



Delft University of Technology

Sparse DDK

A Data-Driven Decorrelation Filter for GRACE Level-2 Products

Qian, Nijia; Chang, Guobin ; Ditmar, Pavel; Gao, Jingxiang ; Wei, Zhengqiang

DOI

[10.3390/rs14122810](https://doi.org/10.3390/rs14122810)

Publication date

2022

Document Version

Final published version

Published in

Remote Sensing

Citation (APA)

Qian, N., Chang, G., Ditmar, P., Gao, J., & Wei, Z. (2022). Sparse DDK: A Data-Driven Decorrelation Filter for GRACE Level-2 Products. *Remote Sensing*, 14(12), Article 2810. <https://doi.org/10.3390/rs14122810>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



Article

Sparse DDK: A Data-Driven Decorrelation Filter for GRACE Level-2 Products

Nijia Qian ^{1,2} , Guobin Chang ^{1,*}, Pavel Ditmar ², Jingxiang Gao ¹ and Zhengqiang Wei ¹

¹ School of Environment Science and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China; nijiaqian@cumt.edu.cn (N.Q.); jxgao@cumt.edu.cn (J.G.); zhengqiangwei@cumt.edu.cn (Z.W.)

² Department of Geoscience and Remote Sensing, Delft University of Technology, 2628 CN Delft, The Netherlands; p.g.ditmar@tudelft.nl

* Correspondence: guobinchang@cumt.edu.cn; Tel.: +86-150-2202-7167

Abstract: High-frequency and correlated noise filtering is one of the important preprocessing steps for GRACE level-2 products before calculating mass anomaly. Decorrelation and denoising kernel (DDK) filters are usually considered as such optimal filters to solve this problem. In this work, a sparse DDK filter is proposed. This is achieved by replacing Tikhonov regularization in traditional DDK filters with weighted L1 norm regularization. The proposed sparse DDK filter adopts a time-varying error covariance matrix, while the equivalent signal covariance matrix is adaptively determined by the Gravity Recovery and Climate Experiment (GRACE) monthly solution. The covariance matrix of the sparse DDK filtered solution is also developed from the Bayesian and error-propagation perspectives, respectively. Furthermore, we also compare and discuss the properties of different filters. The proposed sparse DDK has all the advantages of traditional filters, such as time-varying, location inhomogeneity, and anisotropy, etc. In addition, the filtered solution is sparse; that is, some high-degree and high-order terms are strictly zeros. This sparsity is beneficial in the following sense: high-degree and high-order sparsity mean that the dominating noise in high-degree and high-order terms is completely suppressed, at a slight cost that the tiny signals of these terms are also discarded. The Center for Space Research (CSR) GRACE monthly solutions and their error covariance matrices, from January 2004 to December 2010, are used to test the performance of the proposed sparse DDK filter. The results show that the sparse DDK can effectively decorrelate and denoise these data.

Keywords: GRACE; DDK filter; L1-norm regularization; mass anomaly



Citation: Qian, N.; Chang, G.; Ditmar, P.; Gao, J.; Wei, Z. Sparse DDK: A Data-Driven Decorrelation Filter for GRACE Level-2 Products. *Remote Sens.* **2022**, *14*, 2810. <https://doi.org/10.3390/rs14122810>

Academic Editor: Deodato Tapete

Received: 13 April 2022

Accepted: 9 June 2022

Published: 11 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Global climatic and environmental problems, such as sea-level rise, polar ice sheet and glacier melting, drought, and flood disasters, are closely related to the mass redistribution at the Earth's surface [1,2]. The GRACE project has been successfully applied to retrieve the mass transportation within/between different earth systems, including solid earth, cryosphere, ocean, terrestrial water, etc. However, there are always inevitable high-frequency and stripping errors in unconstrained GRACE level-2 products. These are caused by an anisotropic mission sensitivity associated with a specific orbit and set-up of the mission. As a result, filtering of unconstrained GRACE level-2 solutions is a requisite step before calculating mass transportation.

DDK filters represent a class of de-stripping filters that can reduce both high-frequency and stripping errors effectively. Those filters are designed from a regularization/inversion perspective [3–7]. They are different from empirical decorrelation filters [8–10], which also find wide applications. There are also many other filters, such as [11–18], but these are out of the scope of this study since we mainly focus on DDK-like filters in this work. Usually, two types of input are necessary to design a DDK filter, namely the noise covariances and

the signal covariances. Several different variants of DDK filters have been proposed in the literature, and they are distinguished by their specific selections of these two inputs.

A full error covariance matrix is vital for a filter to be able to de-stripe a solution. This is emphasized in [3], where the DDK filter is designed from the perspective of Wiener filtering. Note that the Wiener-filter perspective was also followed earlier in [19]. However, a diagonal error covariance matrix was employed there rather than a full one, which is exactly the reason why that filter could not destripe solutions efficiently [19]. At the time when the DDK filtering method was proposed, the full error covariance matrices were generally inaccessible for most users, and hence, the computation of an approximate full error covariance matrix from publicly available data was proposed in [5]. It is fortunate that, nowadays, the full error covariance matrices are becoming available from more and more processing centers [6].

The signal covariance matrices are often different for different variants of DDK filters, either in their structures or in the sources of input information. With regard to structures, a diagonal signal covariance matrix in the spectral domain was defined in [4–6], while a diagonal structure in the spatial domain was followed in [3]. The spectral-domain covariance matrix propagated from a diagonal spatial-domain covariance matrix is generally full. It is stated that neglecting the off-diagonal elements of such a full covariance matrix can result in visible differences in the filtered solutions compared to the case when the off-diagonal elements are preserved. However, we think this difference does not necessarily mean that using the full signal covariance matrix in the spatial domain results in a better performance. Concerning the sources of input information, the signal covariances can be obtained from prior geophysical models [4,5] or from the data themselves [3]. The latter approach can be called data-driven or data adaptive; it can be viewed as a variance and covariance component estimation. One of the main potential merits of data adaptive methods is that they cannot be misled by inadequate prior models. On the other hand, it is found in [6] that there is no significant difference between the results obtained, no matter whether the signal covariance matrix is from a prior model or from the data themselves.

In this work, we propose a new DDK scheme for filtering GRACE level-2 geopotential Spherical Harmonic (SH) coefficients. The proposed DDK filter assumes a diagonal structure of the signal covariance matrix in the spectral domain, which is similar to [4–6]. The proposed filter is also data adaptive, similar to [3]. The filter is obtained simply by replacing the L2 norm (of the SH coefficients vector) in Tikhonov regularization with the weighted L1 norm. A linear regression problem constrained with L1-norm regularization is nothing but the famous Lasso problem [20], widely known in the statistical community [21]. This is still a convex problem [22], and the convexity of the L1-norm is exactly the reason why it is widely used compared to other sparse regularizations. The applications cover many different fields, such as statistics, machine learning, and signal processing. It is pointed out that the proposed approach can be understood as a DDK filter with diagonal signal covariance matrices (in the spectral domain) automatically determined from the data. This interpretation follows from viewing the L1-norm as a diagonally weighted L2-norm, which will be introduced further in the next section. In addition, the proposed sparse DDK filter employs the monthly time-varying full error covariance matrices, which have been made available by more and more processing centers. This time-varying covariance information has been proved to be superior to a stationary one [6].

As a byproduct, the filtered solutions (namely, the geopotential SH coefficients) obtained with the proposed approach are sparse. This is the reason why we call our method the sparse DDK filter. A sparse vector is a vector with some of its elements being exactly zero. The reason why the solution can be sparse is briefly explained below. The data to be processed (filtered), namely the SH coefficients, are residual variables, which are obtained by subtracting the SH coefficients provided by processing centers from their corresponding reference values, e.g., a long-term mean or linear trend. Therefore, it is not unnatural that some of the residual SH coefficients are zeros or have small magnitudes. This means that vectors composed of these SH coefficients are compressible. Then, it is feasible to employ

a sparse model (namely the sparse filtered SH solution) to fit such data. Furthermore, a sparse model is also beneficial in terms of model generalization, just because it is simple, and a simpler model is often better from the viewpoint of model selection or statistical learning [21,23]. Note that besides the filtered solution mentioned above, a DDK filter itself, namely the filtering matrix F_0 defined in Equation (5) in the following section, is also sparse (or can be made sparse), with some coefficients equal to zero (in the spectral domain). This sparsity is due to the (approximate) independence between different orders/parities [4]. This (approximate) independence is exactly the starting point to design the empirical decorrelation filters [8], as well as the starting point to simplify the DDK filters [4].

The paper is organized as follows. After this introduction section, the methodology is presented in Section 2, including an introduction to the Lasso problem, the algorithm employed to iteratively compute the filtered geopotential SH coefficients, the method to select appropriate hyperparameters (regularization parameters), and a variance/covariance analysis. In Section 3, the relevant regularization strategies are discussed and compared, and the properties of the proposed sparse DDK filter are discussed, with comparisons to other filtering approaches. Data analysis is presented in Section 4, including signal and noise evaluation for both global and regional cases, followed by the conclusion and outlook in Section 5.

2. Materials and Methods

Let us denote the monthly residual geopotential SH coefficients (with a long-term mean removed) as a vector x . The corresponding time-varying full error covariance matrix, which is assumed to be available, is denoted as Q . The task is designing a DDK-type filter, denoted as F_0 , and computing the filtered solution as $\hat{x} = F_0 x$. No other data are needed. The other input needed to design a DDK filter, namely the signal covariance matrix (in the spectral domain here), denoted as S , is implicitly obtained from the available data, namely x together with Q .

2.1. A Brief Introduction to the Lasso Problem

Assume there is a measurement model expressed as follows:

$$y = B\beta + e, \quad (1)$$

where y , B , β and e are the measurement vector, design matrix, parameter vector to be estimated, and noise vector, respectively. The covariance matrix of measurement noise e is denoted as Q . For the sparse L1-norm regularized case, the estimation is defined as the following:

$$\hat{\beta} = \underset{\hat{\beta}}{\operatorname{argmin}} \left[\left(y - B\hat{\beta} \right)^T Q^{-1} \left(y - B\hat{\beta} \right) + \mu \|\hat{\beta}\|_1 \right], \quad (2)$$

where μ is a regularization parameter. The estimate $\hat{\beta}$ is nothing but the Lasso solution [20]. Prior constraints on the solution are imposed by including the regularization term $\mu \|\hat{\beta}\|_1$, which is similar to that in the conventional DDK filters [5]. However, the L1 norm is employed in this regularization, which is different from the L2 norm in the Tikhonov regularization used in conventional DDK filters. The fitting error term $\left(y - B\hat{\beta} \right)^T Q^{-1} \left(y - B\hat{\beta} \right)$ implies the Gaussian distribution of measurement errors. The Lasso solution can be viewed as a Bayesian solution with a Laplace prior rather than a Gaussian prior implied in conventional DDK filters [24]. The relevance is within the GRACE SH coefficients, making it non-Gaussian, so we adopt a Laplace prior to constrain the magnitudes of the parameters rather than constraining its energy, as implied by a Gaussian prior. It should be noted the regularization term $\mu \|\hat{\beta}\|_1$ may appear in the form of $\mu \|W\hat{\beta}\|_1$ (with W being regularization matrix) in some textbooks [25], and a problem involving $\mu \|W\hat{\beta}\|_1$ is usually called weighted Lasso. As we will show in the following subsection, a weighted Lasso problem

can be converted into a standard one as presented in (2). For simplicity, we only show the standard Lasso in this subsection.

The Lasso solution in Equation (2) is calculated with an efficient algorithm called FISTA [26], which will be briefly introduced in the following subsection. The regularization coefficient μ provides a tradeoff between the fitting term $(\mathbf{y} - \mathbf{B}\hat{\boldsymbol{\beta}})^T \mathbf{Q}^{-1}(\mathbf{y} - \mathbf{B}\hat{\boldsymbol{\beta}})$ (needed to extract signal) and the smoothing term $\|\hat{\boldsymbol{\beta}}\|_1$ (needed to suppress noise in the measurement vector $\mathbf{y}$). An appropriate value for μ is selected objectively according to a tailored variant of Generalized Cross-Validation (GCV), which is presented in the following subsection. The solution is sparse so that some elements of $\hat{\boldsymbol{\beta}}$ are exactly zero; this means that the Lasso method is also playing the role of model coefficient selection, in an automatic way. The tuning of the regularization parameter in Lasso implies that the L1-norm regularization can produce solutions with different degrees of smoothing/denoising and then perform model coefficient selection automatically (according to GCV), namely to balance fitting and smoothing appropriately. This property can facilitate the appropriate separation of signal and noise in DDK filtering, that is, an automatic extraction of the real mass anomaly signal from noisy monthly solutions.

2.2. A Sparse DDK Filter with Weighted Lasso

The sparse DDK filter is designed based on a weighted Lasso, a special variant of the so-called adaptive Lasso in [27]. In order to approximate the attenuation of gravity field signal with increasing degree, we adopt a power law to construct a weighting matrix, namely $\bar{\mathbf{W}}^{-1} = \bar{\mathbf{R}} = \text{diag}[l^{-p}]$, with l being the SH degree and the power index p taken as four empirically [5]. This weighting matrix is then introduced to better approximate the mass anomaly signal (in the spectral domain), so the sparse DDK filtering solution with weighted Lasso is a tailored version of Equation (2):

$$\hat{\mathbf{x}}_0 = \underset{\hat{\mathbf{x}}}{\operatorname{argmin}} \left[(\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{Q}^{-1}(\mathbf{x} - \hat{\mathbf{x}}) + \mu \|\bar{\mathbf{W}}\hat{\mathbf{x}}\|_1 \right]. \quad (3)$$

where the design matrix $\mathbf{B}$ in Equation (2) is an identity matrix here. The weighted Lasso solution $\hat{\mathbf{x}}_0$ is exactly the filtered solution of our sparse DDK filter. We can simply calculate this solution by solving the following standard Lasso problem:

$$\hat{\mathbf{x}}_0 = \bar{\mathbf{W}}^{-1} \hat{\mathbf{z}}_1 = \bar{\mathbf{R}} \left\{ \underset{\hat{\mathbf{z}}}{\operatorname{argmin}} \left[(\mathbf{x} - \bar{\mathbf{R}}\hat{\mathbf{z}})^T \mathbf{Q}^{-1}(\mathbf{x} - \bar{\mathbf{R}}\hat{\mathbf{z}}) + \mu \|\hat{\mathbf{z}}\|_1 \right] \right\}. \quad (4)$$

The standard Lasso problem involved in the weighted Lasso solution in Equation (4) can be abstracted as Equation (2). The matrix $\mathbf{B}$, the measurement vector $\mathbf{y}$, and the parameter vector $\boldsymbol{\beta}$ in Equation (2) correspond to the weighting matrix $\bar{\mathbf{R}}$, the unfiltered SH coefficient vector $\mathbf{x}$ and the transformed unfiltered SH coefficient vector $\hat{\mathbf{z}} = \bar{\mathbf{W}}\hat{\mathbf{x}}$ in Equation (4), respectively. For the sparse DDK filtered SH solution $\hat{\mathbf{x}}_0$ in Equation (4), we firstly solve a standard Lasso problem for the transformed variable $\hat{\mathbf{z}}_1 = \bar{\mathbf{W}}\hat{\mathbf{x}}$, and then transform the solution back to the original variable, namely $\hat{\mathbf{x}}_0 = \bar{\mathbf{R}}\hat{\mathbf{z}}_1$. The FISTA algorithm for the standard Lasso solution $\hat{\mathbf{z}}_1$, and the tailored GCV criterion for the selection of the hyperparameter μ , are introduced in the following Section 2.3.

After obtaining the sparse DDK filtered solution with Equation (4), it is also meaningful to construct an equivalent DDK filtering matrix (or an equivalent filter) from this solution so as to analyze the properties of a sparse DDK filter. This is realized by viewing the L1-norm regularized solution in sparse DDK filter as a weighted L2-norm regularized solution. To be specific, let $\mathbf{W}_0^{-1} = \mathbf{R}_0 = \text{diag}[\|\hat{\mathbf{z}}_1\|]$, then $\|\hat{\mathbf{z}}_1\|_1 = \hat{\mathbf{z}}_1^T \mathbf{W}_0 \hat{\mathbf{z}}_1$. This can be proved as follows. $\|\hat{\mathbf{z}}_1\|_1$ is the sum of the absolute values of all elements in vector $\hat{\mathbf{z}}_1$. In $\hat{\mathbf{z}}_1^T \mathbf{W}_0 \hat{\mathbf{z}}_1$, $\hat{\mathbf{z}}_1^T \mathbf{W}_0$ is a vector with all elements being ± 1 , with the sign of each element being the same as that of the corresponding element in $\hat{\mathbf{z}}_1$ (because $\mathbf{W}_0^{-1} = \text{diag}[\|\hat{\mathbf{z}}_1\|]$). Then, $\hat{\mathbf{z}}_1^T \mathbf{W}_0 \hat{\mathbf{z}}_1$ is also the sum of the absolute values of all the elements in vector $\hat{\mathbf{z}}_1$. This completes the proof. The

Lasso problem given by Equation (4) can be viewed as a weighted L2-norm regularization, and accordingly, the equivalent DDK filtered solution of the proposed sparse DDK can be represented as follows:

$$\hat{x}_0 = \bar{R} \left\{ \underset{\hat{z}}{\operatorname{argmin}} \left[(x - \bar{R}\hat{z})^T Q^{-1} (x - \bar{R}\hat{z}) + \mu \hat{z}^T W_0 \hat{z} \right] \right\} = \bar{R} \left(\bar{R} Q^{-1} \bar{R} + \mu W_0 \right)^{-1} \bar{R} Q^{-1} x = F_0 x. \quad (5)$$

where the equivalent DDK filtering matrix (in the spectral domain) of sparse DDK is defined accordingly as follows:

$$F_0 = \bar{R} \left(\bar{R} Q^{-1} \bar{R} + \mu W_0 \right)^{-1} \bar{R} Q^{-1} = I - \mu Q (\mu Q + \bar{R} R_0 \bar{R})^{-1}. \quad (6)$$

This equivalent DDK filtering matrix is nothing but the one constructed based on Tikhonov regularization such as a conventional DDK filtering matrix. Using this equivalent filtering matrix, we can also obtain the equivalent sparse DDK filtered solution $\hat{x}_E = F_0 x$. Note that this equivalent solution is strictly speaking non-sparse because the elements corresponding to zero-elements in the original sparse solution $\hat{x}_0$ are very small (close to zeros). Therefore, in terms of spatial resolution, this equivalent filtered solution does not present performance degradation.

Next, we also develop the signal covariance matrix of the sparse DDK filtered solution $\hat{x}_0$. Assuming $\hat{\sigma}_0^2$ is a posterior estimate of the variance component to scale the error covariance matrix Q , (how to calculate $\hat{\sigma}_0^2$ is detailed in Section 2.4), then Equation (6) can be written as:

$$\begin{aligned} F_0 &= \bar{R} \left(\bar{R} (\hat{\sigma}_0^2 Q)^{-1} \bar{R} + \mu W_0 \right)^{-1} \bar{R} (\hat{\sigma}_0^2 Q)^{-1} = \left((\hat{\sigma}_0^2 Q)^{-1} + \mu \bar{R}^{-1} W_0 \bar{R}^{-1} \right)^{-1} (\hat{\sigma}_0^2 Q)^{-1} \\ &= \left(Q^{-1} + \mu \hat{\sigma}_0^2 \bar{R}^{-1} W_0 \bar{R}^{-1} \right)^{-1} Q^{-1} = \left(Q^{-1} + S^{-1} \right)^{-1} Q^{-1} \end{aligned} \quad (7)$$

Therefore, the corresponding posterior signal covariance matrix after equivalent DDK filtering is readily available as follows:

$$S = \left(\mu \hat{\sigma}_0^2 \bar{R}^{-1} W_0 \bar{R}^{-1} \right)^{-1} = \left(\mu \hat{\sigma}_0^2 \right)^{-1} \bar{R} R_0 \bar{R}, \quad (8)$$

Clearly this posterior signal covariance matrix is an adjusted version of the prior one R_0 . Let us review the proposed sparse DDK filter. The sparse filtered solution is calculated based on Equation (4), using the numerical algorithm developed in the following subsection. In order to better compare the conventional DDK filters and the sparse DDK filter, an equivalent DDK filtering matrix F_0 of the latter, in Equation (6), is developed from the obtained filtered solution, as well as its equivalent DDK signal covariance matrix in Equation (8).

2.3. Computation Algorithm

In general, no analytical solution is available for a standard Lasso problem in Equation (2), due to the nonlinearity introduced by the L1-norm regularization. Many algorithms have been proposed to solve this kind of problem numerically. In this work, a highly efficient method, called FISTA, is employed [26]. This algorithm has also been used in our previous studies [28–30]. It is a proximal gradient algorithm, which is further accelerated by the Nesterov momentum method [31]. In FISTA, we conduct the following evaluations iteratively until convergence:

$$\begin{aligned} \beta_k &= \mathcal{T}_{\lambda\mu} \left(\vartheta_k - 2\mu B^T Q^{-1} B \vartheta_k + 2\mu B^T Q^{-1} y \right) \\ s_{k+1} &= \frac{1 + \sqrt{1 + 4s_k^2}}{2} \\ \vartheta_{k+1} &= \beta_k + \frac{s_k - 1}{s_{k+1}} (\beta_k - \beta_{k-1}) \end{aligned} \quad (9)$$

The iterations are initialized as $\beta_0 = \vartheta_1 = 0$ and $s_1 = 1$. The iterations are terminated when $\frac{\|\beta_k - \beta_{k-1}\|_2}{\|\beta_{k-1}\|_2} \leq \delta$ with the threshold $\delta = 10^{-5}$ predefined empirically. The soft thresholding operator $\mathcal{T}_\alpha(x)$ is element-wise; namely, the j th element reads as $[\mathcal{T}_\alpha(x)]_j = \begin{cases} x_j - \alpha, & \text{if } x_j > \alpha \\ x_j + \alpha, & \text{if } x_j < -\alpha \\ 0, & \text{otherwise} \end{cases}$. The step size coefficient λ is chosen as $\lambda = \frac{1}{2\lambda_{\max}(\mathbf{B}^T \mathbf{Q}^{-1} \mathbf{B})}$, where $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue. The value β_k in the last iteration is taken as the final estimate, namely $\hat{\beta} = \beta_k$. Note that in Equation (9), both $2\mu \mathbf{B}^T \mathbf{Q}^{-1} \mathbf{B}$ and $2\mu \mathbf{B}^T \mathbf{Q}^{-1} \mathbf{y}$ do not change across iterations, so they just need to be calculated once and stored.

We follow a trial-and-error strategy to select an appropriate hyperparameter, namely the regularization parameter μ in Equation (2). For each trial value of μ , we calculate the following GCV value:

$$\text{GCV}(\mu) = \frac{m \text{RSS}(\mu)}{(m - p)^2}. \quad (10)$$

In the above, the Residual Sum Squares (RSS) term is calculated as $\text{RSS}(\mu) = [\mathbf{y} - \mathbf{B}\hat{\beta}(\mu)]^T \mathbf{Q}^{-1} [\mathbf{y} - \mathbf{B}\hat{\beta}(\mu)]$. The length of the measurement vector $\mathbf{y}$ is denoted as m , and the number of nonzero elements in the solution is denoted as p . We deliberately introduce μ as an argument in the corresponding variables above in order to explicitly show the dependence of these variables on μ . The chosen μ is the one that minimizes GCV, namely $\hat{\mu} = \underset{\mu}{\text{argmin}} \text{GCV}(\mu)$. Equation (10) is the well-known GCV [32,33]; however, it is tailored to the Lasso problem. This tailoring is permitted by the following property of the Lasso problem revealed by [34]: the number of the nonzero parameters in the Lasso solution is an unbiased and asymptotically consistent estimate of the effective number of free parameters of the Lasso estimation problem.

2.4. Variance and Covariance Analysis

The error covariance matrix $\mathbf{Q}$ is often inaccurate, overestimating the data accuracy rather than underestimating it. We can simply scale the covariance matrix with a scalar value to account for this problem. This scalar is often called a variance component, which is denoted as σ^2 . Therefore, ideally, we should replace $\mathbf{Q}$ with $\sigma^2 \mathbf{Q}$. However, the introduction of this variance component does not affect the solution as shown in the above because it has already been absorbed into the tunable regularization parameter. Taking the equivalent DDK filtering matrix F_0 in Equation (6) as an example, the error covariance matrix $\mathbf{Q}$ therein can be flexibly scaled with a regularization parameter (without using a variance component). However, in some accuracy assessments, this variance component is needed. The variance component and the regularization parameters can be estimated simultaneously with variance component estimation approaches [35,36]. However, we estimate the regularization parameter and the variance component separately. This is rational because the variance component does not affect the estimation of the SH parameters. After an appropriate regularization parameter is selected, as discussed in the above, we treat it as known; we then provide two posterior estimates for the variance component. The variance component is estimated for a model shown in Equation (3).

The first estimate is as follows:

$$\hat{\sigma}_1^2 = \frac{\text{RSS} + \text{Reg}}{m + n - m} = \frac{\text{RSS} + \mu \|\bar{\mathbf{W}}\hat{\mathbf{x}}\|_1}{m}, \quad (11)$$

where Reg represents the regularization term, n is the length of the filtered solution $\hat{\mathbf{x}}$. This estimate is based on the assumption that the constraints implied by the regularization term are viewed as pseudo measurements. Therefore, besides the m measurements, we also have n pseudo measurements. Then, we can further view the regularization solution as a least-squares solution using both measurements and pseudo measurements. Therefore, the above general formula is nothing but the standard posterior estimator of the single

variance component in a least-squares adjustment. When applying this general formula to our specific DDK filtering, $m = n$.

The second estimate is as follows:

$$\hat{\sigma}_2^2 = \frac{\text{RSS}}{m - p}. \quad (12)$$

where p represents the number of nonzero elements in the filtered solution $\hat{x}$. This estimate is obtained by considering the findings in [34], which is already introduced above. We failed to find some relations between the two theoretically. We will compare them in our numerical studies presented in the following.

In addition to the above variance component, the error covariance matrix of the filtered variables may also be needed in some situations. Two approaches can be followed to provide such a covariance matrix. The first is obtained from a Bayesian viewpoint, namely viewing the filtered solution as the maximum a posteriori probability (MAP) estimate. We rearrange the original parameter vector as $x = \begin{bmatrix} x_A \\ x_B \end{bmatrix}$, with the upper part being nonzero

and the lower part being zero in the solution. Accordingly, we have $\hat{x} = \begin{bmatrix} \hat{x}_A \\ \hat{x}_B \end{bmatrix} = \begin{bmatrix} \hat{x}_A \\ 0 \end{bmatrix}$. The uncertainty of the solution is as follows:

$$\begin{cases} \text{Prob}(x_B = 0) = \text{Prob}(x_B = \hat{x}_B) = 1 - \text{Prob}(x_B \neq 0) = 1 \\ x_A \sim N(\hat{x}_A, \hat{\sigma}^2 \bar{R} Q_{\hat{z}_A \hat{z}_A} \bar{R}^T) \end{cases}. \quad (13)$$

where $Q_{\hat{z}\hat{z}} = (\bar{R}^T Q^{-1} \bar{R})^{-1} = \begin{bmatrix} Q_{\hat{z}_A \hat{z}_A} & Q_{\hat{z}_A \hat{z}_B} \\ Q_{\hat{z}_B \hat{z}_A} & Q_{\hat{z}_B \hat{z}_B} \end{bmatrix}$ denotes the covariance matrix of the transformed solution $\hat{z} = [\hat{z}_A \hat{z}_B] = [\hat{z}_A 0]$, $N(\cdot)$ denotes a normal (Gaussian) distribution. This is correct up to the second-order terms in the expression for the logarithm of the posterior density. The reader can find a detailed derivation in Supplementary Text S1 in the Supplementary Material file.

In the second approach, the covariance matrix is obtained simply according to the error propagation law (EPL):

$$Q_{\hat{x}\hat{x}} = \hat{\sigma}^2 F_0 Q F_0^T. \quad (14)$$

This kind of covariance, which accounts for formal errors only, tends to overestimate the accuracy. One of the reasons is that only the commission error is propagated while the omission error is neglected in this propagation. In assessing the accuracy of a functional calculated from a filtered SH model, we should better use the covariance in Equation (13). If the filtered SH model is treated as a prior in another SH modeling problem, we should better use the covariance Equation (14) to better represent the data-only statistical information.

3. Properties of the Proposed Sparse DDK Filter

3.1. A Discussion on Relevant Regularization Strategies

The regularizations mentioned in this work can be classified into three groups, namely the L0-norm, the L1-norm, and the L2-norm regularizations. Their relations are briefly discussed here. Figure 1 intuitively summarizes the discussion.

The L2-norm regularization: The signal covariance in the DDK filter in [5] can be viewed as an application of Kaula's rule of thumb in its generalized version. The degree variance of the signal follows a power-law model, but with a power index empirically chosen to fit a prior geophysical model. Kaula's rule of thumb is viewed as having a physical background by many. This is true only when we consider Kaula's rule of thumb in its generalized version. The specific power index in the generalized Kaula's rule of thumb can depend on the time, the reference model, and other factors. If this power index is not tailored, the results can be inferior even to naïve Tikhonov regularization, namely with the identity matrix taken as the regularization matrix [37,38].

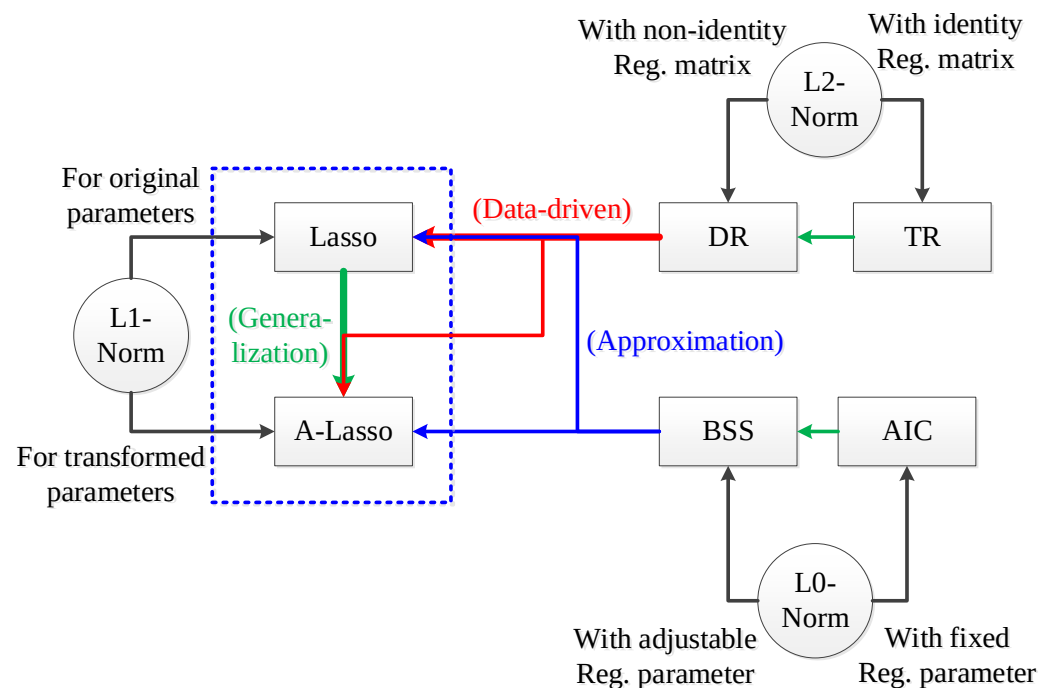


Figure 1. Relations between different regularization schemes. Reg. = regularization; BSS = best subset selection [39]; AIC = Akaike information criterion [40]; Lasso = least absolute shrinkage and selection operator [20]; A-Lasso = adaptive Lasso [27]; TR = Tikhonov regularization [41]; DR = DDK regularization [4,5]. “ $A \rightarrow (\text{Data-driven})B$ ” = B can be viewed as a data-driven version of A ; “ $A \rightarrow (\text{Approximation})B$ ” = B can be viewed as an approximated version of A ; “ $A \rightarrow (\text{Generalization})B$ ” = B can be viewed as a generalized version of A . The schemes in the dotted blue box are the focus of the current study.

The L0-norm regularization: Simply by setting the regularization parameter equal to two in the BSS problem, we can obtain the Akaike information criterion (AIC) [40]. For this, just recall that the L0-norm (corresponding to BSS) of a vector is exactly the number of nonzero elements in this vector; and the nonzero elements exactly define the selected model (corresponding to AIC). Therefore, we can say that the model selection by AIC can be viewed as a special case of the BSS problem, or that the BSS can be viewed as a generalization of model selection with AIC. Both BSS and AIC are combinatorically non-convex, which is the reason why people resort to approximations.

The L1-norm regularization: One relation between the L1-norm and the L0-norm regularizations is that the adaptive Lasso with L1-norm regularization can be viewed as a convex approximation to the non-convex BSS with L0-norm regularization, as already mentioned in above. In fact, the weighted Lasso employed in this work can also be viewed as an approximation to BSS [42]. One relation between the L1-norm and the L2-norm regularizations is that the former can be viewed as a data-adaptive version of the latter, as is already emphasized above. This adaptation stands out automatically from a purely statistical definition, as follows from the Lasso theory [20].

3.2. A Summary of the Properties of the Proposed Sparse DDK Filters

In a nutshell, the proposed DDK filter is time- and space-varying, azimuthally anisotropic, data-adaptive (independent of prior models), and sparse. The proposed filter is compared to other DDK-type filters in terms of these properties. The candidate filters included in this comparison are shown in Table 1. Some of the properties, e.g., spatial homogeneity and azimuthal isotropy, are more meaningful in the original spatial domain (namely, on the sphere) than in the transformed domain (namely, the spectral domain). Therefore, it is beneficial to give the corresponding space-domain filtering kernels of the spectral-domain filtering matrices. For any relevant vector and matrix, we always stick to the ordering rule

of the degrees/orders in the SH coefficients vector $\mathbf{x}$. Let the vector constituting the fully normalized spherical harmonics at location P be denoted as $\mathbf{y}_P$. Then, the filtering kernel is as follows:

$$f_P^{P'} = \mathbf{y}_P^T \mathbf{F}_0 \mathbf{y}_P. \quad (15)$$

Table 1. Properties of different decorrelations filters. DR: DDK regularization [4,5]; ANS: anisotropic non-symmetric [3]; VADER: time variable decorrelation [6]; Spar: sparse DDK filter, proposed herein.

Properties	DDK	ANS	VADER	Spar
Filter Time Variable?	No/Yes	Yes	Yes	Yes
Error Covariances Time Variable?	No/Yes	Yes	Yes	Yes
Signal Covariances Time Variable?	No	No	Yes	Yes
Spatially Inhomogeneous?	Yes	Yes	Yes	Yes
Anisotropic?	Yes	Yes	Yes	Yes
Independent of Prior Models?	No	Yes	Yes	Yes
Filter Sparse?	Yes/No	No	Yes	Yes/No
Solution Sparse?	No	No	No	Yes

Therefore, the values of a gravity functional at location P before and after the filtering are related as follows:

$$\hat{T}_P = \int_{\Omega} f_P^{P'} T_{P'} d\omega'. \quad (16)$$

The properties of the filters under consideration can be summarized as follows:

First, the proposed filter is time-variable, namely varying from month to month. This property is shared with almost all the other filters. The temporal variations of a filter are due to variations of the error covariances and/or the signal covariances. Almost all DDK-type filters can employ time-variable error covariances, though some stationary-filtered solutions (e.g., DDK1, DDK2) are still released. The signal covariances in [3–5] are time-invariable, while the signal covariances in [6] and in this work are time-variable.

Second, the proposed filter is spatially inhomogeneous, i.e., with filtering kernels varying from location to location. This property is shared by all DDK-type filters. This is due to the location-dependent properties of the error covariances. Note that a spatially-inhomogeneous filter is non-symmetric [3].

Third, the proposed filter is anisotropic, i.e., the filtering kernels are azimuth-dependent. The anisotropy corresponds to the dependence of a diagonal filtering matrix on order in the spectral domain. Anisotropy is required from any filter that is able to destripe a solution, simply because the striping errors are anisotropic. Therefore, this property is shared by all DDK-type filters.

Fourth, the proposed filter is data-adaptive, being independent of prior models. This property is shared with [3,6]. The adaptivity is mainly due to the data-driven estimates of signal covariances.

Finally, the proposed filter is solution-sparse, which is shared by none of the others. Note that a filter itself is also sparse, i.e., some of its coefficients in the spectral domain are zeros. It is also possible to make the proposed filter sparse, following the approximations in [4].

It is also interesting to check the other type of decorrelation filters, namely the empirical decorrelation (ED) filters [8,9], in terms of the properties considered here. First, the ED filter is apparently anisotropic, which is necessary for destriping. It is time-variable, because it is constructed from the time-variable data, namely the SH coefficients to be filtered. From this, it also follows that it is data-adaptive. The error and signal covariance matrices are not relevant in ED filters. In our understanding, a complete neglect of covariances, especially the error covariances, is a disadvantage of ED filters. We see this as a waste of available information. ED filters are also spatially inhomogeneous and sparse (in terms of filter coefficients). However, they are generally not sparse in terms of the resulting solutions.

3.3. Demonstration of the Properties of the Sparse DDK Filter

The CSR RL05 solutions supplied with full error covariance matrices (max degree 60) for April 2004 and April 2010 are picked up to demonstrate the properties of the proposed sparse DDK filter. The covariance matrices are readily available for download from <http://download.csr.utexas.edu/outgoing/grace> (accessed on 12 April 2022). The reason we do not use the latest RL06 data is because, currently, CSR only provides the full error covariance matrices of RL05 solutions. We use the three-monthly solutions to devise their corresponding sparse DDK filters, with regularization parameters selected as 5.0×10^6 , 7.5×10^5 , and 2.5×10^6 , respectively. Figure 2 presents their normalized smoothing kernels in the east–west and north–south directions, where the kernel center is located at different latitudes (0° , 30° , 60°) along the 0° meridian. The column-wise comparison demonstrates that sparse DDK is time-variable, whereas the row-wise comparison proves its spatial inhomogeneity. On the other hand, the E–W and N–S directions show different filtering strengths, which implies that the sparse DDK filter is anisotropic. Furthermore, we note the negative side-lobes, which help to reduce signal leakage, when compared to the monotonic Gaussian filtering [5].

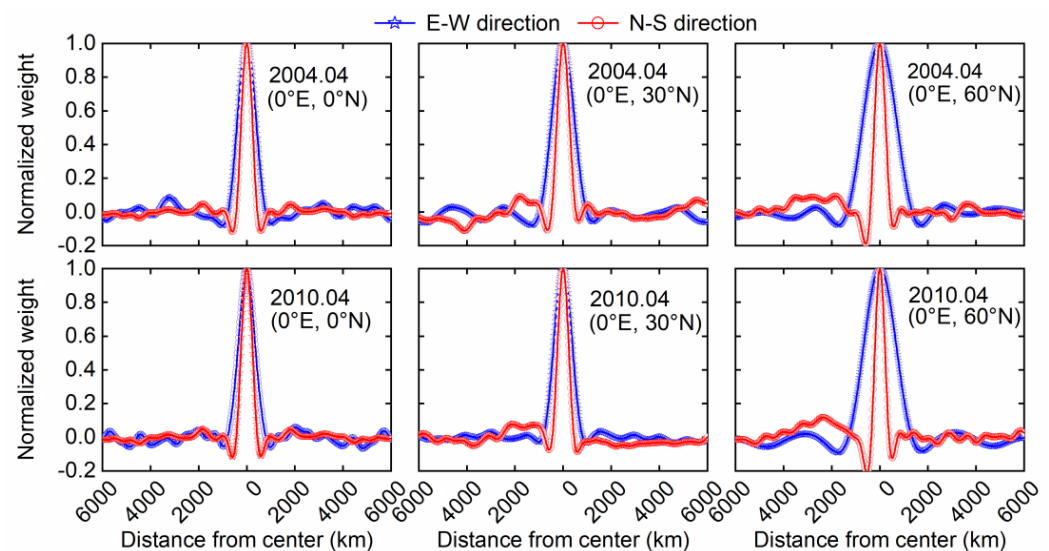


Figure 2. Cross-sections of normalized smoothing kernel of the proposed filter. Red: SN direction, Blue: WE direction. Top: April 2004; bottom: April 2010. **Left:** at (0°E , 0°N); **middle:** at (0°E , 30°N); **Right:** (0°E , 60°N).

Next, the monthly solutions under consideration are filtered using the corresponding sparse DDK filters presented above. Figure 3 presents the distribution of zero and nonzero elements in the filtered monthly solutions, clearly showing a sparsity of the filtered solutions. It is well-known that gravity signals are mainly concentrated in low-degree and low-order terms. We can observe in Figure 3 that those terms are well retained, while the high-degree and high-order terms, which are mainly dominated by noise, are set to zero. This result completely satisfies the original intention of filter design: to suppress high-degree noise while preserving low-degree signal. In addition, we also notice that the maximum degrees of the filtered solutions are lower than that of unfiltered ones (i.e., 60), indicating that the sparse DDK filter reduces the spatial resolution of the solutions. However, this reduction in the nominal resolution is beneficial. Let us explain this using the following simple example. Assume that a high-degree/order SH coefficient before filtering is equal to 1, with 0.1 of it being the signal and 0.9 being noise. Considering the fact that it is impossible for a filter to completely separate the signal from noise at the same spectral point, it is more reasonable to discard this coefficient than to retain it. In this regard, L1-norm regularization can perform this discarding or truncation flexibly while the L2-norm Tikhonov regularization cannot. Furthermore, it has been proved in [25,43] that the

L1-norm minimization may yield better results than the L2-norm minimization considering the SH noise is non-Gaussian due to relevance therein. It is similar to the simple truncation of the spherical harmonic series; however, this sparse regularization-based “truncation” is data-driven or adaptive and, hence, wiser. It can be seen that there are also some zero coefficients at low or medium degrees/orders. This should not be a surprise, recalling that the coefficients represent a residual signal with a reference signal removed. These zero coefficients merely mean that the corresponding terms have been approximated by the reference model with sufficient accuracy. In this regard, we can expect an even sparser solution if a better reference model is used.

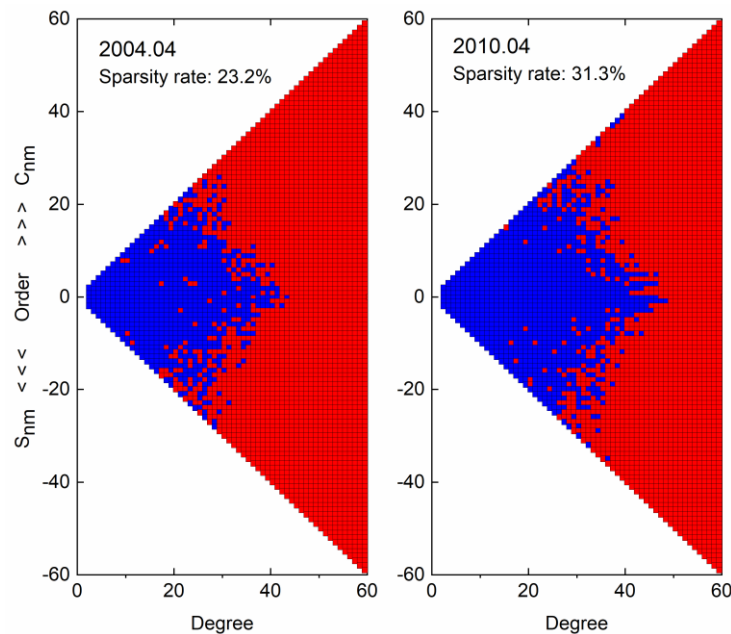


Figure 3. Distribution of nonzero SH coefficients (blue squares) in the solutions obtained with the proposed filter. **Left:** April 2004; **Right:** April 2010. The sparsity rate is defined as the ratio between the numbers of nonzero parameters and all parameters.

4. Results

The 7-year time-series of GRACE monthly solutions with corresponding covariance matrices [44], ranging from January 2004 to December 2010, is used for evaluating the performance of the proposed sparse DDK filter. The regularization parameter μ of each month, selected by GCV, ranges from 5×10^5 to 7.5×10^6 . The following seven filters are compared: Gaussian 300-km smoother (denoted as G300), Swenson and Wahr (SW) empirical decorrelation filter (denoted as Swen), SW filtering + Gaussian 300-km smoothing (denoted as S300), DDK2, DDK3, DDK4 filters and the proposed sparse DDK (denoted as Spar) filter. The three variants of DDK filters, namely DDK2, DDK3 and DDK4, are selected for comparison since we find that their filtering strengths are comparable to that of the sparse DDK. In addition, the CSR RL06 v2.0 mascon solutions are also included for a comparison, considering they do not need filtering and can derive well-localized mass anomaly estimates [12,45]. We calculate the filtered mass anomaly using the filtered solutions with the C_{20} coefficient replaced with that derived from satellite laser ranging (SLR) [46], and with the degree-1 coefficients estimated by [47] added. The glacial isostatic adjustment (GIA) effect is corrected using the ICE-6G_D model [48].

In the following Section 4.1, we compare the filtered mass anomalies of the seven filters for both global and local cases and give their uncertainty estimates. In Section 4.2, we show the signal retaining rates and noise levels of each filtered solutions series. Finally, we evaluate the sparsity of the sparse DDK filtered solution and its signal retainment in polar areas.

4.1. Filtered Mass Anomalies Analysis

The mass anomaly, namely surface density ξ_P , at a location P , in units of EWH, is computed as follows [49]:

$$\xi_P(\theta, \lambda) = \frac{a\rho_{ave}}{3} \frac{2l+1}{1+k_n} \sum_{n=2}^{L_{max}} \sum_{m=0}^n [\Delta C_{nm} \cos(m\lambda) + \Delta S_{nm} \sin(m\lambda)] \bar{P}_{nm}(\cos \theta) = \mathbf{g}_p^T \mathbf{x}, \quad (17)$$

where $\mathbf{g}_p$ is the spectral conversion matrix, which converts the spherical harmonic coefficients vector $\mathbf{x}$ into the surface mass anomaly.

We still select monthly solutions from April 2004 and April 2010 for illustration. The global distribution of EWHs before and after filtering is shown in Figure 4 for the selected months. It is observed that in all seven filtered solutions, we can identify significant geophysical signals. The G300 smoothing can effectively reduce noise, but some residual stripes still exist, especially in ocean areas. The Swen empirical decorrelation can completely de-stripe the solution; however, there is still some high-frequency noise at middle and low latitudes. The S300 filter can handle both high-frequency noise and stripes efficiently, but there is some signal attenuation in high latitudes such as Greenland and Antarctica. In DDK2, DDK3, DDK4, and sparse DDK filtered solutions, the suppression of high-frequency noise and stripes is similar visually. We also notice that the smoothing effect of the sparse DDK is closer to that of the DDK3 filter.

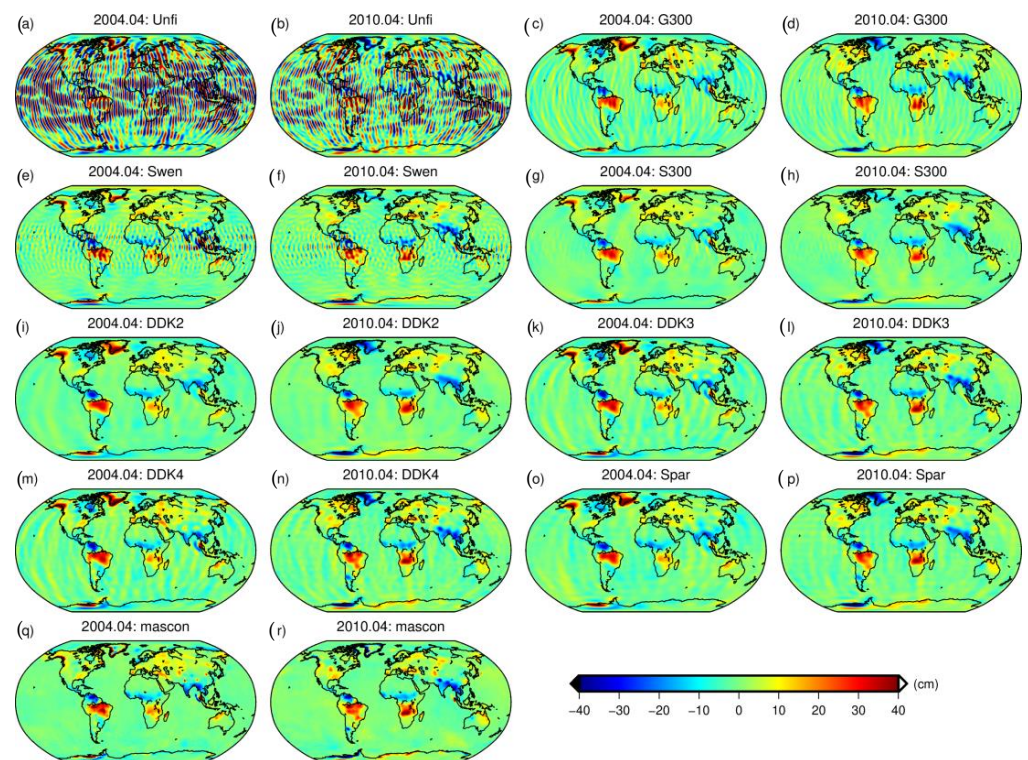


Figure 4. The filtered monthly solutions (in terms of EWH) of 7 filters for April 2004 (the first and the third column) and April 2010 (the second and the fourth column). The solutions before filtering (the first row) and CSR RL06 mascon solutions (the fifth row) are also presented for comparison.

To further evaluate the filtered local mass anomalies of the sparse DDK filter, we selected six study areas: Amazon Basin, Congo Basin, Ganges Basin, Yangtze Basin, Alaskan glaciers (defined in Figure S1 in Supplementary Material file), and Greenland. The surface

mass densities shown in Equation (17) are integrated over a study area to calculate the total mass anomaly ζ in this area, as follows:

$$\zeta = \int_{\Omega} \zeta_P d\omega_P = \int_{\Omega} \mathbf{g}_P^T \hat{\mathbf{x}} d\omega_P = \left[\int_{\Omega} \mathbf{g}_P^T d\omega_P \right] \hat{\mathbf{x}} \approx \left[\sum_i \mathbf{g}_i^T \Delta_i \right] \mathbf{x} = \mathbf{h}^T \hat{\mathbf{x}}. \quad (18)$$

In the above, Δ_i denotes the area of the i -th grid cell in the sum-approximation of the integral. Accordingly, the formal variance of the total mass anomaly is computed through error propagation, as follows:

$$\sigma_{\zeta}^2 \approx \mathbf{h}^T \mathbf{Q}_{\hat{\mathbf{x}}} \mathbf{h}. \quad (19)$$

where $\mathbf{Q}_{\hat{\mathbf{x}}}$ denotes the covariance matrix of the filtered solution.

In the following, we will compare the formal error StD (standard deviation) of the filtered regional mass anomaly of the sparse DDK and the conventional DDK filters (DDK2, DDK3 and DDK4). For the sparse DDK filter, there are four options to calculate $\mathbf{Q}_{\hat{\mathbf{x}}}$ in Equation (19), namely with Equation (11)/Equation (13), with Equation (11)/Equation (14), with Equation (12)/Equation (13), or with Equation (12)/Equation (14). For easy reference, we call the corresponding variance-covariance estimates $\mathbf{Q}_{\hat{\mathbf{x}}}$ “MAP with σ_1 ”, “EPL with σ_1 ”, “MAP with σ_2 ” and “EPL with σ_2 ”, respectively. For the conventional DDK filters, the posterior estimate of the variance component is computed according to the leftmost equation of Equation (11). The formal covariance matrix of a filtered solution from an error propagation point of view is given by Equation (14), while that from a Bayesian point of view is as follows:

$$\mathbf{Q}_{\hat{\mathbf{x}}} = \hat{\sigma}^2 \left(\mathbf{Q}_{\text{DDK}}^{-1} + \mu \mathbf{S}_{\text{DDK}}^{-1} \right)^{-1} = \hat{\sigma}^2 [\mathbf{F}_{\text{DDK}} \mathbf{Q}_{\text{DDK}}] \quad (20)$$

where $\mathbf{F}_{\text{DDK}}$ is the block-diagonal filtering matrix, directly provided in conventional DDK filters. The diagonal signal covariance matrix $\mathbf{S}_{\text{DDK}}$, as a scaled variant of the matrix $\mathbf{W}$ introduced in Section 2.2, is usually provided in the form of a given power index and regularization parameter. The involved time-invariable error covariance matrix $\mathbf{Q}_{\text{DDK}}$ is recovered from $\mathbf{F}_{\text{DDK}}$ and $\mathbf{S}_{\text{DDK}}$, based on the following relations:

$$\mathbf{F}_{\text{DDK}} = \left(\mathbf{Q}_{\text{DDK}}^{-1} + \mu \mathbf{S}_{\text{DDK}}^{-1} \right)^{-1} \mathbf{Q}_{\text{DDK}}^{-1} \quad (21)$$

Figure 5 shows the regional total mass anomaly series (the left panel) for the selected six study areas, as well as their differences with respect to the mascon solutions (cf. the right panel), of which the root mean square differences (RMSD) are presented in Table 2. In general, all the filtered solutions produce similar mass anomaly estimates, and they agree well with the CSR RL06 v2.0 mascon solution. The mass anomaly differences (with mascon as reference) of the sparse DDK filtered solution are as close as those of the other DDK filters. In Alaskan glaciers and Greenland, the S300 filter and Swen filter show significant differences from the other filtered solutions and mascon solution because of stronger signal smoothing by the involved empirical decorrelation filter at middle and high latitudes. In Greenland, the Spar filtered solutions differ from the mascon solution (but are consistent with the other DDK solutions), which may result from differences in leakage errors. This proves the proposed sparse DDK does not show severe significant degradation, although the high-degree terms are forced to zeros when moving to high latitudes.

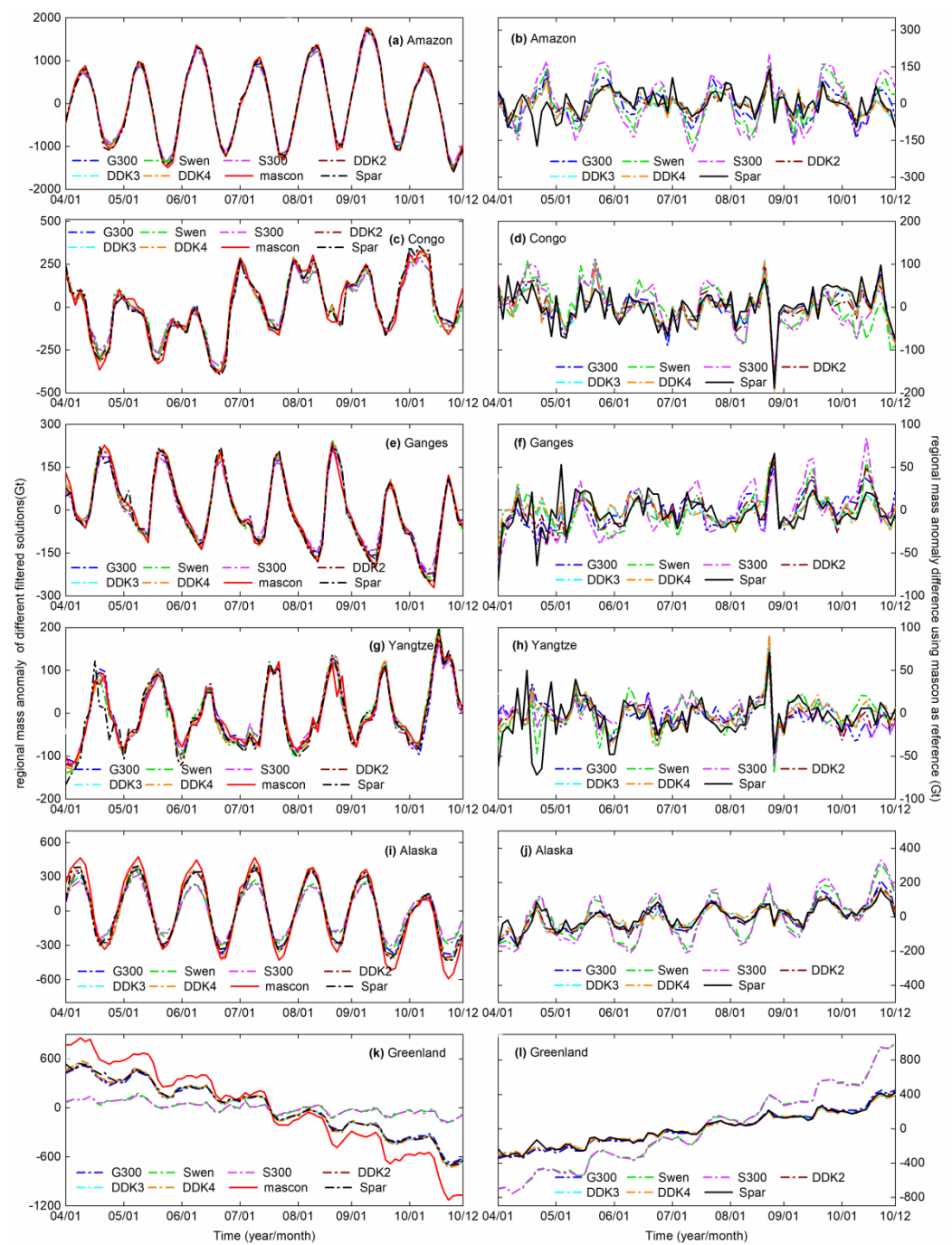


Figure 5. Variations of filtered regional mass anomalies vs. time (**left panel**), and the corresponding differences with respect to the mascon solutions (**right panel**). From top to bottom are results for Amazon Basin, Congo Basin, Ganges Basin, Yangtze Basin, Alaskan glaciers, and Greenland. No leakage correction is made for the filtered solutions.

Table 2. The RMSD of the regional total mass anomaly of different filtered solutions with respect to mascon solutions. The results are calculated from the right panel of Figure 5.

Filter	Amazon	Congo	Ganges	Yangtze	Alaska	Greenland
G300	66.5	42.0	19.5	18.9	79.0	212.8
Swen	76.3	48.7	20.9	20.2	120.6	457.1
S300	99.7	47.8	26.6	17.1	136.3	456.3
DDK2	44.1	39.4	19.1	18.7	69.7	200.5
DDK3	44.6	41.1	16.2	19.7	62.3	189.3
DDK4	45.2	41.3	16.2	20.2	62.1	185.4
Spar	52.3	42.5	21.1	25.2	63.1	191.7

Figure 6 presents the formal error StD of regional mass anomalies for sparse DDK, DDK2, DDK3, and DDK4 filters. The following observations are drawn. First, since the variance component σ_1 in Equation (11) is almost the same size as σ_2 in Equation (12) (the left panel in Figure 5), we can adopt either of them to scale the error covariance matrix of the filtered solution. Second, the error StDs derived from EPL in Equation (14) are always lower than those from MAP in Equation (13). This is because EPL only propagates measurement errors, while modelling errors are neglected. In other words, the uncertainty introduced by prior information (e.g., from regularization) is ignored by EPL. Third, the error StDs in the sparse DDK filter are much larger than those in DDK2, DDK3, and DDK4 filters. This may be due to missing some covariance components for DDK2, DDK3, and DDK4 when we recover the covariance matrix Q_{DDK} in Equation (21) from the block-diagonal F_{DDK} and signal covariance matrix S_{DDK} . Finally, it is difficult to assess whether the formal error StD is too optimistic or pessimistic because we do not know the true regional mass anomalies. However, we still made an approximate verification using the mascon mass anomalies as the truth. Table 3 lists the mean formal error StDs of four calculation schemes for the six areas. Compared with RMSD of sparse DDK in Table 2, we find that the formal error StDs from both MAP and EPL are comparable to those derived from the differences between sparse DDK solutions and mascon solutions in four cases (Amazon, Congo, Ganges, Yangtze) out of six; the discrepancies are small. For Alaskan glaciers and Greenland, the formal error StDs are still over-optimistic. This may be due to smaller signal leakage of the mascon solutions than that of sparse DDK filtered solutions in these regions, thus making the mass anomaly differences between them greater, which cannot be captured by the formal errors.

Table 3. The mean formal error StD of regional mass anomalies calculated from the left panel of Figure 6.

Indicator	Amazon	Congo	Ganges	Yangtze	Alaska	Greenland
MAP with σ_1	67.8	50.8	16.9	20.5	38.0	26.5
EPL with σ_1	60.2	42.5	12.1	17.0	28.3	23.2
MAP with σ_2	67.5	50.6	16.8	20.4	37.8	26.4
EPL with σ_2	60.0	42.4	12.1	16.9	28.2	23.1

4.2. Signal and Noise Level Analysis for Filtered Solutions

First, the signal and noise levels of the filtered solutions in the spectral domain are assessed. Figure 7 presents the geoid degree-variances of the filtered solutions for the three selected months. According to [50,51], the degree-variances of the first 30 degrees mainly reflect the signal amplitude of the time-varying gravity field, while those after degree 30 mainly show the noise level. In our experience, such a statement is probably a bit too optimistic. It would be fairer to say that the range between degrees 20 and 30 is a transition zone, where contributions of noise and signal are similar. From this viewpoint, the S300 filter retains the least time-varying gravity signal, though its noise level is also low. The other filters do not show much difference in signal preservation (the first

30 degrees), but this is converse for noise suppression (after degree 30). It is impressive that compared with the other six filters, the sparse DDK can retain comparable gravity signals before the first 30 degrees while suppressing the noise after 30 degrees significantly. This mainly benefits from the high-degree and high-order sparsity brought by the weighted L1-norm regularization.

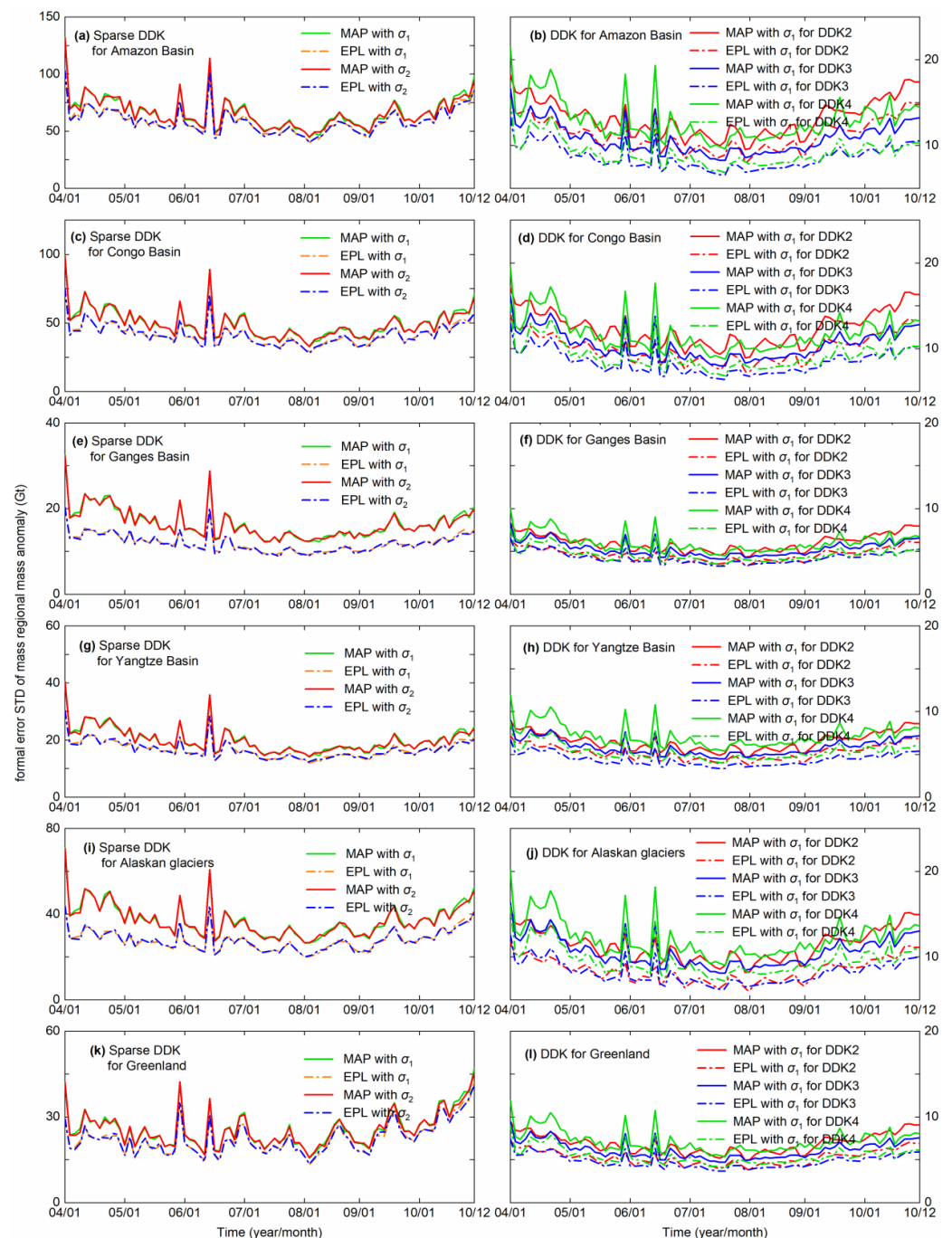


Figure 6. Formal error StD of regional mass anomaly versus time. Left panel: sparse DDK; right panel: DDK2, DDK3 and DDK4. From top to bottom are results for Amazon Basin, Congo Basin, Ganges Basin, Yangtze Basin, Alaskan glaciers, and Greenland. Note that the error covariance matrices Q_{DDK} in DDK1, DDK2 and DDK3 are recovered from block-diagonal filtering matrix F_{DDK} and signal covariance matrix S_{DDK} .

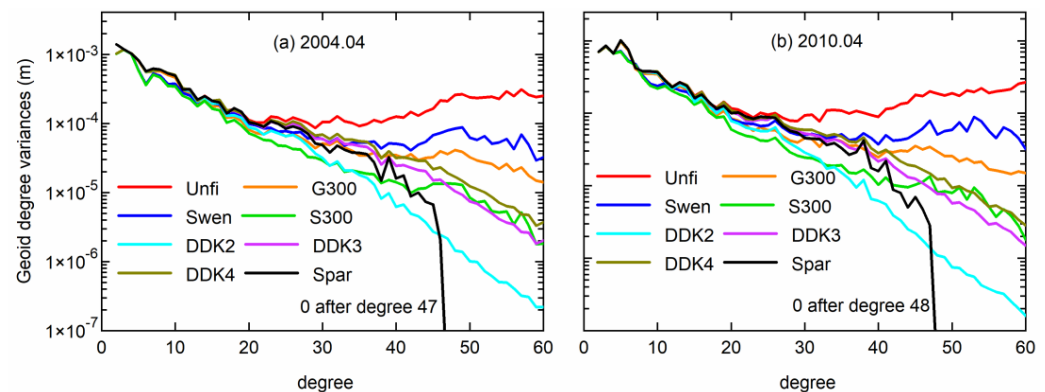


Figure 7. The geoid degree-variance of filtered monthly solutions for: (a) April 2004, (b) April 2010. The high-degree variances of sparse DDK filtered solutions are strictly equal to zero.

Then, the signal and noise levels of filtered solutions are quantitatively analyzed in the spatial domain, both at selected locations and globally. In doing so, we derive a mass anomaly time-series at the Earth's surface, taking into account the Earth's oblateness [52]. The selected locations are five lakes: the Caspian Sea, Ladoga, Nasser, Tanganyika, and Victoria. The mean mass anomalies over those lakes (in terms of EWH) are compared with water level variations extracted from altimetry data. Annual variations are eliminated in the course of the comparison since they may contain noticeable nuisance signals (e.g., nuisance hydrological signals in the GRACE data and steric water level variations in altimetry data). For a similar reason, long-term linear trends are eliminated in the case of the Ladoga and Nasser lakes. Technically, the comparison is performed using a least-squares estimation of unknown coefficients $C_1, \dots, C_5, E$, and, optionally, C_6 , in the following equations:

$$C_1 + C_2 \sin \omega t_i + C_3 \cos \omega t_i + C_4 \sin 2\omega t_i + C_5 \cos 2\omega t_i + \{C_6 t_i\}_{\text{optional}} + Eh(t_i) = d(t_i) \quad (22)$$

where $\omega = \frac{2\pi}{1\text{yr}}$; t_i is the time in the middle of the i -th month, $h(t)$ is the altimetry-based time series of water level variations in the given lake, and $d(t)$ is the GRACE-based time series of mean mass anomalies in the same lake. The post-fit residuals are interpreted as noise in GRACE mass anomaly time-series and reported in terms of noise StD.

The scaling factor E reflects the signal damping in a given GRACE-based time series. It is used to compute the quantity ε , which is called thereafter “relative signal retaining”:

$$\varepsilon = E/E_{\text{exp}}. \quad (23)$$

The “expected signal retaining” E_{exp} in this expression reflects the signal damping caused solely by the truncation of the spherical harmonic series at $L_{\text{max}} = 60$. Thus, the relative signal retaining ε contains information about the effect of filtering in the context of a particular lake. In the case of ideal noise-free input data, ε is equal to 1 if filtering is not applied and reduces as the applied (low-pass) filter becomes increasingly aggressive. Further details concerning this comparison can be found in [53].

The obtained estimates of the noise StD and relative signal retaining for each of the eight solution time-series are shown per lake in Figure 8. Furthermore, Table 4 (the 2nd and 3rd column) reports the mean values of the estimated noise StD and weighted mean values of the relative signal retaining, respectively, which are obtained by averaging over all five lakes. One can see that in most cases, the more aggressive a filter is (i.e., the lower the relative signal retaining), the lower is the resulting noise level. For instance, the S300 filter, which is apparently the most aggressive one, yields the lowest noise StD and the lowest relative signal retaining, whereas the DDK4 filter, being the least aggressive one, results in a relatively high noise level (when the mean values over the five lakes are considered). The sparse DDK filter looks more aggressive than DDK3 but less aggressive than DDK2. To be

specific, the mean values of both relative signal retaining and noise StD are in-between the corresponding values of the DDK2 and DDK3 filters.

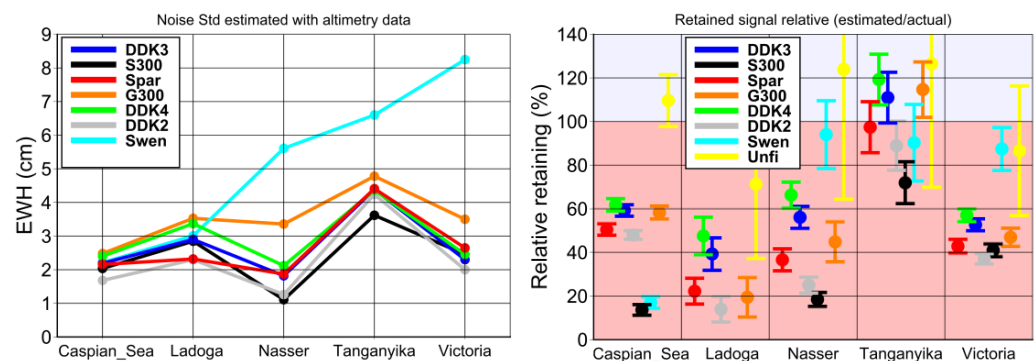


Figure 8. Assessment of the time-series of mean mass anomalies per lake extracted from the eight GRACE solution time-series under consideration: noise standard deviation (**left**) and relative signal retaining (**right**).

Table 4. The second and third columns summarize the results of the lake tests, showing the mean values of the estimated noise StD and weighted mean values of the relative signal retaining, respectively. Both quantities are the result of averaging over all five considered lakes. The last column reports the global estimates of noise standard deviations obtained with the VCE method.

GRACE Solutions	Lake Tests		Global
	Relative Signal Retaining (%)	Noise StD (cm EWH)	Noise StD (cm EWH)
Unfi	110.5 ± 10.4	18.3	23.6
G300	59.0 ± 2.4	3.8	3.8
Swen	26.8 ± 2.6	5.3	5.7
S300	24.1 ± 1.7	2.6	1.3
DDK2	44.4 ± 1.7	2.8	1.6
DDK3	62.0 ± 2.0	3.3	2.4
DDK4	66.7 ± 2.2	3.5	2.8
Spar	50.8 ± 2.1	3.1	2.3

We also estimate the accuracy of a filtered solution time-series globally at the nodes of a global equiangular $1^\circ \times 1^\circ$ grid. Noise StD is estimated at each grid node separately, using the procedure proposed by [13]. That is, a regularized time series is computed at each node by minimizing the following objective function:

$$\Phi[x] = \frac{1}{\sigma_n^2} \sum_i (x_i - d_i)^2 + \frac{1}{\sigma_s^2} \Omega[x], \quad (24)$$

where d_i is the original mass anomaly in the i -th month, x_i is the regularized mass anomaly in the same month (to be estimated), σ_n^2 is an unknown noise variance, σ_s^2 is an unknown signal variance, and $\Omega[x]$ is a regularization functional. The goal of the regularization functional in Equation (24) is to ensure that the regularized mass anomaly time-series $x(t)$ at each grid node is sufficiently realistic so that the deviations of the data from those time-series can be exploited for an accurate estimation of noise variances. A point of special concern is an excessive damping of the original time series, which may result in an overestimation of noise levels. To mitigate this effect, we have applied a tailored regularization, which takes into account that mass anomaly time-series are typically characterized by an annual periodicity and/or long-term trends. Originally, a regularization of this type was introduced by [13], who proposed to minimize year-to-year differences between the values of the first time-derivative of the mass anomaly time series. It was proved that such a

regularization functional does not penalize annual periodic signals and linear trends at all so that at least these signal components are not subject to any damping. In this study, we have adopted a modified variant of that functional, which minimizes year-to-year differences between the values of the second time-derivative of the mass anomaly time series. Assuming that this time series is represented as a continuous function of time $x(t)$ in the time interval $[t_{\text{beg}}; t_{\text{end}}]$, the exploited regularization functional can be written as:

$$\Omega[x] = \int_{t_{\text{beg}}+1\text{yr}}^{t_{\text{end}}} [\ddot{x}(t) - \ddot{x}(t-1\text{yr})]^2 dt \quad (25)$$

where $\ddot{x}(t)$ stands for the second derivative of $x(t)$. This type of regularization was called a “minimization of Month-to-month Year-to-year Triple Differences” (MYTD) in [53], where it was originally proposed. This type of regularization yields better results than the minimization of year-to-year differences between the values of the first time-derivative, which was proposed in the aforementioned publication by [13]. The unknown noise variance σ_n^2 and signal variance σ_s^2 are estimated in parallel with the minimization of the objective function, using the Variance Component Estimation (VCE) method [36]. Only the former value is of interest in this study since it can directly be used to estimate the noise StD σ_n at the given grid node.

The obtained estimates of noise StD for the eight solution time-series under consideration are shown as maps in Figure 9. The global RMS values of the obtained estimates are reported in the last column of Table 4. One can see that the global estimates of noise level rank the solutions basically in the same way as the lake tests. For instance, the lowest noise level is observed after applying a relatively aggressive S300 filter, whereas the DDK4 filter yields a higher noise level than any other filter of the “DDK” type. The proposed sparse DDK filter results in a noise level that is lower than after the DDK3 filtering but higher than after the DDK2 filtering.

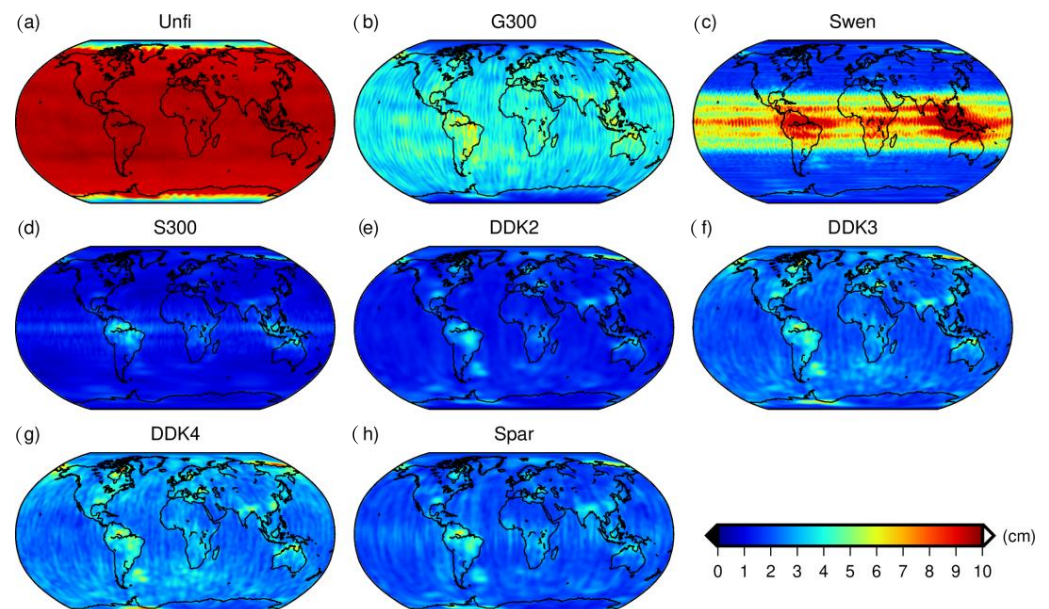


Figure 9. Noise StD for the eight filtered solution time-series under consideration estimated at the nodes of a global $1^\circ \times 1^\circ$ grid with the VCE method.

5. Discussion

Figure 10 shows the sparsity ratio and the maximum SH degree of sparse DDK filtered solutions from January 2004 to December 2010. The sparsity ratios are between 14.1% and

36.5%, while the maximum degrees are between 34 and 52. This sparsity significantly suppresses the high-degree and high-order noise, as presented in Figure 7.

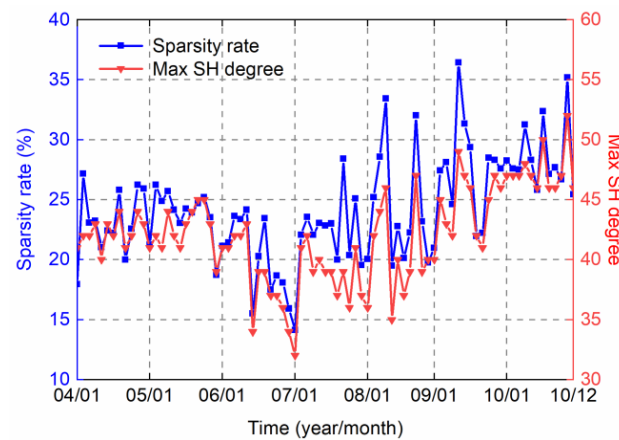


Figure 10. The sparsity rate (left axis) and maximum SH degree (right axis) of sparse DDK filtered solutions from January 2004 to December 2010. The sparsity rate is defined as the percentage of the number of nonzero elements to total number. Note that the degree 0 and 1 terms are not considered. The maximum SH degree is the highest degree of nonzero elements in each sparse DDK filtered solution.

It is well-known that GRACE can sense mass anomalies near the poles much better than in the rest of the world. This implies that a monthly solution complete to a relatively high degree is needed in order to fully exploit information contained in the SH coefficients when mass anomalies in polar areas are estimated. However, our sparse DDK filter may lose some of the high-frequency signals when discarding the high-degree noises. Therefore, it is necessary to check the signal loss of the filtered solutions in polar areas. The six study areas in Section 4.1, located in low, middle, and high latitudes, are taken as examples here. Considering the noise (or stripes) has been largely eliminated visually, as shown in Figure 4, we can approximately assume that the filtered solution mainly consists of the signal. In the following, the retained signals within these areas are reported as the annual amplitude of the filtered solutions.

The average annual amplitude for the six areas is summarized in Table 5. We find that the DDK-type filters always provide higher annual amplitudes than the other filters. In addition, the values of the mascon solutions are similar to those of the DDK filters. In most cases (except Greenland), the average annual amplitudes of sparse DDK filtered solutions are still between those of DDK2 and DDK3, while the largest amplitudes are shown by the DDK4 solutions. This reveals that all the DDK filters show similar behavior. In addition, we also notice that the annual amplitude of sparse DDK filtered solutions tends to become weaker at higher latitudes. Over Greenland, for instance, they are by 5~17% weaker (4.0 versus 4.3, 4.7, 4.8, and 4.2 cm), compared to the other DDK filtered solutions and G300 smoothing, but still stronger than those of Swen and S300 filters.

Table 5. The regional average annual amplitude of filtered mass anomaly solutions in six areas of different latitudes. The results are based on data from January 2004 to December 2010 and presented in EWH (cm).

Region	G300	Swen	S300	DDK2	DDK3	DDK4	Mascon	Spar
Amazon	21.4	23.0	20.6	22.7	23.2	23.3	23.0	23.2
Congo	12.1	13.3	11.6	12.8	13.2	13.3	13.0	13.1
Ganges	12.8	13.4	12.0	13.4	13.9	14.0	13.7	13.7
Yangtze	5.4	5.5	4.6	5.1	5.8	5.9	5.5	5.4
Alaskan glaciers	10.2	8.3	7.5	10.2	11.0	11.2	12.1	10.6
Greenland	4.2	2.9	2.2	4.3	4.7	4.8	6.7	4.0

6. Conclusions

Focusing on decorrelation and denoising for GRACE time-variable gravity field solutions, the contributions of this work include: (1) An alternative filter called sparse DDK is devised by using weighted L1-norm regularization; (2) the conventional mainstream filters are checked and compared from their properties, including time-variability, location-inhomogeneity, anisotropy, independence from the prior model, sparsity, etc., and the relevant regularization strategies, including L0, L1 and L2 regularizations are discussed. The sparse DDK filter ensures that the obtained solutions are sparse, i.e., that most of the high-degree and high-order SH coefficients are zeros. This means that strong noise, which may originate from these coefficients, is suppressed. The performance of the proposed sparse DDK filter and other mainstream filters was compared using real GRACE level-2 data. The results show that the proposed sparse DDK can remove stripes and high-frequency noise efficiently. The annual amplitudes and noise StD based on both point-wise estimates and on average values for selected regions show that the sparse DDK filter yields signal and noise levels that are comparable to those from the traditional DDK filters. The time-series of total mass anomalies in the selected research areas also indicate that the sparse DDK filter can recover mass anomalies similarly to other DDK filters. Finally, the following two points should be noted:

First, the filtering matrix F in sparse DDK can also be simplified into block-diagonal form, as proposed by [4], which can be achieved by retaining only the elements that correspond to SH coefficients of the same order in the error covariance matrix. On the other hand, one can also modify the conventional DDK filter, which is based on a stationary error covariance matrix and signal covariance matrix, into a sparse DDK filter by directly replacing the Tikhonov regularization with the weighted L1-norm regularization. In addition, the regularization parameter involved in sparse DDK are selected by GCV, which is believed to correspond to the optimal tradeoff between signal retainment and noise suppression. Of course, one can also adjust the regularization parameter of sparse DDK freely if they need a different filtering strength.

Second, for the sparse DDK filter, high-degree and high-order sparsity means that the maximum degree of the filtered solution is reduced, implying that the spatial resolution of the filtered solution is reduced. However, this reduction does not cause explicit loss of mass anomaly signal, especially at middle and low latitudes (5~17% signal loss in terms of annual amplitude for high latitudes taking Greenland as an example). Note that the GRACE 60-degree monthly solutions are just adopted in this work as an example. Considering that many data centers nowadays release higher 90-degree (e.g., GFZ, JPL) or 96-degree (e.g., CSR, Tongji) monthly time-varying gravity field solutions, it can be expected that after filtering with sparse DDK, the maximum degree of the filter solutions can reach approximately 60 to 80, which can significantly improve the loss of spatial resolution compared with using 60-degree data. In this case, compared to reserving higher-degree noise, it is worthwhile to discard higher-degree tiny signals, and the negative effects of high-degree signal loss can also be further reduced.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs14122810/s1>, Text S1: Derivation for VCM of Lasso solution; Figure S1: Definition for Alaskan glaciers.

Author Contributions: Conceptualization, G.C.; methodology, G.C. and N.Q.; software, N.Q. and P.D.; validation, P.D., Z.W.; formal analysis, N.Q.; investigation, N.Q.; resources, J.G.; data curation, N.Q.; writing—original draft preparation, N.Q., G.C. and P.D.; writing—review and editing, P.D.; visualization, N.Q. and P.D.; supervision, J.G.; project administration, G.C. and J.G.; funding acquisition, G.C. and J.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work is sponsored partially by the National Natural Science Foundation of China (Grant Nos. 42074001, 41974026, and 41774005), partially by the China Postdoctoral Science Foundation (Grant Nos. 2019M652010 and 2019T120477), partially by the Postgraduate Research and Practice Innovation Program of Jiangsu Province (Grant No. KYCX21_2292) and partially by the Postgraduate Innovation Program of China University of Mining and Technology (Grant No. 2021WLKXJ098).

Data Availability Statement: The CSR GRACE monthly gravity solutions and error covariance matrices used in the study are available via <http://download.csr.utexas.edu/outgoing/grace> (accessed on 12 April 2022). The time-series of water level variations in lakes are based on Jason-1, Jason-2, and Jason-3 satellite altimetry data (courtesy of the USDA/NASA G-REALM program at https://ipad.fas.usda.gov/cropexplorer/global_reservoir/) (accessed on 12 April 2022). The GRACE MATLAB Toolbox (GRAMAT) software is used for the calculation of mass anomaly [54]. The basin boundaries are provided at <http://hydro.iis.u-tokyo.ac.jp/~taikan/TRIPDATA/TRIPDATA.html> (accessed on 12 April 2022). Geometry of the considered lakes was generated with Generic Mapping Tools (GMT) software [55], which makes use of the Global Self-consistent, Hierarchical, High-resolution Geography database (GSHHG) at <https://www.soest.hawaii.edu/pwessel/gshhg> (accessed on 12 April 2022). GMT software was also used to create the plots.

Acknowledgments: Jurgen Kusche is acknowledged for insightful discussions on regularization methods in gravity data processing during Chang's visit to the University of Bonn. The support provided by the China Scholarship Council (CSC) during Qian's visit to Delft University of Technology is also acknowledged. We would like to thank Wei Feng for providing the GRACE MATLAB Toolbox (GRAMAT) [54], Roelof Rietbroek for making his codes on DDK filtering available and Taikan Oki for providing basin boundaries. Gratitude is also extended to the Centre for Space Research (CSR) for providing GRACE monthly solutions and corresponding covariance matrices.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ran, J.; Ditmar, P.; Klees, R.; Farahani, H.H. Statistically optimal estimation of Greenland Ice Sheet mass variations from GRACE monthly solutions using an improved mascon approach. *J. Geod.* **2018**, *92*, 299–319. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Shugar, D.; Jacquemart, M.; Shean, D.; Bhushan, S.; Upadhyay, K.; Sattar, A.; Schwanghart, W.; McBride, S.; de Vries, M.V.W.; Mergili, M. A massive rock and ice avalanche caused the 2021 disaster at Chamoli, Indian Himalaya. *Science* **2021**, *373*, 300–306. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Klees, R.; Revtova, E.; Gunter, B.C.; Ditmar, P.; Oudman, E.; Winsemius, H.C.; Savenije, H.H.G. The design of an optimal filter for monthly GRACE gravity models. *Geophys. J. Int.* **2008**, *175*, 417–432. [\[CrossRef\]](#)
4. Kusche, J.; Schmidt, R.; Petrovic, S.; Rietbroek, R. Decorrelated GRACE time-variable gravity solutions by GFZ, and their validation using a hydrological model. *J. Geod.* **2009**, *83*, 903–913. [\[CrossRef\]](#)
5. Kusche, J. Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models. *J. Geod.* **2007**, *81*, 733–749. [\[CrossRef\]](#)
6. Horvath, A.; Murböck, M.; Pail, R.; Horwath, M. Decorrelation of GRACE time variable gravity field solutions using full covariance information. *Geosciences* **2018**, *8*, 323. [\[CrossRef\]](#)
7. Save, H.; Bettadpur, S.; Tapley, B.D. Reducing errors in the GRACE gravity solutions using regularization. *J. Geodesy* **2012**, *86*, 695–711. [\[CrossRef\]](#)
8. Swenson, S.C.; Wahr, J. Post-processing removal of correlated errors in GRACE data. *Geophys. Res. Lett.* **2006**, *33*. [\[CrossRef\]](#)
9. Duan, X.; Guo, J.; Shum, C.K.; Van Der Wal, W. On the postprocessing removal of correlated errors in GRACE temporal gravity field solutions. *J. Geodesy* **2009**, *83*, 1095–1106. [\[CrossRef\]](#)
10. Zhang, Z.-Z.; Chao, B.F.; Lu, Y.; Hsu, H.-T. An effective filtering for GRACE time-variable gravity: Fan filter. *Geophys. Res. Lett.* **2009**, *36*. [\[CrossRef\]](#)
11. Crowley, J.W.; Huang, J. A least-squares method for estimating the correlated error of GRACE models. *Geophys. J. Int.* **2020**, *221*, 1736–1749. [\[CrossRef\]](#)
12. Prevost, P.; Chanard, K.; Fleitout, L.; Calais, E.; Walwer, D.; van Dam, T.; Ghil, M. Data-adaptive spatio-temporal filtering of GRACE data. *Geophys. J. Int.* **2019**, *219*, 2034–2055. [\[CrossRef\]](#)
13. Ditmar, P.; Tangdamrongsub, N.; Ran, J.; Klees, R. Estimation and reduction of random noise in mass anomaly time-series from satellite gravity data by minimization of month-to-month year-to-year double differences. *J. Geodyn.* **2018**, *119*, 9–22. [\[CrossRef\]](#)
14. Piretzidis, D.; Sra, G.; Karantaidis, G.; Sideris, M.G.; Kabirzadeh, H. Identifying presence of correlated errors using machine learning algorithms for the selective de-correlation of GRACE harmonic coefficients. *Geophys. J. Int.* **2018**, *215*, 375–388. [\[CrossRef\]](#)
15. Belda, S.; Garcia-Garcia, D.; Ferrandiz, J.M. On the decorrelation filtering of RL05 GRACE data for global applications. *Geophys. J. Int.* **2015**, *200*, 173–184. [\[CrossRef\]](#)

16. Zhan, J.-G.; Wang, Y.; Shi, H.-L.; Chai, H.; Zhu, C.-D. Removing correlative errors in GRACE data by the smoothness priors method. *Chin. J. Geophys. -Chin. Ed.* **2015**, *58*, 1135–1144. [\[CrossRef\]](#)
17. Wang, F.; Shen, Y.; Chen, T.; Chen, Q.; Li, W. Improved multichannel singular spectrum analysis for post-processing GRACE monthly gravity field models. *Geophys. J. Int.* **2020**, *223*, 825–839. [\[CrossRef\]](#)
18. Eom, J.; Seo, K.-W.; Lee, C.-K.; Wilson, C.R. Correlated error reduction in GRACE data over Greenland using extended empirical orthogonal functions. *J. Geophys. Res. -Solid Earth.* **2017**, *122*, 5578–5590. [\[CrossRef\]](#)
19. Sasgen, I.; Martinec, Z.; Fleming, K. Wiener optimal filtering of GRACE data. *Studia Geophys. Geod.* **2006**, *50*, 499–508. [\[CrossRef\]](#)
20. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Society. Ser. B Methodol.* **1996**, *58*, 267–288. [\[CrossRef\]](#)
21. Hastie, T.; Tibshirani, R.; Wainwright, M. *Statistical Learning with Sparsity: The Lasso and Generalizations*; CRC Press: London, UK, 2015.
22. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.
23. Burnham, K.P.; Anderson, D.R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*; Springer: New York, NY, USA, 2002.
24. Qian, N.; Chang, G. Optimal filtering for state space model with time-integral measurements. *Measurement* **2021**, *176*, 109209. [\[CrossRef\]](#)
25. Amiri-Simkooei, A. Formulation of L1 Norm Minimization in Gauss-Markov Models. *J. Surv. Eng.* **2003**, *129*, 37–43. [\[CrossRef\]](#)
26. Beck, A.; Teboulle, M. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *Siam J. Imaging Sci.* **2009**, *2*, 183–202. [\[CrossRef\]](#)
27. Zou, H. The adaptive Lasso and its oracle properties. *J. Am. Stat. Assoc.* **2006**, *101*, 1418–1429. [\[CrossRef\]](#)
28. Chang, G.; Qian, N.; Chen, C.; Gao, J. Precise instantaneous velocimetry and accelerometry with a stand-alone GNSS receiver based on sparse kernel learning. *Measurement* **2020**, *159*, 107803. [\[CrossRef\]](#)
29. Qian, N.; Chang, G.; Gao, J. Smoothing for continuous dynamical state space models with sampled system coefficients based on sparse kernel learning. *Nonlinear Dyn.* **2020**, *100*, 3597–3610. [\[CrossRef\]](#)
30. Yang, L.; Chang, G.; Qian, N.; Gao, J. Improved atmospheric weighted mean temperature modeling using sparse kernel learning. *GPS Solut.* **2021**, *25*, 28. [\[CrossRef\]](#)
31. Parikh, N.; Boyd, S. Proximal algorithms. *Found. Trends Optim.* **2014**, *1*, 127–239. [\[CrossRef\]](#)
32. Golub, G.H.; Heath, M.T.; Wahba, G. Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics* **1979**, *21*, 215–223. [\[CrossRef\]](#)
33. Craven, P.; Wahba, G. Smoothing noisy data with spline functions. *Numer. Math.* **1979**, *31*, 377–403. [\[CrossRef\]](#)
34. Zou, H.; Hastie, T.; Tibshirani, R. On the “degrees of freedom” of the lasso. *Ann. Stat.* **2007**, *35*, 2173–2192. [\[CrossRef\]](#)
35. Kusche, J. A Monte-Carlo technique for weight estimation in satellite geodesy. *J. Geod.* **2003**, *76*, 641–652. [\[CrossRef\]](#)
36. Koch, K.R.; Kusche, J. Regularization of geopotential determination from satellite data by variance components. *J. Geod.* **2002**, *76*, 259–268. [\[CrossRef\]](#)
37. Xu, P.; Fukuda, Y.; Liu, Y. Multiple Parameter Regularization: Numerical Solutions and Applications to the Determination of Geopotential from Precise Satellite Orbits. *J. Geod.* **2006**, *80*, 17–27. [\[CrossRef\]](#)
38. Qian, N.; Chang, G.; Gao, J.; Pan, C.; Yang, L.; Li, F.; Yu, H.; Bu, J. Vehicle’s Instantaneous Velocity Reconstruction by Combining GNSS Doppler and Carrier Phase Measurements Through Tikhonov Regularized Kernel Learning. *IEEE Trans. Veh. Technol.* **2021**, *70*, 4190–4202. [\[CrossRef\]](#)
39. Breiman, L. Better subset regression using the nonnegative garrote. *Technometrics* **1995**, *37*, 373–384. [\[CrossRef\]](#)
40. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control.* **1974**, *19*, 716–723. [\[CrossRef\]](#)
41. Tikhonov, A.N.; Arsenin, V.Y. *Solutions of Ill-Posed Problems*; John Wiley & Sons: New York, NY, USA, 1977.
42. Donoho, D.L. For most large underdetermined systems of linear equations the minimal L1-norm solution is also the sparsest solution. *Commun. Pure Appl. Math.* **2006**, *59*, 797–829. [\[CrossRef\]](#)
43. Khodabandeh, A.; Amiri-Simkooei, A. Recursive algorithm for L1 norm estimation in linear models. *J. Surv. Eng.* **2011**, *137*, 1–8. [\[CrossRef\]](#)
44. Save, H.; Bettadpur, S.; Tapley, B.D. High-resolution CSR GRACE RL05 mascons. *J. Geophys. Res. Solid Earth* **2016**, *121*, 7547–7569. [\[CrossRef\]](#)
45. Watkins, M.M.; Wiese, D.N.; Yuan, D.-N.; Boening, C.; Landerer, F.W. Improved methods for observing Earth’s time variable mass distribution with GRACE using spherical cap mascons. *J. Geophys. Res.-Solid Earth* **2015**, *120*, 2648–2671. [\[CrossRef\]](#)
46. Cheng, M.; Tapley, B.D.; Ries, J.C. Deceleration in the Earth’s oblateness. *J. Geophys. Res. Solid Earth* **2013**, *118*, 740–747. [\[CrossRef\]](#)
47. Swenson, S.; Chambers, D.; Wahr, J. Estimating geocenter variations from a combination of GRACE and ocean model output. *J. Geophys. Res. Solid Earth* **2008**, *113*, B8. [\[CrossRef\]](#)
48. Peltier, W.R.; Argus, D.F.; Drummond, R. Comment on “An Assessment of the ICE-6G_C (VM5a) Glacial Isostatic Adjustment Model” by Purcell et al. *J. Geophys. Res.-Solid Earth* **2018**, *123*, 2019–2028. [\[CrossRef\]](#)
49. Wahr, J.; Molenaar, M.; Bryan, F. Time variability of the Earth’s gravity field: Hydrological and oceanic effects and their possible detection using GRACE. *J. Geophys. Res.-Solid Earth* **1998**, *103*, B12. [\[CrossRef\]](#)
50. Luthcke, S.B.; Rowlands, D.D.; Lemoine, F.G.; Klosko, S.M.; Chinn, D.; McCarthy, J.J. Monthly spherical harmonic gravity field solutions determined from GRACE inter-satellite range-rate data alone. *Geophys. Res. Lett.* **2006**, *33*. [\[CrossRef\]](#)

51. Chen, Q.; Shen, Y.; Chen, W.; Zhang, X.; Hsu, H. An improved GRACE monthly gravity field solution by modeling the non-conservative acceleration and attitude observation errors. *J. Geod.* **2016**, *90*, 503–523. [[CrossRef](#)]
52. Ditmar, P. Conversion of time-varying Stokes coefficients into mass anomalies at the Earth's surface considering the Earth's oblateness. *J. Geod.* **2018**, *92*, 1401–1412. [[CrossRef](#)]
53. Ditmar, P. How to quantify the accuracy of mass anomaly time-series based on grace data in the absence of knowledge about true signal? *J. Geod.* **2022**. *under review*.
54. Feng, W. GRAMAT: A comprehensive Matlab toolbox for estimating global mass variations from GRACE satellite data. *Earth Sci. Inform.* **2019**, *12*, 389–404. [[CrossRef](#)]
55. Wessel, P.; Smith, W.H.; Scharroo, R.; Luis, J.; Wobbe, F. Generic mapping tools: Improved version released. *Eos Trans. Am. Geophys. Union* **2013**, *94*, 409–410. [[CrossRef](#)]