, some basics about ARMA processes are provided. Section 5 is devoted to the fast solution of a Toeplitz system of linear equations. The approach pursued assumes that an ARMA model of the colored noise is available. It is shown that the solution of the Toeplitz system of linear equations reduces to two filtering operations. The transfer function of two filters is the inverse of the transfer function associated with the ARMA model. In Sect. 6, an ARMA model identification procedure is proposed, which provides a best-fitting ARMA model from an estimated noise power spectral density (PSD) function. In Sect. 7, sample results of numerical experiments based on simulated GOCE data are presented.

---

## 1 Introduction

This paper is prompted by the new Gravity Field and Ocean Circulation Explorer (GOCE) mission, which will be launched in 2006. This European Space Agency (ESA) mission aims to develop a new model of the Earth's gravity field from a combination of satellite-to-satellite tracking (SST) and satellite gravity gradiometry (SGG) observations. What makes this estimation

## 2 Large LS problems

It is assumed that, after linearization, the relation between the (reduced) observations and the (reduced) unknown parameters can be written as a standard Gauss–Markov model (see e.g. Grafarend and Schaffrin 1993)

$$E\{y\} = Ax, \quad \text{rank}(A) = r, \quad D\{y\} = C, \quad \text{rank}(C) = N \quad (1)$$

where  $y \in \mathcal{R}^N$  denotes the real stochastic vector of observations, which is described by the first moment  $E\{y\}$  (expectation value) and the second moment  $D\{y\}$  (covariance matrix  $C$ ). The term  $x \in \mathcal{R}^r$  denotes the non-stochastic unknown parameter vector. It is related to the expectation  $E\{y\}$  by the non-stochastic design matrix  $A \in \mathcal{R}^{N \times r}$ . The LS estimator of  $x$  is

$$\hat{x} = (A^T C^{-1} A)^{-1} z \quad (2)$$

with the normal equation matrix  $A^T C^{-1} A$  and the right-hand-side vector  $z = A^T C^{-1} y$ .

In large LS problems (i.e. where  $N$  and  $r$  are very large, e.g. of the order of  $10^8$  and  $10^5$ , respectively) with colored observation noise (i.e. with a non-diagonal covariance matrix), the computation of the LS estimator  $\hat{x}$  poses numerical problems for two reasons.

1. The explicit computation of the design matrix  $A$  and the normal matrix  $A^T C^{-1} A$  may be impossible due to execution time and/or computer memory constraints. The memory problem could be solved by storing the matrices on a hard disk; however, this is often not a real option, because the input and output operations may take too long.
2. The covariance matrix associated with the colored noise may be fully populated. Usually, the covariances between observations decrease with increasing intervals of time (so-called lags), so the covariance matrix has a diagonal-dominant structure. Neglecting covariances after some time lag may further simplify the structure of the covariance matrix. Nevertheless, computation of the inverse matrix  $C^{-1}$  may still be impractical if the number of observations is of the order of  $10^8$ .

The first problem can be overcome when the normal equations are solved iteratively. A commonly adopted method is conjugate gradients with pre-conditioning (PCCG) (see e.g. Hestenes and Stiefel 1952). The common form of this algorithm is displayed in Fig. 1.

The major advantage of the PCCG method is that convergence is guaranteed for symmetric positive-definite matrices. Moreover, convergence may be achieved after  $i \ll r$  iterations, if a suitable pre-conditioner is found. From a computational point of view, the most important feature of the PCCG method is that the normal matrix does not have to be accessed directly, but only implicitly, which means that only the product of the normal matrix with a vector has to be computed. This implies that the computation of the vector  $q_i$  (see Fig. 1)

$$\begin{aligned} x_0 &= \text{initial guess (usually, } x_0 = 0) \\ r_0 &= A^T C^{-1} y - A^T C^{-1} A x_0 \\ w_0 &= N_{\text{pre}}^{-1} r_0; \quad \rho_0 = r_0^T w_0 \\ d_0 &= w_0 \\ \text{For } i &= 0, 1, 2, \dots \text{ until } r_i^T r_i < \varepsilon \\ & \quad q_i = A^T C^{-1} A d_i \\ & \quad \alpha_i = \frac{\rho_i}{d_i^T \cdot q_i} \\ & \quad x_{i+1} = x_i + \alpha_i d_i \\ & \quad r_{i+1} = r_i - \alpha_i q_i \\ & \quad w_{i+1} = N_{\text{pre}}^{-1} r_{i+1} \\ & \quad \rho_{i+1} = r_{i+1}^T \cdot w_{i+1} \\ & \quad d_{i+1} = w_{i+1} + \frac{\rho_{i+1}}{\rho_i} d_i \end{aligned}$$

**Fig. 1.** Common form of the method of preconditioned conjugate gradients (PCCG). The pre-conditioner  $N_{\text{pre}}$  is a suitable approximation of the normal matrix  $A^T C^{-1} A$

can be decomposed into three operations, which are also implicit in the above-mentioned sense: first compute  $u_i = A d_i$ , then  $v_i = C^{-1} u_i$ , and finally  $q_i = A^T v_i$ . Consequently, only the *application* to some vector of  $A$ ,  $C^{-1}$ , and  $A^T$ , respectively, has to be computed at every iteration step. This has three advantages: (1) it accelerates the LS estimation by applying fast algorithms for the computation of the vectors  $u_i$  and  $q_i$  (see Ditmar et al. submitted); (2) it saves a lot of memory since the design matrix  $A$  and the normal matrix  $A^T C^{-1} A$  are not computed explicitly; and (3) it opens the door for the efficient handling of colored observation noise using the filtering approach to be discussed in Sect. 5.

## 3 Colored noise and Toeplitz systems

It has been shown in the previous section that when the PCCG method is used to solve the normal equations, the problem of taking colored noise into account reduces to the application of  $C^{-1}$  to some vector  $u_i$  (note also that the computation of the initial residual  $r_0$  requires a similar operation). This application can be written as the solution of a system of linear equations  $C v_i = u_i$ , which has to be performed for each iteration  $i$ . From now on, the index  $i$  is omitted in order to simplify the notation and the solution of a system of linear equations

$$C v = u \quad (3)$$

is considered, with the  $N \times N$  covariance matrix  $C$ . It is assumed that  $C$  is positive definite and Toeplitz. The latter property holds if the noise is a realization of a stationary random process and if there are no gaps in the data.

A wealth of literature has been devoted to the solution of a Toeplitz system of linear equations. Maybe the most popular algorithm is due to Levinson (1946). It has a complexity of  $O(N^2)$  floating-point operations (flops). A number of methods have been proposed based on LU and QR factorizations (see e.g. Brent et al. 1980). They require  $O(N \log^2 N)$  flops, but are known to be more efficient than the Levinson algorithm only for very large  $N$ . Strang (1986) proposed an algorithm based on the PCCG method. His idea was expanded upon by many authors (see e.g. Chan and Ng 1995). The complexity can be reduced to  $O(N \log N)$  flops provided that a suitable pre-conditioner is chosen under certain conditions on the Toeplitz operator.

The solution of Toeplitz systems of linear equations has been addressed several times in the geodetic literature. Colombo (1979), Eren (1980, 1982), and Bottoni and Barzaghi (1993) discussed fast solution techniques for Toeplitz systems in the context of LS collocation. The complexity of the algorithms proposed by Colombo (1979) and Eren (1980, 1982) is of  $O(N^2)$  or  $O(N \log^s N)$ ,  $s \geq 2$ . Bottoni and Barzaghi (1993) use a PCCG method to solve a system of linear equations with a symmetric block Toeplitz matrix with Toeplitz blocks. The application of that matrix to a vector, to be done once per iteration, is performed using the Fast Hartley Transform. The complexity of the algorithm is of  $O(N \log N)$  per iteration. A Jacobi pre-conditioner is used to limit the number of iterations. Schuh (1996) proposes time-domain and frequency-domain approaches, which require  $O(N \log N)$  flops, provided that the Toeplitz matrix  $C$  is band limited or can be approximated by a band-limited Toeplitz matrix, and the bandwidth is small compared with the number of observations. A very similar approach has been proposed in mathematical literature by Jain (1978). Schuh (2000) uses a low-order ARMA model, which approximates the given inverse noise PSD function. The philosophy behind his approach is different from that of the other methods. It should not be seen as a fast method to solve a Toeplitz system of linear equations, but rather as a de-correlation method. The observations are de-correlated using the low-order ARMA filter. Correspondingly, the columns of the design matrix are also filtered. Therefore, for each iteration of the PCCG algorithm,  $r$  columns of the design matrix are filtered. This approach requires  $O(rN)$  flops, where  $r$  is the number of columns of the design matrix (i.e. the number of unknown parameters). According to Schuh (Pers. commun. 2002), the performance has recently been improved to essentially  $O(N)$  flops by avoiding explicit filtering of the columns of the design matrix.

The approach pursued in this paper is not a de-correlation method, but a fast method to solve a Toeplitz system of linear equations by exploiting an ARMA representation of the colored noise. It is shown that the numerical complexity is also of  $O(N)$ , provided that an ARMA model of the colored noise is available.

#### 4 ARMA filters

The pursued approach exploits the fact that any autocovariance function that asymptotically tends to zero can be approximated arbitrarily well by the autocovariance function of an ARMA process (see e.g. Brockwell and Davis 1991). Since the main aim is the fast solution of the Toeplitz system  $Cv = u$ , the colored noise is interpreted as one realization of an ARMA process. The corresponding ARMA model is then exploited to solve the system  $Cv = u$  efficiently.

The process  $\{\xi_n, n \in \mathcal{Z}\}$  is said to be an ARMA( $p, q$ ) process if  $\{\xi_n\}$  is stationary and if for every  $n \in \mathcal{Z}$

$$\xi_n = - \sum_{k=1}^p a_{p,k} \xi_{n-k} + \sum_{i=0}^q b_{q,i} \varepsilon_{n-i}, \quad b_{q,0} = 1, \quad n \in \mathcal{Z} \quad (4)$$

where  $\{\varepsilon_n\}$  is a white-noise process with a zero mean and a variance of  $\sigma_\varepsilon^2$ , and  $\mathcal{Z}$  denotes the set of integer values (see e.g. Brockwell and Davis 1991; Stoica and Moses 1997). Often, the input sequence is assumed to be Gaussian, i.e. normally distributed. The terms  $\{a_{p,k} : k = 1, \dots, p\}$  and  $\{b_{q,i} : i = 0, \dots, q\}$  are the model parameters, and the pair  $(p, q)$  describes the order of the ARMA( $p, q$ ) process. There are two important special cases: (1) if  $p = 0$ , the process is a *moving-average* process of order  $q$  [i.e. MA( $q$ )]; and (2) if  $q = 0$ , the process is purely *autoregressive* of order  $p$  [i.e. AR( $p$ )]. Equation (4) can be written symbolically in the more compact form

$$A_p(z) \xi_n = B_q(z) \varepsilon_n, \quad n \in \mathcal{Z} \quad (5)$$

where  $A_p(z)$  and  $B_q(z)$  are the  $p$ th and  $q$ th-degree polynomials

$$A_p(z) = 1 + \sum_{k=1}^p a_{p,k} z^k \quad (6)$$

$$B_q(z) = 1 + \sum_{i=1}^q b_{q,i} z^i \quad (7)$$

The term  $z$  in Eq. (5) has to be interpreted as a shift operator, defined by  $z^s \xi_n = \xi_{n-s}$  for arbitrary  $s \in \mathcal{Z}$ . The polynomials  $A_p(z)$  and  $B_q(z)$  are often referred to as the AR and MA polynomials, respectively.

When solving Toeplitz systems by exploiting ARMA filters, the properties of causality and invertibility of the ARMA process are required. The ARMA( $p, q$ ) process  $\{\xi_n\}$  is *causal* if, for all  $n \in \mathcal{Z}$ , it can be represented as

$$\xi_n = \sum_{j=0}^{\infty} h_j \varepsilon_{n-j} = H(z) \varepsilon_n \quad (8)$$

with

$$H(z) = \sum_{j=0}^{\infty} h_j z^j = \frac{B_q(z)}{A_p(z)}, \quad |z| \leq 1 \quad (9)$$

Equation (9) is *invertible* if, for all  $n \in \mathcal{Z}$ , the inverse process can be represented as

$$\varepsilon_n = \sum_{j=0}^{\infty} l_j \zeta_{n-j} = L(z) \zeta_n \quad (10)$$

with

$$L(z) = \sum_{j=0}^{\infty} l_j z^j = \frac{A_p(z)}{B_q(z)}, \quad |z| \leq 1$$

A sufficient and necessary condition for causality and invertibility of an ARMA( $p, q$ ) process is that the zeros of  $A_p(z)$  and  $B_q(z)$ , respectively, are strictly outside the unit circle in the complex-number domain.

In the terminology of filters, Eq. (8) describes a linear time-invariant causal filter driven by the white-noise process  $\{\varepsilon_n\}$ . The transfer function of this filter is the rational polynomial  $H(z)$  [Eq. (9)]. Analogously, the invertibility allows  $\{\varepsilon_n\}$  to be expressed as output of a linear time-invariant causal filter driven by the colored-noise process  $\{\zeta_n\}$ . Its transfer function is the inverse of  $H(z)$  [cf. Eq. (9)].

The PSD function of the process  $\{\zeta_n\}$  [Eq. (4)], is given by

$$P_{\text{ARMA}}(f) = \sigma_e^2 \Delta t \frac{|B_q(\exp(-j2\pi f \Delta t))|^2}{|A_p(\exp(-j2\pi f \Delta t))|^2}, \quad -\frac{1}{2\Delta t} \leq f \leq \frac{1}{2\Delta t} \quad (11)$$

with  $j = \sqrt{-1}$ ;  $f$  is the frequency in units of Hz and  $\Delta t$  is the sampling interval in the time domain in units of seconds. The inverse discrete Fourier transform of Eq. (11) provides a sufficiently accurate approximation of the autocovariance sequence (ACS) of the process  $\{\zeta_n\}$  provided that the PSD is sampled sufficiently well.

## 5 ARMA filters and Toeplitz systems

Here it is assumed that the colored noise is modeled as a causal invertible ARMA( $p, q$ ) process. In Sect. 6, a method is proposed to obtain such a model from a noise PSD estimate. Moreover, it is assumed that the vector  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)^T$  contains  $N$  random numbers that represent a white-noise realization. This vector can be filtered by the linear shift-invariant causal filter with transfer function  $H(z)$  [Eq. (9)]. The output of the filter is a vector  $\xi = (\xi_1, \dots, \xi_N)^T$ . The  $q$  ‘past’ input values  $\varepsilon_{1-q}, \dots, \varepsilon_0$  and the  $p$  ‘past’ output values  $\xi_{1-p}, \dots, \xi_0$  are unknown. If they are set equal to zero, the filter output  $\xi$  can be written as

$$\xi = G\varepsilon, \quad G := A^{-1}B$$

where

$$A = \text{toeplitz}(1, a_{p,1}, a_{p,2}, \dots, a_{p,p}, 0, \dots, 0)$$

$$B = \text{toeplitz}(1, b_{q,1}, b_{q,2}, \dots, b_{q,q}, 0, \dots, 0)$$

are  $N \times N$  lower triangular matrices. The inverse of  $A$  exists since the filter is causal. The covariance matrix of the vector  $\varepsilon$  is  $\sigma_e^2 I$ , where  $I$  is the  $N \times N$  unit matrix.

Thus, the covariance matrix of the vector  $\xi$  is  $\tilde{C} = \sigma_e^2 G G^T$ . In general, the matrix  $\tilde{C}$  is not Toeplitz. It is only an approximation to the Toeplitz covariance matrix  $C$  of the colored observation noise. The difference is caused by the initialization of the filter, when past input and output values are set equal to zero. However, if the filter is causal and invertible, only the elements of the  $m \times m$  left upper submatrix of  $\tilde{C}$ , with some  $m \leq p + q$ , may differ significantly from the corresponding submatrix of the covariance matrix  $C$ . If  $N \gg p + q$ , most elements of  $\tilde{C}$  are very close to the elements of the covariance matrix  $C$ .

The matrix  $G$  is a lower triangular Toeplitz matrix (Gardner 1988; Hayes 1996). Since every Toeplitz matrix is also per-symmetric (i.e. symmetric about the anti-diagonal), it holds that  $G^T = X G X$ , where  $X$  is the exchange matrix (i.e. the matrix that is equal to the unit matrix but with the columns in reversed order). Therefore, the matrix  $\tilde{C}$  can be written as

$$\tilde{C} = \sigma_e^2 G X G X$$

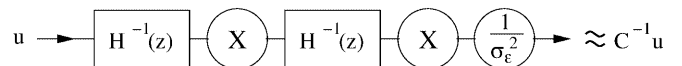
This allows the approximate solution of  $Cv = u$  to be written as

$$\tilde{v} = \frac{1}{\sigma_e^2} X G^{-1} X G^{-1} u$$

where the equality  $X = X^{-1}$  has been used. Note that the application of the matrix  $X$  to a vector means ‘flipping’ this vector (i.e. the last element of the vector becomes the first element of the flipped vector, the second last element becomes the second element, etc). Thus, the approximation  $\tilde{v}$  of  $v$  can be obtained in the following five steps (cf. Fig. 2):

- Step 1. Filter  $u$  [filter with transfer function  $H^{-1}(z)$ ].
- Step 2. Flip result of step 1 (apply exchange matrix  $X$ ).
- Step 3. Filter result of step 2 [filter with transfer function  $H^{-1}(z)$ ].
- Step 4. Flip result of step 3 (apply exchange matrix  $X$ ).
- Step 5. Scale result of step 4 with  $1/\sigma_e^2$ .

The complexity of one filter operation with an ARMA filter of order  $(p, q)$  is  $(p + q)N$  flops. Thus, the overall complexity of computing  $\tilde{v}$  is  $2(p + q)N$  flops. The order  $(p, q)$  of the ARMA filter is independent of  $N$ . According to the authors’ experience, many noise PSD functions of measurement sensors are described very well by low-order ARMA models. For instance, the expected noise PSD function of the GOCE gradiometer can be represented by an ARMA model of order  $p + q < 100$ . Therefore, it is reasonable to assume that when dealing with large LS problems  $p + q \ll N$ . Therefore, the complexity of the presented algorithm is essentially



**Fig. 2.** Approximate computation of  $C^{-1}u$  using ARMA filtering.  $H(z)$  is the transfer function of the ARMA filter [Eq. (9)] and  $X$  is the exchange matrix

$O(N)$ , provided that an ARMA model of the colored noise is available.

## 6 ARMA model identification using the noise PSD function

The procedure of handling colored noise presented in Sect. 5 relies on an ARMA representation of that colored noise. Such a representation is usually obtained from a noise realization. The corresponding procedure is known as model identification and belongs to the standard problems in time-series analysis. The reader is referred to, for example, Kay and Marple (1981), Brockwell and Davis (1991), Choi (1992), and Stoica and Moses (1997).

Unfortunately, there are situations where a noise realization may not be available. For instance, in the framework of ESA's GOCE project, the prime industrial contractor (ALENIA Spazio) has only provided information about the expected performance of the gradiometer on board GOCE in terms of an estimated noise PSD function. Time-series model identification from an estimated noise PSD function has not yet been considered in the literature. Therefore, a procedure for ARMA model identification from an estimated noise PSD function is developed here. However, recall that a noise PSD function contains much less information than a noise realization, and the pursued method of model identification should only be applied if a noise realization is not available. For more details on building the optimal filter, the reader is referred to Klees and Broersen (2002).

In this section, the notation ARMA( $p, q$ ) is used when we do not want to distinguish between AR, MA, or ARMA models. If both  $p \neq 0$  and  $q \neq 0$ , the notation ARMA( $p, q$ ) is used.

### 6.1 Long AR model, residual variance and order selection

The starting point of the model identification from an estimated noise PSD function is the ACS. It can be computed by an inverse discrete Fourier transform of the sampled noise PSD function according to

$$r_n = \Delta f \sum_{k=0}^{N_f-1} P_k \exp(i2\pi kn/N_f), \quad 0 \leq n \leq L \quad (12)$$

where  $r_n$  is the autocovariance for lag  $n$ ,  $P_k$  is the noise PSD function sampled at frequency  $f_k = k\Delta f$ ,  $N_f$  is the number of samples in the frequency domain (i.e. the length of the noise time series used in the estimation of the PSD function),  $\Delta f = (1/N_f\Delta t)$  is the frequency resolution,  $\Delta t$  is the sampling interval in the time domain, and  $L$  is the maximum lag to be computed.  $N_f$  must be known. The parameter  $L$  should not exceed  $N_f/2$  because it holds that

$$E\{r_n\} = \frac{N_f - n}{N_f} R_n + \frac{n}{N_f} R_{N_f-n}, \quad 0 \leq n \leq L \leq N_f - 1$$

where  $R_n$  is the true autocovariance for lag  $n$ . This shows that the autocovariance for lag  $n$  computed by an inverse discrete Fourier transform is the weighted sum of the true autocovariances  $R_n$  and  $R_{N_f-n}$ . That is, the estimated autocovariance is always biased and often deformed by tapering and windowing operations applied to the periodogram estimate of the PSD function. There are no strict rules on how to choose  $L$ . It was found empirically that  $L$  may be much less than the upper bound  $N_f/2$  if it is verified that it remains greater than the selected AR order  $\hat{p}_{\text{AR}}$  (cf. Sect. 6.2.1) and the required intermediate order  $M$  of the long AR model  $C_M(z)$ , which is used for the MA and ARMA computations (cf. Sects. 6.2.2. and 6.2.3).

The next step is to transform the ACS  $\{r_n\}$  to a long AR model,  $C_L(z)$ , of order  $L$ . The simplest way to do this is to solve the Yule–Walker equations [see e.g. Stoica and Moses 1997]

$$r_m + \sum_{i=1}^L c_{L,i} r_{m-i} = 0, \quad m = 1, \dots, L \quad (13)$$

This computation can easily be performed with the Levinson–Durbin recursion in  $O(L^2)$  flops (cf. Kay and Marple 1981). The long AR model  $C_L(z)$  can be seen as an intermediate stage to estimate and select AR, MA, and ARMA models.

Model identification includes the selection of type and order of a model. AR, MA, and ARMA are the types of model to be considered here. Order selection is based on the reduction of the logarithm of the residual variance as a function of the model order, with an additional penalty for every estimated model parameter (see e.g. Choi 1992). The residual variance is the variance of  $\hat{\varepsilon}_n = \hat{A}_p/\hat{B}_q \xi_n$ , where  $\hat{A}_p$  and  $\hat{B}_q$  denote the AR and the MA part, respectively, of the estimated AR–MA model. When only the noise PSD function is available, but not a realization of the random process  $\{\xi_n\}$ , the residual variance cannot be computed. However, it is possible to compare the different AR, MA, and ARMA models with the very long AR model  $C_L(z)$ . When  $A_p(z)$  and  $B_q(z)$  denote the corresponding parts of the true AR–MA process, the random process  $\{\eta_n\}$  is defined by

$$\hat{A}_p(z)\xi_n = \hat{B}_q(z)\eta_n$$

This random process can be related to the unknown white-noise process  $\{\varepsilon_n\}$ , defined by

$$A_p(z)\xi_n = B_q(z)\varepsilon_n$$

using the long AR model  $C_L(z)$ . Assuming  $C_L(z) \approx A_p(z)/B_q(z)$  gives

$$C_L(z)\xi_n \approx \varepsilon_n$$

thus

$$\eta_n = \frac{\hat{A}_p(z)}{\hat{B}_q(z)} \xi_n \approx \frac{\hat{A}_p(z)}{\hat{B}_q(z)C_L(z)} \varepsilon_n$$

That is, the random process  $\{\eta_n\}$  can approximately be expressed as filtered version of the unknown white-noise

process  $\{\varepsilon_n\}$ . The ratio of the output and input variances of this filter,  $\sigma_\eta^2/\sigma_\varepsilon^2$ , can be computed as the power gain of the filter with MA part  $\hat{A}_p(z)$  and AR part  $\hat{B}_q(z)C_L(z)$ , i.e. without knowledge of  $\eta_n$  and  $\varepsilon_n$ . This ratio can be substituted in any order selection criterion that is based on the logarithm of the residual variance.

The order selection criterion GIC (Broersen 2000) is preferred, because it is known to often perform better than other selection criteria proposed in the literature, in particular for small  $N_f$ . Replacing the residual variance by the ratio derived before, the GIC criterion reads

$$\text{GIC} = \ln \left( \frac{\sigma_\eta^2}{\sigma_\varepsilon^2} \right) + 3 \frac{m+1}{N_f} \quad (14)$$

where  $m = \hat{p}$  for an estimated AR( $\hat{p}$ ) model,  $m = \hat{q}$  for an estimated MA( $\hat{q}$ ) model, and  $m = \hat{p} + \hat{q}$  for an estimated ARMA( $\hat{p}, \hat{q}$ ) model. The model with the smallest criterion GIC [Eq. (14)] is the selected model.

## 6.2 Model identification

Model identification from an estimated noise PSD function is based on the computation of a set of AR, MA, and ARMA models up to a maximum order, which has to be chosen by the user. For each model type, the selected model order is the one that gives the smallest GIC [Eq. (14)]. Once a single AR, MA, and ARMA model has been selected, the model with the smallest value of the criterion of Eq. (14) is the final single selected model.

The procedures for estimating AR, MA, and ARMA models are different. These are the subjects of Sects. 6.2.1, 6.2.2, and 6.2.3.

### 6.2.1 AR estimation

Given the ACS [Eq. (12)], AR( $p$ ) models are estimated for  $p = 1, 2, \dots, p_{\max}$  by solving the Yule–Walker equations with the Levinson–Durbin recursion. For each candidate order  $p$ , the residual variance ratio  $\sigma_\eta^2/\sigma_\varepsilon^2$  is computed as the gain of the ARMA model with  $\hat{A}_p(z)$  as the MA part and  $C_L(z)$  as the AR part. The order selection uses the criterion in Eq. (14) with  $m = p$ .

### 6.2.2 MA estimation

Given the ACS [Eq. (12)], MA( $q$ ) models are estimated for  $q = 1, 2, \dots, q_{\max}$ . First, a long intermediate AR model  $C_M(z)$  is computed for each order  $q$  by solving the Yule–Walker equations. The order  $M$  is selected equal to  $M = 2\hat{p}_{\text{AR}} + q$ , where  $\hat{p}_{\text{AR}}$  is the order of the best AR model (Sect. 6.2.1). Then, Durbin’s MA method (Durbin 1959) is used to estimate the MA( $q$ ) model parameters,  $\hat{B}_q(z)$ . Finally, for each candidate order  $q$  the residual variance ratio  $\sigma_\eta^2/\sigma_\varepsilon^2$  is computed as the gain of the ARMA model with MA part 1 and AR part  $\hat{B}_q(z)C_M(z)$ . The order selection uses the criterion in Eq. (14) with  $m = q$ .

### 6.2.3 ARMA estimation

ARMA model identification is often performed using Durbin’s second method of ARMA parameter estimation (Durbin 1960). This method requires as input an initial solution of the AR part of the ARMA model and a long intermediate AR model. Durbin (1960) computes the initial solution of the AR part from a given noise realization. Thus, another strategy has to be developed if only an estimated noise PSD function is available. In this section  $A_p(z)$  and  $B_q(z)$  denote the AR and the MA part, respectively, of the true ARMA model.  $\hat{A}_p(z)$  and  $\hat{B}_q(z)$  denote the AR and MA part, respectively, of the ARMA model to be computed.

Four different methods are proposed with which to obtain the initial estimate of the AR part of the ARMA model. They are denoted long AR, long MA, long COV, and long RINV. All four use a long AR model  $C_M(z)$  as an intermediate model. The order of that model is selected equal to  $M = 3\hat{p}_{\text{AR}} + (\hat{p} + \hat{q})$ , where  $\hat{p}_{\text{AR}}$  is the order of the best AR model (Sect. 6.2.1). The model  $C_M(z)$  is computed from the ACS [Eq. (12)] as solution of the Yule–Walker equations using the Levinson–Durbin recursion. Since  $C_M(z) \approx A_p(z)/B_q(z)$ , a very high-order AR model must be used when the zeros of  $B_q(z)$  are near the unit circle. In this case, the parameters  $\{c_{M,k} : k = 1, \dots, M\}$  of  $C_M(z)$  will not rapidly decay with increasing  $k$ . This will usually be the case of interest, for if the zeros of  $B_q(z)$  are far from the unit circle, they will have negligible effect upon the PSD. Then, building an ARMA model is no longer necessary, as an AR model will suffice.

*Long AR.* This method is based on an idea of Graupe et al. (1975), where an ARMA( $\hat{p}, \hat{q}$ ) model with  $\hat{A}_p(z)$  and  $\hat{B}_q(z)$  as the AR and the MA part, respectively, is approximated by an AR model  $C_M(z)$  according to

$$\frac{\hat{B}_q(z)}{\hat{A}_p(z)} \approx \frac{1}{C_M(z)}, \quad \text{or } \hat{B}_q(z)C_M(z) \approx \hat{A}_p(z) \quad (15)$$

Initial MA parameters can be found with

$$c_{M,m} + \sum_{i=1}^{\hat{q}} \hat{b}_{q,i} c_{M,m-i} \approx 0, \quad m = \max(\hat{p} + 1, \hat{q}), \dots, M$$

This system of linear equations reflects the fact that the coefficients at degrees  $\hat{p} + 1, \dots, M$  in the polynomial  $\hat{B}_q(z)C_M(z)$  must be close to zero. Note that for  $M > \hat{p} + \hat{q}$ , the system is overdetermined, and the parameters are obtained by LS. Finally, the initial AR parameters are obtained from

$$\hat{A}_p(z) \approx C_M(z)\hat{B}_q(z) \quad (16)$$

*Long MA.* This method uses an estimate for the impulse response that is derived from the intermediate long AR model  $C_M(z)$  according to

$$\frac{\hat{B}_q(z)}{\hat{A}_p(z)} \approx \frac{1}{C_M(z)} = G(z) \quad (17)$$

with

$$G(z) = 1 + \sum_{i=1}^{\infty} g_i z^i$$

An approximation of the impulse response  $G(z)$  is obtained by simple long division or by a filter operation with a delta pulse as the input signal. The length  $M_0$  of the impulse response can be chosen freely, much greater than  $M$ . Thereafter, the initial AR parameters  $\hat{A}_p(z)$  are found as LS solutions of the overdetermined system of linear equations

$$g_m + \sum_{i=1}^{\hat{p}} \hat{a}_{p,i} g_{m-i} \approx 0, \quad m = \max(\hat{p}, \hat{q} + 1), \dots, M_0 \quad (18)$$

where it is assumed that the impulse response is practically zero at  $M_0$ .

*Long COV.* The third method to find initial estimates for the AR part  $\hat{A}_p(z)$  of the ARMA( $\hat{p}, \hat{q}$ ) model uses the given ACS [Eq. (12)]. The initial estimate is obtained as LS solution of the extended Yule-Walker equations (see e.g. Stoica and Moses 1997)

$$r_m + \sum_{i=1}^{\hat{p}} \hat{a}_{p,i} r_{m-i} \approx 0, \quad m = \hat{q} + 1, \dots, M \quad (19)$$

*Long RINV.* The fourth method uses inverse autocovariances, i.e. the autocovariances of a time-series model where the AR and MA parts have been exchanged (cf. Priestley 1981). The inverse autocovariances are computed with the parameters of the intermediate long AR model  $C_M(z)$  as

$$R_{\text{inv}}(k) = \sum_{i=0}^{M-k} c_{M,i} c_{M,i+k}, \quad k = 0, 1, \dots, M$$

Furthermore,  $R_{\text{inv}}(k) = 0$  for  $k > M$  and  $R_{\text{inv}}(k) = R_{\text{inv}}(-k)$ . The initial estimates for the MA parameters are calculated as LS solution of the overdetermined system of linear equations

$$R_{\text{inv}}(0) + \sum_{i=1}^{\hat{q}} \hat{b}_{q,i} R_{\text{inv}}(m-i) \approx 0, \quad m = \hat{p} + 1, \dots, M \quad (20)$$

From this initial MA estimate, the estimation of the initial AR parameters continues with Eq. (16).

Once the initial AR part  $\hat{A}_p(z)$  has been found, an algorithm similar to Durbin's (1959) MA method is used to compute the final MA part (step 1). Then, this MA model is used to improve the AR part (step 2). This procedure has to be applied to all four AR initial estimates: the true ARMA process and its AR approximation are given by

$$A_p(z)\xi_n = B_q(z)\varepsilon_n \quad (21)$$

and

$$C_M(z)\xi_n = \hat{\varepsilon}_n \quad (22)$$

respectively. Substitution of  $\xi_n$  from Eq. (21) in Eq. (22), and replacement of the true polynomials by estimated polynomials  $\hat{A}_p(z)$  and  $\hat{B}_q(z)$ , gives

$$\frac{C_M(z)}{\hat{A}_p(z)} \hat{B}_q(z) \approx 1 \quad (23)$$

Equation (23) is used to find the MA parameters  $\hat{B}_q$  using the MA method of Durbin (1959). Thereafter, the final AR parameters  $\hat{A}_p(z)$  are obtained with Eq. (16). The computation of the MA part (step 1) and the AR part (step 2) can be iterated if desired. Iteration may give an improved model if the initial AR estimate is very poor.

The residual variance ratio needed in order selection is computed as the gain of the ARMA filter with MA part  $\hat{A}_p(z)$  and AR part  $\hat{B}_q(z)C_M(z)$ .

So far, ARMA model identification has been restricted to orders  $(\hat{p}, \hat{p} - 1)$  so as to limit the number of candidate models to be estimated. For each of the four proposed methods, ARMA candidate models of order  $(\hat{p}, \hat{p} - 1)$ ,  $\hat{p} = 1, \dots, \hat{p}_{\text{max}}$  are computed. The order is selected using the criterion in Eq. (14) with  $m = \hat{p} + \hat{q}$ . Finally, the best of the four is selected as the one with the smallest criterion in Eq. (14).

According to our experience, all four methods (long AR, long MA, long COV, and long RINV) always have to be used to find the best ARMA model. It was found that sometimes at least one of the four methods provided a solution that depends on the highest candidate ARMA order, which is not desirable. Moreover, in some simulations, numerical problems occurred in particular in the long AR method and the long COV method, and no useful initial estimates were obtained. On the other hand, in all simulations one of the four methods provided a suitable model. This supports the existing empirical evidence that at least one of the four methods performs well for every type of data.

## 7 Computational experiments

In this section, the results of several computational experiments will be presented and discussed. The experiments have been designed to demonstrate the performance of the proposed approach in large LS problems with colored observation noise. In addition, the effect of using a very simple AR-MA representation of the colored noise, instead of the best-fitting AR-MA representation, is investigated. The choice of a simple model may be motivated by computational considerations. It was shown in Sect. 5 that the numerical complexity increases in proportion to the product of the length of the filter and the number of observations. Thus, in large LS problems, it may be reasonable to reduce computational costs by using short filters. All computations were done on an SGI Origin 3800 parallel computer with eight processing elements. The software package GOCESOF (Ditmar and Klees 2002) was used to estimate the potential coefficients by LS.

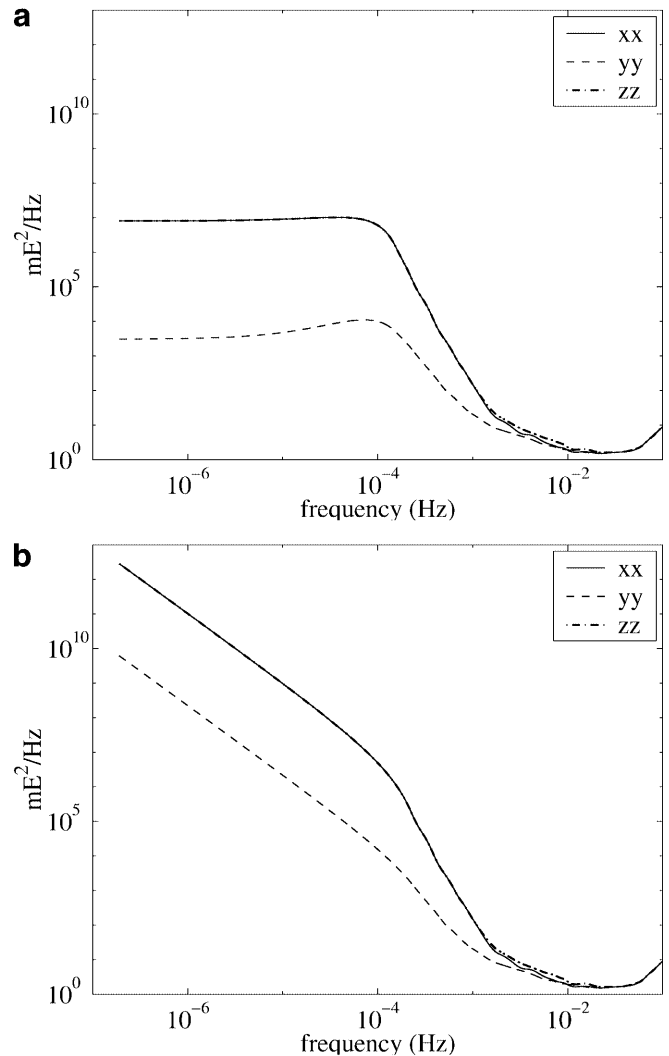
The large LS problem considered in this paper is the estimation of spherical harmonic coefficients from noisy gravity gradients. The following set-up has been chosen: along a realistic 60-day repeat GOCE orbit with a mean altitude of 268.5 km, and a mean inclination of  $96.8^\circ$ , the three diagonal components of the tensor of gravity gradients were generated with a sampling rate of 5 seconds. The difference between the OSU91A gravity field model (Rapp et al. 1991) complete up to degree and order 300 and the GRS80 Somigliana–Pizetti model (Moritz 1980) defines the disturbing potential that has been used in the simulation and that has to be estimated from the observations by LS.

The simulated gravity gradients were deliberately corrupted by colored noise according to the noise PSD functions for the  $xx$ -, the  $yy$ -, and the  $zz$ -component published in European Space Agency (ESA) (1999). Each of the curves was first discretized by 44 samples and then fitted by a cubic spline. Below  $10^{-4}$  Hz, which is the minimum frequency considered in ESA (1999), the spline was extrapolated. Two different extrapolations were used: a natural cubic spline and a  $1/f^2$  increase of the noise power. The former yields flat PSDs below  $10^{-4}$  Hz; the latter results in a strongly increasing noise power with decreasing frequency. The corresponding PSD functions for the three diagonal tensor components are shown in Fig. 3.

Next, appropriate ARMA models were determined to represent the given noise PSDs according to the procedure proposed in Sect. 6: AR(150), ARMA(33,32), and AR(163) for the  $xx$ -, the  $yy$ -, and the  $zz$ -component, respectively. Later on, these models will be referred to as the best ARMA models.

In the first experiment, it is demonstrated how the LS solution is distorted if the correlations among the observations are simply neglected. The results of the inversion in terms of geoid height errors are shown in Fig. 4. The first conclusion is that when the colored noise is not modeled, geoid height errors of the order of meters occur. Moreover, a direct comparison of the two geoid height error plots shown in Fig. 4 makes the effect of a strong noise power below a frequency of  $10^{-4}$  Hz visible (cf. Fig. 3b); such noise yields geoid height error patterns that resemble the satellite ground tracks. If the noise power at low frequencies does not increase significantly, as in Fig. 3a, geoid height errors are significantly smaller (cf. top panel in Fig. 4). This demonstrates that, for further studies, it is important to obtain information about the noise behavior of the gradiometer at low frequencies. Of course, the poor performance of the gradiometer at low frequencies may be compensated when satellite-to-satellite tracking observations are included in the estimation process, since both types of observations are complementary (cf. Ditmar and Klees 2002). Nevertheless, the measured gravity gradients may carry valuable information even at these low frequencies and proper weighting may yield improved gravity field solutions.

Figure 4 also indicates that it is difficult to predict how long-wavelength errors along the orbit are mapped geographically due to the superposition of the satellite's

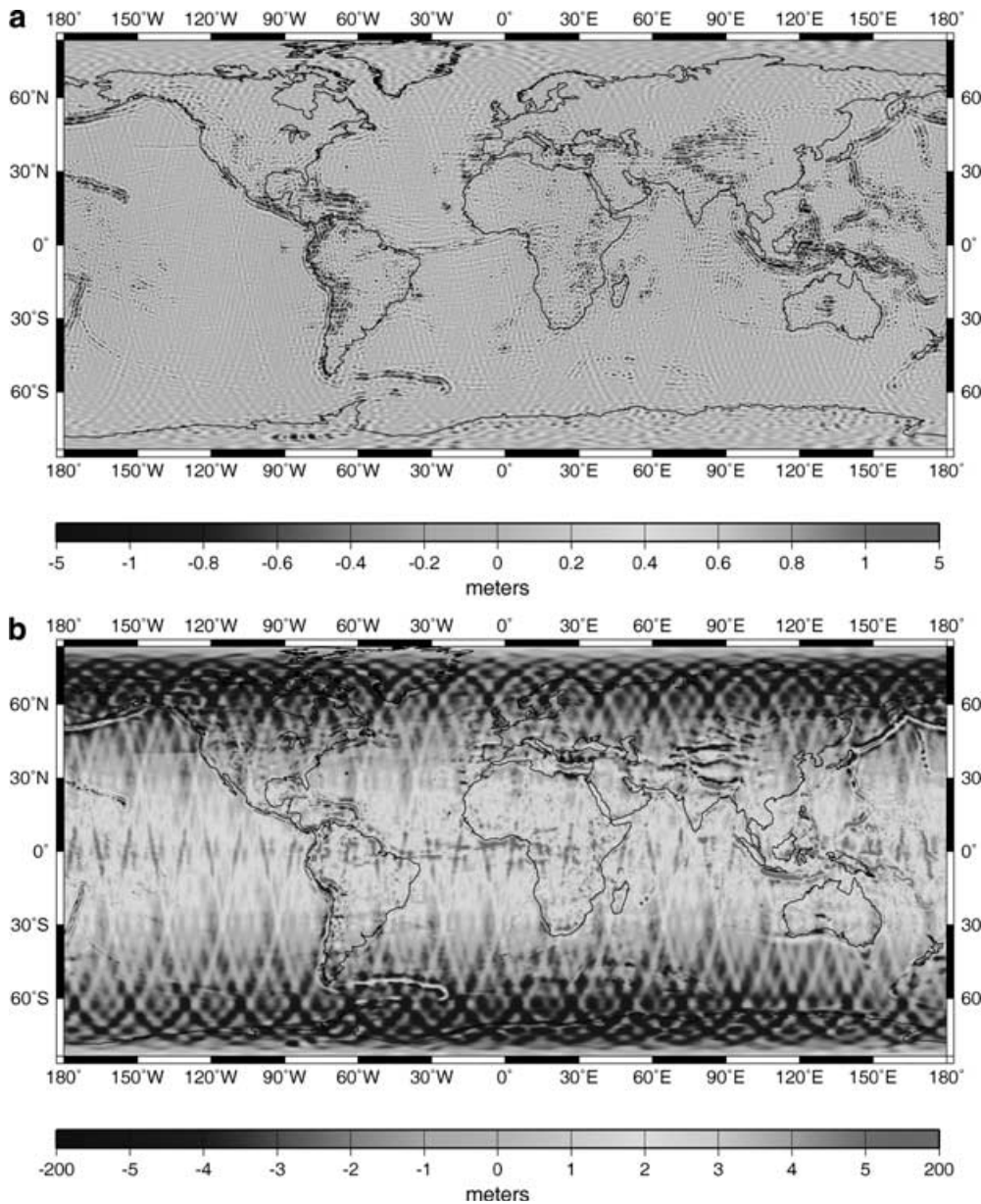


**Fig. 3a, b.** Noise PSD functions of the three diagonal gravity gradients: the PSDs correspond to the spectra of the total gravity-gradient measurement error budget for the diagonal components of the GOCE gradiometer as published in ESA (1999). Cubic splines interpolate the discretized PSDs. Below  $10^{-4}$  Hz the cubic splines were extrapolated to simulate longer time series. **a** natural cubic spline extrapolation. The total noise variances are  $\sigma_{xx}^2 = 2271.6 \text{ mE}^2$ ,  $\sigma_{yy}^2 = 4.0 \text{ mE}^2$ , and  $\sigma_{zz}^2 = 2271.6 \text{ mE}^2$  ( $1 \text{ E} = 10^{-9} \text{ s}^{-2}$ ). **b**  $1/f^2$  extrapolation with first-order continuity. The total noise variances are  $\sigma_{xx}^2 = 1\,766\,105.4 \text{ mE}^2$ ,  $\sigma_{yy}^2 = 3844.0 \text{ mE}^2$ , and  $\sigma_{zz}^2 = 1\,766\,105.4 \text{ mE}^2$ . Note that the  $1/f^2$  extrapolation yields PSDs with much more power at low frequencies compared with the natural cubic spline extrapolation

motion along the orbit and the Earth's rotation. Finally, it is noted that in both solutions the  $yy$  tensor components contribute the most to the LS solution because the total noise power is much lower [ $4.0 \text{ mE}^2$  (top) and  $3844.0 \text{ mE}^2$  (bottom) in Fig. 4] than the noise power of the  $xx$ - and  $zz$ -components [each  $2271.6 \text{ mE}^2$  (top) and  $1\,766\,105.4 \text{ mE}^2$  (bottom) in Fig. 4].

In the next experiment, the best AR–MA models representing the colored observation noise shown in Fig. 3 were used in the LS solution according to the procedure proposed in Sect. 6. The solutions for both PSDs are shown in Fig. 5; the solution presented in the



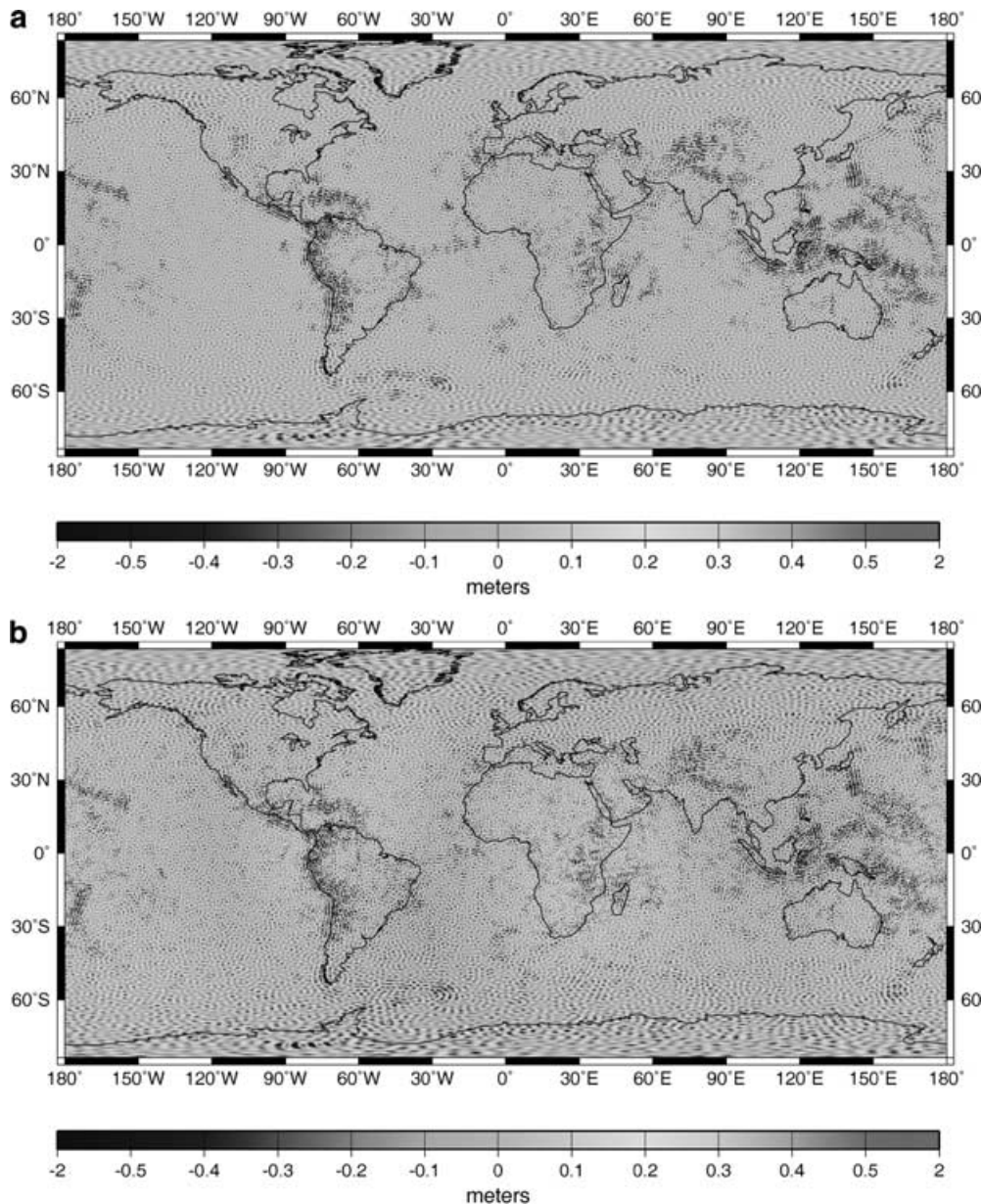


**Fig. 4a, b.** Map of geoid height errors below  $\pm 80^\circ$  latitude. The potential coefficients were estimated without modeling the colored observation noise shown in Fig. 3. *Top:* colored observation noise according to Fig. 3a. The RMS geoid height error in the  $\pm 80^\circ$  latitude band is  $\pm 0.32$  m; the maximum geoid height error is 4.9 m. *Bottom:* colored observation noise according to Fig. 3b. The RMS geoid height error in the  $\pm 80^\circ$  latitude band is  $\pm 2.4$  m; the maximum geoid height error is 16.6 m

top and bottom panels uses the PSD functions shown in Fig. 3a and Fig. 3b, respectively. The former is slightly better in terms of the global RMS (0.17 m versus 0.19 cm) due to the lower noise power at low frequencies. The geographical distribution of the errors does not show significant differences. The largest errors appear in areas with strong signal variations, such as the Andes. Two sources contribute to this effect: first, the noise at high frequencies limits the amount of detail that can be recovered from noisy gravity gradients. Secondly, Tikhonov regularization, which was applied in both cases, yields over-smoothed solutions in areas with larger signal variations. Note that, in all solutions, the regularization parameter was empirically chosen to yield the smallest RMS geoid height error in the  $\pm 80^\circ$  latitude band.

The next question addressed is the effect on the gravity field solution of a 'suboptimal' filter. 'Suboptimal' means that the colored noise is not represented by

the best ARMA model, but by a much simpler one, such as a model of lower order. For each diagonal component of the gravitational tensor, a simple ARMA(2,1) model was estimated. The results are shown in Fig. 6. Some interesting conclusions can be drawn from the two maps of geoid height errors. First of all, a direct comparison of the top panels of Figs. 5 and 6 shows that there are no significant differences between the two solutions. The RMS difference is only 0.05 m, and the maximum difference is 0.32 m. This means that if the noise power at frequencies below  $10^{-4}$  Hz does not increase, very simple ARMA filters are sufficient to model the colored observation noise. The situation is different if the noise power at frequencies below  $10^{-4}$  Hz increases. A comparison of the bottom panels of Figs. 5 and 6 shows that the simple ARMA(2,1) models of the colored noise introduce strong long wavelength errors in the geoid, which superimpose the larger local geoid errors in areas with strong signal variations. This is due to the poor



**Fig. 5a, b.** Map of geoid height errors below  $\pm 80^\circ$  latitude. The best AR–MA model for each tensor component was used to model the colored observation noise shown in Fig. 3. *Top:* noise PSD functions shown in Fig. 3a. The RMS geoid height error in the  $\pm 80^\circ$  latitude band is  $\pm 0.17$  m; the maximum geoid height error is 2.1 m. *Bottom:* noise PSD functions shown in Fig. 3b. The RMS geoid height error in the  $\pm 80^\circ$  latitude band is  $\pm 0.19$  m; the maximum geoid height error is 2.0 m

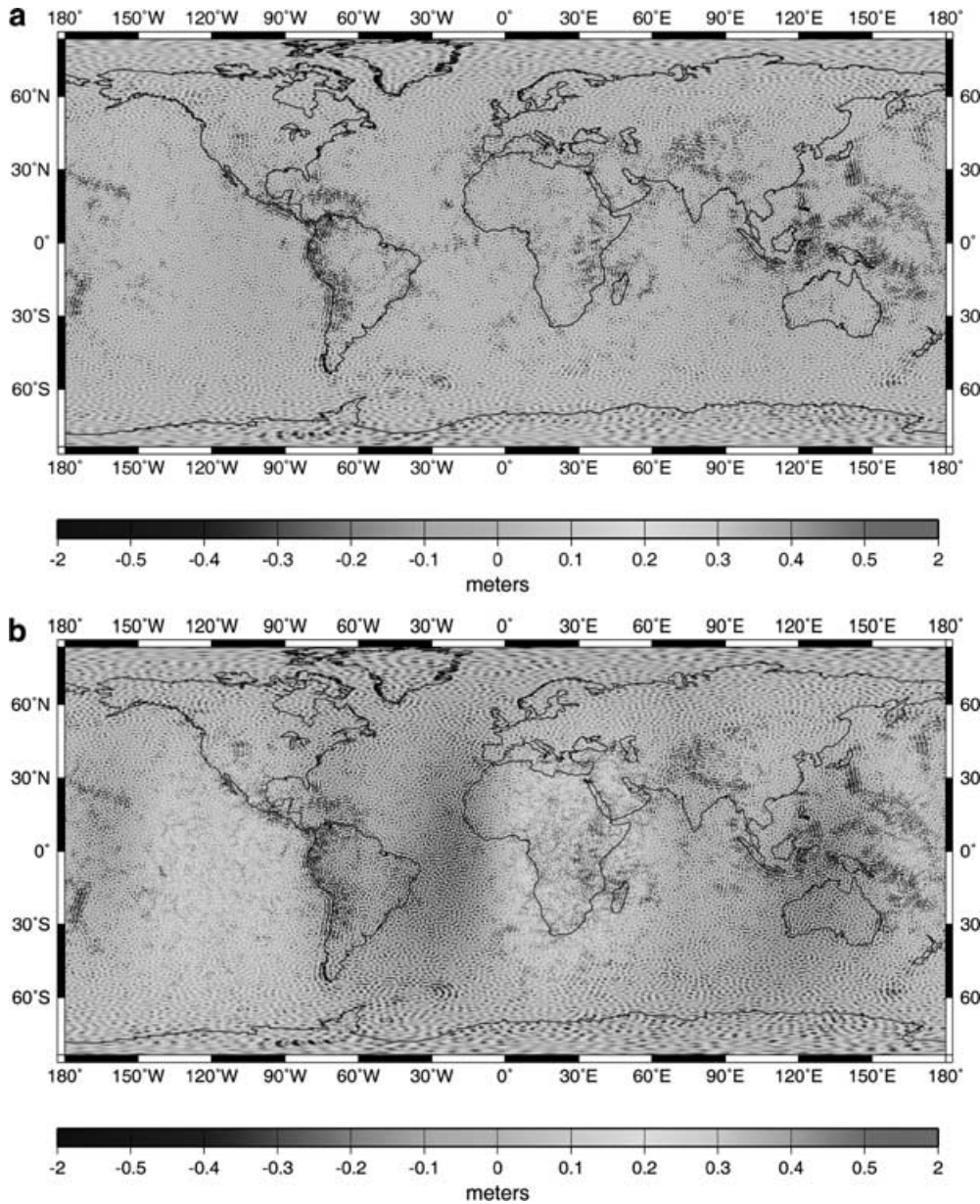
performance of these short filters at low frequencies. Therefore, proper noise modeling becomes more important in the inversion of SGG data if much noise power is concentrated at low frequencies. However, this may not be necessary in combined SGG/SST solutions.

The last aspect to be addressed is the performance of the filter operation in terms of the wall-clock time. Table 1 compares the overall wall-clock time for the LS estimation of the potential coefficients from gravity gradients and the contribution of the filter operation. It demonstrates that the proposed filter approach is very fast and contributes to the overall wall-clock time by only a few percent. This contribution is reduced even further when a simple ARMA(2,1) model is used. However, since in the GOCESoft software the contribution of the filter operation to the overall wall-clock time is minor, the use of simple filters is not necessary

from a computational point of view. The number of iterations increases noticeably when filtering is applied. There are two reasons for this. First, edge effects caused by the initialization of the filters with zero values have been suppressed by deleting the edges after filtering. Practically, this means that the orbit has been made non-repeat, which reduces the performance of the pre-conditioner used in the computations. Second, the pre-conditioner was defined using the noise model published by ESA (1999). Differences between this noise model and the one represented by the ARMA models may have reduced the performance of the pre-conditioner.

## 8 Conclusions and future work

The proposed procedure can be applied to all large LS problems in the presence of stationary colored



**Fig. 6a, b.** Map of geoid height errors below  $\pm 80^\circ$  latitude. A simple ARMA(2,1) approximation of the colored observation noise shown in Fig. 3 was used to design the whitening filter. *Top:* noise PSD functions shown in Fig. 3a. The RMS geoid height error in the  $\pm 80^\circ$  latitude band is  $\pm 0.17$  m; the maximum geoid height error is 2.1 m. *Bottom:* noise PSD functions shown in Fig. 3b. The RMS geoid height error in the  $\pm 80^\circ$  latitude band is  $\pm 0.21$  m; the maximum geoid height error is 2.2 m

**Table 1.** Performance of the filter operation in terms of wall-clock time. ‘PSD 1’ and ‘PSD 2’ refer to the PSD function shown in Fig. 3a and Fig. 3b, respectively

	No filtering	Best AR–MA filter	ARMA(2,1) filter
<i>PSD 1</i>			
Number of iterations	26	124	104
Total wall-clock time (s)	5262 (100%)	22 691 (100%)	18 469 (100%)
Filtering (s)	0	833 (3.7%)	170 (0.9%)
<i>PSD 2</i>			
Number of iterations	19	41	95
Total wall-clock time (s)	3371 (100%)	7676 (100%)	16 672 (100%)
Filtering (s)	0	168 (2.2%)	156 (0.9%)

observation noise. This includes LS collocation. Maximum efficiency is obtained when the normal equations are solved iteratively. Then, extra efforts to properly take colored noise into account require only  $O(N)$  flops. Moreover, the extra computer memory required is practically negligible. Accurate handling of colored noise is especially important if the noise intensity varies

significantly (i.e. by orders of magnitude) as a function of frequency.

However, there are some questions and situations that have not yet been addressed. First, the initialization of the filters by zero values causes edge effects, i.e. local distortions in the geoid below the location of the first and last few observations. The edge effects have been

suppressed so far by simply deleting edge values in the data vector after filtering. Of course, this is a very crude approach. Using real data, this may be not acceptable because time series of observations acquired by sensors on moving platforms often experience data gaps and spikes. A very recent example is the accelerometer on board CHAMP. Hence, a certain modification of the presented approach is necessary. This is the subject of a forthcoming paper (Klees and Ditmar, submitted).

Another aspect to be addressed is the fact that no information about the noise characteristic may be available. Then, the proposed approach cannot be used at all, because it relies on either a given realization of the noise or an estimate of the noise power spectrum. One solution to this problem might be variance component estimation techniques in combination with a suitable parameterization of the autocovariance function. This subject will be addressed in Kusche (submitted).

*Acknowledgments.* This research was supported by the Netherlands Space Research Organization (SRON) under grant eo-018 and the Netherlands Centre for High-Performance Computing (SARA) under grant SG-027. The Centre for High-Performance Computing (HP $\pi$ C) at Delft University of Technology provided additional computing resources. The authors also thank Jose van den IJssel from the Delft Institute for Earth-Oriented Space Research (DEOS) for providing the GOCE orbit data. The valuable comments of three anonymous reviewers and of the editor Will Featherstone are also acknowledged.

## References

- Bottoni GP, Barzaghi R (1993) Fast collocation. *Bull Géod* 67: 119–126
- Brent R, Gustavson F, Yun D (1980) Fast solution of Toeplitz systems of equations and computation of Pad $\epsilon$  approximants. *J Algorithm* 1: 259–295
- Brockwell PJ, Davis RA (1991) *Time series: theory and methods*. 2nd edn. Springer, New York
- Broersen PMT (2000) Finite sample criteria for autoregressive order selection. *IEEE Trans Sig Proc* 48: 3550–3558
- Chan RH, Ng MK (1995) Conjugate gradient methods for Toeplitz systems. Tech rep TR-CS-95-07, Department of Computer Science, Faculty of Engineering and Information Technology, The Australian National University, Canberra
- Choi BS (1992) *ARMA model identification*. Springer, New York
- Colombo OL (1979) Optimal estimation from data regularly sampled on sphere with applications in geodesy. Rep 291, Department of Geodetic Science, The Ohio State University, Columbus
- Ditmar P, Klees R (2002) A method to compute the Earth's gravity field from SGG/SST data to be acquired by the GOCE satellite. Delft University Press (DUP) Science, Delft
- Ditmar P, Klees R, Kostenko F (accepted) Fast and accurate inversion of satellite gravity gradiometry data. *J Geod*
- Durbin J (1959) Efficient estimation of parameters in moving average models. *Biometrika* 46: 306–316
- Durbin J (1960) The fitting of time series models. *Rev Inst Int Stat* 28: 233–243
- Eren K (1980) Spectral analysis of GEOS-3 altimeter data and frequency domain collocation. Rep 297, Department of Geodetic Science, The Ohio State University, Columbus
- Eren K (1982) Toeplitz matrices and frequency domain collocation. *Manuscr Geod* 7: 85–118
- European Space Agency (1999) Gravity field and steady-state ocean circulation mission. Reports for mission selection – the four candidate Earth Explorer Core Missions. ESA SP-1233 (1), European Space Agency, Noordwijk
- Gardner WA (1988) *Statistical spectral analysis – a nonprobabilistic theory*. Prentice Hall, Englewood Cliffs, NJ
- Grafarend EW, Schaffrin B (1993) Ausgleichsrechnung in linearen Modellen. BI Wissenschaftsverlag, Mannheim
- Graupe D, Krause DJ, Moore JB (1975) Identification of autoregressive moving-average parameters of time series. *IEEE Trans Automat Contr* AC-20: 104–107
- Hayes MH (1996) *Statistical digital signal processing and modeling*. John Wiley, New York
- Hestenes MR, Stiefel E (1952) Methods of conjugate gradients for solving linear systems. *J Res Nat Bur Stand* 49: 40–436
- Jain AK (1978) Fast inversion of banded Toeplitz matrices by circular decompositions. *IEEE Trans Acoust Speech Sig Proc* 26: 121–126
- Kay SM, Marple SL (1981) Spectrum analysis – a modern perspective. *Proc IEEE* 69: 1380–1419
- Klees R, Broersen P (2002) How to handle colored noise in large least-squares problems? – Building the optimal filter. Delft University Press (DUP) Science, Delft
- Klees R, Ditmar P (submitted) How to handle colored noise in large least-squares problems in the presence of data gaps. Submitted to Proc V Hotine–Marussi Symposium, Matera, Italy
- Kusche J (submitted) Estimating covariance parameters in gravity downward continuation. Submitted to Proc V Hotine–Marussi Symposium, Matera, Italy
- Levinson N (1946) The Wiener RMS (root mean square) error criterion in filter design and prediction. *J Math Phys* 25: 261–278
- Moritz H (1980) Geodetic reference system 1980. *Bull Géod* 54: 395–405
- Priestley MB (1981) *Spectral analysis and time series*. Academic Press, New York
- Rapp R, Wang Y, Pavlis N (1991) The Ohio State 1991 geopotential and sea surface topography harmonic coefficient models. Rep 410, Department of Geodetic Science and Surveying, The Ohio State University, Columbus
- Schuh WD (1996) Tailored numerical solution strategies for the global determination of the Earth's gravity field. Folge 81, Mitteilungen der geodätischen Institute der Technischen Universität Graz
- Schuh WD (2000) WP 3: scientific data processing algorithms. In: Suenkel H (ed) From Eötvös to MilliGal. Final rep ESA/ESTEC contract no. 13392/98/NL/GD, Graz, pp 105–156
- Stoica P, Moses R (1997) *Introduction to spectral analysis*. Prentice Hall, Upper Saddle River, NJ
- Strang G (1986) A proposal for Toeplitz matrix calculations. *Stud Appl Math* 74: 171–176