Remote Sensing of Environment 115 (2011) 187–204

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

An independent component analysis filtering approach for estimating continental hydrology in the GRACE gravity data

Frédéric Frappart [a,*], Guillaume Ramillien [b,c], Marc Leblanc [d], Sarah O. Tweed [d], Marie-Paule Bonnet [a,e], Philippe Maisongrande [f,g]

[a] Université de Toulouse, UPS, OMP, LMTG, 14 Avenue Edouard Belin, 31400 Toulouse, France
[b] Université de Toulouse, UPS, OMP, DTP, 14 Avenue Edouard Belin, 31400 Toulouse, France
[c] CNRS, OMP, DTP, 14 Avenue Edouard Belin, 31400 Toulouse, France
[d] Hydrological Sciences Research Unit, School of Earth and Environmental Sciences, James Cook University, Cairns, Queensland, Australia
[e] IRD, OMP, LMTG, 14 Avenue Edouard Belin, 31400 Toulouse, France
[f] Université de Toulouse, UPS, OMP, LEGOS, 14 Avenue Edouard Belin, 31400 Toulouse, France
[g] CNES, 18 Avenue Edouard Belin, 31400 Toulouse, France

ARTICLE INFO

Article history:
Received 30 March 2010
Received in revised form 20 August 2010
Accepted 21 August 2010

Keywords:
Filtering technique
Gravimetry from space
Hydrology
Independent component analysis

ABSTRACT

An approach based on Independent Component Analysis (ICA) has been applied on a combination of monthly GRACE satellite solutions computed from official providers (CSR, JPL and GFZ), to separate useful geophysical signals from important striping undulations. We pre-filtered the raw GRACE Level-2 solutions using Gaussian filters of 300, 400, 500-km of radius to verify the non-Gaussianity condition which is necessary to apply the ICA. This linear inverse approach ensures to separate components of the observed gravity field which are statistically independent. The most energetic component found by ICA corresponds mainly to the contribution of continental water mass change. Series of ICA-estimated global maps of continental water storage have been produced over 08/2002–07/2009. Our ICA estimates were compared with the solutions obtained using other post-processing of GRACE Level-2 data, such as destriping and Gaussian filtering, at global and basin scales. Besides, they have been validated with *in situ* measurements in the Murray–Darling Basin. Our computed ICA grids are consistent with the different approaches. Moreover, the ICA-derived time series of water masses showed less north–south spurious gravity signals and improved filtering of unrealistic hydrological features at the basin-scale compared with solutions obtained using other filtering methods.

1. Introduction

Continental water storage is a key component of global hydrological cycles and plays a major role in the Earth's climate system via controls over water, energy and biogeochemical fluxes. In spite of its importance, the total continental water storage is not well-known at regional and global scales because of the lack of *in situ* observations and systematic monitoring of the groundwaters (Alsdorf & Lettenmaier, 2003).

The Gravity Recovery and Climate Experiment (GRACE) mission provides a global mapping of the time-variations of the gravity field at an unprecedented resolution of ~400 km and a precision of ~1 cm in terms of geoid height. Tiny variations of gravity are mainly due to redistribution of mass inside the fluid envelops of the Earth (i.e., atmosphere, oceans and continental water storage) from monthly to decade timescales (Tapley et al., 2004).

Pre-processing of GRACE data is made by several providers (University of Texas, Centre for Space Research — CSR, Jet Propulsion Laboratory — JPL, GeoForschungsZentrum — GFZ and Groupe de Recherche en Géodésie Spatiale — GRGS) which produce residual GRACE spherical harmonic solutions that mainly represent continental hydrology as they are corrected from known mass transfers using *ad hoc* oceanic models (*i.e.*, Toulouse Unstructured Grid Ocean model 2D — T-UGOm 2D) and atmospheric reanalyses from National Centers for Environmental Prediction (NCEP) and European Centre for Medium Weather Forecasting (ECMWF). Unfortunately these solutions suffer from the presence of important north–south striping due to orbit resonance in spherical harmonics determination and aliasing of short-time phenomena which are geophysically unrealistic.

Since its launch in March 2002, the GRACE terrestrial water storage anomalies have been increasingly used for large-scale hydrological applications (see Ramillien et al., 2008; Schmidt et al., 2008 for reviews). They demonstrated a great potential to monitor extreme hydrological events (Andersen et al., 2005; Seitz et al., 2008; Chen et al., 2009), to

* Corresponding author. Tel: +33 5 61 33 26 63; fax: +33 5 61 33 25 60.
E-mail addresses: frederic.frappart@lmtg.obs-mip.fr (F. Frappart), guillaume.ramillien@dtp.obs-mip.fr (G. Ramillien), marc.leblanc@jcu.edu.au (M. Leblanc), sarah.tweed@jcu.edu.au (S.O. Tweed), marie-paule.bonnet@ird.fr (M.-P. Bonnet), philippe.maisongrande@cnes.fr (P. Maisongrande).

0034-4257/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2010.08.017

estimate water storage variations in the soil (Frappart et al., 2008), the aquifers (Rodell et al., 2007; Strassberg et al., 2007; Leblanc et al., 2009) and the snowpack (Frappart et al., 2006, in press), and hydrological fluxes, such as basin-scale evapotranspiration (Rodell et al., 2004a; Ramillienet al., 2006a) and discharge (Syed et al., 2009).

Because of this problem of striping that limits geophysical interpretation, different post-processing approaches for filtering GRACE geoid solutions have been proposed to extract useful geophysical signals (see Ramillien et al., 2008; Schmidt et al., 2008 for reviews). These include the classical isotropic Gaussian filter (Jekeli, 1981), various optimal filtering decorrelation of GRACE errors (Han et al., 2005; Seo & Wilson, 2005; Swenson & Wahr, 2006; Sasgen et al., 2006; Kusche, 2007; Klees et al., 2008), as well as statistical constraints on the time evolution of GRACE coefficients (Davis et al., 2008) or from global hydrology models (Ramillien et al., 2005). However, these filtering techniques remain imperfect as they require input non-objective *a priori* information which are most of the time simply tuned by hand (*e.g.* choosing the cutting wavelength while using the Gaussian filtering) or based on other rules-of-thumb.

We propose another post-processing approach of the Level-2 GRACE solutions by considering completely objective constraints, so that the gravity component of the observed signals is forced to be uncorrelated numerically using an Independent Component Analysis (ICA) technique. This approach does not require *a priori* information except the assumption of statistical independence of the elementary signals that compose the observations, *i.e.*, geophysical and spurious noise. The efficiency of ICA to separate gravity signals and noise from combined GRACE solutions has previously been demonstrated on one month of Level-2 solutions (Frappart et al., 2010). In this paper, we use this new statistical linear method to derive complete time series of continental water mass change.

The first part of this article presents the datasets used in this study: the monthly GRACE solutions to be inverted by ICA and to be used for comparisons, and the *in situ* data used for the validation of our estimates over the Murray–Darling drainage basin (~1 million of km²). This region has been selected for validation because of available dense hydrological observations. The second part outlines the three steps of the ICA methodology. Then the third and fourth parts present results and comparisons with other post-processed GRACE solutions at global and regional scales, and *in situ* measurements for the Murray–Darling Basin respectively. Error balance of the ICA-based solutions is also made by considering the effect of spectrum truncation, leakage and formal uncertainties.

2. Datasets

2.1. The GRACE data

The GRACE mission, sponsored by National Aeronautics and Space Administration (NASA) and Deutsches Zentrum für Luft- und Raumfahrt (DLR), has been collecting data since mid-2002. Monthly gravity models are determined from the analysis of GRACE orbit perturbations in terms of Stokes spherical harmonic coefficients, *i.e.*, geopotential or geoid heights. The geoid is an equipotential surface of the gravity field that coincide with mean sea level. For the very first time, monthly global maps of the gravity time-variations can be derived from GRACE measurements, and hence, to estimate the distribution of the change of mass in the Earth's system.

2.1.1. The Level-2 raw solutions

The Level-2 raw data consist of monthly estimates of geopotential coefficients adjusted for each 30-day period from raw along-track GRACE measurements by different research groups (i.e., CSR, GFZ and JPL). These coefficients are developed up to a degree 60 (or spatial resolution of 333 km) and corrected for oceanic and atmospheric effects (Bettadpur, 2007) to obtain residual global grids of ocean and land signals corrupted by a strong noise. These data are available at: ftp://podaac.jpl.nasa.gov/grace/.

2.1.2. The destriped and smoothed solutions

The monthly raw solutions (RL04) from CSR, GFZ, and JPL were destriped and smoothed by Chambers (2006) for hydrological purposes. These three datasets are available for several averaging radii (0, 300 and 500 km on the continents and 300, 500 and 750 km on the oceans) at ftp://podaac.jpl.nasa.gov/tellus/grace/monthly.

In this study, we used the Level-2 RL04 raw data from CSR, GFZ and JPL, that we filtered with a Gaussian filter for radii of 300, 400 and 500 km, and the destriped and smoothed solutions for the averaging radii of 300 and 500 km over land.

2.2. The hydrological data for the Murray–Darling Basin

In the predominantly semi-arid Murray–Darling Basin, most of the surface water is regulated using a network of reservoirs, lakes and weirs (Kirby et al., 2006) and the surface water stored in these systems represent most of the total surface water (SW) present across the basin. A daily time series of the total surface water storage in the network of reservoirs, lakes, weirs and in-channel storage was obtained from the Murray–Darling Basin Commission and the state governments from January 2000 to December 2008.

In the Murray–Darling Basin, we derived monthly soil moisture (SM) storage values for the basin from January 2000 to December 2008 from the NOAH land surface model (Ek et al., 2003), with the NOAH simulations being driven (parameterization and forcing) by the Global Land Data Assimilation System (Rodell et al., 2004b). The NOAH model simulates surface energy and water fluxes/budgets (including soil moisture) in response to near-surface atmospheric forcing and depending on surface conditions (*e.g.*, vegetation state, soil texture and slope) (Ek et al., 2003). The NOAH model outputs of soil moisture estimates have a 1° spatial resolution and, using four soil layers, are representative of the top 2 m of the soil.

In situ estimates of annual changes in the total groundwater storage (GW) across the drainage basin were obtained from an analysis of groundwater levels observed in government monitoring bores from 2000 to 2008. Compared to earlier estimates by Leblanc et al. (2009), the groundwater estimates presented in this paper provide an update of the *in situ* water level and a refinement of the distribution of the aquifers storage capacity.

Assuming that (1) the shallow aquifers across the Murray–Darling drainage basin are hydraulically connected and that (2) at a large scale the fractured aquifers can be assimilated to a porous media, changes in groundwater storage across the area can be estimated from observations of groundwater levels (*e.g.*, Rodell et al., 2007; Strassberg et al., 2007). Variations in groundwater storage (ΔS_{GW}) were estimated from *in situ* measurements as:

$$\Delta S_{GW} = S_y \Delta H \tag{1}$$

where S_y is the aquifer specific yield (%) and H is the groundwater level (L^{-1}) observed in monitoring bores. Groundwater level data (H) were sourced from State Government departments that are part of the Murray-Darling Basin (QLD; Natural Resources and Mines; NSW; Department of Water and Energy; VIC; Department of Sustainability and Environment; and SA; Department of Water Land and Biodiversity Conservation). Only government observation bores (production bores excluded) with an average saturated zone <50 m from the bottom of the screened interval were selected. Deeper bores were excluded as they can reflect processes occurring on longer time scales (Fetter, 2001). A total of 6183 representative bores for the unconfined aquifers across the Murray–Darling Basin were selected on the basis of construction and monitoring details obtained from the State departments. ~85% (5075) of the selected monitoring bores have a

maximum annual standard deviation of the groundwater levels below 2 m for the study period (2000–2008) and were used to analyze the annual changes in groundwater storage during the period 2000 to 2008. The remaining 15% of the observation bores, with the highest annual standard deviation, were discarded as possibly under the immediate influence of local pumping or irrigation. The potential influence of irrigation on some of the groundwater data is limited because during this period of drought, irrigation is substantially reduced across the basin. Changes in groundwater levels across the basin were estimated using an annual time step as most monitoring bores have limited groundwater level measurements in any year (50% of bores with 5 measures per year). The annual median of the groundwater level was first calculated for each bore and change at a bore was computed as the difference of annual median groundwater level between two consecutive years. For each year, a spatial interpolation of the groundwater level change was performed across the basin using a kriging technique. Spatial averages of annual groundwater level change were computed for each aquifer group.

The Murray–Darling drainage basin comprises several unconfined aquifers that can be regrouped into 3 categories according to their lithology: a clayey sand aquifer group (including the aquifers Narrabri (part of), Cowra, Shepparton, and Murrumbidgee (part of)); a sandy clay aquifer group (including the aquifers Narrabri (part of), Parilla, Far west, and Calivil (part of)); and a fractured rock aquifer group comprising metasediments, volcanics and weathered granite (including the aquifers Murrumbidgee (part of), North central, North east, Central west, Barwon, and Queensland boundary). The specific yield is estimated to range from 5 to 10% for the clayey sand unconfined aquifer group (Macumber, 1999; Cresswell et al., 2003; Hekmeijer & Dawes, 2003a; CSIRO, 2008); from 10 to 15% for the shallow sandy clay unconfined aquifer group (Macumber, 1999; Urbano et al., 2004); and from 1 to 10% for the fractured rock aquifer group (Cresswell et al., 2003; Hekmeijer & Dawes, 2003b; Smitt et al., 2003; Petheram et al., 2003). *In situ* estimates of changes in GW storage are calculated using the spatially averaged change in annual groundwater level across each type of unconfined aquifer group and the mean value of the specific yield for that group using Eq. (1); while the range of possible values for the specific yield was used to estimate the uncertainty.

Groundwater changes in the deep, confined aquifers (mostly GAB and Renmark aquifers) are either due to: 1) a change in groundwater recharge at the unconfined outcrop; 2) shallow pumping at the unconfined outcrop or 3) deep pumping in confined areas for farming (irrigation and cattle industry). GRACE TWS estimates accounts for all possible sources of influence, while GW *in situ* estimates only include those occurring across the outcrop. Total pumping from the deep, confined aquifers was estimated to amount to -0.42 km^{3}yr^{-1} in 2000 (Ife & Skelt, 2004), while groundwater pumping across the basin was -1.6 km^3 in 2002–2003 (Kirby et al., 2006). To allow direct comparison between TWS and *in situ* GW estimates, pumping from the deep aquifers was added to the *in situ* GW time series assuming the -0.42 km^{3}yr^{-1} pumping rate remained constant during the study period.

3. Methodology

3.1. ICA-based filter

ICA is a powerful method for separating a multivariate signal into subcomponents assuming their mutual statistical independence (Comon, 1994; De Lathauwer et al., 2000). It is commonly used for blind signal separation and has various practical applications (Hyvärinen & Oja, 2000), including telecommunications (Ristaniemi & Joutsensalo, 1999; Cristescu et al., 2000), medical signal processing (Vigário, 1997; van Hateren & van der Schaaf, 1998), speech signal processing (Stone, 2004), and electrical engineering (Gelle et al., 2001; Pöyhönen et al., 2003).

Assuming that an observation vector y collected from N sensors is the combination of P ($N \geq P$) independent sources represented by the source vector x, the following linear statistical model can be considered:

$$y = Mx \tag{2}$$

where M is the mixing matrix whose elements m_{ij} ($1 \leq i \leq N$, $1 \leq j \leq P$) indicate to what extent the jth source contribute to the ith observation. The columns $\{m_j\}$ are the mixing vectors.

The goal of ICA is to estimate the mixing matrix M and/or the corresponding realizations of the source vector x, only knowing the realizations of the observation vector y, under the assumptions (De Lathauwer et al., 2000):

1) the mixing vectors are linearly independent,
2) the sources are statistically independent.

The original sources x can be simply recovered by multiplying the observed signals y with the inverse of the mixing matrix also known as the "unmixing" matrix:

$$x = M^{-1}y \tag{3}$$

To retrieve the original source signals, at least N observations are necessary if N sources are present. ICA remains applicable for square or over-determined problems. ICA proceeds by maximizing the statistical independence of the estimated components. As a condition of applicability of the method, non-Gaussianity of the input signals has to be checked. The central limit theorem is then used for measuring the statistical independence of the components. Classical algorithms for ICA use centering and whitening based on eigenvalue decomposition (EVD) and reduction of dimension as main processing steps. Whitening ensures that the input observations are equally treated before dimension reduction.

ICA consists of three numerical steps. The first step of ICA is to centre the observed vector, i.e., to substract the mean vector $m = E\{y\}$ to make y a zero mean variable. The second step consists in whitening the vector y to remove any correlation between the components of the observed vector. In other words, the components of the white vector $\tilde{y}$ have to be uncorrelated and their variances equal to unity. Letting $C = E\{yy^t\}$be the correlation matrix of the input data, we define a linear transform B that verifies the two following conditions:

$$\tilde{y} = By \tag{4}$$

and:

$$E\{\tilde{y}\tilde{y}^t\} = I_P \tag{5}$$

where I_P is identity matrix of dimension $P \times P$.

This is easily accomplished by considering:

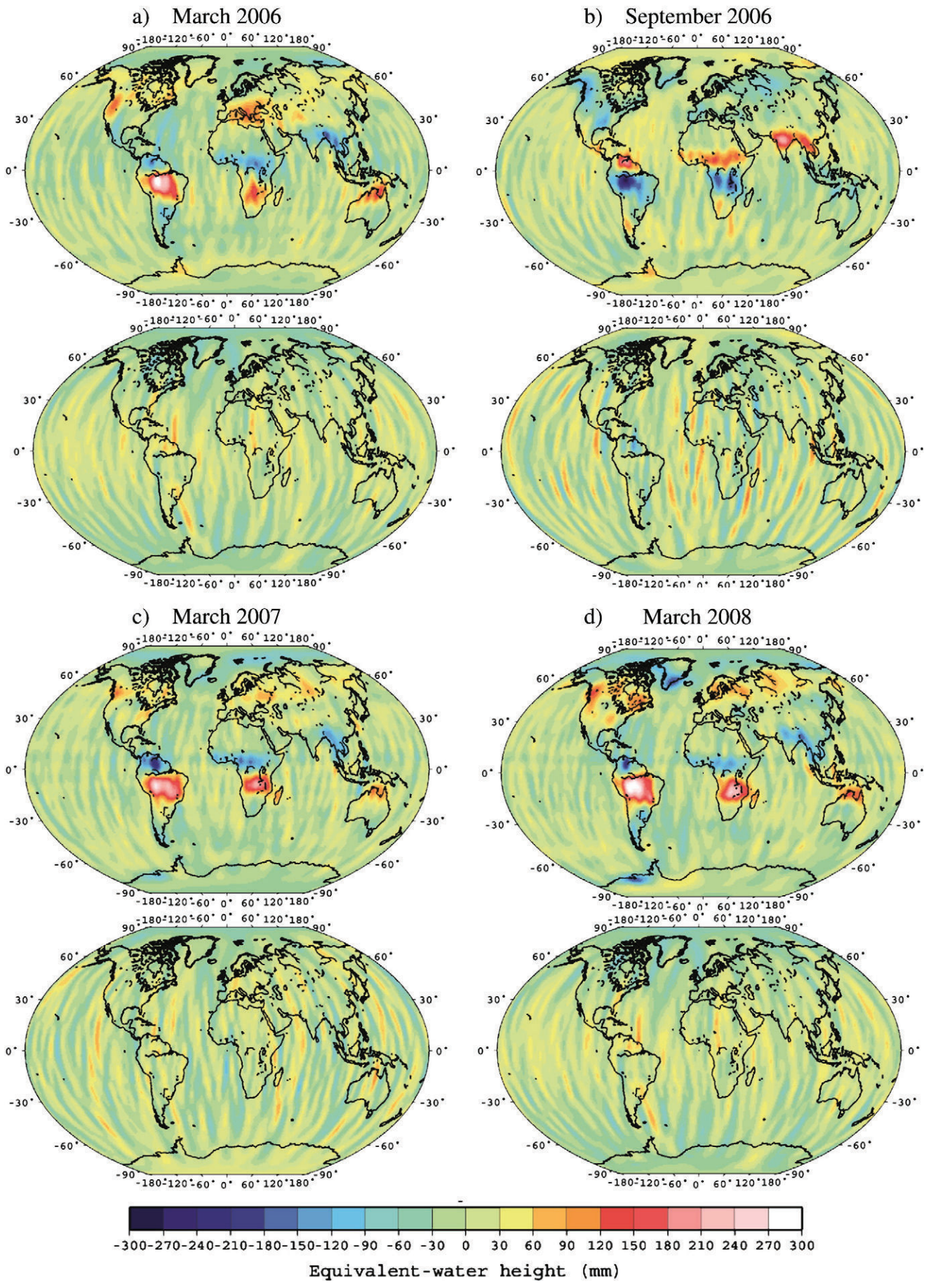
$$B = C^{-\frac{1}{2}} \tag{6}$$

The whitening is obtained using an EVD of the covariance matrix C:

$$C = EDE^t \tag{7}$$

where E is the orthogonal matrix of the eigenvectors of C and D is the diagonal matrix of its eigenvalues. $D = \mathrm{diag}(d_1,\ldots,d_P)$ as a reduction of the dimension of the data to the number of independent components (IC) P is performed, discarding the too small eigenvalues.

For the third step, an orthogonal transformation of the whitened signals is used to find the separated sources by rotation of the joint density. The appropriate rotation is obtained by maximizing the non-


a) March 2006
b) September 2006
c) March 2007
d) March 2008
-300 -270 -240 -210 -180 -150 -120 -90 -60 -30 0 30 60 90 120 150 180 210 240 270 300
Equivalent-water height (mm)

normality of the marginal densities, since a linear mixture of independent random variables is necessarily more Gaussian than the original components.

Many algorithms of different complexities have been developed for ICA (Stone, 2004). The *FastICA* algorithm, a computationally highly efficient method for performing the estimation of ICA (Hyvärinen & Oja, 2000) has been considered to separate satellite gravity signals. It uses a fixed-point iteration scheme that has been found to be 10 to 100 times faster than conventional gradient methods for ICA (Hyvärinen, 1999).

We used the *FastICA* algorithm (available at http://www.cis.hut.fi/projects/ica/fastica/) to unravel the IC of the monthly gravity field anomaly in the Level-2 GRACE products. We previously demonstrated, on a synthetic case, that land and ocean mass anomalies are statistically independent from the north–south stripes using information from land and ocean models and simulated noise (Frappart et al., 2010). Considering that the GRACE Level-2 products from CSR, GFZ and JPL are different observations of the same monthly gravity anomaly and, that the land hydrology and the north–south stripes are the independent sources, we applied this methodology to the complete 2002–2009 time series. The raw Level-2 GRACE solutions present Gaussian histograms which prevent the successful application of the ICA method. To ensure the non-Gaussianity of the observations, the raw data have been preprocessed using Gaussian filters with averaging radii of 300, 400 and 500 km as in Frappart et al. (2010).

3.2. Time series of basin-scale total water storage average

For a given month t, the regional average of land water volume $\delta V(t)$ (or height $\delta h(t)$) over a given river basin of area A is simply computed from the water height δh_j, with $j = 1, 2, \ldots$ (expressed in terms of mm of equivalent water height) inside A, and the elementary surface $R_e^2 \delta\lambda\delta\theta \sin\theta_j$:

$$\delta V(t) = R_e^2 \sum_{j\in A} \delta h_j\left(\theta_j, \lambda_j, t\right) \sin\theta_j \delta\lambda\delta\theta \tag{8}$$

$$\delta h(t) = \frac{R_e^2}{A} \sum_{j\in A} \delta h_j\left(\theta_j, \lambda_j, t\right) \sin\theta_j \delta\lambda\delta\theta \tag{9}$$

where θ_j and λ_j are co-latitude and longitude of the jth point, $\delta\lambda$ and $\delta\theta$ are the grid steps in longitude and latitude respectively (generally $\delta\lambda = \delta\theta$). In practice, all points of A used in Eqs. (8) and (9) are extracted for the eleven drainage basins masks at a 0.5° resolution provided by Oki and Sud (1998), except for the Murray–Darling Basin where we used basin limits from Leblanc et al. (2009).

3.3. Regional estimates of formal error

As ICA provides separated solutions which have Gaussian distributions, the variance of the regional average for a given basin is:

$$\sigma_{formal}^2 = \frac{\sum_{k=1}^{L} \sigma_k^2}{L^2} \tag{10}$$

where σ_{formal} is the regional formal error, σ_k is the formal error at a grid point number k, and L is the number of points used in the regional averaging.

If the points inside the considered basin are independent, this relation is slightly simplified:

$$\sigma_{formal} = \frac{\sigma_k}{\sqrt{L}} \tag{11}$$

3.4. Frequency cut-off error estimates

Error in frequency cut-off represents the loss of energy in the short spatial wavelength due to the low-pass harmonic decomposition of the signals that is stopped at the maximum degree N_1. For the GRACE solution separated by ICA; $N_1 = 60$, thus the spatial resolution is limited and stopped at ~330 km by construction. This error is simply evaluated by considering the difference of reconstructing the remaining spectrum between two cutting harmonic degrees N_1 and N_2, where $N_2 > N_1$ and N_2 should be large enough compared to N_1 (*e.g.*, $N_2 = 300$ in study):

$$\sigma_{truncation} = \sum_{n=0}^{N_2} \xi_n - \sum_{n=0}^{N_1} \xi_n = \sum_{n=N_1}^{N_2} \xi_n \tag{12}$$

using the scalar product

$$\xi_n = \sum_{m=0}^{n} \left(C_{nm}A_{nm} + S_{nm}B_{nm}\right) \tag{13}$$

where A_{nm} and B_{nm} are the harmonic coefficients of the considered geographical mask, and C_{nm} and S_{nm} are the harmonic coefficients of the water masses.

3.5. Leakage error estimates

We define «leakage» as the portion of signals from outside the considered geographical region that pollutes the region's estimates. By construction, this effect can be seen as the limitation of the geoid signals degree in the spherical harmonics representation. For each basin and at each period of time, leakage is simply computed as the average of outside values by using an «inverse» mask, which is 0 and 1 in and out of the region respectively, developed in spherical harmonics and then truncated at degree 60. This method of computing leakage of continental water mass has been previously proposed for the entire continent of Antarctica (Ramillien et al., 2006b), which revealed that the seasonal amplitude of this type of error can be quite important (*e.g.* up to 10% of the geophysical signals). In case of no leakage, this average should be zero (at least, it decreases with the maximum degree of decomposition). However, the maximum leakage of continental hydrology remains in the order of the signals magnitude itself.

4. Results and discussion

4.1. ICA-filtered land water solutions

The methodology presented in Frappart et al. (2010) has been applied to the Level-2 RL04 raw monthly GRACE solutions from CSR, GFZ and JPL, preprocessed using a Gaussian filter with a radius of 300, 400 and 500 km, over the period July 2002 to July 2009. The results of this filtering method are presented in Fig. 1 for four different time periods (March and September 2006, March 2007 and March 2008) using the GFZ solutions Gaussian-filtered with a radius of 400 km. Only the ICA-based GFZ solution is presented since, for a specific radius, the ICA-based CSR, GFZ, and JPL solutions only differ from a

Fig. 1. GRACE water storage from GFZ filtered with a Gaussian filter of 400 km of radius. (Top) First ICA component corresponding to land hydrology and ocean mass. (Bottom) Sum of the second and third components corresponding to the north–south stripes. (a) March 2006, (b) September 2006, (c) March 2007, and (d) March 2008. Units are millimeters of EWT.

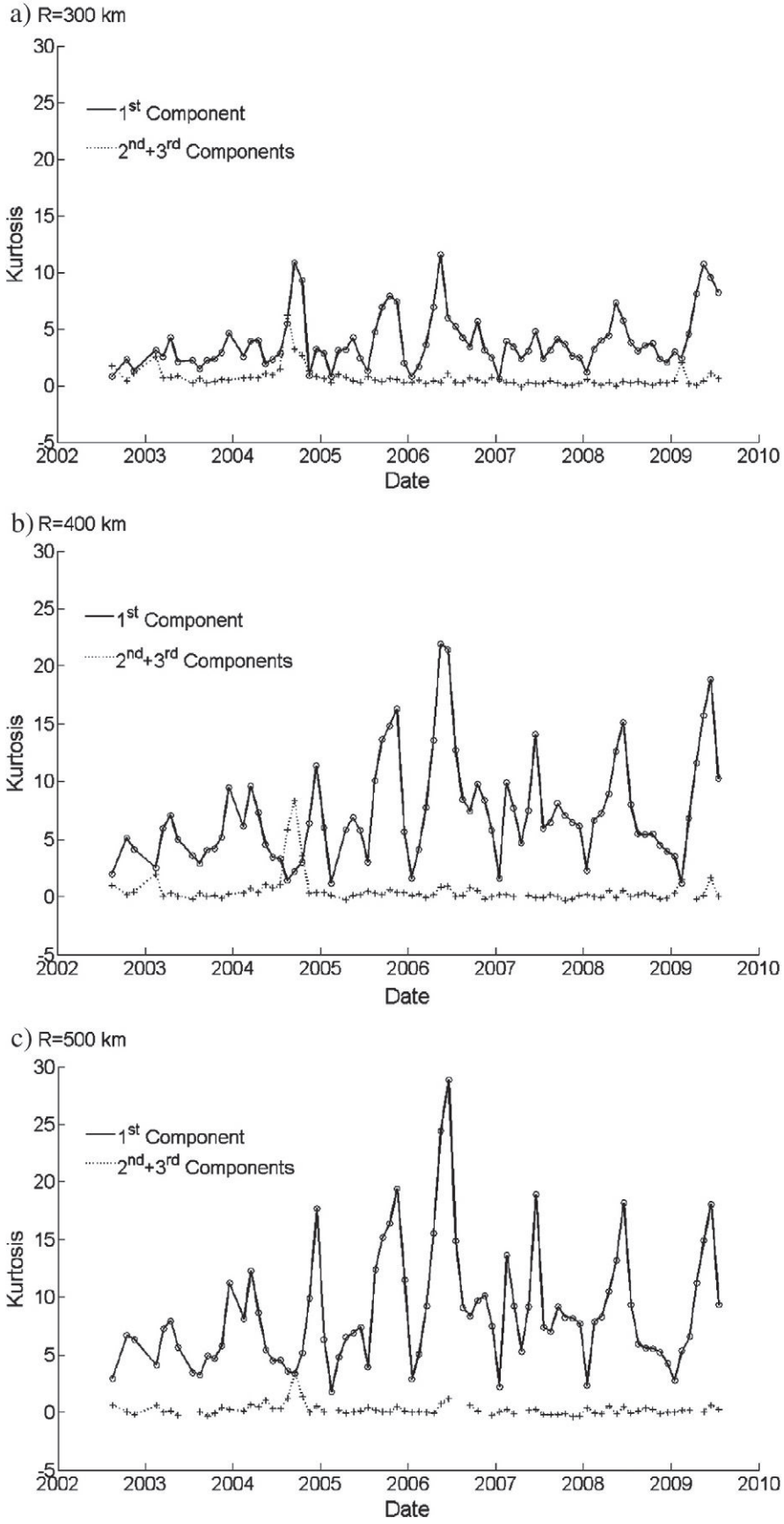


Fig. 2. Time series of the kurtosis of the mass anomalies detected by GRACE after Gaussian filtering for radii of a) 300 km, b) 400 km, and c) 500 km.

scaling factor for each specific component. The ICA-filtered CSR, GFZ, and JPL solutions are obtained by multiplying the *j*th IC with the *j*th mixing vector (Eq. 2). As the last two modes correspond to the north–south stripes, we present their sum in Fig. 1.

The first component is clearly ascribed to terrestrial water storage with variations in the range of ±450 mm of Equivalent Water Thickness (EWT) for an averaging radius of 400 km. The larger water mass anomalies are observed in the tropical regions, *i.e.*, the Amazon, the Congo, the Ganges and the Mekong Basins, and at high latitudes in the northern hemisphere. The components 2 and 3 correspond to the north–south stripes due to resonances in the satellite's orbits. They are smaller than the first component by a factor of 3 or 4 as previously found (Frappart et al., 2010).

The FastICA algorithm was unable to retrieve realistic patterns and/or amplitudes of TWS-derived from GRACE data preprocessed using a Gaussian filter with a radius 300 km for several months (02/2003, 06 to 11/2004, 02/2005, 07/2005, 01/2006, 01/2007, and 02/2009). Some of these dates, such as the period between June and November 2004, correspond to deep resonance between the satellites caused by an almost exact repeat of the orbit, responsible for a significantly poorer accuracy of the monthly solutions (Chambers, 2006). As ICA is based on the assumption of independence of the sources, if the sources

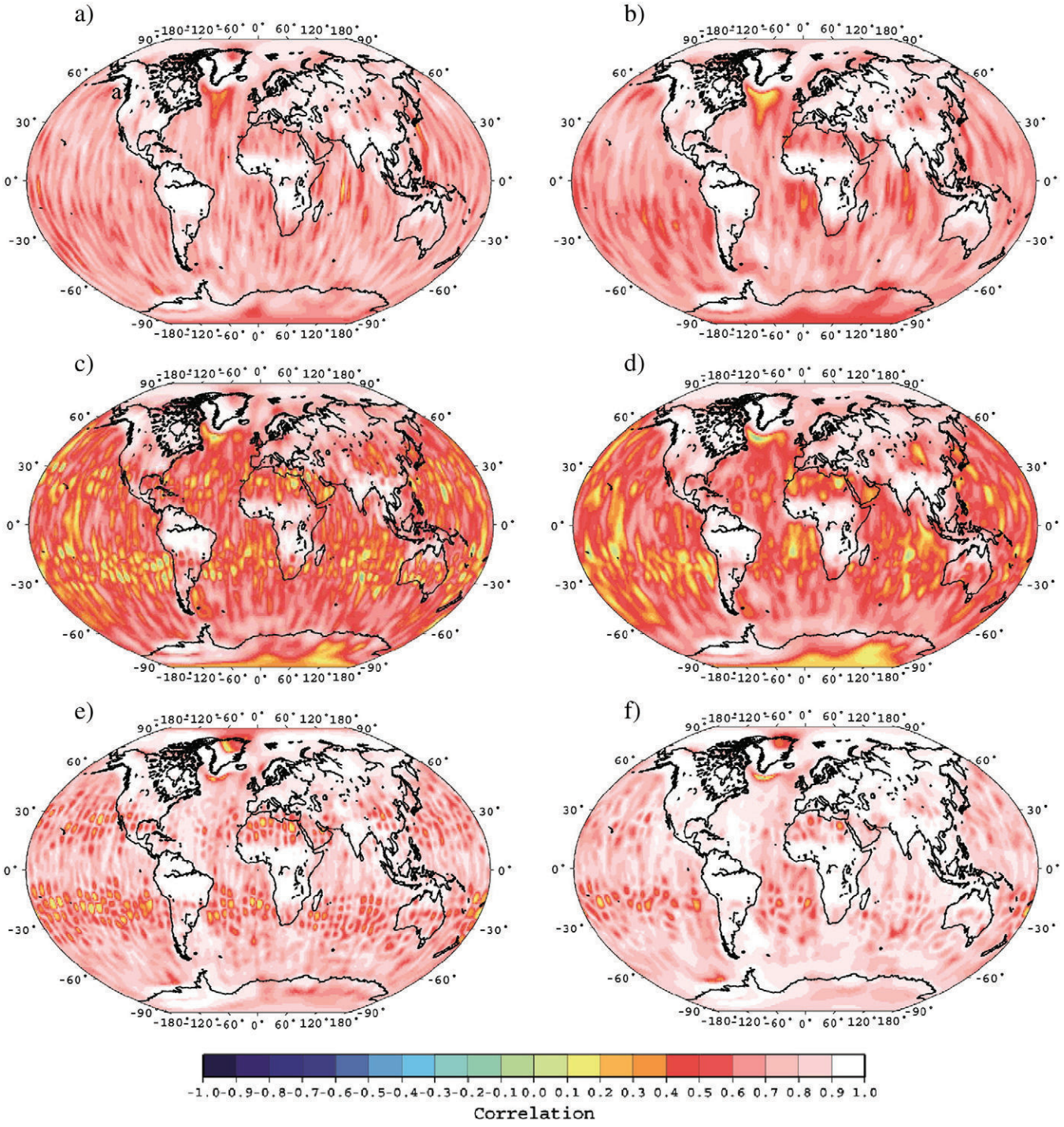


Fig. 3. Correlation maps over the period 2003–2008 between the ICA-filtered TWS and the Gaussian-filtered TWS. Left column: ICA400–G400 (a: CSR, c: GFZ, and e: JPL). Right column: ICA500–G500 (b: CSR, d: GFZ, and f: JPL).

exhibit similar statistical distribution, the algorithm is unable to separate them.

A classical measure of the peakiness of the probability distribution is given by the kurtosis. The kurtosis K_y is dimensionless fourth moment of a variable y and classically defined as:

$$K_y = \frac{E\{y^4\}}{E\{y^2\}^2} \tag{14}$$

If the probability density function of y is purely Gaussian, its kurtosis has the numerical value of 3. In the following, we will consider the excess of kurtosis ($K_y - 3$) and refer to the kurtosis as it is commonly done. So a variable y will be Gaussian if its kurtosis remains close to 0.

The time series of the kurtosis of the sources separated using ICA are presented in Fig. 2 for different radii of Gaussian filtering (300, 400 and 500 km) of GRACE mass anomalies. The kurtosis of the sum of the 2nd and 3rd ICs, corresponding to the north–south stripes, is most of

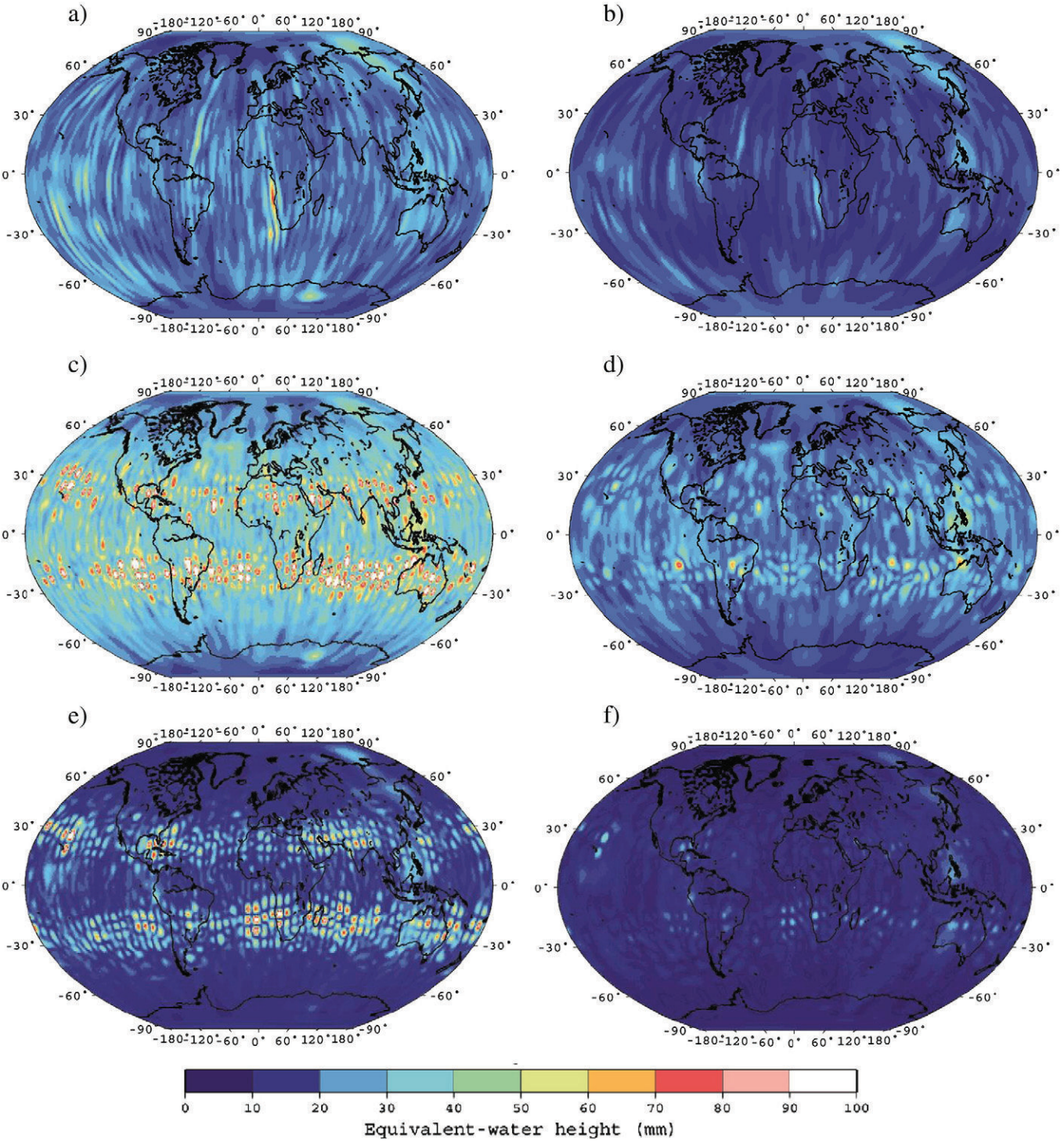


Fig. 4. RMS maps over the period 2003–2008 between the ICA-filtered TWS and the Gaussian-filtered TWS. Left column: ICA400–G400 (a: CSR, c: GFZ, and e: JPL). Right column: ICA500–G500 (b: CSR, d: GFZ, and f: JPL).

the time, close to 0; that is to say that the meridian oriented spurious signals is almost Gaussian. Almost equal values of the kurtosis for the 1st IC and the sum of the 2nd and 3rd ICs can be observed for several months. Most of the time, they correspond to time steps where the algorithm is unable to retrieve realistic TWS (02/2003, 08/2004, 11/2004, 02/2005, 01/2006, 01/2007, and 02/2009).

We also observed that the number of time steps with only one IC (the outputs are identical to the inputs, *i.e.*, no independent sources are identified and hence no filtering was performed) increases with the radius of the Gaussian filter (none at 300 km, 2 at 400 km, and 7 at 500 km).

In the following, as the ICA-derived TWS with a Gaussian prefiltering of 300 km, exhibits an important gap of 6 months in 2004, we will only consider the solutions obtained after a pre-processing with a Gaussian filter for radii of 400 and 500 km (ICA400 and ICA500).

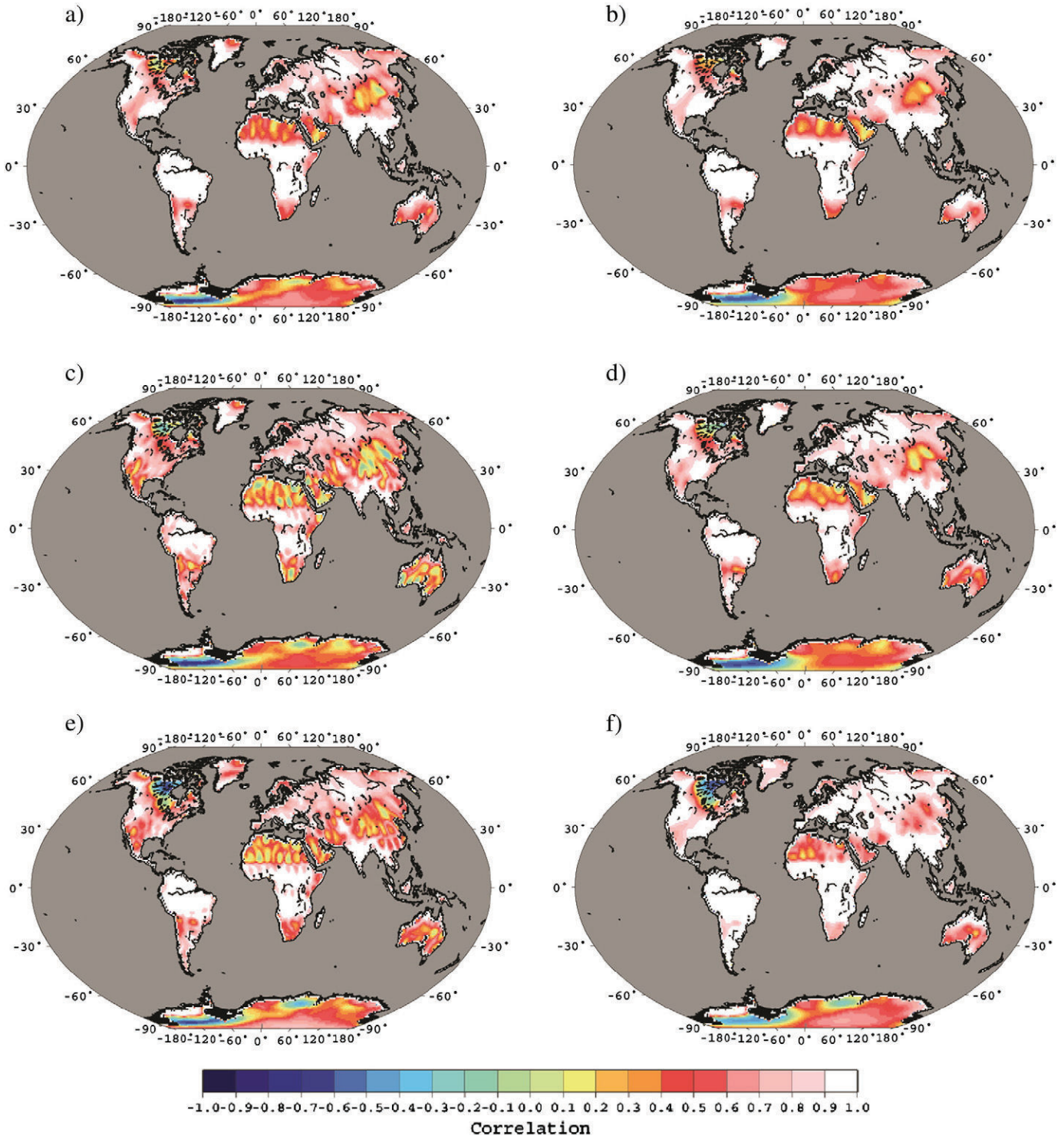


Fig. 5. Correlation maps over the period 2003–2008 between the ICA-filtered TWS and the destriped and smoothed TWS. Left column: ICA400–DS300 (a: CSR, c: GFZ, and e: JPL). Right column: ICA500–DS500 (b: CSR, d: GFZ, and f: JPL).

4.2. Global scale comparisons

Global scale comparisons have been achieved with commonly-used GRACE hydrology preprocessing: the Gaussian filter (Jekeli, 1981) and the destriping method (Swenson & Wahr, 2006) for several smoothing radii.

4.2.1. ICA versus Gaussian-filtered solutions

Advantages of extracting continental hydrology using ICA after a simple Gaussian filtering have to be demonstrated for the complete period of availability of the GRACE Level-2 dataset, as it was for one period of GRACE Level-2 data in Frappart et al. (2010). Numerical tests of comparisons before and after ICA have been made to show full

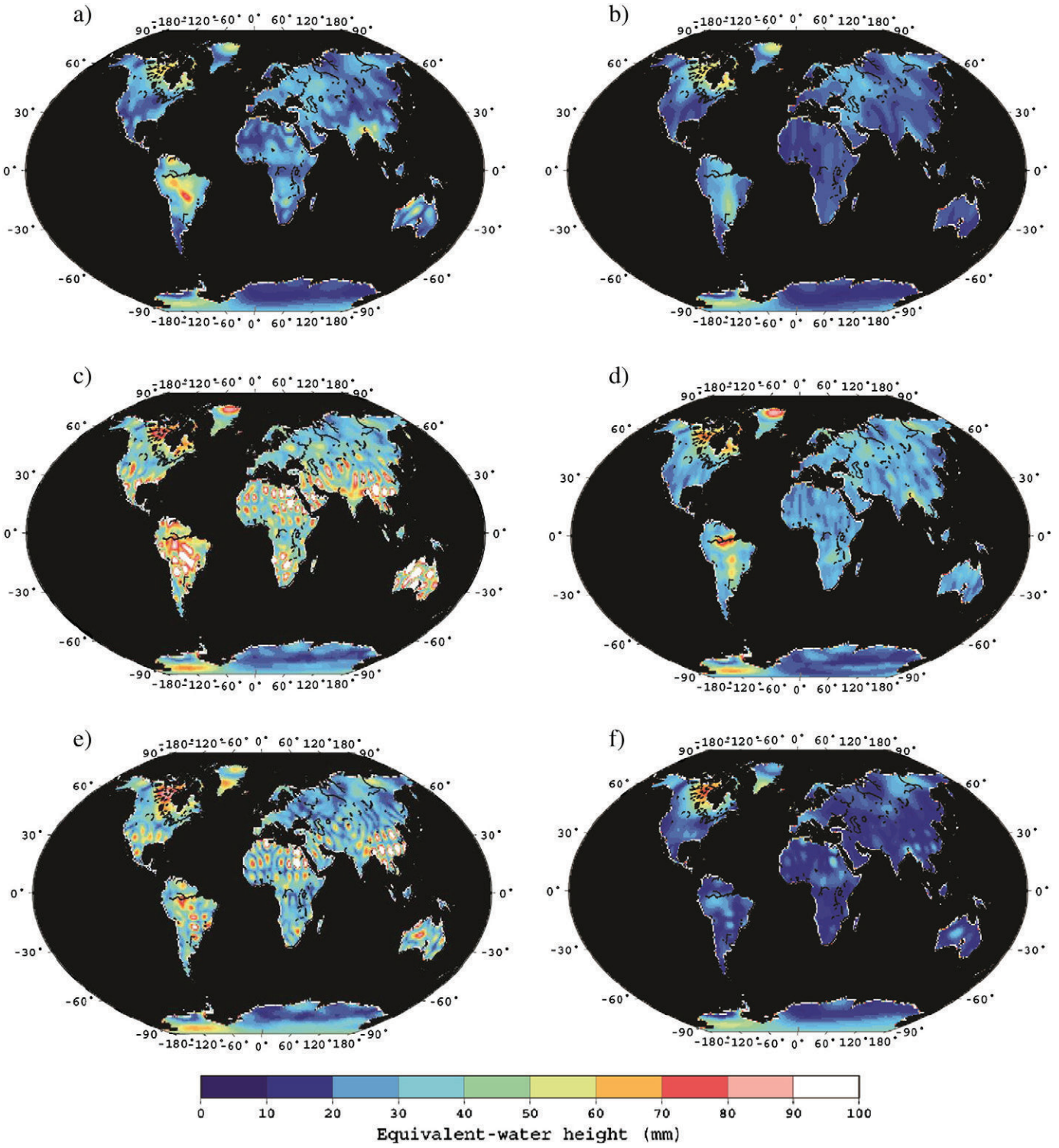


Fig. 6. RMS maps over the period 2003–2008 between the ICA-filtered TWS and the destriped and smoothed TWS. Left column: ICA400–DS300 (a: CSR, c: GFZ, and e: JPL). Right column: ICA500–DS500 (b: CSR, d: GFZ, and f: JPL).