Geophysical Journal International

Geophys. J. Int. (2009) **179,** 1499–1515 doi: 10.1111/j.1365-246X.2009.04355.x

Evaluation of GRACE filter tools from a hydrological perspective

S. Werth, A. Güntner, R. Schmidt* and J. Kusche†

Deutsches GeoForschungsZentrum GFZ, Telegrafenberg, 14473 *Potsdam, Germany. E-mail: swerth@gfz-potsdam.de*

Accepted 2009 August 4. Received 2009 July 22; in original form 2008 March 12

SUMMARY

Approximately seven years of time-variable gravity data from the satellite mission Gravity Recovery and Climate Experiment (GRACE) are available to quantify present-day mass variations on and near the Earth's surface. Mass variations caused by the continental water cycle are the dominant signal component after subtracting contributions from atmosphere and oceans. This makes hydrology a primary area of application of GRACE data. To derive water storage variations at the scale of large river basins, appropriate filter techniques have to be applied to GRACE gravity fields given in a global spherical harmonic representation. A desirable filter technique minimises both GRACE data error and signal leakage across the border of the region of interest. This study evaluates the performance of six widely used filter methods (isotropic filters, anisotropic filters and decorrelation methods) and their parameter values to derive regionally averaged water mass variations from GRACE data. To this end, filtered time series from GRACE for 22 of the world's largest river basins were compared to continental water mass variations from a multimodel mean of three global hydrological models (WGHM, GLDAS and LaD). Filter-induced biases for seasonal amplitudes and phases of water storage variations, as well as satellite and leakage error budgets, were quantified for each river basin and explained in terms of storage variations in and around the basin. The optimum filter types and filter parameters were identified for each basin. The best results were provided by a decorrelation method that uses GRACE orbits for the filter design. Our ranking between all filter types and parameters depended on the geographical location, shape and signal characteristics of the specific river basin. Based on a multicriterial evaluation of satellite and leakage error, as well as an error assessment of the hydrological data, the filter selection and parameter optimisation results were shown to be reliable for 17 river basins. The results serve as a guideline for the optimal filtering of GRACE global spherical harmonic coefficients for hydrological applications.

Key words: Time series analysis; Time variable gravity; Hydrology.

1 INTRODUCTION

Global monthly gravity field solutions from the US–German satellite gravity mission Gravity Recovery and Climate Experiment (GRACE) trace mass redistributions close to the Earth's surface (Reigber *et al.* 2005; Tapley *et al.* 2004). Assumptions as defined by Chao (2005) and Wahr *et al.* (1998) enable the transformation of gravity variations into time series of global maps of surface mass anomalies. Due to their integrative nature, global coverage and previously unrivalled accuracy, GRACE-derived mass variations give insight into processes within the Earth's subsystems. This knowledge helps to improve understanding and modelling of geophysical mass transfers within the Earth's system. For example, recent studies have considered estimations of ice mass loss within polar regions (Chen *et al.* 2008; Wouters *et al.* 2008), observations of oceanic circulations (Dobslaw & Thomas 2007), components of the continental water cycle (Boronina & Ramillien 2008; Swenson *et al.* 2008; Niu *et al.* 2007) and interactions between these subsystems (Chambers *et al.* 2007; Ramillien *et al.* 2008).

Water mass variations within the continental hydrological cycle are a major signal recovered from the GRACE gravity data after removal of signals from tides, atmosphere and oceans. Numerous studies show an overall good agreement between variations of total continental water storage (TWS) from global hydrological models and from GRACE, especially for large river basins (for a recent overview see Güntner 2009). Ramillien *et al.* (2005) and Schmidt *et al.* (2006) compared output from the WaterGAP Global Hydrology Model (WGHM) and the Land Dynamics (LaD) model to GRACE data and found a good general correspondence. Syed *et al.* (2008) confirmed this agreement for the Global Land Data

*Now at: EADS Astrium GmbH, Munich, Germany.
†Now at: Bonn University, Bonn, Germany.

Assimilation System (GLDAS). Schmidt *et al.* (2008b) found similar dominant seasonal and inter-annual TWS periods for GRACE and simulated data of GLDAS, LaD and WGHM for the Amazon, the Ganges and the Mississippi river basins. However, the degree of agreement between GRACE and hydrological models clearly varies with the region and river basin of interest (e.g. Ramillien *et al.* 2005). Except for a few regions (e.g. Swenson *et al.* 2006), there is a lack of independent TWS observation data at large spatial scales that are consistent with GRACE. Thus, in spite of the uncertainties inherent in hydrological models, simulation data are currently the only way to evaluate TWS variations from GRACE for large areas.

A method of deriving time series of total regionally-averaged mass variations from global GRACE gravity fields represented as coefficients of spherical harmonics (SH) has been suggested by Wahr *et al.* (1998). GRACE measurement and processing errors, which are often referred to as satellite errors, mostly distort SH coefficients of high resolution. One way to reduce noise in the monthly solutions is to constrain (or regularise) the coefficients (e.g. towards a mean field) in the course of GRACE data processing (e.g. Save *et al.* 2008; Watkins *et al.* 2008; Lemoine *et al.* 2007). But the need for regularisation characteristics vary widely between different scientific applications of GRACE data (Kusche 2007). As a result, unconstrained solutions are mostly published by the processing centres, making the application of a post-processing filter technique indispensable. Filtering aims at the suppression of noisy high-resolution coefficients of the gravity field, that is, smoothing the original data to a lower spatial resolution. Furthermore, decorrelation techniques can be applied to remove striping artefacts of GRACE gravity data, which can be interpreted as realisation of anisotropically correlated noise in the coefficients (Swenson & Wahr 2006). In this paper, both smoothing and decorrelation techniques are subsumed under the term filtering.

As a drawback, filtering implies the leakage of signals from outside the region of interest into the resulting time series as well as the non-unity weighting of the signal variability inside the region of interest (Swenson & Wahr 2002; Klees *et al.* 2007). Hereafter, both effects will be referred to as leakage. According to Swenson & Wahr (2002), leakage varies between different filter types. Klees *et al.* (2007) listed three simplified cases of leakage scenarios: (1) the signal (TWS anomaly) outside the area of interest is of the same sign as the signal inside, (2) the signal outside is zero, and (3) the signal outside is of a different sign than the signal inside. Using a Gaussian smoother of different filter widths, Klees *et al.* (2007) concluded that the first case would lead to the lowest total leakage error and the third case to the highest leakage error. Since the signal intensities outside and inside of a river basin vary widely between different regions due to varying hydrological characteristics, the leakage error also depends on the region of interest.

Consequently, in order to select an appropriate filter method, the user has to balance between remaining satellite errors and the spatial resolution (i.e. leakage error), and has to find an optimal balance, specifically, for each river basin they intend to analyse. Filter types developed so far differ in their assumptions on signal-noise properties of the true GRACE-derived mass variations. Some studies evaluate specific filter types. For example, Swenson & Wahr (2002) developed two anisotropic methods and compared them with an isotropic Gaussian filter by evaluation of signal leakage. Han *et al.* (2005) showed that anisotropic smoothing is necessary to consider the degree and order dependence of GRACE coefficient errors. Seo *et al.* (2006) described error reductions within GRACE water-mass variations when using a time-dependent noise-minimizing filter instead of the Gaussian method. Schrama *et al.* (2007) determined the radius of the Gaussian smoother with an empirical orthogonal function (EOF) analysis and by comparing it with GPS load measurements. A recent decorrelating filter method and an overview of several GRACE filter techniques was given by Kusche (2007). In order to compensate for the effect of amplitude damping by filtering, Velicogna & Wahr (2006) introduced a scaling factor to recover the full hydrological signal in time series of TWS variations. Similarly, Chen *et al.* (2007) used a scale factor to readjust amplitude damping effects caused by the Gaussian filtering of GRACE data relative to TWS from GLDAS simulations.

An optimised spatial resolution of GRACE data by use of an adequate filter algorithm is especially crucial for hydrological studies, where a separation of water-mass variations of different river basins is of high interest for water balance studies. The transport of water masses can be concentrated to small regions like the river network and its inundation areas with low signal correlations to other mass transport processes. In addition, the hydrological signal of interest is composed of mass variations in several water storage compartments of the continental water cycle (such as snow, surface water or groundwater), which differ in their modes of temporal variability or spatial correlation lengths (Güntner *et al.* 2007). Hence, particular hydrological features have to be considered when selecting appropriate GRACE filter techniques with small leakage and satellite errors for applications in continental hydrology. Nevertheless, a comprehensive evaluation of various filters from this perspective is missing in the literature so far. In particular, the following questions arise. (1) Which filter is optimal for which scale, location or shape of a river basin of interest? (2) Which filter is superior for which regional signal properties, that is, for different sources of water mass variations in continental hydrology? (3) What are the filter properties in terms of TWS amplitude falsification and phase shifts? (4) Which filter removes striping artefacts sufficiently well? In this study, we address the first three questions from the perspective of hydrological applications such as water balance analysis or hydrological modelling. To this end, we evaluate GRACE filter methods by simulating data from global hydrological models, which at present provide the only alternative data set of TWS variations for large areas.

2 METHODS AND DATA

To evaluate different filter types, time series of continental water storage variations from GRACE (Section 2.2) were evaluated using the three global hydrological models WGHM, GLDAS and LaD (Section 2.3). In the absence of alternative observation data at the relevant scale, hydrological model data were considered the most realistic information on continental water mass variations. An analysis of differences between the models was undertaken to uncover their uncertainties (Section 3.1). To reduce model-specific errors in the evaluation data set, an ensemble mean of the three models was used for the filter analysis. It was assumed that the reduced GRACE signal used in this study is governed by hydrological processes, and that the GRACE data are corrupted by satellite errors but not by other geophysical processes. The different filter methods (Section 2.1) were applied to compute time series of water storage variations for selected large river basins, after converting the hydrological fields into a spherical harmonic representation. To assure consistency, GRACE and model data were filtered in the same way. The similarity of measured and modelled TWS time series was evaluated by a correspondence criterion, which is described

Table 1. Overview of the tested filter methods I–VI.

Method	Variable parameter	Reference
I	r_g	Jekeli (1981)
II	$\Delta_{\max}$	Swenson & Wahr (2002)
III	G_l, σ_0	Swenson & Wahr (2002)
IV	f	Seo *et al.* (2006)
V	w_a, n_e	Swenson & Wahr (2006)
VI	x	Kusche (2007)

in Section 2.4. Computations were repeated for each filter method with varying filter parameters. The optimal filter method of deriving water mass variations was selected for each river basin from the maximum of the correspondence criteria, which is expected when the total error (leakage and satellite error) in the filtered GRACE data is minimal.

2.1 Filter methods

In this study, six post-processing filter methods for derivation of regionally averaged water mass variations from GRACE's global gravity field solutions were evaluated. The smoothing of the gravity field can be interpreted by a weighted spatial averaging for a region of every point on the globe in order to reduce noise that disturbs the signal components on higher spatial scales. A short description of the isotropic (degree dependent) filters, the anisotropic (degree and order dependent) filters and the two anisotropic decorrelation methods used in our study, is given below. For details on the filter methods, the reader should refer to the respective original publications. For each filter method, the parameters that define the degree of smoothing strength are explained below (see a list in Table 1).

(I) The widely used isotropic Gaussian filter was proposed by Jekeli (1981) as a way of smoothing out the Earth's gravity field. Its weighting function is derived from the Gaussian probability density function, which has its highest weight in the centre and diverges to zero with increasing distance from the kernel. The form parameter of the symmetric bell-shaped weighting function may be expressed as filter width r_g (eq. 59 in Jekeli 1981). The variable r_g represents the radius at which the filter weighting function declines to 50 per cent of its maximum value, and it is used to tune this degree-dependent smoothing method.

(II) Another filtering method was developed by Swenson & Wahr (2002) and applied with degree- and order-dependency for this analysis. The idea behind this filter design is to apply less smoothing to GRACE coefficients with relatively small errors that are relevant to a signal within the region of interest. Hence, a spherical harmonic representation of the basin function is used to compute the filter weights. No direct assumption about the signal is introduced. The user may tune this method by deciding for a total maximum satellite error of basin average $\Delta_{\max}$ (eq. 45 in Swenson & Wahr 2002). To approximate this *a priori* fixed maximum satellite error, the filter weights are computed iteratively from a propagation of the smoothed GRACE coefficient errors to the basin average.

(III) Another degree- and order-dependent technique by Swenson & Wahr (2002) minimises the sum of GRACE satellite error and signal leakage. The satellite error is propagated from the GRACE coefficient errors. Signal leakage is estimated by an exponential signal model, which is parameterised by the auto-correlation length G_l and standard deviation σ_0 of the expected geophysical signal (eq. 41 in Swenson & Wahr 2002).

(IV) Seo *et al.* (2006) proposed a time-dynamic filter that optimises the signal-to-noise ratio of each GRACE coefficient individually. We applied the method B_4 of their study, which uses the GRACE SH coefficients themselves as a signal estimate. Seo *et al.* (2006) derived a monthly filter version from the monthly GRACE coefficient errors. For the present study, a static filter was computed from the variance of the monthly coefficient errors. These variances were modified with a dimensionless error factor f, as a means of tuning the filter's degree of smoothing.

(V) Swenson & Wahr (2006) published an empirical decorrelation method that has to be followed by a subsequent application of one of the filter methods explained above. To reduce the correlation between coefficients of the same order but increasing degrees, they fit and remove a quadratic polynomial in a moving window from the coefficients, and they do so separately for even and odd degrees. The moving window is cantered at the coefficient to be filtered. No details on the window size are provided by Swenson & Wahr (2006); therefore, its design orients on Press *et al.* (1992) for the present study. The size of the window has to be decreased (e.g. with a Gaussian function) for increasing degrees in order to avoid too much signal damping. Consequently, one has to define four parameters for the decorrelation process: the initial and the final window size, w_a and w_e, as well as the degree of the first and last coefficient to be filtered, n_a and n_e. For computations represented below, $n_a = 2$ and $w_e = 3$ were fixed. Discrete variations for $w_a = [10, 20, 30, 40, 50]$ and $n_e = [10, 20, 30]$ were tested. Thereafter, a global filter optimisation described by Chen *et al.* (2006), who proposed to maximise the ratio of the spatial signal root mean square (RMS) for ocean versus land, was applied. The three optimised versions of filter V were concluded from a combination of $n_e = [30]$ with $w_a = [10, 20, 30]$, which were used for further investigations.

(VI) Another decorrelation method, by Kusche (2007), makes use of the GRACE orbital geometry and can be interpreted as an anisotropic filter. This method imitates the regularisation of GRACE data processing, using *a priori* diagonal signal and dense error covariance matrices. The latter are derived synthetically from GRACE orbits. The filter's degree of smoothing may be tuned by a regularisation parameter $a = 10^x$ of the signal covariance matrix (eq. 22 in Kusche 2007). Three filter versions with $x = 12$, $x = 13$ and $x = 14$ were applied in this study.

2.2 GRACE data

Monthly basin-averaged surface-mass variations were derived from GRACE-only global gravity field-model time series generated at GFZ German Research Centre for Geosciences (GRACE Level-2 products, version GFZ-RL04, Schmidt *et al.* 2008a). These data were obtained from the GFZ Information System and Data Center for a period ranging from 02/2003 until 07/2007 (excluding unavailable months 06/2003 and 01/2004) up to degree and order 120. They consist of unconstrained gravity fields (Flechtner 2007). Effects of the atmosphere and oceans are removed at the GRACE data centre by applying the appropriate model data. For this study, water mass variations are derived relative to a mean field for the years 2003–2006, and trends were removed from the time series. Coefficients from degree 2 were used, and degree 1 coefficients were set to zero. This is adequate because degree-1 coefficients are also excluded from the hydrological model data used for comparison. The accuracy of GRACE gravity fields varies in time and space. Schmidt *et al.* (2008a) quantified the global average error of derived water mass variations to 13–15 mm of a water mass equivalent column (w.eq.) for a circular area with a radius of 800 km. For filter parameterisations, estimates of GRACE error covariances were taken

from calibrated coefficient errors, which are published together with GFZ-RL04 fields.

2.3 Hydrological data

Continental water-storage data provided by three global hydrological models were used for the analyses.

The WaterGAP Global Hydrology Model (WGHM, Döll *et al.* 2003) is a conceptual global model that simulates the continental water cycles, excluding the regions of Antarctica and Greenland. Modelled water storages include interception, soil water, snow, groundwater and surface water. For this study, data sets were available from 01/2003 until 12/2007 from the most recent version of the model (Hunger & Döll 2008). WGHM was forced by monthly climate data from European Centre for Medium-Range Weather Forecasts (ECMWF) and precipitation data from GPCC (Global Precipitation Climatology Centre). Output were of 0.5° resolution and were calibrated by tuning a runoff coefficient parameter against observed river runoff at 1,235 discharge stations worldwide. Water storage simulated with WGHM has recently been analysed at the global scale by Güntner *et al.* (2007).

The GLDAS (Rodell *et al.* 2004) may incorporate a variety of land-surface models. For this study, the 'National Centers for Environmental Prediction/Oregon State University/Air Force/Hydrologic Research Lab Model' (NOAH, Ek *et al.* 2003) was used. GLDAS was forced by precipitation data from NRL (U.S. Naval Research Laboratory) as well as a number of atmospheric conditions from different sources, such as ECMWF and Global Data Assimilation System (GDAS, Rodell *et al.* 2004). Model tuning was realised by assimilation of skin temperature observations from the Television Infrared Observation Satellite (TIROS, Rodell *et al.* 2004). GLDAS-NOAH represented simulations for snow, canopy and soil water storages covering the period from 03/2000 until 04/2008 on a 0.25°-grid between latitude 60°S and 90°N. The LaD model was developed as a land-surface model by Milly & Shmakin (2002a) to simulate global water and energy balances with International Satellite Land Surface Climatology Project (ISLSCP) data for radiation, precipitation, surface pressure, temperature, humidity and wind speed. The variability of soil, groundwater and snow storages was modelled over all continents, excluding Antarctica and Greenland, with a spatial resolution of 1°. The model was tuned by an adjustment of seven parameters of land properties, for example, surface albedo, thermal conductivity or surface roughness length (Milly & Shmakin 2002b). Validation of the model output was undertaken by observation-based discharge measurements for large river basins (Milly & Shmakin 2002a). For this study, the LaD model version, LadWorld–Gascoyne, was available from 01/1980 to 07/2007.

Model strategies, tuning concepts and input data vary widely between the three models used here. GLDAS and LaD were developed as land-surface models with physically based model equations that describe both water and energy fluxes. The sub-grid variability of hydrological processes within these models is either ignored (LaD) or captured by additional parameters or functions (GLDAS). In contrast, WGHM is a water-balance model with conceptual equations that are a simplified representation of water transport processes on large scales. In a station-based calibration, WGHM parameters that are not directly observable are varied until a sufficient agreement of modelled and observed river discharge is achieved. Similarly, LaD is calibrated by river discharge applying spatially distributed parameters. In contrast, data assimilation in GLDAS denotes the direct integration of spatially distributed satellite measurements as parameter or system states into the model by Kalman-filtering.

In addition, it has to be noted that the three models represent different water storage compartments on the continents. While soil water and snow storage changes are simulated by all models, only WGHM simulates the water transport and storage in surface water bodies and only LaD includes an ice component. Moreover, GLDAS–NOAH does not include groundwater in its model structure. Due to the small variability of canopy interception water, its absence in LaD can be neglected.

Errors in input data, model structure and parameters propagate to errors in the model output. Due to the different concepts and data used by GLDAS, LaD and WGHM, their errors are expected to be of different spatial and temporal characteristics, which are analysed by differences in TWS in Section 3.1.

To reduce uncertainties caused by specific errors of individual models, multimodel ensembles and, in particular, the ensemble mean, are often used in hydrology, oceanography and atmospheric sciences as a more robust estimate of the system state or of forecast fields (e.g. Hagedorn *et al.* 2005; Tebaldi & Knutti 2007; Regonda *et al.* 2006). In this study, comparisons of GRACE with simulated hydrological data was undertaken with a multimodel mean of WGHM, GLDAS and LaD, hereafter named as Average of Three Global Hydrological Models (A3HM). In order to compute the A3HM, global fields of total continental water storage were calculated for each hydrological model by adding up all simulated storage compartments. The TWS data of each model were averaged to monthly means and re-gridded to a common 0.5° resolution. Then, for each month, the mean of the three model data sets was calculated to give the A3HM monthly time series. Antarctica and Greenland were excluded from the analysis. See Table 2 for global signal intensity of A3HM compared to the other models. A3HM data were transformed into time series of spherical harmonic coefficient sets, up to degree and order 150. To ensure consistency of A3HM and GRACE data, monthly basin-average TWS variations around the mean were computed with the same filter methods as the GRACE data. The common period of analysis in this study was 02/2003–07/2007.

For the regional analysis, the 22 biggest river basins worldwide, with catchment areas greater than 730 000 km^2,were selected (Fig. 1). As example basins of different climate zones, the Amazon, the Indus, the Nile and the Ob river basins were analysed in more detail.

Table 2. Global weighted-RMS of TWS variations (in millimeters of a water mass equivalent column) for the global hydrological models WGHM, GLDAS, LaD and the multimodel mean A3HM (col. 2–5) are derived from unfiltered data sets after application of a Gaussian filter with 500 km half-length. Columns 6–8 show relative differences of wRMS values between the hydrological models.

	wRMS [mm]				Relative difference		
	WGHM	GLDAS	LaD	A3HM	(GLDAS-WGHM)/GLDAS	(GLDAS-LaD)/GLDAS	(LaD-WGHM)/LaD
Unfiltered	60.8	97.8	62.4	64.4	0.38	0.36	0.03
Gaussian (500 km)	15.5	20.7	16.0	16.3	0.25	0.23	0.03

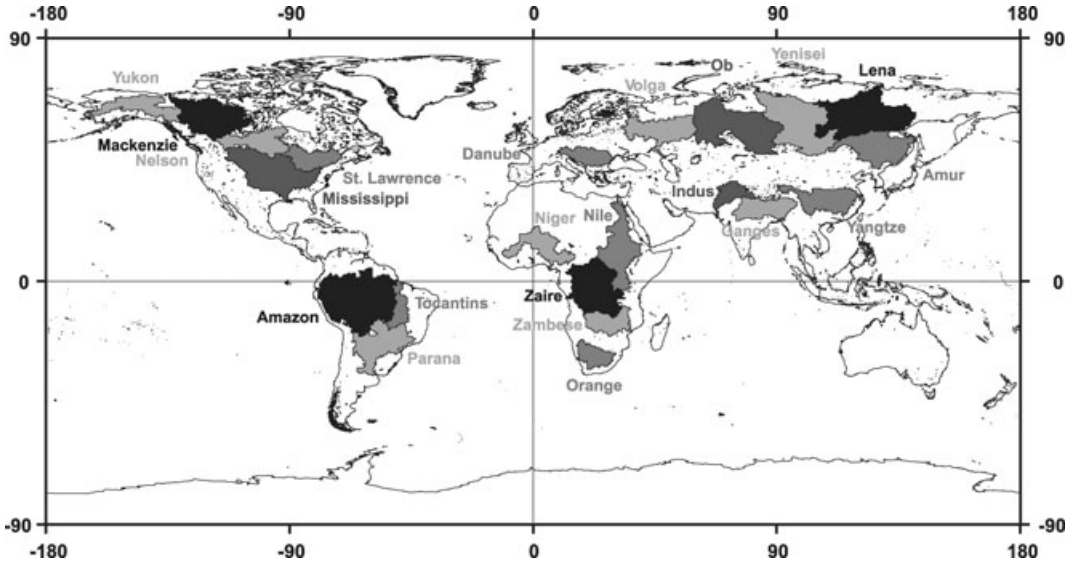

Figure 1. The 22 largest river basins worldwide (with an area greater than 730 000 km^2).

2.4 Correspondence criteria

In hydrology, the Nash–Sutcliffe coefficient (NSC, Nash & Sutcliffe 1970) is a widely used parameter to measure the performance of simulated time series against observations. The coefficient is defined by the sum of squared differences between predicted (P) and observed (O) values, normalised by the sum of squared deviations of the observations to their mean, during the period of interest with n time steps:

$$NSC = 1 - \sum_{i=0}^{n}(O_i - P_i)^2 \cdot \left[\sum_{i=0}^{n}(O_i - \bar{O})^2\right]^{-1}, \qquad (1)$$

where $\bar{O}$ is the mean of the observations over the examined period. NSC ranges from 1 (indicating perfect fit) down to $-\infty$. A value lower than zero denotes that the model is worse than if $\bar{O}$ was used as a predictor. Therefore, results with values <0 were discarded in this study. NSC not only evaluates consistency in phase, like the correlation coefficient (CC), but also in amplitude and absolute level of simulated versus observed time series. This is demonstrated in Fig. 2 by comparing two sine waves that only differ either in phase (x-axis in Fig. 2a) or in amplitude (x-axis in Fig. 2b). In this study, NSC was used as a correspondence criterion to evaluate several filter techniques by comparing measured (filtered GRACE data) and simulated (filtered modelled data) time series.

For stronger smoothing, the amplitudes of seasonal TWS variations usually are reduced more strongly, due to an increasing leakage effect. At the same time, the satellite error would decrease in GRACE time series, while it is zero for any filter parameter in the hydrological time series. If the modelled hydrological data comprehend no simulation error, and if they represent the only remaining seasonal signal in GRACE, the leakage error would be the same in both time series, and they would become more similar to each other for stronger smoothing. This may misleadingly cause higher NSC values for higher filter parameter values. Therefore, a measure of the leakage effect was introduced by weighting NSC with an attenuation factor w, which accounts for strong signal attenuation due to filtering. Hence, w was computed from the summed squared difference ($\epsilon^* = \sum_{i=0}^{n}(P_i^* - P_i)^2$) between the filtered ($P$) and unfiltered ($P^*$) time series of simulated hydrological data normalised

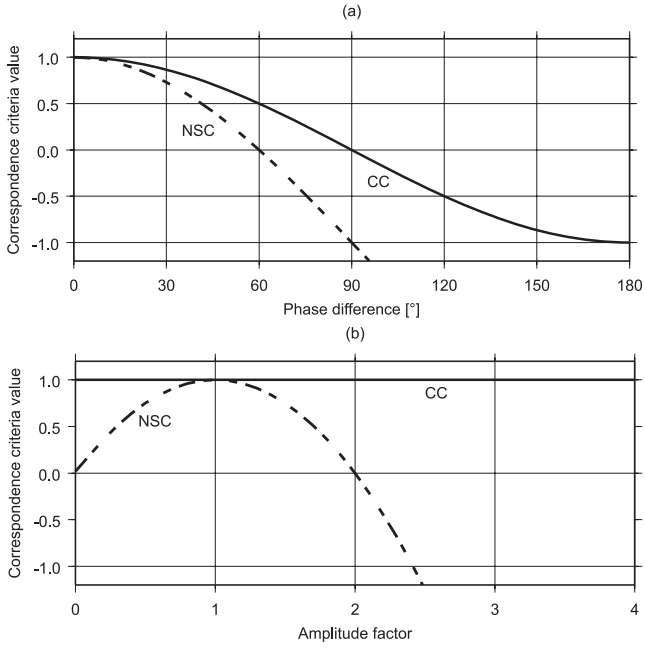

Figure 2. Nash–Sutcliffe coefficient (NSC) versus correlation coefficient (CC) for two sine waves that differ in (a) phase or (b) amplitude.

by the squared sum of the unfiltered time series ($\sigma^* = \sum_{i=0}^{n} P_i^{*2}$). Finally, we get

$$wNSC = \left(1 - \frac{\epsilon^*}{\sigma^*}\right) * NSC. \qquad (2)$$

To evaluate the reliability of the results obtained with *wNSC*, an alternative correspondence measure, the *index of agreement* (see Willmot 1984) was used. This measure also evaluates phase and amplitude differences between modelled and observed time series.

3 RESULTS AND DISCUSSION

3.1 Uncertainties of the hydrological model data

Since GRACE provides the only large-scale observation data of continental water storage change, global hydrological models provide the only means of evaluating GRACE methods for the estimation of TWS variability. In this respect, errors and differences in hydrological models need to be carefully considered. In Section 2.3, it was shown that model structure, forcing data and strategies for parameter tuning, differ considerably between WGHM, GLDAS and LaD. Maps of TWS variability for the hydrological models (Fig. 3) expose the consequences of these different concepts.

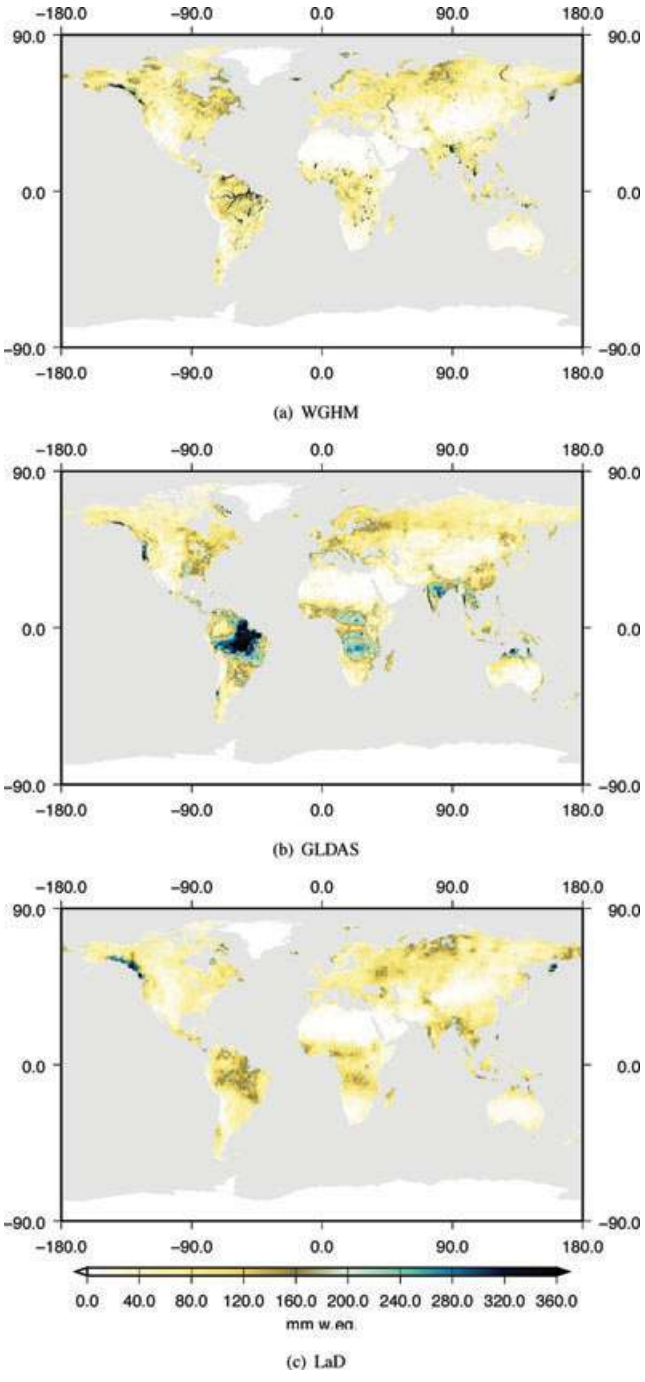

Figure 3. RMS of monthly variability of TWS from (a) WGHM, (b) GLDAS and (c) LaD during 2003–2006 (unfiltered).

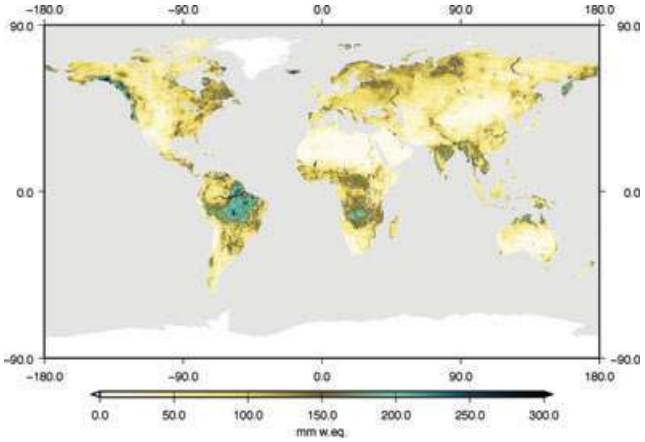

Figure 4. RMS of monthly differences of TWS variability between WGHM, GLDAS and LaD during 2003–2006 (unfiltered).

The spatial distribution of the TWS variability in WGHM tends to exhibit linear patterns, reflecting the presence of the surface water storage compartment in the model, including rivers and their inundation areas (see Fig. 3a). In contrast, TWS variability from GLDAS and LaD is more gradually distributed in space (Fig. 3b and c) in line with larger correlation lengths of soil–water storage (and groundwater for LaD), which dominates TWS in these models. Furthermore, GLDAS amplitudes of TWS variations are larger than those of the other two models. Thus, for the 0.5°-cell-wise RMS-differences between the models (Fig. 4), the largest differences occur for GLDAS versus WGHM or LaD, whereas WGHM and LaD are more similar to each other. The differences in the simulated TWS variability between the models may amount to 300 mm w.eq., which is close to the signal magnitude itself. In the difference maps of Fig. 4, the linear patterns caused by surface water storage in WGHM not present in the other models, are obvious again. The largest differences occur within the river basins of the Amazon, Congo, Ganges, Mekong, Yukon, St. Lawrence and Ob rivers. Thus, a main difference in TWS variability between the models can be attributed to the fact that different storage compartments with different spatial characteristics are represented in the models.

However, the cell comparisons between models, as shown in Fig. 4, may be misleading if basin-average water storage variations and water balances are of interest. This is the case when considering the lower resolution GRACE data. Relative differences between models decrease on the river basin or global scale, for example, after computing basin averages or reducing the resolution of TWS data by applying a GRACE-filter method. Relative model differences of global (latitude) weighted RMS of TWS reduce after a Gaussian filtering of 500 km, when compared to unfiltered data (Table 2). Nevertheless, much smaller differences in signal magnitudes between WGHM and LaD, than of both models relative to GLDAS, remain even after global averaging (Table 2).

In contrast, temporal correlations of TWS time series are very high between the hydrological models (Fig. 5). WGHM and LaD are nearly perfectly correlated on all land areas (Fig. 5b), except for a small region in the Himalayas and some linear river courses (e.g. Lena river). The first deviation may be due to differences in the snow algorithms and the latter due to the absence of surface water routing in LaD. This process causes longer residence times and, thus, delayed storage depletion within river basins for WGHM.

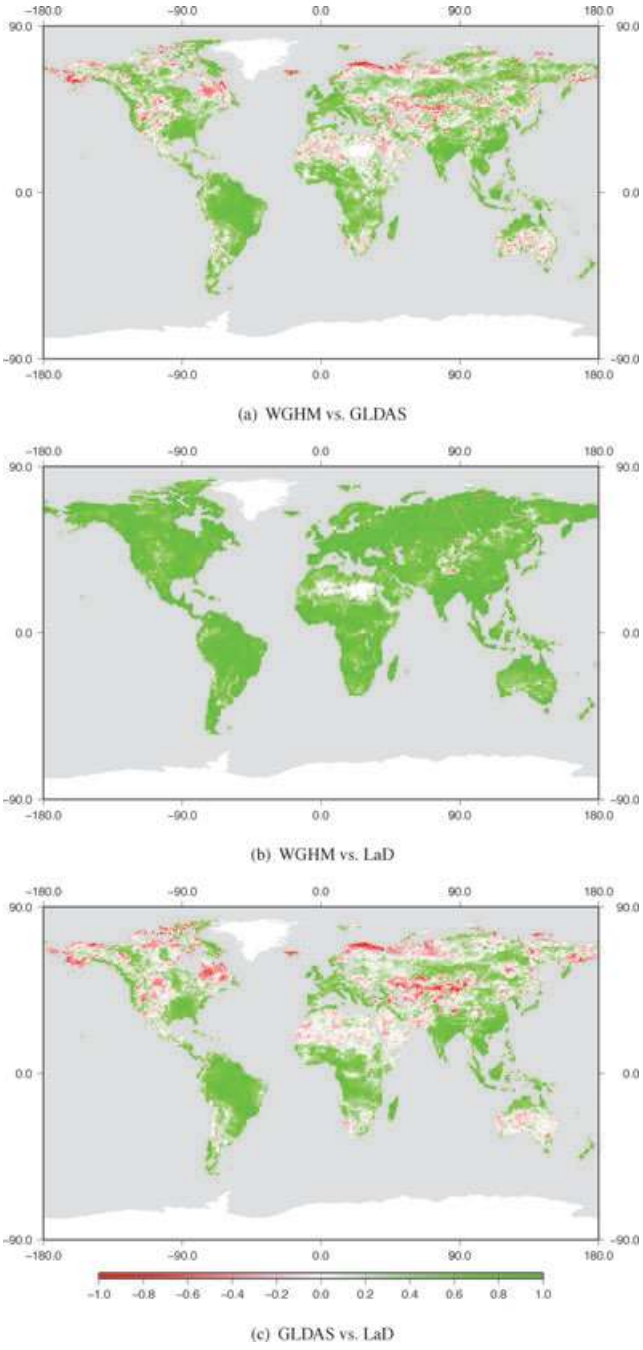

Figure 5. Correlation of monthly TWS between WGHM, GLDAS and LaD during 2003–2006.

But correlation maps for WGHM versus GLDAS (Fig. 5a) and WGHM versus LaD (Fig. 5c) indicate good temporal correlations for the major river basins (e.g. Amazon, Zaire, Ganges, Mississippi, or large parts of Ob, Yenisei and Lena) despite large differences in amplitudes of TWS variations as shown in Fig. 4. Dry areas, such as North Africa, central North America, central Australia and central Asia, are not well correlated in time between the models, but the TWS change signals are very small (compare to Fig. 3) or negligible in these areas. Low correlations for regions with large TWS variability only appear in small areas of Scandinavia, East-Siberia and the northeast of North America.

To conclude, differences of TWS variations between the three global hydrological models are quite large when evaluated at the grid scale. These differences are mainly due to different model structures in terms of water storage components represented in each of the models. In previous studies (e.g. compare Güntner 2009; Werth & Güntner 2008), no model was shown to be most consistent relative to GRACE-derived TWS variations. Nevertheless, global hydrological models represent the most comprehensive and state-of-the-art data on continental water-cycle processes on large scales. Therefore, they are the only data source to evaluate GRACE-derived estimations of TWS variability. Relative differences between the models reduce on the scale of river basins and are relevant for comparisons to GRACE data. Furthermore, the models show a good temporal agreement, especially within regions of large TWS variations. For this study, the model mean A3HM provided a compromise between the three independent model realisations of different concepts. A3HM averaged out particular model errors due to individual model structures and input data sets. Only a few systematic errors that may prevail in all input data sets, such as those due to the generally small number of precipitation stations in specific regions (e.g. parts of Africa, South America or Central Asia), cannot be reduced in this way. In evaluating GRACE filter methods, A3HM currently provides the most adequate estimation of water storage variations on the continents.

3.2 Filter evaluation

3.2.1 Filter effects on seasonal amplitude and phase

Different filter methods cause different GRACE error reduction and leakage effects when applied to different river basins. To understand reasons for such differences, filtered time series with non-decorrelating filter methods and different filter parameters are shown in Figs 6–8 (a–d) for GRACE (top) and A3HM (bottom) derived TWS variations. Examples are given for three river basins (Amazon, Indus, Ob) to illustrate the effects of different climate zones with diverse hydrology and different regimes of TWS variations. The Amazon exhibits a strong signal that dominates northern South America (Fig. 6, dotted time series). The signal of a surrounding area of the Amazon basin (defined by a latitudinal and longitudinal buffer of 8° around the catchment boundaries) exhibits a much smaller signal with a slight phase shift (triangles). The application of filters with a weaker smoothing strength (blue-coloured time series) generates erroneous time series in terms of GRACE and nearly undamped time series in terms of A3HM. Stronger smoothing (pink-coloured time series) leads to higher TWS amplitude attenuation due to the small signal in the surrounding areas. The amplitude damping is stronger for filter I compared to filter IV and most prominent for filter II with small parameter values, as well as for filter III, with very small signal variance parameter values. Phase shifts of the surrounding areas are too small to have a noticeable influence on the seasonal phase of the filtered Amazon time series.

In contrast to the Amazon basin, higher differences between the model-based and GRACE-derived TWS data occur for the Indus river basin. Also, the smaller size of this basin leads to more erroneous GRACE time series for weaker smoothing, and differences between the filter methods become more evident (Fig. 7). The Indus basin is influenced by a strong signal in surrounding regions with opposite seasonal phase. For example, the closely located Ganges River has a strong signal caused by the Indian summer monsoon, whereas Indus water storage variations are more influenced by snow accumulation and melt. Furthermore, the eastern desert in the Indus basin exhibits low TWS variability. Thus, a strong leakage effect

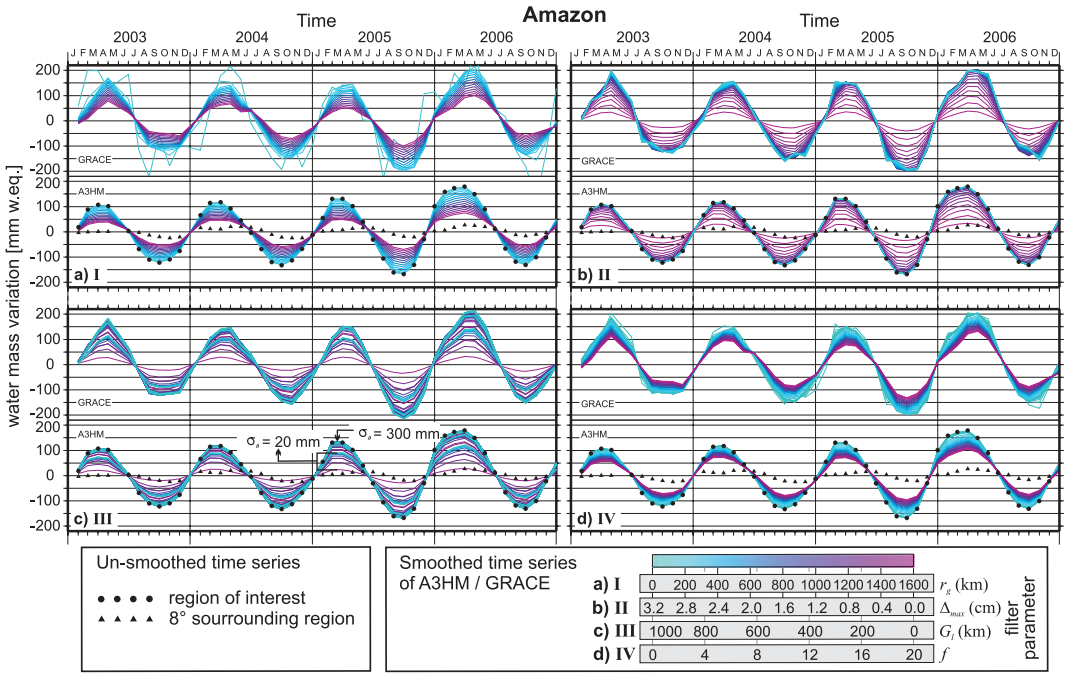

Figure 6. Time series of TWS variations for the Amazon river basin after applying the non-decorrelating filter methods: (a) I, (b) II, (c) III and (d) IV. For different values of the filter parameter, the graphs colour gradually changes from blue (weak smoothing) to pink (strong smoothing). From the two parameters of filter III, G_l is colour-coded and σ_0 graphs, with maximal and minimal values, are exemplarily indicated in subfigure (c). See further explanations in the main text.

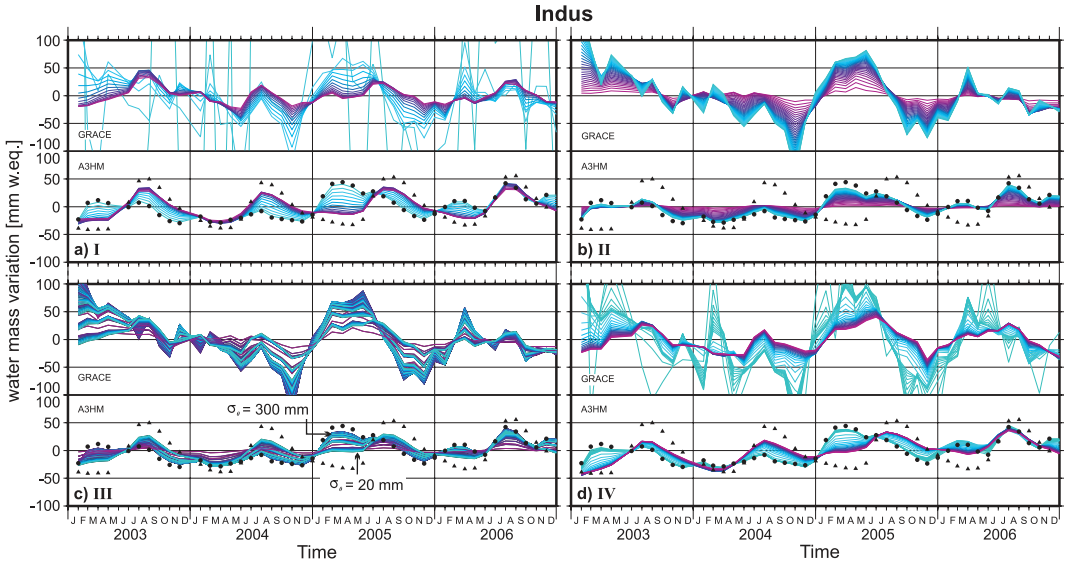

Figure 7. Same as Fig. 6 but for the Indus basin.

of the surrounding areas causes strong phase shifts towards the surrounding signal when using filter IV and I, with strong smoothing. This occurs with filter III, as well, with signal variance parameters (50 or 20 mm w.eq.) that are too small. In contrast, for filters II and III (the latter with signal variance parameters greater than 100), only amplitude damping can be observed.

The Ob basin (Fig. 8) is surrounded by regions with equal phase and similar amplitude (see Fig. 3). Hence, signal leakage is less dominant. The similar hydrological signal characteristics of surrounding river basins (e.g. Volga, Yenisei) balance signal truncation inside the Ob basin. Therefore, filters I and IV cause, overall, very little amplitude damping. Filter IV even exhibits slightly increased amplitudes for some parameters compared to the unfiltered A3HM signal. On the other hand, amplitude damping of filters II and III becomes strong for very small parameters of II and small signal variance parameters of III. For the Ob basin, phase shifts are a negligible filter effect. Compared to the Amazon, the more erroneous GRACE time series for lower smoothing are explainable by other factors, such as the more complex shape, or smaller size, of the Ob river basin.

These three examples show that phase shifts and amplitude attenuation of TWS time series differ between the river basins depending

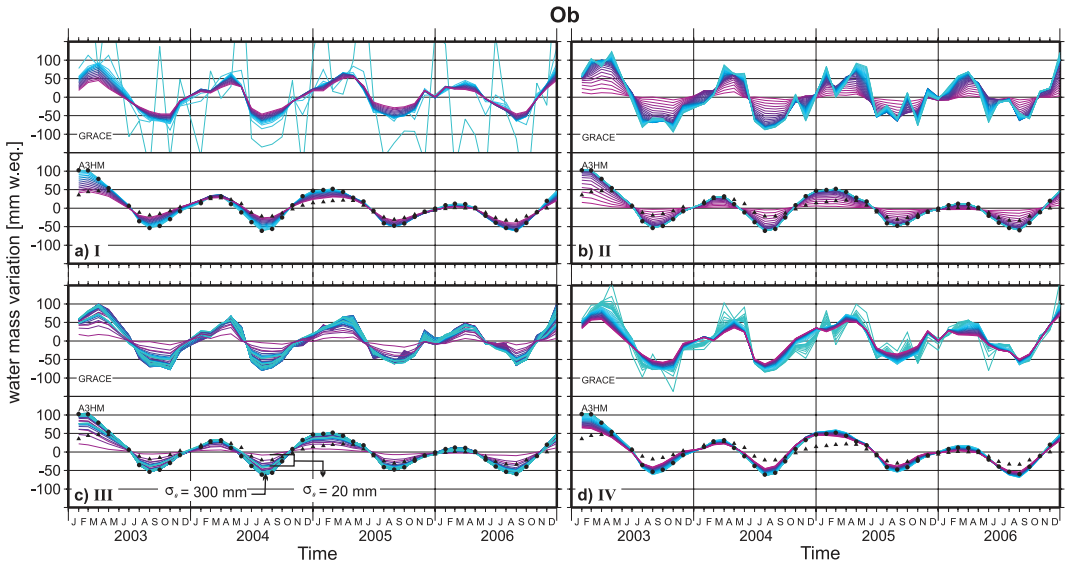

Figure 8. Same as Fig. 6 but for the Ob basin.

on the applied filter methods, the signal properties inside and outside the basin, and the basin size or shape. If phase shifts between the time series outside and inside a basin are of negligible size, the leakage scenarios established by Klees *et al.* (2007) are comprehensible. For example, the small signal of the Amazon's surrounding region biases the TWS time series of the Amazon basin more strongly than does the time series of the Ob basin when influenced by a signal of similar size outside the Ob basin. But when a marked phase shift between the signal outside and inside the basin is present, the three leakage scenarios vary between months because the ratio of the signal outside the basin to the signal inside the basin varies. Therefore, phase shifts occur in the filtered time series of such regions. This applies especially to high (strong) parameters of filters I and IV (smoothing), as shown for the Indus basin. In such cases, the application of amplitude scaling or bias correction factors as proposed by Velicogna & Wahr (2006) and Klees *et al.* (2007) will not recover the hydrological signal after filtering.

A summary of seasonal phases and amplitudes for the 22 river basins and their surrounding areas is given in Table 3. Phase shifts and amplitude differences as an effect of filtering are shown in Table 4. To make them comparable, the different filter methods were parameterised in such a way that they generate the same RMS of monthly satellite error (propagated from the coefficient errors) as a Gaussian smoother of 500 km radius. If phase shifts of the surrounding region are large, and if amplitudes are of similar magnitude as the signal inside the basin, an impact of the phase of the surrounding signal is clearly visible in the filtered time series (e.g. Amur, Indus). A much smaller signal amplitude in the surrounding region compared to the basin itself results in strong amplitude damping (e.g. Ganges, Tocantins, Zambezi). Both effects are simultaneously visible for a few basins (e.g. Amur, Indus, Parana). Also, the size and sign of both effects vary between the filter methods. Some basins (e.g. Amur, Lena, Nelson, Nile) exhibit phase shifts of different signs. For other basins, the size of amplitude damping differs largely between filter methods of equal satellite error reduction (e.g. Danube, Ganges, St. Lawrence, Tocantins). Due to a weak annual signal of the Orange basin (see seasonal amplitude of Orange in Table 3), leakage tends to increase the annual amplitude for this basin. Probable reasons for the different filter effects will be given in the next section.

Table 3. Seasonal amplitude (A, col. 2) and phase (Φ, col. 3) of TWS variations for the 22 river basins, derived from the ensemble model mean A3HM. Respectively, seasonal amplitude difference (ΔA, col. 4) and phase shift (ΔΦ, col. 5) are computed for an 8° surrounding region.

Basin	Basin		Surr. region	
	A [mm]	Φ [day]	ΔA [mm]	ΔΦ [day]
Amazon	136	−23	−119	−32
Amur	10	154	−4	−54
Danube	47	−1	−15	+13
Ganges	125	−182	−75	−3
Indus	14	−65	+27	−126
Lena	26	36	−17	−3
Mackenzie	34	9	−8	−2
Mississippi	26	5	−16	+43
Nelson	23	26	+8	−16
Niger	77	165	−59	+19
Nile	37	−182	−34	−32
Ob	50	−2	−26	−6
Orange	1	−40	+18	+21
Parana	76	−5	−20	−7
St. Lawrence	85	3	−49	+12
Tocantins	230	−15	−129	−17
Volga	73	−3	−41	+1
Yangtse	39	−146	+9	−31
Yenisei	32	9	−16	−9
Yukon	43	9	−17	+1
Zaire	23	8	−19	+180
Zambezi	103	−19	−60	+7

3.2.2 *Correspondence of GRACE to hydrology data*

The different filter methods and smoothing rates were evaluated with the *wNSC* correspondence criteria (explained in Section 2.4) against A3HM data (Table 5). Example results for the Amazon and Ob basins are shown in detail in Fig. 9. Applying the Gaussian filter for the Amazon basin, averaged time series are very sensitive to damping when evaluated by *wNSC*. A radius of 300 km results in the highest wNSC value, that is, the best correspondence of GRACE and A3HM data. This radius is similar to the results of Schrama

Table 4. Filter-induced bias of the seasonal amplitude (ΔA, col. 3–8) and phase (ΔΦ, col. 9–14) for 22 river basins and six filter methods (I-VI). Parameters for each filter were set to give the same propagated satellite error (ϵ_{sat} in col. 2) as a Gaussian smoother of 500 km radius. Results were computed from A3HM data. See Table 3 for seasonal amplitude and phases of the unsmoothed signal.

		ΔA [mm]						ΔΦ [day]					
Basin	ϵ_{sat} [mm]	I	II	III	IV	V	VI	I	II	III	IV	V	VI
Amazon	10	−18	−2	−6	−18	−47	+1	−2	+0	+0	−1	−2	+0
Amur	11	−3	+0	−4	−3	−9	−1	−29	+7	+8	−40	−64	−7
Danube	15	−3	−15	−1	−1	−17	−2	+5	+1	+4	+2	+9	+3
Ganges	14	−27	−19	−22	−43	−49	−6	−3	−1	−2	+1	−3	−2
Indus	18	−2	−7	−7	+2	+7	−5	−94	−41	−67	−91	−117	−29
Lena	10	−3	−1	−7	+6	−8	+0	+1	+1	+6	−9	−5	+3
Mackenzie	12	+5	−1	−3	−9	−1	+2	+0	−1	−1	+4	+1	+1
Mississippi	10	−1	+0	−1	−7	−7	+0	+2	−1	+1	+17	+1	−1
Nelson	12	+1	−4	+1	−5	−4	+1	−3	+2	+1	+6	−14	+4
Niger	14	−8	−9	−15	+2	−23	+2	+1	−1	−2	−1	+2	+0
Nile	15	−4	−4	−1	−17	−9	+0	−4	−7	−4	+8	−5	−5
Ob	10	−3	−1	−7	−2	−24	+1	+0	+0	−1	+2	+5	+0
Orange	19	+3	+7	+5	+4	+5	−1	+21	−167	−171	+11	+20	−119
Parana	14	−10	−31	−29	−16	−20	−11	+2	+9	+8	+7	+3	+4
St. Lawrence	14	−17	−16	−16	−47	−42	−2	+6	−1	+4	+4	+6	+1
Tocantins	22	−42	−145	−27	−45	−89	−22	−4	+2	−3	−4	−6	−4
Volga	12	−12	−7	−9	−10	−33	−2	+1	+0	+2	+4	+4	+1
Yangtse	13	+2	−4	−1	−11	−5	+0	−8	+2	−3	+1	−9	+0
Yenisei	10	+0	+0	+1	+7	−15	+2	−1	+0	−2	−2	−2	−1
Yukon	13	+4	+3	−7	+6	−8	+17	−1	−3	−3	−3	−1	−2
Zaire	13	−1	−1	+0	+4	−8	+1	−5	−2	−3	−19	−5	−5
Zambezi	16	−17	−22	−25	−21	−38	−2	+2	+2	+2	+1	+2	+1

Table 5. Weighted NSC (*wNSC*) evaluation of GRACE filter types with A3HM data: highest *wNSC* for each filter type and corresponding filter parameter values in brackets. Bold *wNSC* values indicate the overall optimal filter method for each basin.

	Weighted Nash–Sutcliffe-Coefficient (wNSC)					
Basin	I (r_g[km])	II (Δ_{max}[mm])	III (σ_0[mm], G_l[km])	IV (f)	V (w_a)	VI (x)
Amazon	0.82 (300)	0.84 (11)	**0.87** (250,300)	0.83 (1)	0.70 (30),II	**0.86** (13)
Amur	0.27 (300)	**0.35** (25)	0.31 (300,100)	0.26 (2)	0.15 (30),I	0.21 (13)
Danube	0.63 (300)	0.69 (27)	0.70 (250,1000)	0.66 (0.6)	0.46 (30),II	**0.75** (12)
Ganges	0.81 (300)	0.81 (17)	0.88 (300,500)	0.77 (1)	0.76 (30),II	**0.91** (12)
Indus	0.15 (400)	0.29 (21)	**0.33** (200,1000)	0.25 (2)	0.11 (30),III	**0.32** (13)
Lena	**0.49** (300)	0.42 (13)	**0.49** (300,1000)	**0.49** (2)	**0.50** (20),IV	**0.49** (12)
Mackenzie	0.60 (400)	0.59 (13)	**0.65** (150,200)	0.60 (1)	0.42 (30),II	0.60 (12)
Mississippi	0.61 (400)	0.59 (13)	0.64 (150,1000)	0.54 (1)	0.60 (30),I	**0.66** (12)
Nelson	0.31 (500)	0.29 (30)	0.33 (200,1000)	0.22 (0.4)	0.30 (30),II	**0.51** (12)
Niger	0.85 (300)	**0.88** (23)	**0.89** (200,200)	0.86 (0.7)	0.78 (30),IV	**0.89** (12)
Nile	0.56 (400)	0.58 (14)	**0.61** (150,900)	0.43 (0.5)	0.57 (30),II	0.59 (13)
Ob	0.76 (300)	0.73 (13)	**0.80** (100,900)	0.74 (4)	0.43 (30),IV	**0.81** (13)
Orange	0.29 (600)	0.09 (41)	0.32 (20,1000)	0.17 (7)	**0.38** (20),I	0.28 (14)
Parana	0.67 (500)	0.48 (16)	0.69 (200,1000)	0.63 (2)	0.58 (30),II	**0.73** (12)
St. Lawrence	**0.37** (200)	0.15 (20)	0.24 (20,1000)	0.24 (0.1)	0.22 (30),I	0.21 (14)
Tocantins	0.78 (400)	0.78 (34)	**0.85** (300,900)	0.80 (0.7)	0.69 (30),II	**0.85** (12)
Volga	0.70 (300)	0.66 (15)	0.75 (100,900)	0.70 (1)	0.50 (30),II	**0.78** (13)
Yangtse	0.74 (400)	0.71 (17)	0.79 (300,700)	0.69 (2)	0.62 (30),III	**0.82** (12)
Yenisei	0.60 (400)	0.57 (12)	0.63 (50,500)	0.63 (16)	0.42 (30),IV	**0.66** (14)
Yukon	0.50 (300)	**0.59** (16)	**0.59** (150,100)	0.52 (1)	0.24 (30),IV	0.57 (12)
Zaire	0.41 (400)	0.47 (12)	0.49 (100,300)	0.47 (2)	0.41 (30),III	**0.51** (13)
Zambezi	0.75 (300)	**0.81** (27)	**0.81** (300,200)	**0.82** (0.7)	0.64 (30),III	**0.82** (12)

et al. (2007), who selected a globally optimal Gaussian filter radius of 275 km by comparison with GPS load measurements. The *wNSC* correspondence rapidly decreases for smaller and higher radii than 300 km. The degree-only dependency of filter method I does not take into account differences in accuracy for coefficients of equal degree but different order. Therefore, method I may either filter coefficients with an acceptable signal-to-noise ratio too strongly or may not sufficiently filter coefficients containing large errors. Thus, basin average values are either affected by signal leakage from surrounding areas or by large errors.The Amazon basin is located close to the ocean (which inheres a signal close to zero) both to its east and to its west. To its north, it borders on the equator (where a

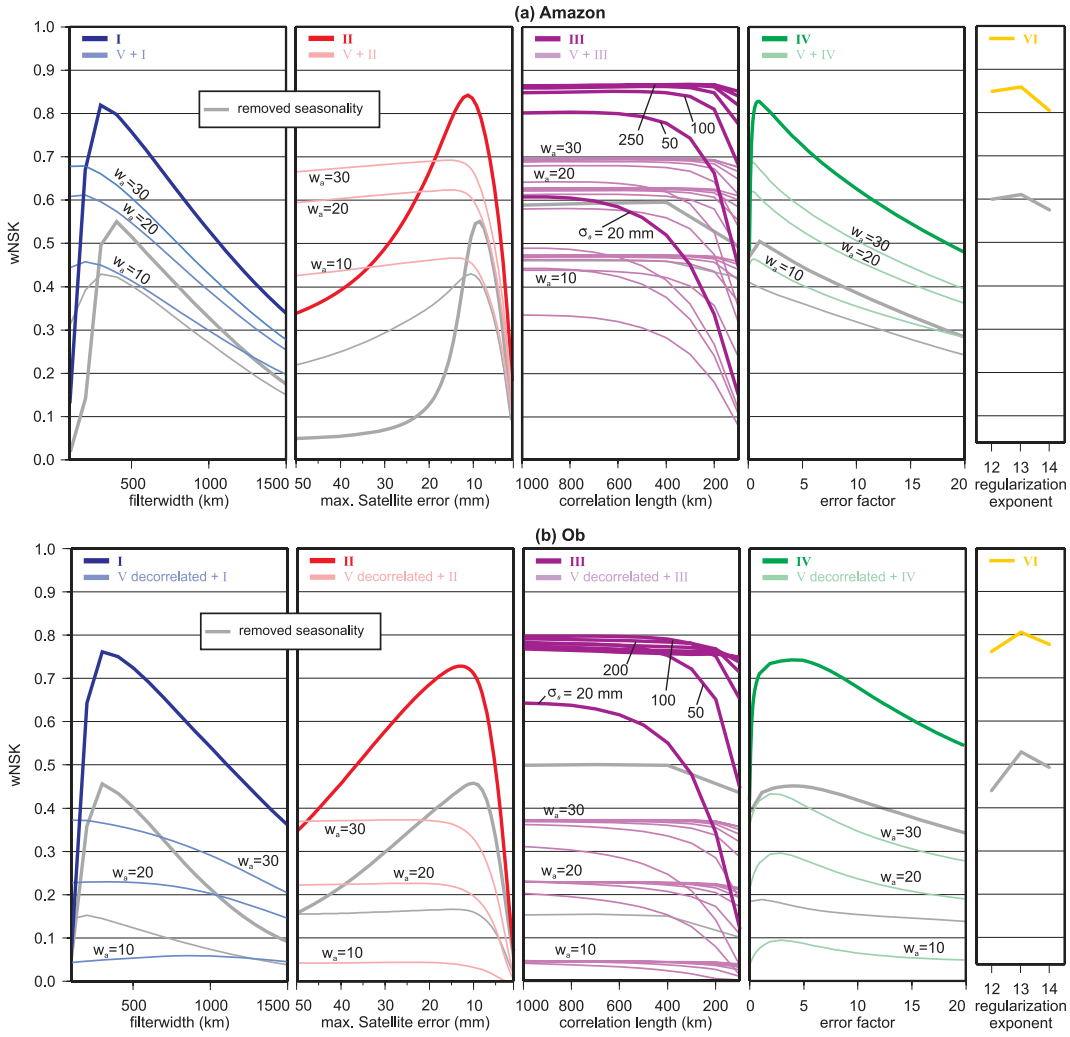

Figure 9. Weighted NSC (*wNSC*) performance of different filter types (I–VI) and grades of smoothing strengths for: (a) the Amazon and (b) the Ob river basins. Bold lines in blue: Gaussian (I), red: basin optimised (II), purple: signal model optimised (III), green: SNR optimised (IV), yellow: decorrelation VI. Light coloured lines: decorrelation V (additionally applied to I-IV) with $n_e = 30$ for all displayed graphs. Grey lines: seasonality removed before computation of *wNSC* for all filters. See a filter description in Section 2.1.

shifted seasonality of water storage occurs further north). Therefore, basin averages for the Amazon are quite sensitive to leaking signals or amplitude damping (as shown previously) for large filter radii. This causes high parameter sensitivity of filter I.

Filter II and IV also exhibit a distinct sensitivity to filter parameters. In the case of filter II, the correspondence to hydrological model data is poor for maximum admitted satellite errors smaller than about 5 mm in both river basins. In this case, parameter values that are too low (i.e. strong smoothing) increase signal leakage (i.e. amplitude damping). Smoothing with filter II, for admitted maximum satellite errors in the range of 8–15 mm, performs well for both basins. Maximum *wNSC* values of II are slightly higher than optimum results of I for the Amazon, and are somewhat lower than optimum results of I for the Ob basin. These results follow from the design of filter II, which preferably preserves coefficients that contain important signals of the examined basin and, thus, reduces signal leakage. Furthermore, the anisotropic design of II (compared to the isotropic Gaussian) distinguishes between the orders of coefficients with the same degree, which enables a finer adjustment of the filter weights. The filter II design is of particular benefit for the Amazon basin, where signal separation from the surrounding areas is important for preventing amplitude damping, as described above. This strategy of filter II is of less advantage for the Ob basin, where leakage of I is compensated by similar signals in surrounding areas (as described previously). This also explains the lower filter parameter sensitivity of II for the Ob basin.

The optimal error factor of the anisotropic filter IV is $f = 1$ for Amazon and $f = 4$ for Ob. This implies that the correlated GRACE coefficient errors are properly estimated within the Amazon basin, whereas they are regionally underestimated within the Ob region. Optimal error factors for IV differ between the river basins because the quality of GRACE coefficient error assessment varies regionally (see also Horwath & Dietrich 2006). The optimised filter IV is nearly as good as each of the other optimised filter types for the Amazon. Again, due to similarly large signal characteristics around the Ob basin, signal damping by leakage is small when using filter IV. The positive leakage for some smoothing rates of that filter increases TWS amplitudes, as shown in the previous section. Therefore, larger parameter values for IV hardly damp regionally averaged time series and only reduces errors. This also leads to a smaller

parameter sensitivity of that filter for the Ob basin than for the Amazon.

By contrast, the anisotropic filter type III is comparatively insensitive in terms of *wNSC* values to parameter variations. Correlation lengths greater than 300 km, and standard deviations greater than 100 mm, provide *wNSC* values that differ less than 0.04 in both river basins. Sensitivity of the standard deviation parameter is higher than that of the correlation length parameter. This confirms Swenson & Wahr (2002), that an exponential signal model is a good approximation for estimating the leakage error, and that it does not strongly depend on the exact estimation of its parameter values, as long as σ_0 and G_l are not too small. Compared to the other filters, III provides the highest *wNSC* results for the Amazon and the second highest for the Ob basin.

For the Amazon basin, decorrelation by V (thin coloured lines in Fig. 9) does not improve the correspondence between filtered GRACE and hydrological model time series for any of the four filter types discussed above. This follows from the low efficiency of V in equatorial regions. Signals of these regions are dominant in near-sectorial coefficients (with similar degree and order), which are corrected incompletely by that method (Swenson & Wahr 2006). Decorrelation with filter V does not give better results than the four nondecorrelating methods for the Ob basin. Outside the equatorial region, improvement by decorrelation filter V only occurs for the Lena, Orange, Mississippi, Parana and St. Lawrence basins.

The alternative decorrelation method VI for $a = 10^{12}$ gives a *wNSC* value close to the filter III optima for the Amazon (black dashed line in Fig. 9). Results of VI for $a = 10^{13}$ are superior to all filter methods for the Ob basin. A low parameter sensitivity of VI is visible for both river basins.

Schaefli & Gupta (2007) showed that the NSC is very sensitive to seasonality. Since seasonality is the most dominant signal in most river basins, *wNSC* was re-computed after removing the seasonal signal from the time series. These results are shown by the grey graphs in Fig. 9 for all filter types, respectively (including the optimal filter V with $w_a = 30$). The *wNSC* values become smaller due to relatively high errors in the small non-seasonal water storage signal of GRACE and the models. But the highest *wNSC* is likely to occur for similar filter parameter values when compared to results that include the seasonal signal. This shows that the optimum filter technique for a specific river basin is more a function of filter properties in combination with the geographical characteristics of the region of interest, than a function of the selected time period or of TWS temporal dynamics. Thus, the results obtained here can be expected to be of broad relevance for hydrological studies.

A summary of filter comparison for all 22 basins is given in Table 5, with the highest *wNSC* value and the corresponding filter parameter value for each filter type. The optimal filter type for each basin is indicated by bold numbers. Among the basins, different filter methods with different filter parameter values appear to be optimal. GRACE basin-average TWS time series with optimum filtering generally have a high correspondence to A3HM, except for Amur, Indus, Nelson, St. Lawrence and Orange. For the latter five basins, the error map (Fig. 4) shows large differences of TWS variability between the hydrological models for the St. Lawrence basin only; but temporal correlations between the models are poor for large areas inside all of the five basins (Fig. 5). Hence, large uncertainties in the hydrological model data may lead to uncertain results for the filter optimisation for these basins. For filter I, the highest *wNSC* occurs for radii from 300 to 400 km for nearly all river basins. This indicates that the spatial resolution of GRACE-derived TWS variations is mostly better than 500 km. The optimised maximum satellite error of filter II tends to be larger than the 10 mm water equivalent. This illustrates the limitations in accuracy of GRACE TWS estimates due to GRACE measurement errors. Furthermore, optimal correlation lengths of filter III vary considerably between the river basins. Lower correlation lengths may be due to the importance of surface water storage concentrated in a small spatial domain, as pointed out by Güntner *et al.* (2007). Filter V provides optimal filter results for only two river basins (Lena, Orange). Parameter optimisation of V is not straightforward, because the four filter parameters (w_a, w_e, n_a and n_e) may have to be adjusted individually for each basin in addition to the parameter of the subsequently applied filter method. Method VI provides the highest *wNSC* for 16 river basins, and its performance is also in the same range as the best alternate filter methods for the remaining basins. The anisotropic decorrelation method of VI seems to efficiently preserve the hydrological signal while reducing GRACE satellite errors. In addition, VI exhibits low filter-parameter sensitivity. This supports the method's strategy of deriving an error covariance matrix from satellite orbits in order to decorrelate the coefficients in the filter process. Finally, the computations were repeated for the Index of Agreement (Willmot 1984). This measure of correspondence between GRACE and hydrological model data generally confirms the results as provided previously (not shown).

For all filter methods, a final estimation of biases of the seasonal amplitude and phase in the TWS time series after application of the optimised filter, is provided in Table 6, based on A3HM data. For both amplitudes and phases, biases are reduced for many river basins in comparison to Table 4, where a standard Gaussian filter, or filter of equivalent smoothing strength, were applied. This indicates a successful optimisation of the filter type and parameter. Large phase shifts remain for the Indus and Orange basin only. For most of the other basins, filters III and VI had the smallest seasonal phase shifts and amplitude damping.

3.2.3 Multicriterial error analysis

For an alternative evaluation of the filter methods with hydrological data, satellite and leakage errors of the different filters were evaluated in a multicriterial way (Fig. 10). The leakage error was derived as an RMS of differences between filtered and unfiltered A3HM time series in monthly TWS variations. The satellite error was derived as an RMS from a monthly propagation of the calibrated coefficient errors into the basin averages. The total error is given by the squared sum of both error components. Hence, the point closest to the origin in Fig. 10 provides the smallest total error and indicates the optimum filter type according to the error budget.

The error budgets of the Amazon (Fig. 10a) and the Ob (Fig. 10d) basin show a well defined ranking between the filter methods. The decorrelation method VI is superior in reducing the total error. The second best error budget is provided by filter III. Furthermore, the more complex the shape (e.g. Nile) or the smaller the size of a river basin (e.g. Indus), the larger is the total error and the smoother are the error graphs in Fig. 10. For the Indus and the Nile basin, leakage for low smoothing rates amounts to several millimetres. Also, for these critical basins' characteristics, filters VI and III, respectively, provide the filter versions with the best error budgets, though versions of method II are located close to the optimum as well. The method V exhibits high leakage errors in the error plots of all basins, which explains the *wNSC* results from above. Hence,

Table 6. Filter-induced bias of the seasonal amplitude (ΔA, col. 2–7) and phase (ΔΦ, col. 8–13) for 22 river basins and the six optimised filter methods (I–VI) as listed in Table 5. Results were computed from A3HM data. See Table 3 for seasonal amplitude and phases of the unsmoothed signal.

	ΔA [mm]						ΔΦ [day]					
Basin	I	II	III	IV	V	VI	I	II	III	IV	V	VI
Amazon	−7	−4	+0	−4	−27	−1	−1	+0	−1	+0	+0	+0
Amur	−2	+0	+0	+0	−8	−1	−11	+3	+5	−12	+23	−7
Danube	−2	−5	−2	−4	−10	−3	+3	+0	+2	+2	+10	+2
Ganges	−13	−12	−5	−13	−25	−4	−1	+0	−1	+0	−3	−1
Indus	−5	−5	−4	−4	+0	−5	−74	−27	−24	−26	−105	−29
Lena	−1	−2	+0	+5	−3	+0	+2	+2	+2	−4	−11	+3
Mackenzie	+5	−2	+1	+1	+2	+2	+0	−1	+0	−1	−2	+1
Mississippi	−1	−1	+0	−1	−3	+0	+1	−1	+0	+3	−1	+0
Nelson	+1	−1	+0	−2	−9	−1	−3	+0	+2	+1	−9	+4
Niger	−2	−2	+0	+5	−9	+2	+1	+0	+0	−1	+1	+0
Nile	−3	−3	−1	−6	−5	−1	−3	−5	−3	+3	−1	−1
Ob	+0	−2	+0	+3	−18	+1	+0	+0	+0	+0	+6	+0
Orange	+4	+0	+1	+4	+4	+2	+21	+11	−181	+12	+18	−138
Parana	−10	−14	−8	−12	−20	−6	+2	+4	+3	+5	+6	+3
St. Lawrence	−4	−6	−37	+0	−42	−14	+2	+0	+4	+0	+6	+3
Tocantins	−31	−21	−10	−16	−55	−7	−3	−2	−2	−1	−4	−3
Volga	−6	−7	−3	−2	−27	−2	+1	+0	+2	+1	+5	+1
Yangtse	+1	−2	+0	−7	−6	+0	−5	+1	+0	+0	−2	+0
Yenisei	+1	−1	+0	+3	−9	+2	−1	−1	−2	+0	−2	−2
Yukon	+10	+4	+6	+8	+1	+13	−1	−3	−3	−1	−1	−2
Zaire	−1	−1	+1	+6	−4	+1	−5	−1	−3	−18	−6	−5
Zambezi	−6	−1	+0	−1	−21	+2	+1	+1	+1	+0	+1	+1

method V is not a generally efficient filter approach for deriving basin-averaged TWS variations from GRACE gravity fields.

In summary, the order of the filters in terms of their total error budget in Fig. 9 closely matches the filter type ranking by the *wNSC*-evaluation (Table 5). For most cases, the decorrelation method VI provides the best error budget. For a similar satellite error reduction in the GRACE data, the leakage error of VI is much smaller compared to the other filter methods. This explains the small seasonal amplitude damping and phase shifts for this method in many river basins (Table 4). A list of TWS satellite, leakage and total error for the 22 river basins after application of the optimised decorrelation method VI, is provided in Table 7. The comparison of these total error values with estimations of seasonal TWS amplitudes from A3HM (Table 3) indicates that the estimation of GRACE-derived seasonal water mass variations is not reliable for the Amur, Indus, Nelson and Orange basins, as the error exceeds the signal magnitudes. This coincides with the small correspondence of GRACE and A3HM-derived time series of TWS variations for these river basins in Table 5.

Besides method III in the Indus basin, the best filter methods found for each river basin in the previous section, by the *wNSC*-evaluation (black circles in Fig. 10), are very close to the minimum satellite and leakage error budget. This result confirms the broader validity of the optimum filter selection procedure.

3.2.4 Sensitivity to errors in amplitude of the hydrological data

In Section 3.1 it was shown that the differences in TWS variations between the hydrological models consist of amplitude differences rather than phase shifts. Consequently, the influence of amplitude errors in the hydrological data on the *wNSC*-evaluation of filter parameter and methods has to be estimated. Therefore, a second *wNSC*-evaluation is undertaken in this section. Ahead of filtering and *wNSC* evaluation, the monthly A3HM grid data are multiplied by a factor of 1.5. This factor is estimated as an average maximum difference between A3HM and GRACE TWS amplitudes. Subsequently, the *wNSC*-evaluation is repeated. Normalised differences of the optimised filter parameters relative to the ones optimised with the original A3HM data (Table 5) are shown in Fig. 11. The results for the Nelson and St. Lawrence basins are excluded because *wNSC* values below zero were obtained and, therefore, no optimised parameter values could be achieved. Differences for the other basins mainly occur for parameters of filter III, which exhibits a low sensitivity for its filter parameter concerning filter performance (see Fig. 9). Parameter selection of VI also shows differences (Amur, Ob, Orange, Zaire), but here as well, the sensitivity of filter performance is low (see Section 3.2.2). Parameter differences of I, II and IV are either zero or are of expected evaluation uncertainties of one or two parameter step sizes (100 km, 2 mm and a factor of 1, respectively). Hence, except for the Nelson and St. Lawrence basins, these results prove that a possible error in the amplitude of the hydrological data would have small effects on the selection of optimal filter parameters by the *wNSC*-evaluation. This robustness of filter parameters is mainly due to the identical filtering of both data sets (GRACE and hydrological data) for the *wNSC* evaluation, in combination with an accounting of leakage errors by the weighting factor w in *wNSC*. This approach prevents a simple fitting of GRACE to hydrological amplitudes, since amplitude damping affects both data sets.

4 CONCLUSIONS

The results show that filter types and their corresponding parameters have to be selected carefully in order to derive basin-averaged time series of water storage variations from GRACE spherical harmonic data. The different smoothing effects of the different filter methods lead to varying balances of satellite and leakage errors in each river basin. We could determine the individual best filter

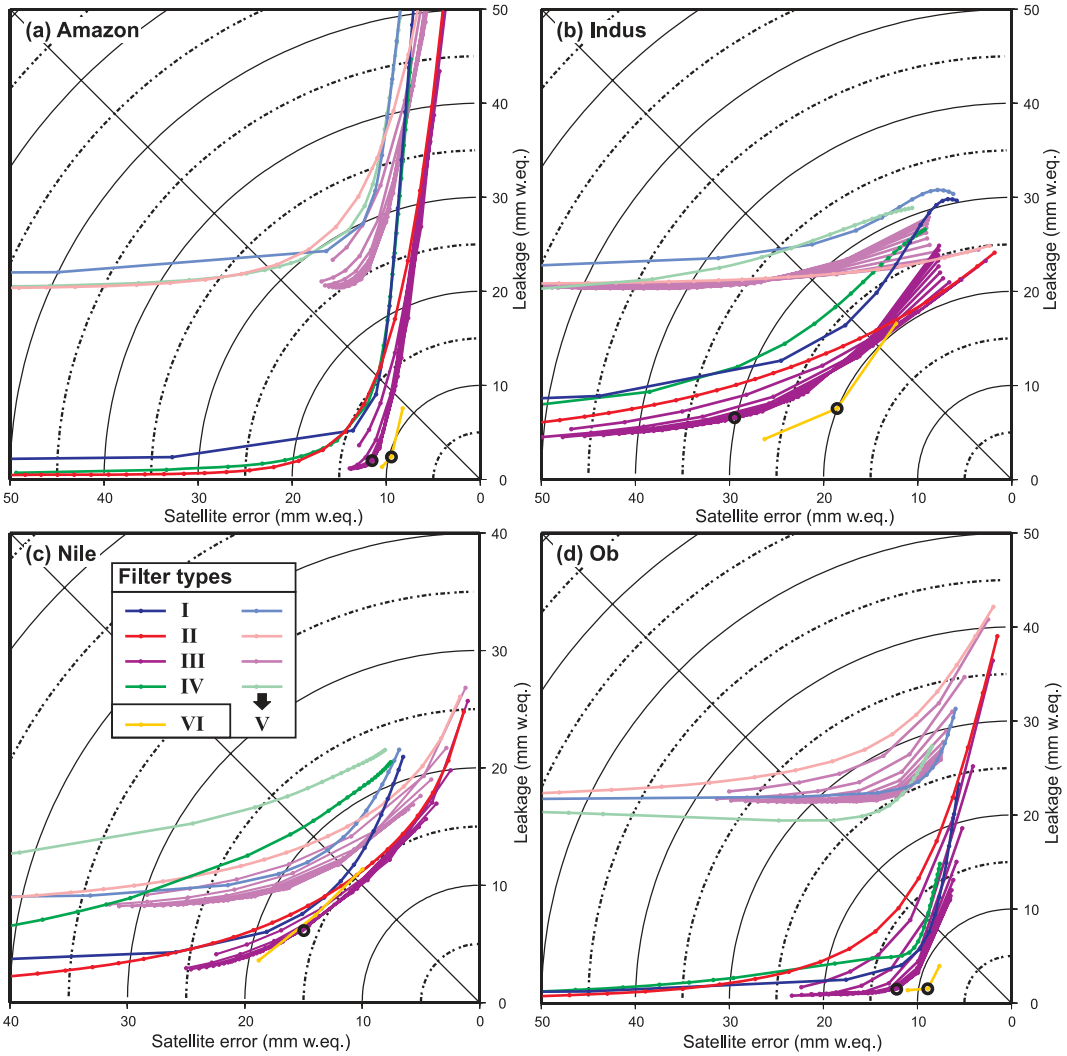

Figure 10. Error budget as hydrological leakage error (RMS differences of filtered and unfiltered A3HM time series) versus GRACE satellite error (RMS of propagated monthly coefficient errors) of TWS variations of : (a) the Amazon, (b) Indus, (c) Nile and (d) Ob river basins. Applied filter methods are shown by lines in blue: Gaussian (I), red: basin optimised (II), purple: signal model optimised (III), green: SNR optimised (IV). Light coloured lines: decorrelation filter V. Yellow line: decorrelation filter VI. Black circles indicate the individual optimised filter parameter from Table 5 and the bold black circles indicate the respective optimal filter method. See a filter description in Section 2.1.

methods for deriving basin-averaged water mass variations for the 22 largest river basins worldwide. When being evaluated by global hydrology, optimal parameters of the individual filter types vary for different basin sizes, shapes, and locations, as well as for signal type and intensity. Filter type VI provides generally good results. The differences of signal characteristics, like seasonal amplitude and phase, inside and outside a region of interest, highly influences the efficiency of a filter method. If phase shifts due to signals outside the river basin affect the TWS estimation, a bias or amplitude correction by a scale factor will not adequately recover the signal. Instead, a previous selection of an optimal filter type is expected to allow for a best possible bias correction. Additionally, for many filter types, the selection of an optimal parameter for the specific location and shape of the basin or process is necessary.

The decorrelation method VI was be the most efficient approach for the set of river basins analysed in this study. Only for basins of generally poor agreement between GRACE and hydrological data (Amur, Orange and St. Lawrence), was there a considerably higher correspondence provided by other filter methods. The usage of GRACE orbit-configurations to design a synthetic error covariance matrix sufficiently reduces the satellite error while preserving most of the hydrological signal for most of the river basins—even if they exhibit a small size or complex signal characteristics (e.g. Danube). To conclude, the general and global adaptability with moderate parameter sensitivity makes method VI the most reliable of the six analysed filter tools.

The isotropic Gaussian filter technique (I), or filter methods with little information on total error characteristics (II, IV, V), does not sufficiently address globally varying conditions for the extraction of basin-averaged TWS variations. Strong smoothing filter versions of method II and III tend to have more pronounced amplitude damping, while filters I and IV generally lead to phase shifts in the time series. Thus, the particular parameter values for these filter methods must be chosen carefully. It was shown that filter I gives acceptable results if the signal around the basin exhibits equal characteristics (Ob, Lena). If the river basin is characterised by a strong signal of TWS variability and is of large size and circuit shape, like the Amazon, leakage effects may be small if the parameter of

Table 7. 22 basin individual standard deviation of monthly satellite and leakage error for decorrelation VI optimised filtering (see Table 5 for respective parameter).

Basin	ϵ_{sat} [mm]	ϵ_{leak} [mm]	ϵ_{ges} [mm]
Amazon	9.4	2.4	9.7
Amur	11.1	2.6	11.4
Danube	21.6	4.5	22.1
Ganges	17.1	4.2	17.6
Indus	18.6	7.6	20.1
Lena	10.0	1.3	10.1
Mackenzie	12.3	2.2	12.5
Mississippi	11.0	1.6	11.1
Nelson	16.2	5.6	17.1
Niger	18.3	2.0	18.4
Nile	14.0	7.5	15.9
Ob	8.9	1.5	9.0
Orange	12.9	4.7	13.7
Parana	16.3	5.8	17.3
St. Lawrence	10.3	16.7	19.6
Tocantins	32.0	12.3	34.3
Volga	11.2	2.6	11.5
Yangtse	14.9	1.4	15.0
Yenisei	8.0	2.8	8.5
Yukon	16.3	11.6	20.0
Zaire	11.9	5.3	13.0
Zambezi	20.7	3.6	21.0

filter I is chosen carefully. But if the signal around the basin is of different characteristics because of such factors as the vicinity of oceans (Ganges, Yukon), deserts (Nile, Indus) or smaller signals in surrounding regions (Danube), leakage may reach high values for filter I. Furthermore, the leakage effect may be time-dependent in cases in which surrounding areas are characterised by a different seasonal water storage regime. In this case, method I is inappropriate. The principle of method II is only advantageous when its parameter values are optimised and the river basin of interest exhibits a complex shape or small size (e.g. Niger, Yukon and Zambezi). Method III can efficiently deal with similar (e.g. Ob) or different (e.g. Nile) signal characteristics outside the area of interest, as well as with small or complex basin shapes (Niger, Zambezi). It provides good filter results for half of the river basins due to its efficient leakage estimation with an exponential signal model. Because of high leakage effects, the decorrelation method V only provides satisfactory filter results in some basins (Orange, Lena). However, when using method V for decorrelation additional smoothing, and therefore parameter optimisation, is necessary. A nonpractical basin-based (instead of global) optimisation of the decorrelation parameter, in addition to the parameter of the superimposed filter method, may lead to improved results for that method. The conclusions above have been supported by reduction of amplitude and phase differences, total error budget maps and an amplitude sensitivity test.

It should be remembered that the results are derived from comparisons with model-based hydrological data, which might contain structural errors within specific basins. Such errors are caused by mis-modelled or missing processes, within any model, erroneous mode-forcing data or parameters. However, global hydrological models provide the only source of alternative TWS data sets for evaluating GRACE data. The ensemble mean A3HM was used as a compromise between three widely used hydrological models (GLDAS, LaD and WGHM) in order to evaluate GRACE filter parameters with the best possible accuracy, at present. Furthermore, the applied models mainly differ in amplitudes of TWS and it is shown that an amplitude error has a small effect on the filter evaluation. Only for the Indus, Amur, St. Lawrence, Orange and Nelson, may filter parameter type and selection be unreliable. A rather low correspondence between modelled and GRACE-derived data sets is due to temporal uncertainty of simulated TWS variability for these basins. The results in terms of optimum filters were shown to be robust, both for seasonal and non-annual TWS dynamics, in river basins. Nevertheless, for other applications, such as those with a focus on spatial patterns or secular trends, another prioritisation of filter methods may be more appropriate.

Hence, the discussion of adequate filter methods will likely continue as long as there is no breakthrough in accuracy for GRACE or GRACE Follow-On gravity field models. Filter types and parameters, as derived in this study, are particularly useful for GRACE data analysis within hydrological applications, such as monitoring of water mass exchange on the continent, studies of inter-annual variability in the hydrosphere, or using reliable water storage data as input for assimilation into large-scale hydrological models. The results obtained here can also be used as a guideline for filter selection for areas that were not specifically considered in this study.

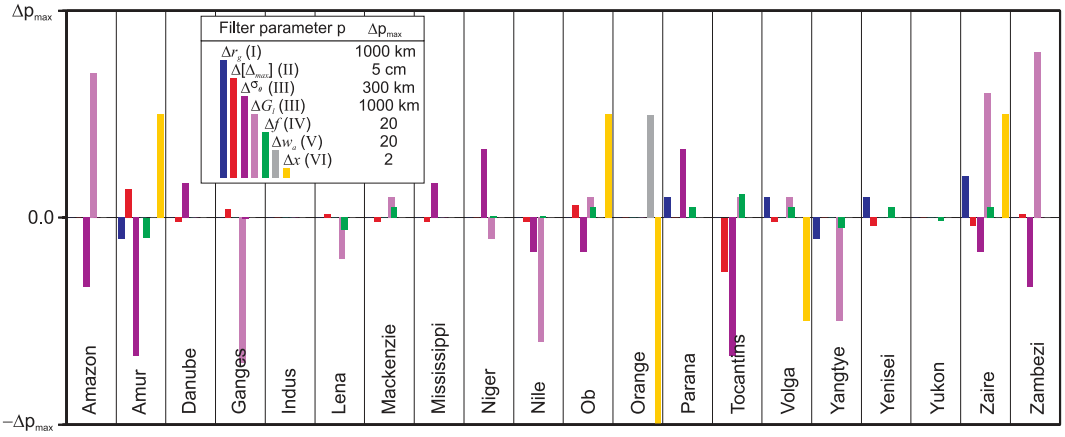

Figure 11. Parameter differences for Gaussian (I, blue), basin optimised (II, red), signal model optimised (III, light and dark purple), SNR optimised (IV, green), decorrelation V (grey) and decorrelation VI (black) filter methods for 20 river basins. Parameter differences are derived from *wNSC*-evaluations with A3HM versus a modified A3HM version of one-and-a-half times increased signal amplitudes: $\Delta p = p(A3HM) - p(1.5*A3HM)$. In the graphic, differences are normalised by the maximum parameter values chosen in this study (see legend).

ACKNOWLEDGMENTS

We wish to thank P.C.D. Milly, M. Rodell, J. Alcamo and P. Döll for providing the LaD, GLDAS and WGHM model data, respectively. Thanks to M. Scheinert for his GravTools software. The German Ministry of Education and Research (BMBF) supported these investigations within the geo-scientific R+D program GEOTECHNOLOGIEN 'Erfassung des Systems Erde aus dem Weltraum'. We also thank two anonymous reviewers for their constructive comments, which helped to improve the manuscript.

REFERENCES

Boronina, A. & Ramillien, G., 2008. Application of AVHRR imagery and GRACE measurements for calculation of actual evapotranspiration over the Quaternary aquifer (Lake Chad basin) and validation of groundwater models, *J. Hydrol.,* **348**(1-2), 98–109.

Chambers, D.P., Tamisiea, M.E., Nerem, R.S. & Ries, J.C., 2007. Effects of ice melting on GRACE observations of ocean mass trends, *Geophys. Res. Lett.,* **34,** L05610, doi:10.1029/2006GL029171.

Chao, B.F., 2005. On inversion for mass distribution from global (time-variable) gravity field, *J. Geodyn.,* **39**(3), 223–230.

Chen, J., Wilson, C., Famiglietti, J. & Rodell, M., 2007. Attenuation effect on seasonal basin-scale water storage changes from GRACE time-variable gravity, *J. Geodyn.,* **81**(4), 237–245.

Chen, J.L., Wilson, C.R. & Seo, K.W., 2006. Optimized smoothing of Gravity Recovery and Climate Experiment (GRACE) time-variable gravity observations, *J. geophys. Res.,* **111,** B06408, doi:10.1029/2005JB004064.

Chen, J.L., Wilson, C.R., Tapley, B.D., Blankenship, D. & Young, D., 2008. Antarctic regional ice loss rates from GRACE, *Earth Planet. Sci. Lett.,* **266**(1-2), 140–148.

Dobslaw, H. & Thomas, M., 2007. Simulation and observation of global ocean mass anomalies, *J. geophys. Res.,* **112,** C05040, doi:10.1029/2006JC004035.

Döll, P., Kaspar, F. & Lehner, B., 2003. A global hydrological model for deriving water availability indicators: model tuning and validation, *J. Hydrol.,* **270**(1-2), 105–134.

Ek, M.B., Mitchell, K.E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G. & Tarpley, J.D., 2003. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, *J. geophys. Res.,* **108**(D22), 8851.

Flechtner, F., 2007. GFZ Level-2 processing standards document for Level-2 product Release 0004, Rev. 1.0, GRACE 327-743 (GR-GFZ-STD-001), Deutsches GeoForschungsZentrum (GFZ), Potsdam, Germany.

Güntner, A., 2009. Improvement of global hydrological models using GRACE data, *Surv. geophys.,* **29**(4-5), 375–397.

Güntner, A., Stuck, J., Werth, S., Döll, P., Verzano, K. & Merz, B., 2007. A global analysis of temporal and spatial variations in continental water storage, *Water Resour. Res.,* **43,** W05416, doi:10.1029/2006WR005247.

Hagedorn, R., Doblas-Reyes, F.J. & Palmer, T.N., 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting–I. Basic concept, *Tellus A,* **57**(3), 219–233.

Han, S.C., Shum, C.K., Jekeli, C., Kuo, C.Y., Wilson, C. & Seo, K.W., 2005. Non-isotropic filtering of GRACE temporal gravity for geohysical signal enhancement, *Geophys. J. Int.,* **163**(1), 18–25.

Horwath, M. & Dietrich, R., 2006. Errors of regional mass variations inferred from GRACE monthly solutions, *Geophys. Res. Lett.,* **33,** L07502, doi:10.1029/2005GL025550.

Hunger, M. & Döll, P., 2008. Value of river discharge data for global-scale hydrological modeling, *Hydrol. Earth Syst. Sci.,* **12**(3), 841–861.

Jekeli, C., 1981. Alternative methods to smooth the Earth's gravity field, Tech. Rep. 327, Department of Geodetic Science and Surveying, Ohio State Univ., Columbus, OH.

Klees, R., Zapreeva, E.A., Winsemius, H.C. & Savenije, H.H.G., 2007. The bias in GRACE estimates of continental water storage variations, *Hydrol. Earth Syst. Sci.,* **11**(4), 1227–1241.

Kusche, J., 2007. Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models, *J. Geodesy,* **81**(11), 733–749.

Lemoine, J.-M., Bruinsma, S., Loyer, S., Biancale, R., Marty, J.-C., Perosanz, F. & Balmino, G., 2007. Temporal gravity field models inferred from GRACE data, *Adv. Space Res.,* **39**(10), 1620–1629.

Milly, P.C. & Shmakin, A.B., 2002a. Global modeling of land water and energy balances. Part I: the Land Dynamics (LaD) model, *J. Hydrometeorol.,* **3**(3), 283–299.

Milly, P.C. & Shmakin, A.B., 2002b. Global modeling of land water and energy balances. Part II: land-characteristic contributions to spatial variability, *J. Hydrometeorol.,* **3**(3), 301–310.

Nash, J.E. & Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part 1—a discussion of principles, *J. Hydrol.,* **10**(3), 282–290.

Niu, G.Y., Seo, K.W., Yang, Z.L., Wilson, C., Su, H., Chen, J. & Rodell, M., 2007. Retrieving snow mass from GRACE terrestrial water storage change with a land surface model, *Geophys. Res. Lett.,* **34,** L15704, doi:10.1029/2007GL030413.

Press, W.H., Teukolsky, S.T., Vetterling, W.T. & Flannery, B.P., 1992. *Numerical recipes in FORTRAN: The art of scientific computing* (2nd ed.), Cambridge Univ. Press, Cambridge.

Ramillien, G., Frappart, F., Cazenave, A. & Güntner, A., 2005. Time variations of land water storage from an inversion of 2 years of GRACE geoids, *Earth Planet. Sci. Lett.,* **235**(1-2), 283–301.

Ramillien, G., Bouhours, S., Lombard, A., Cazenave, A., Flechtner, F. & Schmidt, R., 2008. Land water storage contribution to sea level from GRACE geoid data over 2003–2006, *Global Planet. Change,* **60**(3-4), 381–392.

Regonda, S.K., Rajagopalan, B., Clark, M. & Zagona, E., 2006. A multimodel ensemble forecast framework: application to spring seasonal flows in the Gunnison River basin, *Water Resour. Res.,* **42,** W09404, doi:10.1029/2005WR004653.

Reigber, C., Schmidt, R., Flechtner, F., König, R., Meyer, U., Neumayer, K.H., Schwintzer, P. & Zhu, S.Y., 2005. An Earth gravity field model complete to degree and order 150 from GRACE: EIGEN-GRACE02S, *J. Geodyn.,* **39**(1), 1–10.

Rodell, M. *et al.*, 2004. The Global Land Data Assimilation System, *B. Am. Meteorol. Soc.,* **85**(3), 381–394.

Save, H., Bettadpur, S. & Tapley, B.D., 2008. The use of regularization for global GRACE solutions, in *Proceeding of the GRACE Science Team Meeting, San Francisco,* CSR-GR-08-01, Center for Space Research, The University of Texas at Austin, Texas.

Schaefli, B. & Gupta, H.V., 2007. Do Nash values have value?, *Hydrol. Process.,* **21**(15), 2075–2080.

Schmidt, R. *et al.*, 2006. GRACE observations of changes in continental water storage, *Global Planet. Change,* **50**(1-2), 112–126.

Schmidt, R., Flechtner, F., Meyer, U., Neumayer, K.H., Dahle, C., König, R. & Kusche, J., 2008a. Hydrological signals observed by the GRACE satellites, *Surv. geophys.,* **29**(4-5), 319–334.

Schmidt, R., Petrovic, S., Güntner, A., Barthelmes, F., Wünsch, J. & Kusche, J., 2008b. Periodic components of water storage changes from GRACE and global hydrology models, *J. geophys. Res.,* **113,** B08419, doi:10.1029/2007JB005363.

Schrama, E.J.O., Wouters, B. & Lavalle, D., 2007. Signal and noise in Gravity Recovery and Climate Experiment (GRACE) observed surface mass variations, *J. geophys. Res.,* **112,** B08407, doi:10.1029/2006JB004882.

Seo, K.W., Wilson, C.R., Famiglietti, J.S., Chen, J.L. & Rodell, M., 2006. Terrestrial water mass load changes from Gravity Recovery and Climate Experiment (GRACE), *Water Resour. Res.,* **42,** W05417, doi:10.1029/2005WR004255.

Swenson, S. & Wahr, J., 2002. Methods for inferring regional surface-mass anomalies from Gravity Recovery and Climate Experiment (GRACE) measurements of time-variable gravity, *J. geophys. Res.,* **107**(B9), doi:10.1029/2001JB000576.

Swenson, S. & Wahr, J., 2006. Post-processing removal of correlated errors in GRACE data, *Geophys. Res. Lett.,* **33,** L08402, doi:10.1029/2005GL025285.

Swenson, S., Yeh, P.J.F., Wahr, J. & Famiglietti, J.S., 2006. A comparison of terrestrial water storage variations from GRACE with *in situ* measurements from Illinois, *Geophys. Res. Lett.,* **33,** L16401, doi:10.1029/2006GL026962.

Swenson, S., Famiglietti, J., Basara, J. & Wahr, J., 2008. Estimating profile soil moisture and groundwater variations using GRACE and Oklahoma Mesonet soil moisture data, *Water Resour. Res.,* **44,** W01413, doi:10.1029/2007WR006057.

Syed, T.H., J.S., F., Rodell, M., Chen, J. & Wilson, C.R., 2008. Analysis of terrestrial water storage changes from GRACE and GLDAS, *Water Resour. Res.,* **44,** W02433, doi:10.1029/2006WR005779.

Tapley, B.D., Bettadpur, S., Watkins, M. & Reigber, C., 2004. The gravity recovery and climate experiment: mission overview and early results, *Geophys. Res. Lett.,* **31,** L09607, doi:10.1029/2004GL019920.

Tebaldi, C. & Knutti, R., 2007. The use of the multi-model ensemble in probabilistic climate projections, *Philos. T. R. Soc. A,* **365**(1857), 2053–2075.

Velicogna, I. & Wahr, J., 2006. Measurements of time-variable gravity show mass loss in Antarctica, *Science,* **311**(5768), 1754–1756.

Wahr, J., Molenaar, M. & Bryan, F., 1998. Time variability of the Earth's gravity field: hydrological and oceanic effects and their possible detection using GRACE, *J. geophys. Res.,* **108**(B12), 30 205–30 229.

Watkins, M.M., Yuan, D.N., Kuang, D., Bertiger, W., Byun, S., Lu, W. & Kuizinga, G.L., 2008. JPL L-2 GRACE solutions: harmonics, mascons, iteration and constaints, in *Proceeding of the GRACE Science Team Meeting, San Francisco,* CSR-GR-08-01, Center for Space Research, The University of Texas at Austin, Texas.

Werth, S. & Güntner, A., 2008. Intercomparison of global hydrological models in terms of water storage, in *AGU Fall Meeting, December 2008,* San Francisco, California.

Willmot, C.J., 1984. On the evaluation of model performance in physical geography, in *Spatial Statistics and Models,* pp. 443–460, D. Reidel, Dordrecht, the Netherlands.

Wouters, B., Chambers, D. & Schrama, E.J., 2008. GRACE observes small-scale mass loss in Greenland, *Geophys. Res. Lett.,* **35,** L20501, doi:10.1029/2008GL034816.