

Denoising Satellite Gravity Signals by Independent Component Analysis

F. Frappart, G. Ramillien, P. Maisongrande, and M.-P. Bonnet

Abstract—Independent component analysis (ICA) is a blind separation method based on simple assumptions of the independence of sources and the non-Gaussianity of observations. An approach based on ICA is used here to extract hydrological signals over land and oceans from the polluting stripping noise due to orbit repetitiveness and present in the gravity anomalies detected by the Gravity Recovery and Climate Experiment (GRACE) satellites. We took advantage of the availability of monthly level-2 GRACE solutions from three official providers (i.e., CSR, JPL, and GFZ) that can be considered as different observations of the same phenomenon. The efficiency of the methodology is demonstrated on a synthetic case. Applied to one month of GRACE solutions, it allows for clearly separating the total water storage change from the meridional-oriented spurious gravity signals on the continents but not on the oceans. This technique gives results equivalent to the destriping method for continental water storage.

Index Terms—Filtering technique, gravimetry from space, hydrology, independent component analysis (ICA).

I. INTRODUCTION

THE Gravity Recovery and Climate Experiment (GRACE) geodetic mission provides a global mapping of the time variations of the gravity field [1]. So far, raw GRACE orbit observations are used to produce ten-day and monthly geoids in terms of Stokes coefficients (i.e., dimensionless spherical harmonics of the geopotential) up to degree 60–150. Solutions are corrected from atmospheric and oceanic mass variations using *ad hoc* reanalysis and models, but these data, which represent mainly continental hydrology and ocean masses (plus other unmodeled phenomena), contain meridian-oriented spurious gravity signals or stripes. This stripping is mainly due to resonances at degree 15, 30 . . . and insufficient forward modeling (principally ocean tides and atmospheric and oceanic nontidal mass variations). Several strategies for filtering GRACE solutions have been proposed to remove these north–south-oriented features and extract useful hydrological signals on continents. They include classical isotropic Gaussian filter [2], various optimal filtering decorrelations of GRACE errors [3]–[7], and statistical constraints on the time evolution of GRACE coefficients [8] or from global hydrology models [9]. Neverthe-

less, these approaches are insufficient as the results still suffer from aliasing, leakage, and noise contamination. We undertake a different strategy. Instead of comparing the GRACE solutions from official providers [Center for Space Research (CSR), GeoForschungsZentrum (GFZ), and Jet Propulsion Laboratory (JPL)], we assume that any CSR, GFZ, and JPL solution contains joint information. We combine GRACE solutions from these groups and use an independent component analysis (ICA) approach to extract signals that are statistically independent as the mixture of space and time decorrelated geophysical signal and noise, without subjective *a priori* constraints. ICA is a fast linear method which has been successfully applied in different domains like separating independent sources from microphone records [10]. Our objective is to separate the different independent components (i.e., the sources according to ICA terminology), combining the level-2 GRACE products from CSR, GFZ, and JPL (i.e., the observations) and expecting the ICA-estimated components to match interpretable geophysical signals. This approach is first tested on simulated GRACE observations and then applied to real GRACE products. We compared our results with those from the empirical destriping and smoothing method [11].

II. ICA

ICA is a powerful method for multivariate data analysis, particularly for blind signal separation, and has various practical applications, including communications, medical signal processing, speech signal processing, and so on [12]–[15]. A typical example of ICA is isolating speech signals from recordings of people talking simultaneously in a room (“cocktail party problem” [16]), under the assumption of no time delays or echoes. An important consideration is that if N sources are present, at least N observations (e.g., from microphones or receivers) are needed to retrieve the original source signals, with ICA remaining applicable for square or overdetermined problems. ICA proceeds by maximizing the statistical independence of the estimated components. As a first step of the method, non-Gaussianity of the input signals must be verified. The central limit theorem is used for measuring the statistical independence of the components. Typical algorithms for ICA use centering and whitening based on eigenvector decomposition and reduction of dimension as main processing steps. Whitening ensures that the input observations are equally treated before dimension reduction.

Many algorithms of different complexities have been developed for ICA [10]. For separating satellite gravity signals, we consider the *FastICA* algorithm, which is a computationally

Manuscript received October 1, 2009; revised October 23, 2009. Date of publication February 5, 2010; date of current version April 29, 2010. This work was supported in part by the fondation Sciences et Techniques pour l’Aéronautique et l’Espace in the framework of the CYMENT Project.

The authors are with the Université de Toulouse, Observatoire Midi-Pyrénées, 31400 Toulouse, France (e-mail: frederic.frappart@lmtg.obs-mip.fr; guillaume.ramillien@ntp.obs-mip.fr; philippe.maisongrande@cnes.fr; marie-paule.bonnet@ird.fr).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LGRS.2009.2037837

highly efficient method for performing the estimation of ICA [14]. It uses a fixed-point iteration scheme that has been found to be 10–100 times faster than conventional gradient methods for ICA [11].

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 , considering a linear statistical model, it can be written as

$$y = Mx \quad (1)$$

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

ICA aimed at estimating 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 following assumptions [15].

- 1) The mixing vectors are linearly independent.
- 2) The sources are statistically independent.

The original sources x 's can be simply recovered by multiplying the observed signals y 's with the inverse of the mixing matrix also known as the “unmixing” matrix

$$x = M^{-1}y. \quad (2)$$

ICA consists of three numerical steps. The first step of ICA is to center the observed vector, i.e., to remove its mean $m = E\{y\}$. 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 unity. Letting $C = E\{yy^t\}$ be the correlation matrix of the input data, we define a linear transform B that verifies the following two conditions:

$$\tilde{y} = By \quad (3)$$

$$E\{\tilde{y}\tilde{y}^t\} = I_P \quad (4)$$

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

This is easily accomplished by considering

$$B = C^{-\frac{1}{2}}. \quad (5)$$

A reduction of the dimension of the data to the number of independent components P , discarding the too small eigenvalues obtained by principal component analysis, is performed simultaneously as the whitening.

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

We apply a linear ICA through the *FastICA* algorithm made available at <http://www.cis.hut.fi/projects/ica/fastica/> to unravel the independent components of the monthly gravity field anomaly in the level-2 GRACE products. We assumed that the GRACE level-2 products from CSR, GFZ, and JPL are the dif-

ferent observations of the same monthly gravity anomaly, the land hydrology and ocean mass anomalies, and the north–south stripes, the independent sources.

III. DATA SET

A. Model Outputs Used to Generate Synthetic Signal

1) *Land Hydrology Signal From WGHM Model*: The WaterGAP Global Hydrology Model (WGHM) was specifically designed to estimate river discharge for water resource assessments [18]. It computes $0.5^\circ \times 0.5^\circ$ gridded time series of monthly runoff, snow depth, soil water within the root zone, ground water and surface water storage in rivers, lakes, and wetlands, and river discharge. Here, we used the total water storage (TWS) grid for March 2006 resampled at 1° to represent the continental part of the water cycle.

2) *OBP From ECCO Model*: Ocean bottom pressure (OBP) data represent the vertically integrated mass of the atmosphere and ocean above a peculiar point of the seafloor. In this study, we use the $1^\circ \times 1^\circ$ OBP grid for March 2006 from the estimating the circulation & climate of the ocean (ECCO) model with data assimilation of sea surface topography from altimetry (ECCO-JPL). The OBP data are available at <http://ecco.jpl.nasa.gov/external/>.

We remove from the TWS (OBP, respectively) grid the time average TWS (OBP, respectively) map between January 2003 and December 2007 to derive anomaly grids, thus being consistent with the processing of GRACE data over land and oceans.

B. GRACE Products

The GRACE mission, sponsored by NASA and DLR, has been collecting data since mid-2002.

- 1) Raw level-2 products: 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 60° (or a spatial resolution of 333 km) and corrected for oceanic and atmospheric effects [1], [19] 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) The monthly raw solutions (RL04) from CSR, GFZ, and JPL were destriped and smoothed by Chambers [11] for hydrological purposes. These three data sets 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>.

IV. RESULTS

A. Simulated GRACE Data

In order to validate the ICA method for separation of signals in GRACE products, we make a synthetic data set by adding different levels of noise to reference observations and inverting the results. March 2006 was chosen because continental water

mass amplitudes were particularly strong and thus recognizable (i.e., ~ 600 mm of equivalent water thickness (EWT) over the Amazon basin).

We generated a simulated GRACE observation by adding three complementary components: land hydrology, ocean, and noise signals. The land hydrology component comes from the TWS, expressed in terms of EWT, of the WGHM model. We used OBP from the ECCO model to simulate the ocean mass anomaly. To be consistent with the spatial resolution of GRACE, TWS and OBP signals were filtered with a 200–300-km Gaussian low-pass filter. The noise added in our simulations was obtained as the difference between level-1 and level-2 land water solutions from CNES/GRGS [20] over the same monthly time period and normalized from 10% to 100% of the maximum of the synthetic ocean and land signals.

To generate three synthetic sets of observations, similar to the GRACE products from CSR, GFZ, and JPL, for instance, as input for the ICA method, we multiply the noise signal by 0.9, 1, and 1.1 before adding the land hydrology and ocean components. The simulated GRACE observations present a clearly non-Gaussian histogram. Each input grid is converted in a vector of 64 800 ($= 360 \times 180$ grid points) values sorted according to their geographical coordinates for both land and oceans. Hence, we can apply the ICA method to these simulated GRACE observations.

We performed the ICA for levels of noise from 10% to 100%. These levels are defined as the ratio between the maximum of amplitude of the noise and the maximum of amplitude of the simulated GRACE signal. We present the results for a level of noise of 100% (Fig. 1, top). The *FastICA* method allows for successfully isolating the hydrological signal over land and ocean from the noise whatever the level of noise. The results of the ICA can be seen in the middle and bottom panels of Fig. 1. The synthetic GRACE input is corrupted by a strong noise whose maxima are of the same intensity as those of the hydrological signal. After applying the ICA method, two clearly separated modes appear: the hydrological signal and noise. The synthetic data correspond to an overdetermined case, that is to say, a case where the number of observations y is greater than the number of sources x . Both spatial patterns and maximum intensity are well separated. Numerical comparisons between the synthetic hydrological signal and its estimate by the ICA method are presented in Table I for increasing levels of noise. These differences logically increase with the noise. In all cases, the mean bias amplitude is lower than 0.5 mm of EWT and slightly negative. The maximum rms is lower than 3 mm of EWT, and the maximum difference reaches 35.5 mm, that is to say, 0.5% (6%, respectively) for a reference hydrological signal amplitude of 600 mm.

This preliminary test of using ICA for separating synthetic signals is successful, as the structures of the hydrological models are fully recovered with realistic amplitudes.

B. Real GRACE Data

We now apply the ICA method to the real GRACE data for separating useful hydrological signals from uncorrelated noise for a given monthly period. We consider three 1° global

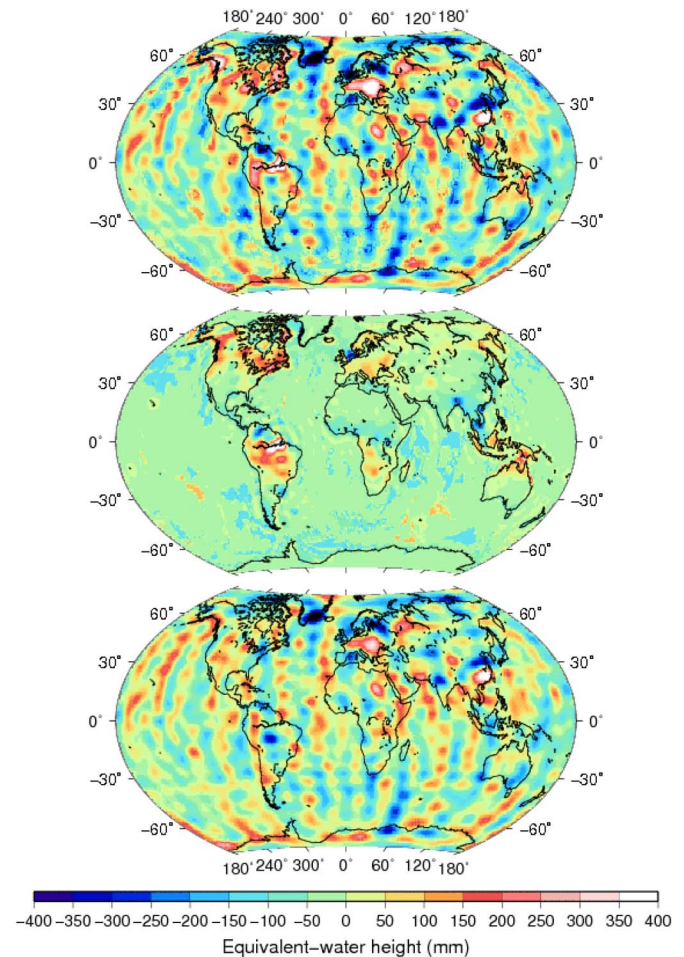


Fig. 1. (Top) Synthetic GRACE signal for a level of noise of 100%. (Middle) First ICA component corresponding to the land and ocean signals. (Bottom) Second ICA component associated with the north-south stripes. Units are millimeters of EWT.

TABLE I
COMPARISON BETWEEN THE RESULTS OF THE ICA METHOD AND THE SYNTHETIC DATA FOR INCREASING LEVELS OF NOISE

Noise level (%)	20	30	40	50	60	70	80	90	100
Mean Bias (mm)	-0.1	-0.13	-0.17	-0.2	-0.24	-0.28	-0.31	-0.35	-0.38
RMS (mm)	1.54	1.64	1.78	1.95	2.13	2.33	2.54	2.76	2.98
Maximum Difference (mm)	9.46	10.92	14.44	17.95	21.46	24.97	28.49	32.00	35.51

grids corresponding to the CSR, GFZ, and JPL level-2 GRACE solutions for March 2006 as three independent observations of the Earth's gravity field variations. Our starting assumption is that the level-2 GRACE solution is a linear combination of different spatially independent gravity sources, as long as this resulting mixing does not have a pure Gaussian distribution. Unfortunately, the histogram of the raw GRACE solutions is purely Gaussian as noisy stripes dominate the observation. We preprocessed the level-2 GRACE solutions using Gaussian filters with different radius from 300 to 900 km. When the GRACE solutions are Gaussian filtered with averaging radii ranging from 300 to 500 km, the distributions of the observations are sufficiently non-Gaussian to allow the use of the ICA method. For averaging radii lower than 300 km (or greater than

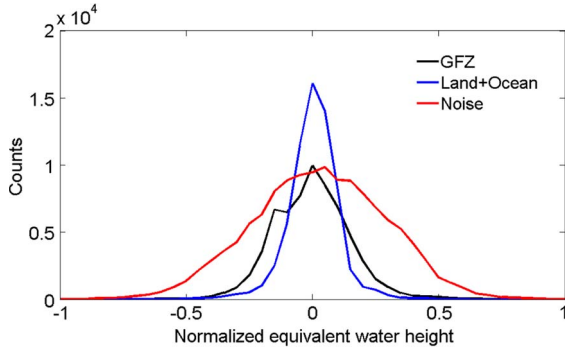


Fig. 2. Normalized histograms of the GRACE water storage from GFZ filtered with a Gaussian filter of 400 km of radius for March 2006 and of the corresponding ICA-estimated independent components.

500 km), the signal is dominated by the north–south stripes (respectively, the smoothing of the GRACE solutions is too important); thus, the corresponding histogram is Gaussian. As a consequence, the ICA method is unable to isolate the hydrological signal over land and oceans from noise in these cases.

We applied *FastICA* to Gaussian prefiltered GRACE solutions for averaging radii of 300, 400, and 500 km. As ICA is linear, the ICA-based CSR, GFZ, and JPL solutions only differ from a scaling factor for each specific component. The ICA-filtered CSR, GFZ, and JPL solutions are obtained by multiplying the j th independent component with the j th mixing vector (1). As the last two modes exhibit similar histograms corresponding to the north–south stripes, we present their sum in Fig. 2. The distribution of each independent component is clearly separated. We chose to present in Fig. 3 the independent the Gaussian-filtered using a radius of 400 km level-2 GFZ solution and the independent components corresponding to ICA-filtered GFZ solutions. The first component is clearly related to continental hydrology with variations in the range of $+450/-300$ mm ($350/-200$ mm, $300/-150$ mm) of EWT for an averaging radius of 300 km (respectively, 400 and 500 km) and to ocean signals with variations in the range of ± 230 mm (± 180 mm, ± 150 mm) of EWT for an averaging radius of 300 km (respectively, 400 and 500 km; see Table II). The larger water mass anomalies are observed in the Amazon, Congo, and Ganges basins over land, and in the Mediterranean Sea and the Gulf of Carpentaria by northern Australia over ocean (Fig. 3, middle). Components 2 and 3 (we present their sum in Fig. 3, right bottom) correspond to the north–south stripes, i.e., resonances in the satellite’s orbits. They are smaller by a factor of three or four (Table II).

C. Comparisons With Destriped and Smoothed GRACE Data

The first component has been compared to the destriped and smoothed level-2 GRACE data processed by Chambers [11] for each data center. The results of this comparison are presented in Table III over land and oceans for radii of 300 and 500 km. A good agreement between the ICA estimates and the destriped and smoothed GRACE data is observed over the continents and increases with the averaging radius. This result is consistent with the fact that the dynamics of

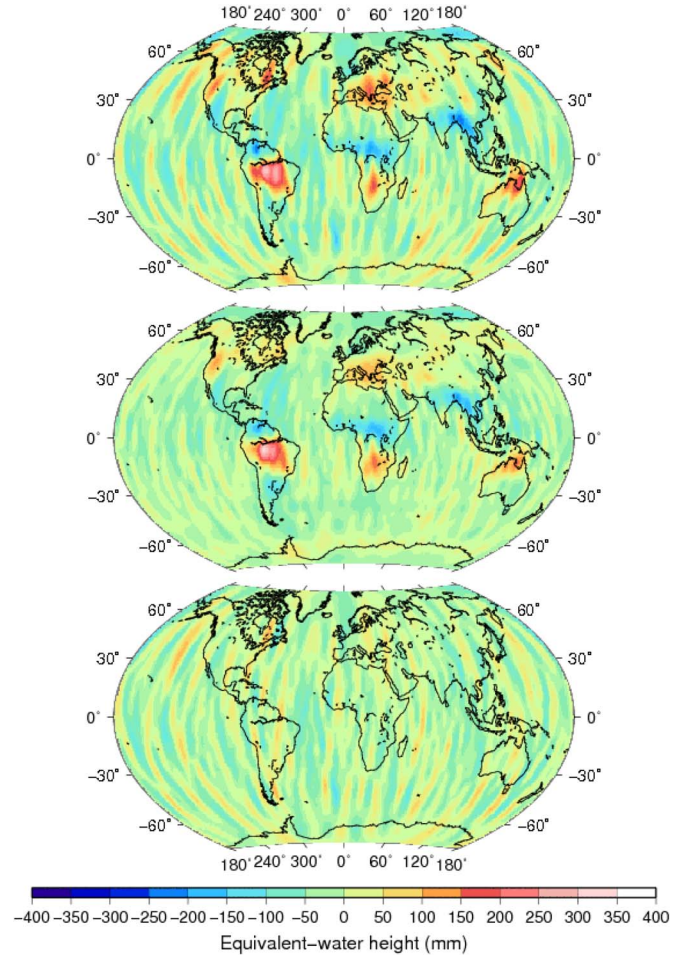


Fig. 3. (Top) GRACE water storage from GFZ filtered with a Gaussian filter of 400 km of radius for March 2006. (Middle) First ICA component corresponding to land hydrology and ocean mass. (Bottom) Sum of the second and third components corresponding to the north–south stripes. Units are millimeters of EWT.

the noise in the prefiltered observations logically decreases as the averaging radius increases. Table II presents the main characteristics of the ICA estimates for 300-, 400-, and 500-km cases and compares them to the characteristics of the destriped and smoothed solutions. For a 300-km averaging radius, the ICA estimates present a stronger signal and the same variability as destriped and smoothed solutions, whereas for an averaging radius of 500 km, signal amplitudes are similar, but the ICA-based solutions exhibit a lower variability. For an averaging radius of 400 km, the ICA estimates are very similar to the destriped and smoothed products at 500 km (same level of variability, correlation of 0.9) but exhibit a stronger signal.

Over the oceans, the agreement between the ICA estimates and the destriped and smoothed products is lower (Table III). This is mainly due to the fact that the GRACE signal over the oceans is generally weaker than the land signals by factors of two or three, both on seasonal and interannual time scales and also because the mean monthly ocean model was not added back. Moreover, as the signal over the oceans is very different from one GRACE raw product to another, the ICA approach is unable to well separate the geophysical signal from the north–south stripes.

TABLE II
CHARACTERISTICS OF THE FIRST MODE OF THE ICA ESTIMATES FOR AVERAGING RADII OF 300,
400, AND 500 km AND DESTRIPED AND SMOOTHED PRODUCTS AT 300 AND 500 km

	ICA filtered estimates									Destriped and smoothed products					
	300 km			400 km			500 km			300 km			500 km		
	CSR	GFZ	JPL	CSR	GFZ	JPL	CSR	GFZ	JPL	CSR	GFZ	JPL	CSR	GFZ	JPL
Amplitude max (mm)	444	408	427	356	339	326	300	289	275	403	359	344	320	294	278
Std (mm)	61	56	59	52	50	48	47	46	43	61	61	59	53	53	50

TABLE III
COMPARISON BETWEEN ICA ESTIMATES AND DESTRIPED AND SMOOTHED GRACE DATA
OVER LAND AND OCEANS FOR AVERAGING RADII OF 300 AND 500 km

	Land						Ocean					
	300 km			500 km			300 km			500 km		
	CSR	GFZ	JPL	CSR	GFZ	JPL	CSR	GFZ	JPL	CSR	GFZ	JPL
RMS (mm)	37.6	37.5	37.3	21.1	22.4	20.0	40.5	39.8	39.9	20.5	22.6	20.7
Correlation	0.8	0.8	0.8	0.9	0.9	0.9	0.4	0.4	0.4	0.6	0.5	0.5

V. CONCLUSION

Our study is the first attempt to destripe satellite gravity information using an ICA-based strategy, only taking the statistical properties of the observations into account and without any extra information. After having been successfully tested on synthetic noisy data (rms error lower than 0.5% for a maximum amplitude of noise equal to the maximum of the continental hydrology signal), ICA has been applied to separate the statistically independent components of the Earth's gravity field by combining monthly level-2 land water solutions processed by the three official providers of GRACE products (CSR, GFZ, and JPL). The main component corresponds to the water mass anomaly on continents and oceans. The second and third ICA estimates are pure noise of less amplitude, to be cancelled from the GRACE signals for geophysical interpretation. On real GRACE data, this approach requires the use of a preprocessing filter to verify the assumption of non-Gaussianity of the observations. This method gives equivalent results as the destripping method on the continents and even less smoothed results. On the oceans, as the signal is weaker and the input is quite different, ICA is unable to extract correctly the geophysical signal from the noise. Our study demonstrates the ability of ICA to discriminate efficiently geophysical components from the total measured gravity field, as continental hydrology change, and even in the presence of important uncorrelated noise. In the future, this efficient linear technique offers an interesting alternative to classical filtering, for example, it would be considered for cleaning an entire time series of GRACE geoids from polluting noise.

REFERENCES

- [1] B. D. Tapley, S. Bettadpur, J. C. Ries, P. F. Thompson, and M. Watkins, "GRACE measurements of mass variability in the Earth system," *Science*, vol. 305, no. 5683, pp. 503–505, Jul. 2004.
- [2] J. Wahr, S. Swenson, and I. Velicogna, "The accuracy of GRACE mass estimates," *Geophys. Res. Lett.*, vol. 33, no. 6, p. L06401, Mar. 2006. DOI: 10.1029/2005GL0253052005.
- [3] S.-C. Han, C. K. Shum, C. Jekeli, C.-Y. Kuo, C. Wilson, and K.-W. Seo, "Non-isotropic filtering of GRACE temporal gravity for geophysical signal enhancement," *Geophys. J. Int.*, vol. 163, no. 9, pp. 18–25, Oct. 2005.
- [4] K.-W. Seo and C. R. Wilson, "Simulated estimation of hydrological loads from GRACE," *J. Geod.*, vol. 78, no. 7/8, pp. 442–456, Mar. 2005.
- [5] I. Sasgen, Z. Martinec, and K. Fleming, "Wiener optimal filtering of GRACE data," *Stud. Geophys. Geod.*, vol. 50, no. 4, pp. 499–508, Oct. 2006.
- [6] S. Swenson and J. Wahr, "Post-processing removal of correlated errors in GRACE data," *Geophys. Res. Lett.*, vol. 33, no. 8, p. L08402, Apr. 2006. DOI: 10.1029/2005GL025285.
- [7] J. Kusche, "Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models," *J. Geod.*, vol. 81, no. 11, pp. 733–749, Nov. 2007.
- [8] J. L. Davis, M. E. Tamisiea, P. Elósegui, J. X. Mitrovica, and E. M. Hill, "A statistical filtering approach for Gravity Recovery and Climate Experiment (GRACE) gravity data," *J. Geophys. Res.*, vol. 113, no. B4, p. B04410, Apr. 2008. DOI: 10.1029/2007JB005043.
- [9] G. Ramillien, F. Frappart, A. Cazenave, and A. Güntner, "Time variations of land water storage from an inversion of 2 years of GRACE geoids," *Earth Planet. Sci. Lett.*, vol. 235, no. 1/2, pp. 283–301, Jun. 2005.
- [10] J. V. Stone, *Independent Component Analysis: A Tutorial Introduction*. Cambridge, MA: MIT Press, 2004.
- [11] D. P. Chambers, "Evaluation of new GRACE time-variable gravity data over the ocean," *Geophys. Res. Lett.*, vol. 33, no. 17, p. L17603, Sep. 2006. DOI: 10.1029/2006GL027296.
- [12] P. Comon, "Independent component analysis: A new concept?" *Signal Process.*, vol. 36, no. 3, pp. 287–314, Apr. 1994.
- [13] J. H. van Hateren and A. Van Der Schaaf, "Independent component filters of natural images compared with simple cells in primary visual cortex," in *Proc. Biol. Soc.*, Mar. 1998, vol. 265, no. 1394, pp. 359–366.
- [14] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, no. 4/5, pp. 411–430, May/Jun. 2000.
- [15] L. De Lathauwer, B. De Moor, and J. Vandewalle, "An introduction to independent component analysis," *J. Chemometrics*, vol. 14, no. 3, pp. 123–149, May/Jun. 2000.
- [16] S. Haykin and Z. Chen, "The cocktail party problem," *Neural Comput.*, vol. 17, no. 9, pp. 1875–1902, Sep. 2005.
- [17] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, May 1999.
- [18] P. Döll, F. Kaspar, and B. Lehner, "A global hydrological model for deriving water availability indicators: Model tuning and validation," *J. Hydrol.*, vol. 270, no. 1/2, pp. 105–134, Jan. 2003.
- [19] S. Bettadpur, "CSR level-2 processing standards document for level-2 product release 0004," GRACE 327–742, 2007. Rev. 3.1.
- [20] J.-M. Lemoine, S. Bruinsma, S. Loyer, R. Biancale, J.-C. Marty, F. Pérosanz, and G. Balmino, "Temporal gravity field models inferred from GRACE data," *Adv. Space Res.*, vol. 39, no. 10, pp. 1620–1629, Oct. 2007.