

Journal of
Applied Remote Sensing

RemoteSensing.SPIEDigitalLibrary.org

**Uncertainties of the Gravity
Recovery and Climate Experiment
time-variable gravity-field solutions
based on three-cornered hat method**

Vagner G. Ferreira
Henry D. C. Montecino
Caleb I. Yakubu
Bernhard Heck

Uncertainties of the Gravity Recovery and Climate Experiment time-variable gravity-field solutions based on three-cornered hat method

Vagner G. Ferreira,^{a,*} Henry D. C. Montecino,^{b,c} Caleb I. Yakubu,^a and Bernhard Heck^d

^aHohai University, Jiangning Campus, School of Earth Sciences and Engineering, Institute of Surveying and Engineering, 8 Focheng Xilu, Nanjing 211100, China

^bUniversity of Concepción, Department of Geodetic Sciences and Geomatics, 0201 Juan Antonio Coloma, Los Ángeles 4451032, Chile

^cFederal University of Paraná, Sector of Earth Sciences, Department of Geomatics, 210 Avenida Coronel Francisco Heráclito dos Santos, Jardim das Américas, Curitiba 81.531-990, Brazil

^dKarlsruhe Institute of Technology, Geodetic Institute, 7 Englerstraße, Karlsruhe 76128, Germany

Abstract. Currently, various satellite processing centers produce extensive data, with different solutions of the same field being available. For instance, the Gravity Recovery and Climate Experiment (GRACE) has been monitoring terrestrial water storage (TWS) since April 2002, while the Center for Space Research (CSR), the Jet Propulsion Laboratory (JPL), the GeoForschungsZentrum (GFZ), and the Groupe de Recherche de Géodésie Spatiale (GRGS) provide individual monthly solutions in the form of Stokes coefficients. The inverted TWS maps (or the regionally averaged values) from these coefficients are being used in many applications; however, as no ground truth data exist, the uncertainties are unknown. Consequently, the purpose of this work is to assess the quality of each processing center by estimating their uncertainties using a generalized formulation of the three-cornered hat (TCH) method. Overall, the TCH results for the study period of August 2002 to June 2014 indicate that at a global scale, the CSR, GFZ, GRGS, and JPL presented uncertainties of 9.4, 13.7, 14.8, and 13.2 mm, respectively. At a basin scale, the overall good performance of the CSR was observed at 91 river basins. The TCH-based results were confirmed by a comparison with an ensemble solution from the four GRACE processing centers. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JRS.10.015015](https://doi.org/10.1117/1.JRS.10.015015)]

Keywords: generalized N -cornered hat method; measurement error; terrestrial water storage; uncertainties.

Paper 15579 received Aug. 14, 2015; accepted for publication Feb. 2, 2016; published online Feb. 25, 2016.

1 Introduction

The twin satellite mission Gravity Recovery and Climate Experiment (GRACE) was launched on March 17, 2002. In orbit for the past 13 years, the satellites have been providing global models of the time-varying gravity field of the Earth. As is well known, the composition and structure of the Earth, including the distribution of the atmosphere and water mass on and below the Earth's surface (e.g., rivers, lakes, soil moisture storage, groundwater storage, and ice including snow), are reflected by the Earth's gravity field.¹ In this regard, the sensitivity of the GRACE satellite mission to detecting changes in the continental water cycle has clearly been demonstrated over the last years (see, e.g., Ref. 2 and references therein). GRACE has been widely used for estimating terrestrial water storage anomalies (TWSA) and changes, for instance, in the Amazon,³

*Address all correspondence to: Vagner G. Ferreira, E-mail: vagnergf@hhu.edu.cn

Yangtze,⁴ Volta,⁵ Nile,⁶ and other river basins. Additional studies from GRACE are related to monitoring the groundwater withdrawal in India⁷ and China,⁸ contributions of glaciers and ice caps to sea-level rise,⁹ and evaluation of naturally and anthropogenically induced variations in water availability in Africa.¹⁰ In this work, however, the aim is to assess the quality of the GRACE-derived TWSA fields inverted from the level-2 products that were calculated by four different centers. Three of these centers are identified in the mission proposal as the GRACE Science Data System (SDS), while the other is the solution provided by the Groupe de Recherche de Géodésie Spatiale (GRGS) at the Centre National d'Études Spatiales (CNES).^{11,12} The three SDS centers are the University of Texas Center for Space Research (CSR),¹³ the NASA Jet Propulsion Laboratory (JPL),¹⁴ and the Deutsches GeoForschungsZentrum (GFZ).¹⁵

Error estimates for space-borne sensors often rely on ground truth validation;¹⁶ therefore, in the absence of such data, the assessment of the uncertainties of satellite-based products is a challenge. This applies especially to regions such as Africa and South America for which data are limited. This is even worse for the GRACE-derived TWSA fields, for which there are no direct TWSA measurements to validate the satellite-derived measurements. Uncertainties of GRACE-derived TWSA fields have been estimated by fitting and removing a constant and an annual cycle from the monthly values of each Stokes coefficient.^{16,17} For instance, based on 22 monthly GRACE gravity field solutions, Wahr et al.¹⁷ have estimated the global area weighted root mean square error (RMSE) of TWSA at 21 mm when filtered with a 750 km Gaussian smoothing radius. Additionally, they have reported that the errors decrease as the radius increases, falling from 38 mm at 500 km to 15 mm at 1000 km. Recently, Sakumura et al.¹⁸ removed the majority of the signals attributed to seasonal, subseasonal (semiannual and 161-day cycle), and linear variations as a means of assessing the four different data centers (i.e., CSR, GFZ, JPL, and the GRGS). These authors have reported an RMSE of 12.3 mm for CSR, 11.3 mm for GFZ, 14.4 mm for JPL, and 13.0 mm for GRGS at a global scale. At a basin scale, irrespective of the size and relative signal amplitude, the solutions all vary from the ensemble mean (composed from CSR, JPL, GFZ, and GRGS) on the order of 10 to 15 mm.¹⁸

Efforts have also been made to assess the GRACE-derived TWSA fields; for example, combining measurements and/or models in the context of the water-budget equations^{19–23} or in terms of groundwater estimations.^{8,24–26} Additionally, direct comparisons with hydrology models,^{27,28} remotely sensed rainfall data,^{29,30} and ocean-bottom pressure products³¹ have been reported. In practice, however, the data considered to represent the truth in these validations are subject to measurement errors and/or spatial scale mismatching. Nevertheless, these regional comparisons show typical values of <25 mm, which appears to indicate the level of accuracy of GRACE solutions in terms of TWSA. While the estimation of GRACE errors could be improved by using the full covariance matrix³² and errors in the background models,³³ a comparison of the GRACE measurements with true observations of the target quantity would be more interesting for validation purposes. However, as mentioned in the previous paragraph, there are no such *in situ* data for assessing GRACE-derived TWSA fields. The three-cornered hat (TCH) method can offer an alternative approach to quantify the uncertainty of each GRACE processing center. Originally applied to assessing the relative precision of oscillators and timing devices,³⁴ this method has recently been applied to the fields of geodesy^{35–38} and hydrology.^{39,40}

The TCH method is based on a difference approach relying on the removal of common signals from the observations (i.e., the truth), then providing the uncertainties that reflect the measurement errors.³⁸ According to Chin et al.,³⁵ the method is algebraically simple to apply, at least to the three datasets that have statistically independent measurement error processes. However, the generalized formulation of the TCH method, as proposed by Premoli and Tavella,⁴¹ considers the correlations among the noises of the time series by minimizing the global correlation among them. In this investigation, the generalized formulation of the TCH method, taking into account the correlations, is applied to assess the relative quality of GRACE-derived TWSA fields from the three SDS processing centers and the GRGS solution. Although using the mean of the four (or more) centers is recommended (see, e.g., Ref. 24), it is still necessary to assess the individual uncertainties of each processing center. This is important to guide the GRACE processing centers and help the end users by confirming that the accuracy and precision of the estimates lie within a required specification.

2 Methodology and Data

2.1 Three-Cornered Hat Method

In the absence of a reference dataset, the TCH method can be used to estimate the relative uncertainties of the GRACE-derived TWSA from different sources if at least three products are available. To estimate the uncertainty in the TWSA datasets, consider the time series of the available products stored as $\{X_i\}_{i=1,2,\dots,N}$, where i corresponds to each solution center (that is, $N = 4$, three products from the GRACE SDS processing centers, CSR, JPL, and GFZ, and one from the GRGS solution), and split each time series as

$$X_i = S + \varepsilon_i \quad \forall i = 1, \dots, N, \quad (1)$$

where S is the true signal and ε_i is a zero-mean white noise process (representing the measurement error³⁵), here representing the noise deviation of the GRACE processing center i . Since no true estimate of S is available, the differences between $N - 1$ processing centers and one center designated as the reference (chosen arbitrarily) can be computed as³⁶

$$Y_{iN} \equiv X_i - X_N = \varepsilon_i - \varepsilon_N \quad \forall i = 1, \dots, N - 1, \quad (2)$$

where X_N is the reference time series. The CSR-derived TWSA time series was selected as the reference series. Nevertheless, the results are independent of the special choice of a particular GRACE processing center (see, e.g., Refs. 36 and 40).

The samples of the $N - 1$ solution centers' differences [Eq. (2)] are concatenated in an $M \times (N - 1)$ matrix as

$$Y = [Y_{1N} \quad Y_{2N} \quad \dots \quad Y_{(N-1)N}], \quad (3)$$

where each row contains a monthly observation (here, $M = 143$; i.e., 143 months from August 2002 to June 2014, after making provision for the missing months as shown in Sec. 3.2), and each column represents the difference between each time series and the reference. The covariance matrix S of the series of differences is computed as

$$S = \text{cov}(Y), \quad (4)$$

where $\text{cov}(\bullet)$ is the covariance operator. A generic element of S represents either a variance estimate (for $i = j$) or a covariance estimate (for $i \neq j$). The $N \times N$ covariance matrix (Allan covariance matrix) of the individual noises R , whose elements are the unknowns of the problem and should be determined, is related to S by⁴²

$$S = H^T \cdot R \cdot H \quad \text{with} \quad H = \begin{bmatrix} I \\ -\mathbf{u}^T \end{bmatrix}, \quad (5)$$

where I is the $(N - 1) \times (N - 1)$ identity matrix and \mathbf{u} is the $(N - 1)$ vector $[1 \ 1 \ \dots \ 1]^T$.

Equation (5) can be rewritten as⁴³

$$S = [I - \mathbf{u}] \begin{bmatrix} \hat{R} & \mathbf{r} \\ \mathbf{r}^T & r_{NN} \end{bmatrix} \begin{bmatrix} I \\ -\mathbf{u}^T \end{bmatrix}, \quad (6)$$

where \hat{R} is the $(N - 1) \times (N - 1)$ submatrix, \mathbf{r} is the $(N - 1)$ vector $[r_{1N} r_{2N} \dots r_{N-1,N}]^T$ grouping the covariance estimates that involve the N 'th time series, and r_{NN} is the variance of the N 'th series (i.e., the reference series). The partitioning of R as in Eq. (6) is necessary to solve the underdetermined problem in Eq. (5) by isolating the N free parameters (that is, \mathbf{r} and r_{NN} —specifically, $r_{1N}, r_{2N}, \dots, r_{N-1,N}$, and r_{NN}). Once the free parameters have been estimated, the solution for the other unknown elements of R (that is, the elements of \hat{R}) is given by

$$\hat{R} = S - r_{NN}[\mathbf{u}\mathbf{u}^T] + \mathbf{u}\mathbf{r}^T + \mathbf{r}\mathbf{u}^T. \quad (7)$$

To determine the N free parameters, a suitable objective function can be defined, which must always fulfill the positive definiteness of R . This important constraint on the solution domain for

the free parameters is valid if and only if $\det(R) > 0$, as shown by Tavella and Premoli.⁴³ However, Premoli and Tavella⁴¹ have pointed out that this condition is not sufficient to determine a unique solution for R . They have proposed an optimal choice criteria for the free parameters based on the minimization of the global correlation among the noises of the involved time series, maintaining the positive definiteness of R . Galindo and Palacio⁴⁴ have suggested the constrained minimization problem aiming to reach a unique solution (minimum of the quadratic mean of covariances) using the Kuhn-Tucker theorem, where the objective function F is given by⁴⁴

$$F(\mathbf{r}, r_{NN}) = \sum_{i:i} \frac{r_{ij}^2}{K^2}, \tag{8}$$

with a constraint function⁴⁴

$$G(\mathbf{r}, r_{NN}) \equiv -\frac{r_{NN} - [\mathbf{r} - r_{NN}\mathbf{u}]^T \cdot S^{-1} \cdot [\mathbf{r} - r_{NN}\mathbf{u}]}{K} < 0, \tag{9}$$

where K denotes $K = \sqrt[n-1]{\det(S)}$. The initial conditions were selected to assure that the initial values fulfill the constraints

$$r_{iN}^{(0)} = 0, \quad i < N \quad \text{and} \quad r_{NN}^{(0)} = (2 \cdot \mathbf{u}^T \cdot S^{-1} \cdot \mathbf{u})^{-1}, \tag{10}$$

as proposed by Torcaso et al.⁴⁵ After determining the free parameters ($r_{1N}, r_{2N}, \dots, r_{N-1,N}$, and r_{NN}) by minimizing Eq. (8), the remaining unknown elements of R can be determined by Eq. (7).

2.2 Gravity Recovery and Climate Experiment Derived Terrestrial Water Storage Anomalies

2.2.1 Terrestrial water storage anomaly computations

The shape of the Earth’s gravity field is a geoid, the equipotential surface that best fits, in a least-squares sense, the global mean sea level. Consequently, variations in gravity are equivalent to variations in the geoid. Temporal variations of the geoid, predominantly resulting from changes in water mass (considering that the atmospheric and oceanic contributions have been removed from GRACE measurements during data processing for dealiasing purposes) can be inverted to surface mass anomalies $\Delta\sigma$, which is expressed as⁴⁶

$$\Delta\sigma(\theta, \lambda, t) = \sum_{n=0}^{n_{\max}} \sum_{m=0}^n [\Delta\tilde{C}_{nm}(t) \cos m\lambda + \Delta\tilde{S}_{nm}(t) \sin m\lambda] \bar{P}_{nm}(\cos \theta), \tag{11}$$

in kg/m^2 for a particular month t . \bar{P}_{nm} is the normalized associated Legendre function of degree n and order m , θ is the co-latitude, λ is the longitude, and $\Delta\tilde{C}_{nm}$ and $\Delta\tilde{S}_{nm}$ are the residual surface density coefficients, which are given by⁴⁶

$$\begin{bmatrix} \Delta\tilde{C}_{nm}(t) \\ \Delta\tilde{S}_{nm}(t) \end{bmatrix} = R \frac{\rho_e}{3} \frac{2n+1}{1+k_n} \begin{bmatrix} \Delta\bar{C}_{nm}(t) \\ \Delta\bar{S}_{nm}(t) \end{bmatrix}, \tag{12}$$

where ρ_e is the average density of the Earth, R is the mean radius of the Earth (6371×10^3 m), and k_n is the load Love number for degree n . The residual spherical harmonic coefficients, which describe the functionals of the Earth’s gravity field (also known as Stokes coefficients), $\Delta\bar{C}_{nm}(t)$ and $\Delta\bar{S}_{nm}(t)$, are defined as

$$\begin{bmatrix} \Delta\bar{C}_{nm}(t) \\ \Delta\bar{S}_{nm}(t) \end{bmatrix} = \begin{bmatrix} \bar{C}_{nm}(t) \\ \bar{S}_{nm}(t) \end{bmatrix} - \frac{1}{M} \sum_{t=1}^M \begin{bmatrix} \bar{C}_{nm}(t) \\ \bar{S}_{nm}(t) \end{bmatrix}, \tag{13}$$

where the long-term mean of Stokes coefficients \bar{C}_{nm} and \bar{S}_{nm} is removed from each monthly value to exclude the static gravitational field. In Eq. (13), M is the total number of monthly solutions (134 months from August 2002 till July 2014; nine months are missing).

Equation (11) is used for computing the TWSA for each grid point defined by θ and λ ; thus, one can use the grid for averaging the TWSA over a particular region (e.g., a river basin). Another approach for averaging the TWSA over a river basin is based on the exact averaging kernel, $\vartheta(\theta, \lambda)$, which is a function that describes the shape of the basin as⁴⁷

$$\vartheta(\theta, \lambda) = \begin{cases} 0 & \text{outside the basin} \\ 1 & \text{inside the basin} \end{cases} \quad (14)$$

The function $\vartheta(\theta, \lambda)$ can be expanded in spherical harmonic coefficients, ϑ_{nm}^c and ϑ_{nm}^s , as

$$\begin{Bmatrix} \vartheta_{nm}^c \\ \vartheta_{nm}^s \end{Bmatrix} = \frac{1}{4\pi} \int_0^{2\pi} \int_0^\pi \vartheta(\theta, \lambda) \bar{P}_{nm}(\cos \theta) \begin{Bmatrix} \cos m\lambda \\ \sin m\lambda \end{Bmatrix} \sin \theta d\theta d\lambda. \quad (15)$$

Using ϑ_{nm}^c and ϑ_{nm}^s , which describe $\vartheta(\theta, \lambda)$, the averaged surface mass anomalies $\overline{\Delta\sigma}$ can be expressed by a sum of Stokes coefficients as⁴⁷

$$\overline{\Delta\sigma}(t) = \frac{1}{\Omega} \sum_{n=0}^{n_{\max}} \sum_{m=0}^n [\vartheta_{nm}^c \Delta\tilde{C}_{nm}(t) + \vartheta_{nm}^s \Delta\tilde{S}_{nm}(t)], \quad (16)$$

where Ω , the angular area of the region of interest over the sphere, is equal to the ϑ_{00}^c coefficient. The use of Eq. (16) is recommended because, as observed by Swenson and Wahr,⁴⁷ an average calculated using the exact representation of the basin shape is influenced by the satellite measurement errors, which rapidly increase for high degrees and orders (short wavelengths).

2.2.2 Gravity Recovery and Climate Experiment level 2 products

Gravity field solutions (static and time-variable gravity fields) are obtained from the level 1 data after further processing. As few users can construct their own solutions from the level 1 product, the GRACE mission partner institutions (CSR, JPL, and GFZ) and CNES/GRGS provide monthly solutions in the form of Stokes coefficients known as level 2 (L2) data to some degree and order (d/o), freely available on the Internet (e.g., Ref. 48). Additionally, institutions such as the Delft Institute of Earth Observation and Space Systems, Delft University of Technology, Netherlands, and Tongji University, China, provide independent solutions based on alternative approaches. Each center processes the data using different methodology, which may result in differences in their solutions.²⁸ Additionally, Wouters et al.⁴⁹ have reported that the differences in the approaches of the processing centers are attributed to the background models used (e.g., an *a priori* model of the Earth's gravity field, force models representing luni-solar and third body tides, and gravitational effects of the ocean and atmospheric mass variations), the period over which the orbits are integrated, weighting of the data, etc.

In this study, however, the monthly solutions generated by the SDS processing centers (CSR Release 05,¹³ GFZ Release 05,¹⁵ and JPL Release 05.1¹⁴) and GRGS Release 03-v1¹² were used. The processing center solutions of CSR, GFZ, GRGS, and JPL are computed based on the integration of variational equations (dynamic method), where monthly solutions (GFZ also provides weekly solutions) have been estimated throughout the whole GRACE mission period. Currently, an alternative approach using radial basis functions is under investigation at GFZ with promising results.⁵⁰ Despite the approach being the same, possible differences can be related to, for example, the background models. As such, it is important to determine the noise deviations of each solution relative to an unknown true value (see Sec. 2.1) to assess their strengths and weaknesses. The Stokes coefficients from the four processing centers were considered up to d/o 60 in Eq. (11) (i.e., $n_{\max} = 60$, which can be associated with the shortest resolution as ρ [km] = $(\pi/n_{\max})R$ [km] at the equator). The degree 1 coefficients ($C_{1,0}$, $C_{1,1}$, and $S_{1,1}$), which represent the changes in the geocenter, and the degree 2 coefficient ($C_{2,0}$), which is associated with the oblate shape of the Earth, were not considered here. Since they would be a common-mode signal for the four centers, through Eq. (2), the contribution of these low-degree coefficients is canceled out.

GRACE gravity fields at high-degree coefficients exhibit a high level of noise, which is known as stripes at the spatial domain.⁵¹ Therefore, to obtain coherent results, it is necessary to remove these stripes in postprocessing by reducing correlated errors with a minimal impact on the real signal. These correlations can be reduced by using an *a priori* synthetic model of the observation geometry, as suggested by Kusche in Ref. 52. The filtered coefficients, $\mathbf{x}_{\gamma(a)}$, are obtained by⁵²

$$\mathbf{x}_{\gamma(a)} = (N + aM)^{-1}b = (N + aM)^{-1}N\mathbf{x} = W_a\mathbf{x}, \quad (17)$$

using the weighting factor a to the signal-covariance matrix M [see Eq. (20) of Ref. 52]. In Eq. (17), \mathbf{x} is the vector of unfiltered Stokes coefficients, N is the normal matrix, and $E\{\mathbf{x}\mathbf{x}^T\} = M^{-1}$ is an *a priori* signal covariance matrix. Parameter a can be adjusted to tune the smoothness of the solution. Here, a value of $a = 1 \times 10^{13}$ was adopted, which is called DDK2, with the level of smoothing approximately corresponding to a Gaussian smoothing radius of 340 km. An exhaustive comparison of the suitability of the filter methods available can be found in Ref. 53, and the choice of DDK2 is supported by their results and, additionally, is consistent with the results provided by Sakumura et al. in Ref. 18. Although it is not necessary to filter GRGS monthly solutions,^{11,12} since they have already been stabilized during their generation process, in this study, they have been filtered as the SDS processing centers.

3 Results and Discussion

In this section, the main results from this study are presented. Section 3.1 shows the spectral intercomparisons of L2 products (i.e., Stokes coefficients) of the four processing centers. Section 3.2 shows the results related to global comparisons of the four GRACE processing centers, and Sec. 3.3 shows the results at the basin scale. Finally, Sec. 3.4 shows a summary of the results.

3.1 Spectral Intercomparison of the Gravity Recovery and Climate Experiment Processing Centers

The relative intercomparisons pertaining to the spectral domain of the GRACE SDS processing centers and the GRGS solution in Fig. 1 show the Pearson's linear correlation coefficient (CC) in the Stokes coefficients after applying the DDK2 filter. It is possible that CSR and GFZ are more correlated to each other than, for example, CSR and JPL. CSR and GFZ show a mean correlation for Stokes coefficients $\Delta\bar{C}$ and $\Delta\bar{S}$ of 0.60 and 0.59, respectively, while between CSR and JPL, it is 0.57 and 0.57 for $\Delta\bar{C}$ and $\Delta\bar{S}$, respectively. GFZ and JPL show a relative correlation of 0.53 and 0.52 for $\Delta\bar{C}$ and $\Delta\bar{S}$, respectively. Additionally, the strength of relationships between SDS processing centers and the GRGS solution is moderate. CSR and GRGS show a mean CC of 0.41 and 0.41 for $\Delta\bar{C}$ and $\Delta\bar{S}$, respectively; GFZ and GRGS show a mean CC of 0.41 and 0.42 for $\Delta\bar{C}$ and $\Delta\bar{S}$, respectively; JPL and GRGS show a mean CC of 0.38 and 0.38 for $\Delta\bar{C}$ and $\Delta\bar{S}$, respectively. It is possible to see that for the sectorial ($n = m$) coefficients with d/o higher than 20 and for coefficients with d/o higher than 40, the CC values are weak for SDS centers and GRGS comparison pairs.

Additionally, Fig. 2 shows the Nash-Sutcliffe efficiency (NSE) coefficient [for details, see Eq. (18) of Ref. 54]. The NSE coefficient ranges between $-\infty$ and 1, where the coefficient tends to 1 provided that the phase, amplitude, and mean of the simulated and observed time series agree. Values between 0 and 1 are generally viewed as acceptable levels of performance, whereas values <0 indicate that the mean observed value is a better predictor than the simulated value.⁵⁵ Therefore, results with NSE values <0 , which indicate unacceptable performance, were masked out in black. Similar to the correlation, NSE coefficients decrease with increasing d/o. However, it is possible to see that for the sectorial coefficients, the NSE values are <0 for all comparison pairs. Overall, the SDS processing centers show good NSE values up to degree 50 and order 20 (near zonal coefficients) among themselves. For d/o higher than 40 and 20, respectively, the relative comparison between GRGS and the SDS solutions shows low NSE values, where the majority are <0 . This could be due to amplitude reduction rather than phase shift, since NSE evaluates consistency in phase and amplitude and CC only in phase.

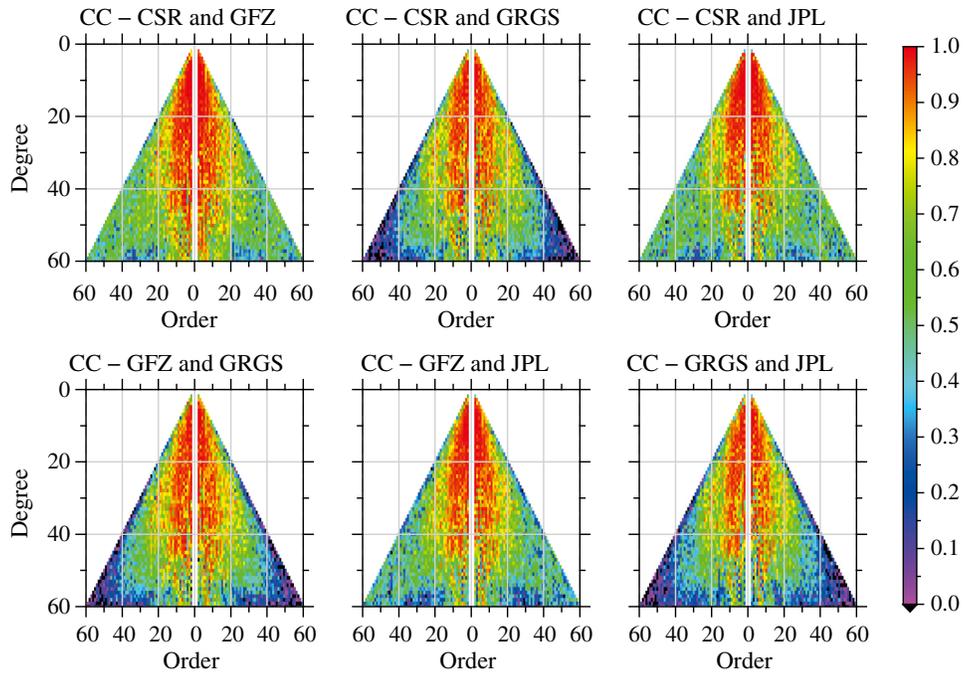


Fig. 1 CC between the GRACE SDS processing centers and the GRGS solutions for each degree and order. For each triangle, the left part shows the $\Delta\bar{C}_{nm}$ coefficients and the right part shows the ΔS_{nm} coefficients.

3.2 Global Comparisons of the Solutions

First, the global grids (1 deg \times 1 deg) of TWSA were computed using Eq. (11) for the period of August 2002 to June 2014, starting from degree 3 and order 0 up to d/o 60. Subsequently, 134 GRACE-derived TWSA monthly fields were resampled for exactly the middle of each month,

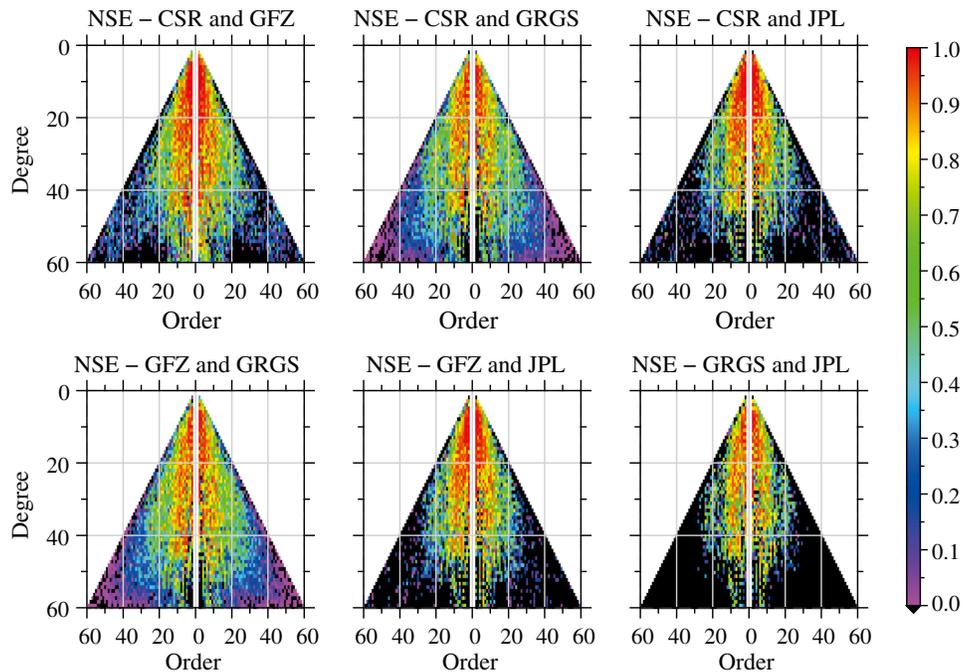


Fig. 2 NSE between the GRACE SDS processing centers and the GRGS solutions for each degree and order. For each triangle, the left part shows the $\Delta\bar{C}_{nm}$ coefficients and the right part shows the ΔS_{nm} coefficients.

and the data of the nine missing months were interpolated. This was necessary since the number of days differs by a few days among the GRACE processing centers and all the datasets must be aligned in time in order to use the TCH method. Following the computations of TWSA fields, a simultaneous fitting of a time series taking into account constant, linear, and periodic variations (annual, semiannual, and 161-day amplitudes and phases) was carried out at each grid point (θ, λ) . This was necessary to verify whether the differences/similarities from the intercomparisons in Sec. 3.1 propagated to the spatial domain in terms of linear trends and annual amplitudes. Overall, the relative comparisons of annual amplitudes and linear trends (see Fig. 3) between every set of two products showed correlations higher than 0.99 and the same patterns. As shown in Fig. 1, larger amplitude coefficients exhibited higher NSE and CC values, which implies that annual and long-term components of TWSA fields derived from these processing centers provide similar results.

The quality of the results published by the four processing centers was assessed using the uncertainties computed from the TCH method. First, Eq. (8) was minimized at each grid point to derive the four ($N = 4$) free parameters, which were used to compute the remaining elements of matrix R using Eq. (7). The diagonal elements of matrix R , which contains the noise variances ($r_{11}, r_{22}, r_{33}, r_{44}$), were then used to compute the noise deviations, which express the quality of each processing center (GFZ, GRGS, JPL, and CSR). Despite the differences between the strategies adopted by each processing center to compute the Stokes coefficients, the GRACE

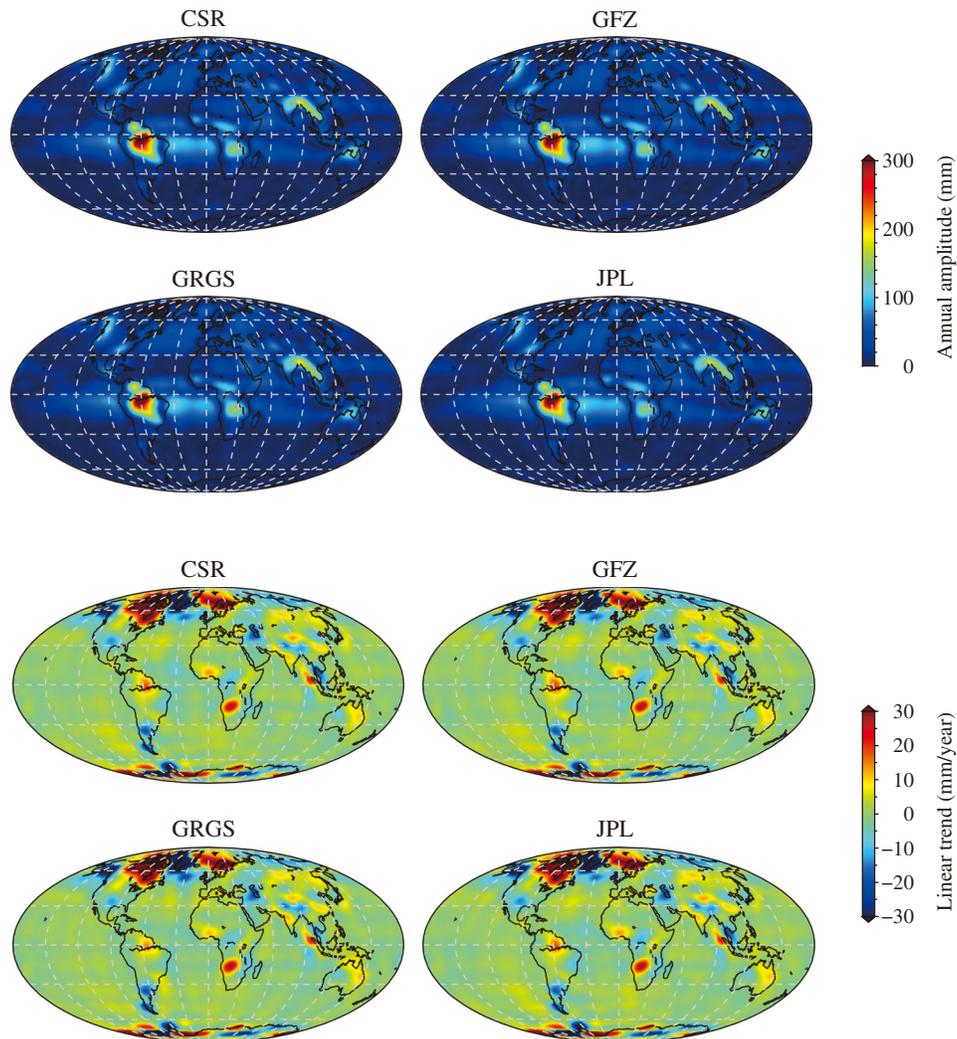


Fig. 3 Annual amplitudes (top panels) and linear trends (bottom panels) of the global fields of TWSA for the period from August 2002 to June 2014.

measurements are the same among the centers. Thus, it is possible to assume that GRACE detects the same geophysical phenomena and the common signals are canceled out by using Eq. (2), making the TCH method able to provide methodology-dependent errors of each processing center.

The results of TCH-derived uncertainties are presented in Fig. 4. It is possible to see that all GRACE processing centers (i.e., CSR, GFZ, GRGS, and JPL) present patterns of uncertainty that resemble the stripes. CSR seems to present the smallest uncertainties relative to the other three solutions, especially over the Greenland, Amazon, and Antarctic regions. The uncertainty map of GFZ shows larger overall error patterns than JPL, which shows higher uncertainties near the poles. Figure 4 also illustrates the uncertainties over the polar regions for the four processing centers; it confirms that the JPL-computed uncertainties are higher over these regions. Wahr et al.¹⁶ have reported that the errors are generally smaller over the poles than at low latitudes, apparently due to denser ground-track coverage near those regions. Furthermore, in Ref. 18, Sakumura et al. have reported that JPL-derived TWSA fields present relatively higher variations with respect to CSR over the North and South Poles. They have linked this deviation to the known motion of the sun relative to the orbital plane, which manifests as error in tides and spacecraft environment, and somehow, it is larger in the JPL solution. Additionally, CSR and JPL calculate the monthly solutions as a deviation from the GIF-48 mean geopotential model.^{13,14} However, there is a slight difference between CSR and JPL on the diurnal and semi-diurnal band of ocean tide contribution, where CSR RL05 uses the GOT4.8 and JPL RL05.1 uses the GOT4.7, which differ only in the harmonics of the S2 tide.¹³

The area-weighted uncertainties for each processing center were computed by considering the uncertainty distribution presented in Fig. 4. The results show that CSR presents a globally averaged (weighted) uncertainty of 9.4 mm, GFZ 13.7 mm, GRGS 14.8 mm, and JPL 13.2 mm. The low performance of GFZ is evident over the mid-latitude and tropic regions (Fig. 4), and

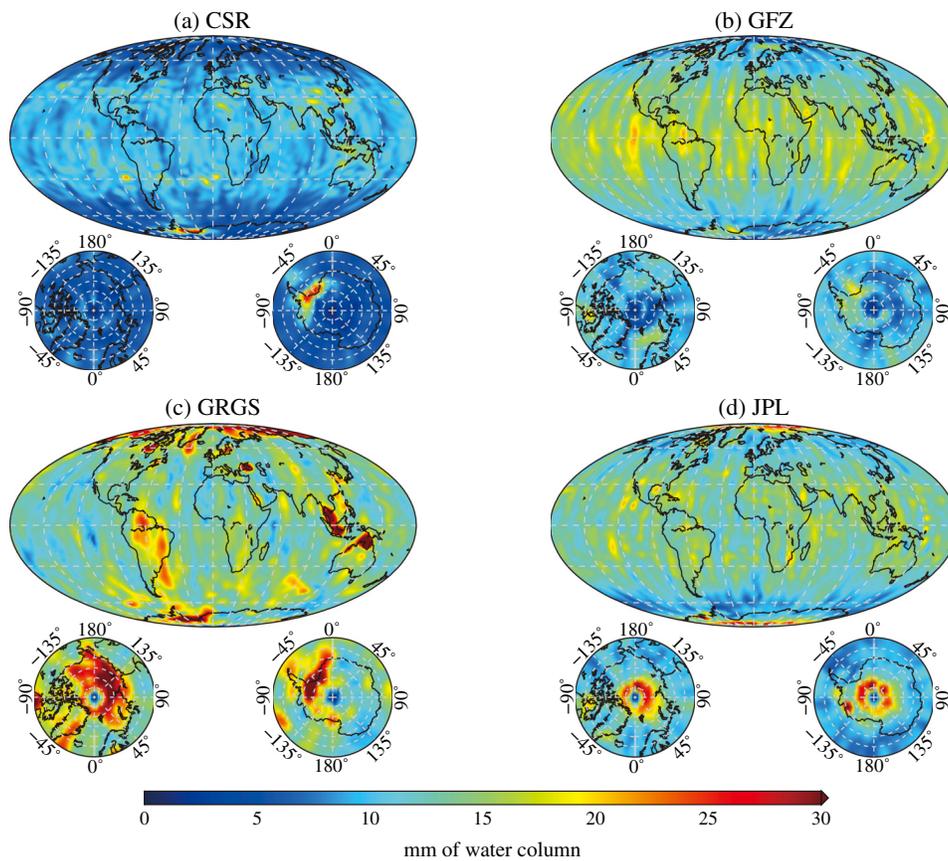


Fig. 4 Uncertainties of the GRACE-derived TWSA of the (a) CSR, (b) GFZ, (c) GRGS, and (d) JPL processing centers.

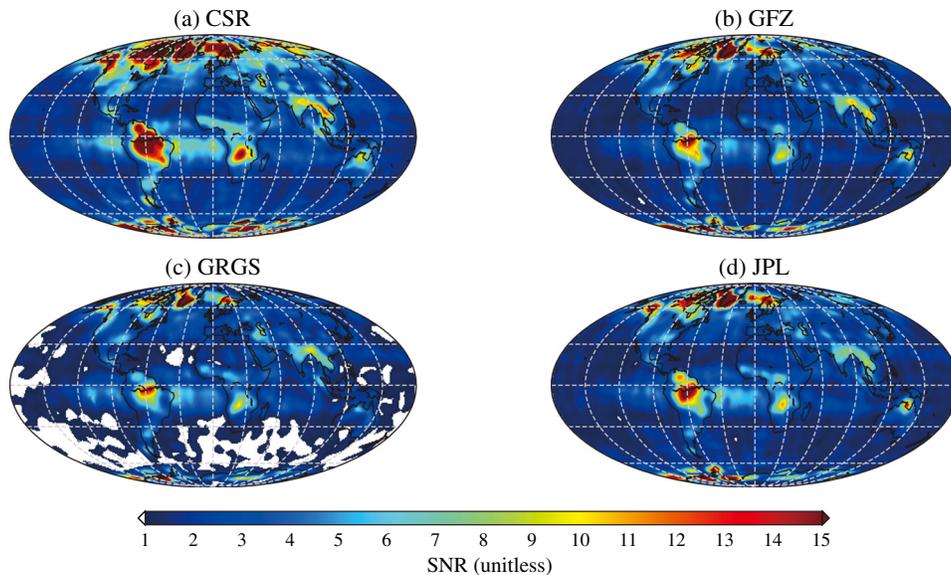


Fig. 5 SNR based on the standard deviation of each time series and the uncertainties based on TCH results (Fig. 4) for the (a) CSR, (b) GFZ, (c) GRGS, and (d) JPL processing centers. The SNR values < 1, which indicate more noise than signal, were masked out in white.

appears to be primarily due to increased striping over the oceans. Chambers and Bonin et al.³¹ have reported that the CSR, GFZ, and JPL time series are virtually the same over the oceans in terms of comparisons based on ocean-bottom pressure. Large uncertainties of GRGS are evident, particularly over regions above the latitudes ± 80 deg, Northern Australia, and Southeast Asia. It is well known that the ensemble mean is effective in reducing the noise in comparison with the members used in its computation, as shown in Ref. 18. Here, the ensemble mean using the CSR, GFZ, GRGS, and JPL series was computed and compared with each individual solution. The standard deviation of the differences was 9.4 mm for CSR, 11.5 mm for GFZ, 12.0 mm for GRGS, and 11.2 mm for JPL, which confirms the results derived by the TCH method, despite the magnitude of the uncertainties and the differences between standard deviation and Allan deviation.

However, the noises are insufficient to describe the relative quality of each processing center. Hence, for each GRACE product (i.e., CSR, GFZ, GRGS, and JPL), the signal-to-noise ratio (SNR) was estimated as the ratio between the standard deviation of the time series of TWSA and the uncertainties presented in Fig. 4; the results are presented in Fig. 5. The spatial patterns of SNRs are different from those of the noise distribution (Fig. 4), indicating higher values in regions with strong hydrological signals (e.g., Amazon basin, Congo/Zambezi basins, and Northwest India) and in regions under significant ice storage changes (e.g., Greenland and Asian high mountains). Area-weighted SNRs were computed for each processing center considering the SNRs presented in Fig. 5. The results show that CSR presents an SNR of 4.2, GFZ of 2.9, GRGS of 2.3, and JPL of 2.9, indicating the overall good performance of the CSR product relative to the other two SDS centers as well as to GRGS. However, over regions of low signal variations, such as the Sahara Desert and Gobi Desert, China, it is apparent that GFZ and JPL show smaller variations relative to their uncertainties, while CSR presents high SNRs over these regions. Again, in comparison with the ensemble mean, the mean of the SNRs shows 3.9 for CSR, 3.2 for GFZ, 2.8 for GRGS, and 3.3 for JPL.

3.3 Basin Scale Comparison of the Solutions

The river basins of the various drainage areas and locations (Fig. 6) were carefully chosen in order to quantify the relative differences of the four data centers at basin scale. The boundaries of the selected 91 basins with areas larger than 100×10^3 km² were obtained from the WRI Major Watersheds of the World Delineation, available at Ref. 56. This choice was based on the fact that

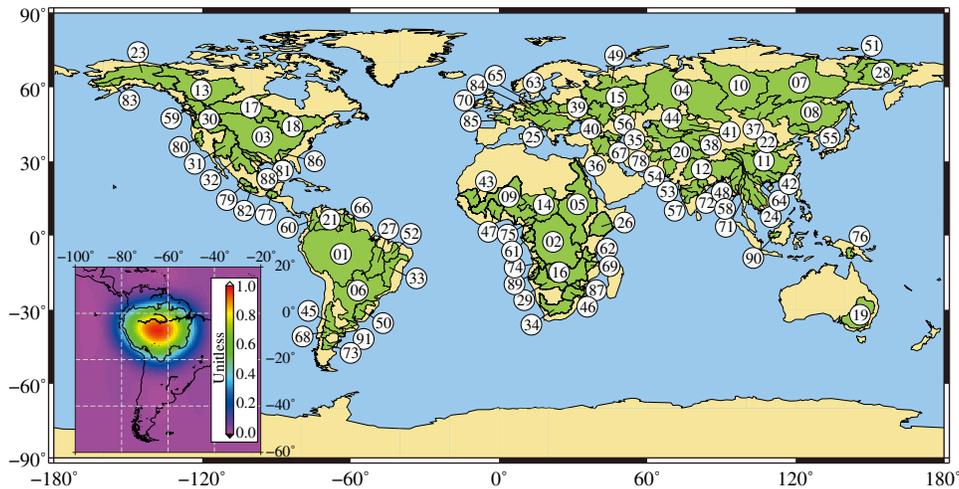


Fig. 6 Watersheds of the world larger than 100,000 km² extracted from the WRI Major Watersheds of the World Delineation. The numbers assigned here to each basin can be used to identify the names of the river basins (Table 1). The embedded panel shows the geographic mask adopted for the Amazon basin after DDK2 smoothing.

basins smaller than the spatial resolution of the GRACE-derived TWSA (~ 333.6 km at the equator, considering the expansion up to d/o 60) could not be adequately captured by GRACE. However, for anyone interested in using GRACE for basins smaller than this spatial resolution, the study of Longuevergne et al. in Ref. 57 is recommended. The basin-averaged TWSA for all 91 basins were computed by applying Eq. (16) for the period of August 2002 to June 2014, starting from degree 3 and order 0 up to d/o 60. The spherical harmonic coefficients ϑ_{nm}^c and ϑ_{nm}^s , defining a mask with the perimeter of the basins, shown in Fig. 6, were computed up to d/o 60 using Eq. (15), then filtered with the DDK2 filter scheme using Eq. (17). An example of the geographic mask is shown in Fig. 6 for the Amazon basin after filtering its spherical harmonic coefficients (ϑ_{nm}^c and ϑ_{nm}^s).

The TCH-based uncertainties are shown in Fig. 7(a), from which it is evident that CSR provides the lowest uncertainties for almost all 91 basins, and GFZ and JPL present similar results, with JPL performing slightly better than GFZ. However, GRGS presents the lowest performance compared to the SDS processing centers over the river basins considered here, particularly over basins 21, 27, 33, 52, 66, and 90. The TCH-based uncertainties of the four processing centers vary with respect to the basin size, which is in agreement with the fact that GRACE is more sensitive and precise for large river basins (see, e.g., Refs. 16, 58, and 59). In order to determine whether the relative basins' hydrological signal variations influence the magnitude of the uncertainties, SNR values for each river basin were computed [Fig. 7(b)]. It is apparent that the scatter of the SNRs is relatively decreasing; that is, it generally depends on the size of the basin. This finding somewhat agrees with the results presented in Ref. 18 (see Fig. 3), where it has been shown that the differences in the solutions are due to random errors rather than signal differences. Regardless of the size of the basin, the amplitude of the hydrological signal also plays an important role in the performance of the GRACE-derived TWSA fields. For example, the SNRs of basins 02 to 11 and 31 to 45 are relatively lower than those of the basins with smaller areas. Overall, all four centers showed high performance in recovering the hydrological variations over the basins considered here (see Table 1 for the names and Fig. 6 for the locations of all 91 basins).

Nevertheless, CSR (blue solid line) noise deviation [Fig. 7(a)] and SNR [Fig. 7(b)], for basins Saskatchewan-Nelson (No. 17), Volta (No. 47), Paranaiba (No. 52), Mississippi (No. 03), Sacramento (No. 80), Tocantins (No. 27), and Sao Francisco (No. 33), present the relatively lowest noise deviations and the largest SNRs by a factor of ~ 2 among SDS processing centers. The GRACE SDS processing centers present the highest SNR values for the Amazon basin (No. 01), because of its size and the strength of its hydrological signal (Table 1), while GRGS presents the highest SNR for the Fraser basin (No. 59). The noise deviation of the Amazon basin is 4.2 mm based on the CSR, 4.4 for JPL solutions, 7.6 mm for GFZ, and

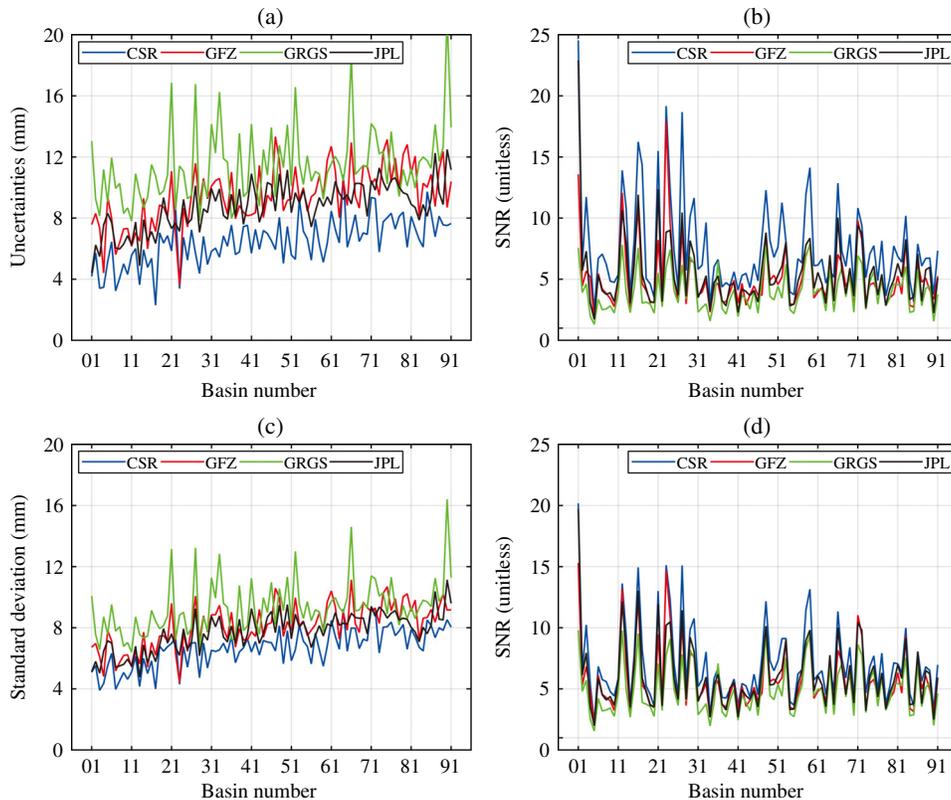


Fig. 7 (a) The TCH-based uncertainties and (b) SNRs for the 91 watersheds. (c) The ensemble mean based standard deviations and (d) the SNRs for the 91 watersheds.

13.0 mm for GRGS (Table 1). The relative comparison between the CSR and GRGS, in terms of SNR, shows that the former presents values approximately three times larger than the latter for the Ob (No. 04), Parnaíba (No. 52), Amazon (No. 01), São Francisco (No. 33), Yukon (No. 23), and Tocantins (No. 27) river basins. Additionally, for the Saskatchewan-Nelson basin (No. 17), the SNR of the CSR is four times larger than that of the GRGS.

Similar to Sec. 3.2, the ensemble solution calculated as the arithmetic mean of TWSA time series derived from the Stokes coefficients provided by the four processing centers was used to assess each ensemble member. Despite the differences between the magnitudes of the TCH and ensemble mean based uncertainties, the results shown in Fig. 7(c) are comparable to those of Fig. 7(a). It is evident that the uncertainties increase overall, while the basin size decreases. Additionally, the figure shows the overall good performance of CSR relative to that of JPL and GFZ. The second-best performance is that of JPL, followed by GFZ. The SNR results shown in Fig. 7(d) are also comparable to those of Fig. 7(b), where, taking into consideration the relative hydrological signal, it can be concluded that GRACE is able to detect changes in the terrestrial water storage of smaller basins.

In order to summarize the finds presented in Table 1 and Fig. 7, a multiple-comparison test was performed to determine the similarities (if any) of the CSR, GFZ, GRGS, and JPL solutions over the 91 river basins. First, a nonparametric Kruskal-Wallis test⁶⁰ at 95% confidence interval was applied to the uncertainties shown in Table 1 to determine whether there were statistically significant differences between the solutions. Figures 8(a) and 8(b) show the box plots that visually present the summary statistics for each solution pertaining to 91 river basins for TCH and ensemble mean based errors, respectively. It is evident that the notches in the GFZ and JPL box plots overlap, and it can be concluded with 95% confidence that the true medians do not differ. Overall, the box plots for TCH and ensemble mean based error panels [Figs. 8(a) and 8(b)] show the same patterns, while GRGS presents outliers. The basins classified as outliers for the GRGS solutions are No. 90, No. 66, No. 21, No. 27, No. 52, and No. 33, for both centers. These six basins are located in regions where GRGS seems to present high uncertainties [Fig. 4(c)] relative

Table 1 Uncertainties and SNRs for the 91 basins larger than $100 \times 10^3 \text{ km}^2$. Each basin location is shown in Fig. 6 and can be identified by the numbers in the first column of the table.

| Basin | | | Uncertainties | | | | SNR | | | |
|-------|---------------------|-------------------------|---------------|------|------|------|------|------|------|------|
| No. | Name | Area (km ²) | CSR | GFZ | GRGS | JPL | CSR | GFZ | GRGS | JPL |
| 01 | Amazon | 6,035,841 | 4.2 | 7.6 | 13.0 | 4.4 | 24.5 | 13.6 | 7.6 | 22.9 |
| 02 | Zaire | 3,679,287 | 5.7 | 8.3 | 9.2 | 6.2 | 6.3 | 4.4 | 3.9 | 5.7 |
| 03 | Mississippi | 3,228,838 | 3.4 | 7.4 | 8.1 | 5.5 | 11.7 | 5.7 | 4.6 | 7.3 |
| 04 | Ob | 3,085,466 | 3.5 | 4.4 | 11.2 | 7.6 | 6.8 | 5.2 | 2.0 | 3.1 |
| 05 | Nile | 3,057,377 | 5.1 | 9.4 | 9.0 | 8.3 | 2.4 | 1.7 | 1.3 | 1.7 |
| 06 | Parana | 2,565,049 | 6.4 | 8.0 | 12.0 | 7.7 | 6.7 | 5.5 | 3.4 | 5.3 |
| 07 | Lena | 2,418,600 | 3.2 | 5.6 | 9.9 | 6.0 | 7.1 | 4.2 | 2.5 | 4.0 |
| 08 | Amur | 2,214,920 | 4.1 | 6.4 | 10.3 | 6.0 | 6.2 | 3.8 | 2.6 | 3.8 |
| 09 | Niger | 2,197,243 | 5.0 | 7.3 | 8.3 | 6.3 | 4.8 | 3.4 | 2.8 | 3.9 |
| 10 | Yenisey | 1,926,905 | 4.3 | 7.3 | 8.7 | 6.9 | 4.7 | 2.8 | 2.2 | 3.2 |
| 11 | Yangtze | 1,709,102 | 5.6 | 6.2 | 7.8 | 6.3 | 5.3 | 5.0 | 3.8 | 5.1 |
| 12 | Ganges-Brahmaputra | 1,662,311 | 6.0 | 7.1 | 10.9 | 7.7 | 13.9 | 12.1 | 7.8 | 10.6 |
| 13 | Mackenzie | 1,537,735 | 3.6 | 6.5 | 9.6 | 4.9 | 10.8 | 6.1 | 3.9 | 7.9 |
| 14 | Lake Chad | 1,497,075 | 5.9 | 9.3 | 8.7 | 7.9 | 3.7 | 2.7 | 2.3 | 3.0 |
| 15 | Volga | 1,393,757 | 4.8 | 6.0 | 9.4 | 6.5 | 8.2 | 6.6 | 4.3 | 6.1 |
| 16 | Zambezi | 1,381,198 | 5.4 | 8.0 | 11.5 | 7.1 | 16.2 | 10.8 | 7.6 | 11.9 |
| 17 | Saskatchewan-Nelson | 1,138,155 | 2.3 | 7.1 | 10.8 | 6.3 | 14.3 | 4.6 | 3.0 | 5.4 |
| 18 | St. Lawrence | 1,055,493 | 7.0 | 8.9 | 9.5 | 7.7 | 5.1 | 4.1 | 3.1 | 4.2 |
| 19 | Murray-Darling | 1,049,014 | 6.4 | 8.7 | 9.8 | 9.3 | 4.6 | 3.2 | 3.0 | 3.1 |
| 20 | Indus | 1,030,597 | 6.8 | 8.1 | 10.8 | 8.0 | 3.5 | 3.1 | 2.2 | 3.1 |
| 21 | Orinoco | 934,465 | 5.9 | 11.1 | 16.9 | 7.3 | 15.5 | 8.2 | 5.5 | 12.4 |
| 22 | Hwang Ho | 881,193 | 8.7 | 6.9 | 8.5 | 7.7 | 2.9 | 3.6 | 2.7 | 3.2 |
| 23 | Yukon | 834,934 | 3.4 | 3.6 | 11.4 | 7.1 | 19.2 | 17.9 | 6.0 | 8.8 |
| 24 | Mekong | 801,880 | 6.7 | 8.7 | 11.1 | 9.2 | 12.3 | 9.6 | 7.4 | 9.0 |
| 25 | Danube | 801,832 | 5.3 | 7.4 | 9.3 | 7.6 | 7.3 | 5.0 | 4.0 | 5.0 |
| 26 | Jubba-Shebelle | 799,531 | 7.2 | 9.6 | 9.5 | 7.9 | 4.3 | 3.3 | 3.1 | 3.7 |
| 27 | Tocantins | 776,785 | 5.7 | 11.6 | 16.8 | 10.1 | 18.7 | 9.2 | 6.2 | 10.5 |
| 28 | Kolyma | 751,157 | 4.4 | 8.8 | 7.8 | 7.1 | 5.7 | 3.0 | 3.4 | 3.6 |
| 29 | Okavango | 692,213 | 6.8 | 10.6 | 10.4 | 8.6 | 10.1 | 6.6 | 6.8 | 8.2 |
| 30 | Columbia | 670,165 | 5.0 | 9.3 | 9.4 | 8.4 | 11.6 | 6.4 | 6.3 | 6.9 |
| 31 | Colorado | 669,314 | 5.8 | 10.1 | 14.1 | 9.9 | 5.8 | 3.7 | 2.3 | 3.5 |
| 32 | Rio Grande | 641,609 | 6.0 | 10.5 | 12.3 | 9.1 | 6.2 | 3.9 | 2.7 | 4.3 |
| 33 | Sao Francisco | 621,622 | 5.4 | 10.6 | 16.2 | 9.9 | 9.6 | 4.9 | 3.0 | 5.3 |

Table 1 (Continued).

| Basin | | Area (km ²) | Uncertainties | | | | SNR | | | |
|-------|------------------|-------------------------|---------------|------|------|------|------|-----|------|------|
| No. | Name | | CSR | GFZ | GRGS | JPL | CSR | GFZ | GRGS | JPL |
| 34 | Orange | 620,401 | 6.8 | 9.2 | 11.9 | 8.2 | 2.9 | 2.3 | 1.6 | 2.4 |
| 35 | Amu Darya | 611,451 | 6.2 | 8.3 | 11.7 | 7.9 | 6.0 | 4.5 | 3.1 | 4.7 |
| 36 | Tigris-Euphrates | 600,305 | 7.5 | 11.0 | 8.0 | 9.6 | 6.6 | 4.6 | 6.4 | 5.3 |
| 37 | Baikal | 588,895 | 5.6 | 8.0 | 8.7 | 8.1 | 4.4 | 3.2 | 2.6 | 3.2 |
| 38 | Tarim (Yarkand) | 580,923 | 5.8 | 8.8 | 13.6 | 9.9 | 4.7 | 3.2 | 2.1 | 2.8 |
| 39 | Dnieper | 534,768 | 7.4 | 8.4 | 9.4 | 7.6 | 4.5 | 4.0 | 3.4 | 4.5 |
| 40 | Don | 487,495 | 7.5 | 8.2 | 9.7 | 8.9 | 5.6 | 4.9 | 4.2 | 4.7 |
| 41 | Ili | 478,974 | 5.7 | 8.2 | 14.1 | 10.9 | 4.1 | 3.0 | 1.9 | 2.3 |
| 42 | Si | 459,251 | 7.3 | 8.4 | 10.4 | 9.5 | 5.3 | 4.7 | 3.7 | 4.2 |
| 43 | Senegal | 434,749 | 6.0 | 11.4 | 8.8 | 7.7 | 5.5 | 2.9 | 3.3 | 4.0 |
| 44 | Syr Darya | 429,847 | 7.2 | 8.0 | 12.5 | 8.6 | 4.3 | 3.7 | 2.6 | 3.7 |
| 45 | Rio Colorado | 418,990 | 7.0 | 9.5 | 10.8 | 10.3 | 6.3 | 4.4 | 4.1 | 4.1 |
| 46 | Limpopo | 413,588 | 6.5 | 9.1 | 13.9 | 10.2 | 4.9 | 3.7 | 2.2 | 3.1 |
| 47 | Volta | 412,481 | 5.9 | 13.3 | 9.4 | 8.5 | 8.4 | 3.7 | 5.0 | 6.0 |
| 48 | Irrawaddy | 386,957 | 8.1 | 11.8 | 12.8 | 10.9 | 12.3 | 8.4 | 7.8 | 8.8 |
| 49 | N. Dvina | 361,615 | 5.0 | 8.5 | 11.3 | 9.2 | 8.2 | 4.9 | 3.6 | 4.5 |
| 50 | Uruguay | 343,288 | 7.5 | 9.2 | 14.1 | 11.2 | 6.8 | 5.3 | 3.3 | 4.6 |
| 51 | Indigirka | 330,261 | 5.6 | 9.3 | 10.1 | 8.4 | 7.6 | 4.6 | 4.4 | 5.3 |
| 52 | Parnaiba | 329,441 | 5.3 | 11.4 | 16.6 | 9.7 | 11.3 | 5.2 | 3.4 | 6.1 |
| 53 | Godavari | 319,823 | 9.2 | 9.2 | 11.6 | 9.0 | 8.1 | 7.9 | 6.2 | 8.0 |
| 54 | Helmand | 317,312 | 7.1 | 10.0 | 10.2 | 9.8 | 3.9 | 3.0 | 2.5 | 2.8 |
| 55 | Liao | 274,646 | 6.6 | 8.4 | 12.1 | 8.7 | 3.7 | 2.9 | 2.2 | 3.0 |
| 56 | Ural | 262,748 | 5.3 | 9.4 | 10.4 | 7.4 | 6.7 | 3.9 | 3.5 | 4.8 |
| 57 | Krishna | 251,617 | 7.9 | 9.6 | 11.1 | 8.4 | 6.3 | 5.0 | 4.4 | 5.8 |
| 58 | Salween | 249,041 | 6.3 | 10.8 | 10.9 | 9.3 | 11.9 | 7.0 | 6.8 | 7.8 |
| 59 | Fraser | 240,935 | 5.1 | 9.4 | 9.2 | 8.8 | 14.1 | 7.8 | 7.9 | 8.4 |
| 60 | Magdalena | 232,208 | 6.4 | 11.8 | 11.0 | 9.4 | 6.1 | 3.5 | 3.7 | 4.2 |
| 61 | Ogooue | 221,217 | 8.5 | 12.7 | 11.6 | 8.9 | 6.2 | 3.9 | 4.2 | 5.6 |
| 62 | Rufiji | 203,544 | 7.3 | 11.1 | 12.0 | 10.4 | 6.6 | 4.3 | 4.1 | 4.6 |
| 63 | Vistula | 183,927 | 6.4 | 8.0 | 11.5 | 9.6 | 4.7 | 3.6 | 2.4 | 3.0 |
| 64 | Song Hong | 171,848 | 10.8 | 10.5 | 10.4 | 9.2 | 5.7 | 6.0 | 5.9 | 7.0 |
| 65 | Oder | 164,965 | 6.1 | 8.8 | 12.4 | 9.5 | 5.2 | 3.4 | 2.3 | 3.2 |
| 66 | Essequibo | 164,857 | 7.0 | 12.9 | 18.6 | 8.9 | 12.9 | 7.0 | 4.9 | 10.0 |

Table 1 (Continued).

| No. | Basin | | Uncertainties | | | | SNR | | | |
|-----|------------------------|-------------------------|---------------|------|------|------|------|-----|------|-----|
| | Name | Area (km ²) | CSR | GFZ | GRGS | JPL | CSR | GFZ | GRGS | JPL |
| 67 | Kura-Araks | 159,531 | 8.2 | 9.5 | 10.9 | 10.3 | 7.3 | 6.3 | 5.9 | 5.8 |
| 68 | Negro | 160,588 | 6.5 | 8.7 | 11.2 | 10.3 | 6.1 | 4.4 | 3.6 | 3.8 |
| 69 | Ruvuma | 152,339 | 7.0 | 11.4 | 11.5 | 10.2 | 8.7 | 5.2 | 5.5 | 5.9 |
| 70 | Rhine | 151,487 | 6.9 | 11.4 | 11.3 | 8.1 | 4.3 | 2.6 | 2.3 | 3.6 |
| 71 | Chao Phraya | 151,450 | 9.4 | 10.2 | 14.2 | 10.5 | 10.8 | 9.8 | 6.9 | 9.4 |
| 72 | Mahanadi | 149,125 | 9.3 | 10.6 | 13.8 | 9.8 | 9.4 | 8.3 | 6.3 | 8.7 |
| 73 | Chobut | 141,152 | 5.8 | 9.5 | 12.2 | 11.3 | 5.1 | 3.1 | 2.5 | 2.6 |
| 74 | Kwanza | 140,896 | 7.8 | 12.3 | 12.4 | 10.4 | 7.1 | 4.5 | 4.8 | 5.3 |
| 75 | Sanaga | 133,895 | 8.0 | 13.1 | 10.2 | 9.8 | 7.7 | 4.7 | 5.6 | 6.1 |
| 76 | Fly | 133,142 | 8.3 | 10.3 | 13.7 | 10.5 | 4.9 | 3.9 | 3.0 | 3.7 |
| 77 | Usumacinta | 126,436 | 7.3 | 11.9 | 11.1 | 10.6 | 8.1 | 4.9 | 5.1 | 5.4 |
| 78 | Hari Rud | 126,310 | 8.1 | 9.9 | 9.4 | 10.0 | 3.6 | 3.0 | 2.8 | 2.9 |
| 79 | Santiago-Lerma-Chapala | 126,031 | 8.4 | 12.2 | 11.2 | 9.7 | 5.1 | 3.5 | 3.5 | 4.2 |
| 80 | Sacramento | 125,338 | 6.1 | 12.8 | 10.1 | 9.6 | 7.7 | 3.8 | 4.6 | 5.0 |
| 81 | Brazos | 121,181 | 8.1 | 10.9 | 11.3 | 8.9 | 6.9 | 5.3 | 4.5 | 6.3 |
| 82 | Balsas | 119,007 | 7.3 | 12.1 | 10.0 | 8.7 | 6.4 | 3.8 | 4.5 | 5.2 |
| 83 | Kuskokwim | 118,428 | 6.6 | 8.1 | 11.7 | 7.9 | 10.2 | 8.3 | 6.0 | 8.2 |
| 84 | Elbe | 115,016 | 6.1 | 10.3 | 12.0 | 8.9 | 4.9 | 2.9 | 2.3 | 3.3 |
| 85 | Loire | 114,547 | 9.7 | 10.0 | 11.8 | 8.1 | 3.1 | 2.7 | 2.4 | 3.6 |
| 86 | Alabama | 113,852 | 8.3 | 10.9 | 11.3 | 9.3 | 7.9 | 6.3 | 5.8 | 7.1 |
| 87 | Save | 113,497 | 6.8 | 8.8 | 14.1 | 12.3 | 6.1 | 4.8 | 2.8 | 3.2 |
| 88 | Colorado River | 112,249 | 8.1 | 11.2 | 11.7 | 9.7 | 6.7 | 5.1 | 4.2 | 5.8 |
| 89 | Cunene | 109,646 | 7.6 | 12.4 | 12.6 | 8.9 | 6.7 | 4.1 | 4.4 | 6.0 |
| 90 | Kapuas | 102,482 | 7.5 | 8.7 | 21.1 | 12.5 | 3.7 | 3.4 | 1.6 | 2.2 |
| 91 | Rio Salado | 102,038 | 7.6 | 10.4 | 13.9 | 11.2 | 7.3 | 5.2 | 3.7 | 5.1 |

to the SDS centers, such as Indonesia (Kapuas River basin, No. 90) and South America (basin Nos. 66, 21, 27, 52, and 33).

Subsequently, a follow-up test was conducted to identify which processing center presented uncertainties resulting from a different distribution. The relative performance pertaining to the 91 river basins by the multiple-comparison test, based on the Tukey-Kramer procedure,⁶¹ is shown in Figs. 8(c) and 8(d) (TCH and ensemble mean based errors, respectively). The test indicates a significant difference between the uncertainties of CSR and those of the other three processing centers. However, there is no significant difference between the uncertainties of GFZ and JPL; therefore, the test does not reject the null hypothesis that the uncertainties of these two solutions obtained from the same distribution. Additionally, the uncertainties of GRGS, for the 91 river basins, originate from a different distribution to those of CSR, GFZ, and JPL, since there is

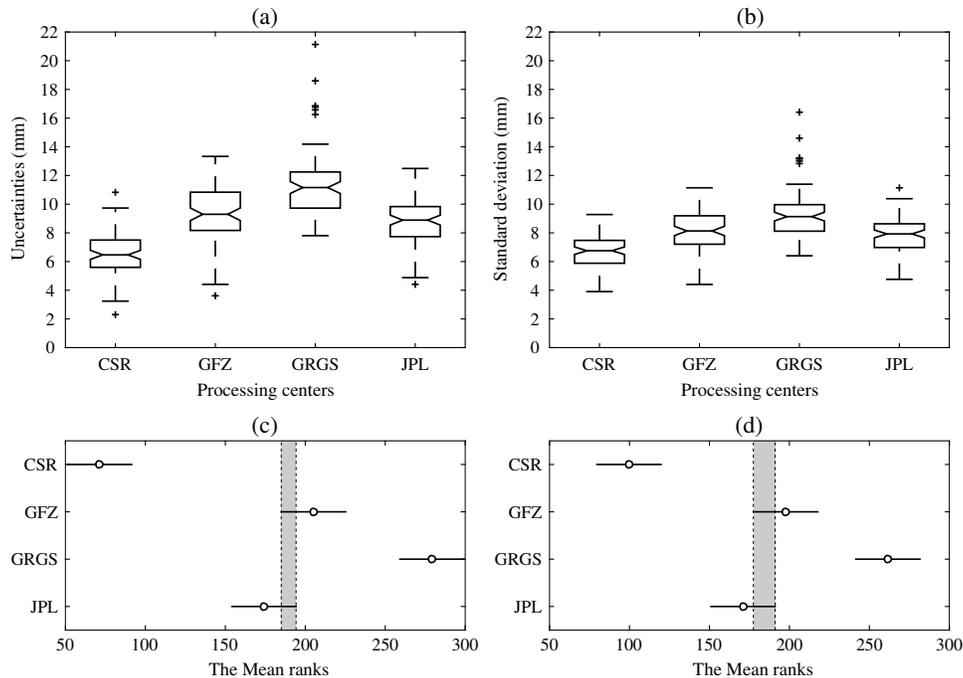


Fig. 8 Box plot summaries of the error estimates based on (a) TCH and (b) the ensemble mean. The performance rankings of the four solutions based on (c) TCH and (d) the ensemble mean for the 91 river basins. The whiskers at panels (a) and (b) indicate the maximum and minimum range of the noise estimates, the box shows the interquartile (first and third quartile) range of variation, the segment inside the box shows the median of the data, and the crosses indicate the outliers. The lower x-axis values in (c) and (d) indicate superior performance.

significant difference among their means. Again, the results from the ensemble solution confirm those computed by the TCH method, despite the magnitude of the estimated uncertainties from both approaches.

3.4 Results Summary

The intercomparison of the Stokes coefficients (Sec. 3.1) from the four GRACE processing centers (CSR, GFZ, GRGS, and JPL) showed an overall good agreement in terms of correlation coefficient between each pair of solutions up to degree 50 and order 40, with the GRGS being less correlated with the SDS processing centers (Fig. 1). Additionally, the Stokes coefficients from the four centers, up to degree 40 and order 20, showed relatively good agreement in terms of the NSE coefficient (Fig. 2). Moreover, the sectorial Stokes coefficients ($n = m$) showed NSE values less than zero for each pair of solutions, which indicates unacceptable performance. However, this relative comparison is unable to validate these products since it generally depends on the choice of one of them as the truth. One option would be, for example, to use an ensemble solution, computed as an arithmetic mean of the four centers, as suggested by Sakumura et al.¹⁸ to assess each ensemble member. As mentioned in Sec. 1, GRACE errors could be estimated by using the full covariance matrix and errors in the background models;^{32,33} yet a comparison of the GRACE-derived TWSA fields with true observations would be more interesting for validation purposes. However, there are no such *in situ* data for directly assessing GRACE-derived TWSA.

The importance of the present study lies in assessing the quality of each processing center using a generalized formulation of the TCH method (Sec. 2.1), which allows the noises of the time series to have a certain degree of correlation (see, e.g., Ref. 36). At a global scale, CSR presents a weighted averaged uncertainty of 9.4 mm and an SNR of 4.2, while GFZ, GRGS, and JPL have uncertainties of 13.7, 14.8, and 13.2 mm, with SNRs of 2.9, 2.3, and 2.9, respectively (Sec. 3.2). In addition to the recognized problem associated with the GRGS over regions above

latitudes ± 82 deg, large uncertainties were also indicated for regions such as Northern Australia, Southeast Asia, and the Amazon basin [see Fig. 4(c)]. Comparisons pertaining to the 91 large river basins exceeding 100×10^3 km² showed that CSR, GFZ, GRGS, and JPL present mean uncertainties of 6.5, 9.4, 11.4, and 8.8 mm (Sec. 3.3). The TCH-based uncertainties were confirmed by comparing the ensemble solution (arithmetic mean of the four centers in terms of TWSA) with each ensemble member. However, TCH-based internal precision makes sense only when all processing centers are considered together. Overall, the results are encouraging and show that the TCH method has the potential to assess the quality of other space-borne sensors that deliver observations of the same target variable.

4 Conclusions

The quality of four GRACE processing centers has been estimated by using a generalized formulation of the TCH method. The overall comparisons of TCH-based uncertainties pertaining to the period from August 2002 to June 2014 showed that the four solutions (CSR, GFZ, GRGS, and JPL) lie within 5 to 15 mm in terms of global grids, as well as basin-averaged values of TWSA. However, during the time span of the comparisons, CSR provides the most precise monthly solution in terms of TWSA at the global and basin scale, though the performance of each processing center differs from basin to basin. Over the polar regions, JPL shows higher uncertainties relative to the other two SDS processing centers, while the performance of GRGS is inadequate for high-latitude regions, Northern Australia, and Southeast Asia. However, the low performance of GRGS over these regions is yet to be investigated. Overall, it can be concluded that the applicability of GRACE for a specific river basin (or a region) would depend on the strength of the signal (i.e., on the annual cycle), rather than on the size of the basin. Although the use of the ensemble mean of the four (or more) centers is recommended from the practical standpoint, the present study contributed with the possibility of properly choosing the weights, so the TWSA ensemble mean series will have a low noise variance. However, studies are needed to determine whether these findings could be impacted in the case in which the assumption of low correlation among the noises of the series fails.

Acknowledgments

Vagner G. Ferreira acknowledges the financial support provided by the National Natural Science Foundation of China (Grant Nos. 41204016 and 41574001) and the Fundamental Research Funds for the Central Universities (Grant No. 2015B21014). He is also grateful for the research opportunity provided by Humboldt-Foundation, Research Alumni Initiative, and benefited by a grant from Karlsruhe Institute of Technology (International Scholars and Welcome Office) as part of the Research in Germany Campaign, funded by the Federal Ministry of Education and Research. Professor Dr.-Ing. Jürgen Kusche is appreciated for providing the DDK filter coefficients. We are grateful for the data used in this study as described in Secs. 2.2.2 and 3.3. We thank Professor Hongjie Xie (associate editor) and the three anonymous reviewers for their constructive comments that helped us to improve the quality of this manuscript.

References

1. B. D. Tapley et al., "GRACE measurements of mass variability in the Earth system," *Science* **305**(5683), 503–505 (2004).
2. J. T. Reager, B. F. Thomas, and J. S. Famiglietti, "River basin flood potential inferred using GRACE gravity observations at several months lead time," *Nat. Geosci.* **7**(8), 588–592 (2014).
3. L. Xavier et al., "Interannual variability in water storage over 2003–2008 in the Amazon basin from GRACE space gravimetry, *in situ* river level and precipitation data," *Remote Sens. Environ.* **114**(8), 1629–1637 (2010).
4. X. Hu et al., "Seasonal water storage change of the Yangtze River basin detected by GRACE," *Sci. China Ser. D* **49**(5), 483–491 (2006).

5. V. G. Ferreira, Z. Gong, and S. A. Andam-Akorful, "Monitoring mass changes in the Volta River basin using GRACE satellite gravity and TRMM precipitation," *Bol. Ciencias Geod.* **18**(4), 549–563 (2012).
6. J. Awange et al., "Water storage changes and climate variability within the Nile Basin between 2002 and 2011," *Adv. Water Resour.* **73**, 1–15 (2014).
7. M. Rodell, I. Velicogna, and J. S. Famiglietti, "Satellite-based estimates of groundwater depletion in India," *Nature* **460**(7258), 999–1002 (2009).
8. W. Feng et al., "Evaluation of groundwater depletion in North China using the gravity recovery and climate experiment (GRACE) data and ground-based measurements," *Water Resour. Res.* **49**(4), 2110–2118 (2013).
9. T. Jacob et al., "Recent contributions of glaciers and ice caps to sea level rise," *Nature* **482**(7386), 514–518 (2012).
10. M. Ahmed et al., "The use of GRACE data to monitor natural and anthropogenic induced variations in water availability across Africa," *Earth–Science Rev.* **136**, 289–300 (2014).
11. S. Bruinsma et al., "CNES/GRGS 10-day gravity field models (release 2) and their evaluation," *Adv. Space Res.* **45**(4), 587–601 (2010).
12. S. Bruinsma et al., "Updated release 3 of the GRACE gravity solutions from CNES/GRGS," in *2014 Fall Meeting, AGU*, pp. 15–19 (2014).
13. S. Bettadpur, "UTCSR level-2 processing standards document," Technical Report GRACE 327-742 (CSR-GR-12-xx), Center for Space Research, The University of Texas at Austin, Austin, Texas (2012).
14. M. M. Watkins and D.-N. Yuan, "JPL level-2 processing standards document for level-2 product release 05.1," Technical Report GRACE 327-744 (v 5.1), Jet Propulsion Laboratory, California Institute of Technology, Pasadena (2014).
15. C. Dahle et al., "GFZ GRACE level-2 processing standards document for level-2 product release 0005," Scientific Technical Report STR12/02—Data, Deutsches GeoForschungsZentrum, Potsdam (2013).
16. J. Wahr, S. Swenson, and I. Velicogna, "Accuracy of GRACE mass estimates," *Geophys. Res. Lett.* **33**(6), L06401 (2006).
17. J. Wahr et al., "Time-variable gravity from GRACE: first results," *Geophys. Res. Lett.* **31**(11), L11501 (2004).
18. C. Sakumura, S. Bettadpur, and S. Bruinsma, "Ensemble prediction and intercomparison analysis of GRACE time-variable gravity field models," *Geophys. Res. Lett.* **41**(5), 1389–1397 (2014).
19. T. H. Syed et al., "Total basin discharge for the Amazon and Mississippi River basins from GRACE and a land-atmosphere water balance," *Geophys. Res. Lett.* **32**(24), L24404 (2005).
20. A. Cesanelli and L. Guarracino, "Estimation of regional evapotranspiration in the extended Salado Basin (Argentina) from satellite gravity measurements," *Hydrogeol. J.* **19**(3), 629–639 (2011).
21. C. Lorenz et al., "Large-scale runoff from landmasses: a global assessment of the closure of the hydrological and atmospheric water balances," *J. Hydrometeorol.* **15**(9), 2111–2139 (2014).
22. S. A. Andam-Akorful et al., "Multi-model and multi-sensor estimations of evapotranspiration over the Volta Basin, West Africa," *Int. J. Climatol.* **35**(10), 3132–3145 (2015).
23. V. G. Ferreira and Z. Asiah, "An investigation on the closure of the water budget methods over Volta Basin using multi-satellite data," in *International Association of Geodesy Symposia*, pp. 1–8, Springer, Berlin, Heidelberg (2015).
24. R. Xiao et al., "Monitoring groundwater variations from satellite gravimetry and hydrological models: a comparison with in-situ measurements in the Mid-Atlantic region of the United States," *Remote Sens.* **7**(1), 686–703 (2015).
25. S. Swenson et al., "Estimating profile soil moisture and groundwater variations using GRACE and Oklahoma Mesonet soil moisture data," *Water Resour. Res.* **44**(1), W01413 (2008).
26. M. Shamsudduha, R. G. Taylor, and L. Longuevergne, "Monitoring groundwater storage changes in the highly seasonal humid tropics: validation of GRACE measurements in the Bengal Basin," *Water Resour. Res.* **48**(2), W02508 (2012).

27. T. H. Syed et al., "Analysis of terrestrial water storage changes from GRACE and GLDAS," *Water Resour. Res.* **44**(2), W02433 (2008).
28. R. Klees et al., "A comparison of global and regional GRACE models for land hydrology," *Surv. Geophys.* **29**(4–5), 335–359 (2008).
29. H. C. Winsemius et al., "Assessment of gravity recovery and climate experiment (GRACE) temporal signature over the upper Zambezi," *Water Resour. Res.* **42**(12), W12201 (2006).
30. J. Awange et al., "On the suitability of the $4^\circ \times 4^\circ$ GRACE mascon solutions for remote sensing Australian hydrology," *Remote Sens. Environ.* **115**(3), 864–875 (2011).
31. D. P. Chambers and J. A. Bonin, "Evaluation of Release-05 GRACE time-variable gravity coefficients over the ocean," *Ocean Sci. Discuss.* **9**(3), 2187–2214 (2012).
32. L. Jensen, R. Rietbroek, and J. Kusche, "Land water contribution to sea level from GRACE and Jason-1 measurements," *J. Geophys. Res. Ocean.* **118**(1), 212–226 (2013).
33. E. Forootan et al., "Comparisons of atmospheric mass variations derived from ECMWF reanalysis and operational fields, over 2003–2011," *J. Geod.* **88**(5), 503–514 (2014).
34. J. E. Gray and D. Allan, "A method for estimating the frequency stability of an individual oscillator," in *28th Annual Symp. on Frequency Control*, Vol. 5805, pp. 243–246, IEEE, Berlin, Germany (1974).
35. T. M. Chin, R. S. Gross, and J. O. Dickey, "Multi-reference evaluation of uncertainty in Earth orientation parameter measurements," *J. Geod.* **79**(1–3), 24–32 (2005).
36. L. Koot, O. Viron, and V. Dehant, "Atmospheric angular momentum time-series: characterization of their internal noise and creation of a combined series," *J. Geod.* **79**(12), 663–674 (2006).
37. P. Valtý et al., "Assessing the precision in loading estimates by geodetic techniques in Southern Europe," *Geophys. J. Int.* **194**(3), 1441–1454 (2013).
38. C. Abbondanza et al., "Three-corner hat for the assessment of the uncertainty of non-linear residuals of space-geodetic time series in the context of terrestrial reference frame analysis," *J. Geod.* **89**(4), 313–329 (2015).
39. D. Long, L. Longuevergne, and B. R. Scanlon, "Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites," *Water Resour. Res.* **50**(2), 1131–1151 (2014).
40. J. L. Awange et al., "Uncertainties in remotely sensed precipitation data over Africa," *Int. J. Climatol.* **36**(1), 303–323 (2016).
41. A. Premoli and P. Tavella, "A revisited three-cornered hat method for estimating frequency standard instability," *IEEE Trans. Instrum. Meas.* **42**(1), 7–13 (1993).
42. F. J. Galindo and J. Palacio, "Post-processing ROA data clocks for optimal stability in the ensemble timescale," *Metrologia* **40**(3), S237–S244 (2003).
43. P. Tavella and A. Premoli, "Estimating the instabilities of N clocks by measuring differences of their readings," *Metrologia* **30**(5), 479–486 (1994).
44. F. J. Galindo and J. Palacio, "Estimating the instabilities of N correlated clocks," in *Proc. of the 31st Annual Precise Time and Time Interval Meeting*, pp. 285–296 (1999).
45. F. Torcaso et al., "Estimating frequency stability and cross-correlations," in *30th Annual Precise Time and Time Interval Meeting*, pp. 69–82 (1998).
46. J. Wahr, M. Molenaar, and F. Bryan, "Time variability of the Earth's gravity field: hydrological and oceanic effects and their possible detection using GRACE," *J. Geophys. Res.* **103**(B12), 30205–30229 (1998).
47. S. Swenson and J. Wahr, "Methods for inferring regional surface-mass anomalies from gravity recovery and climate experiment (GRACE) measurements of time-variable gravity," *J. Geophys. Res.* **107**(B9), 2193 (2002).
48. F. Barthelmes and W. Köhler, "International Centre for Global Earth Models (ICGEM)," *J. Geodesy* **86**(10), 932–934 (2012). <http://icgem.gfz-potsdam.de/ICGEM/>
49. B. Wouters et al., "GRACE, time-varying gravity, Earth system dynamics and climate change," *Rep. Prog. Phys.* **77**(11), 116801 (2014).
50. C. Dahle et al., "Gravity field mapping from GRACE: different approaches-same results?," in *Int. Association of Geodesy Symposia*, pp. 1–11, Springer, Berlin, Heidelberg (2015).
51. S. Swenson and J. Wahr, "Post-processing removal of correlated errors in GRACE data," *Geophys. Res. Lett.* **33**(8), L08402 (2006).

52. J. Kusche, "Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models," *J. Geod.* **81**(11), 733–749 (2007).
53. S. Werth et al., "Evaluation of GRACE filter tools from a hydrological perspective," *Geophys. J. Int.* **179**(3), 1499–1515 (2009).
54. N. Sneeuw et al., "Estimating runoff using hydro-geodetic approaches," *Surv. Geophys.* **35**(6), 1333–1359 (2014).
55. D. N. Moriasi et al., "Model evaluation guidelines for systematic quantification of accuracy in watershed simulations," *Trans. ASABE* **50**(3), 885–900 (2007).
56. FAO, "Geonetwork: find and analyse geo-spatial data," WRI Major Watersheds of the World Delineation, 2006, <http://www.fao.org/geonetwork/srv/en/main.home> (19 June 2015).
57. L. Longuevergne, B. R. Scanlon, and C. R. Wilson, "GRACE hydrological estimates for small basins: evaluating processing approaches on the High Plains Aquifer, USA," *Water Resour. Res.* **46**(11), W11517 (2010).
58. A. Y. Sun, "Predicting groundwater level changes using GRACE data," *Water Resour. Res.* **49**(9), 5900–5912 (2013).
59. J. T. Reager and J. S. Famiglietti, "Characteristic mega-basin water storage behavior using GRACE," *Water Resour. Res.* **49**(6), 3314–3329 (2013).
60. P. E. McKight and J. Najab, "Kruskal-Wallis test," *The Corsini Encyclopedia of Psychology*, I. B. Weiner and W. E. Graighead, Eds., 4th ed., Vol. 2, p. 904, John Wiley & Sons Inc., Hoboken, New Jersey (2010).
61. J. A. Rafter, M. L. Abell, and J. P. Braselton, "Multiple comparison methods for means," *SIAM Rev.* **44**(2), 259–278 (2002).

Vagner G. Ferreira is an associate professor at the School of Earth Sciences and Engineering, Hohai University, China. He received his MSc and PhD degrees in geodetic sciences from the Federal University of Paraná, Brazil, in 2008 and 2011, respectively. He was with the Geodetic Institute of Karlsruhe (GIK), Karlsruhe Institute of Technology (KIT), Germany, during 2009, where he worked on local gravity field modeling. His current research interests include hydro-geodesy.

Henry D. C. Montecino is a PhD candidate at the Graduate Program in Geodetic Science, Federal University of Paraná, Brazil, and a lecturer at the Department of Geodetic Sciences and Geomatics, University of Concepción, Chile. He received his BSc degree from the University of Concepción in 2006 and his MSc degree from the Federal University of Paraná in 2012.

Caleb I. Yakubu is an MSc student at the School of Earth Sciences and Engineering, Hohai University, China. He received his BSc degree in geomatic engineering from the Kwame Nkrumah University of Science and Technology, Ghana, in 2013.

Bernhard Heck is a full professor at KIT, responsible for the chair of physical and satellite geodesy at GIK. His scientific activities include various branches of geodesy, such as mathematical and physical geodesy, analysis of geodetic networks, GNSS positioning and GNSS meteorology, as well as geodynamics.