# A global water cycle reanalysis (2003–2012) merging satellite gravimetry and altimetry observations with a hydrological multi-model ensemble

**Albert I.J.M. van Dijk[1*], Luigi J. Renzullo[2], Yoshihide Wada[3], Paul Tregoning[4]**

[1] {Fenner School of Environment & Society, The Australian National University, Canberra, Australia}

[2] {CSIRO Land and Water, Canberra, Australia}

[3] {Department of Physical Geography, Utrecht University, Utrecht, Netherlands}

[4] {Research School of Earth Sciences, The Australian National University, Canberra, Australia}

Correspondence to: Albert Van Dijk, albert.vandijk@anu.edu.au

## Abstract

We present a global water cycle reanalysis that reconciles water balance estimates derived from the GRACE satellite mission, satellite water level altimetry and off-line estimates from several hydrological models. Error estimates for the sequential data assimilation scheme were derived from available uncertainty information and the triple collocation technique. Errors in four GRACE storage products were estimated to be 11–12 mm over land areas, while errors in monthly storage changes derived from five global hydrological models were estimated to be 17–28 mm. Prior and posterior estimates were evaluated against independent observations of river water level and discharge, snow water storage and glacier mass loss. Data assimilation improved or maintained agreement overall, although results varied regionally. Uncertainties were greatest in regions where glacier mass loss and sub-surface storage decline are both plausible but poorly constrained. We calculated a global water budget for 2003–2012. The main changes were a net loss of polar ice caps (-342 Gt $y^{-1}$) and mountain glaciers (-230 Gt $y^{-1}$), with an additional decrease in seasonal snow pack (-18 Gt $y^{-1}$). Storage increased due to new impoundments (+16 Gt $y^{-1}$), but this was compensated by decreases in

other surface water bodies (-10 Gt $y^{-1}$). If the effect of groundwater depletion (-92 Gt $y^{-1}$) is excluded, sub-surface water storage increased by +110 Gt $y^{-1}$ due particularly to increased wetness in northern temperate regions and in the seasonally wet tropics of South America and southern Africa.

## 1. Introduction

More accurate global water balance estimates are needed, to better understand interactions between the global climate system and water cycle (Sheffield et al., 2012), the causes of observed sea level rise (Boening et al., 2012; Fasullo et al., 2013; Cazenave et al., 2009; Leuliette and Miller, 2009), human impacts on water resources (Wada et al., 2010; 2013), and to improve hydrological models (van Dijk et al., 2011) and initialise water resources forecasts (Van Dijk et al., 2013). The current generation of global hydrological models have large uncertainties arising from a combination of data deficiencies (e.g., precipitation in sparsely gauged regions; poorly known soil, aquifer and vegetation properties) and overly simplistic descriptions of important water cycle processes (e.g. groundwater dynamics, human water resources extraction and use, wetland hydrology and glacier dynamics). Data assimilation (DA) is used routinely to overcome data and model limitations in atmospheric reconstructions or ‘reanalysis’. In hydrological applications, DA has been largely limited to flood forecasting, but new applications are being developed (Liu et al., 2012a), including promising developments towards large-scale water balance reanalyses, alternatively referred to as monitoring, assessment or estimation (van Dijk and Renzullo, 2011).

Here, we undertake a global water cycle reanalysis for the period 2003–2012. Specifically, we attempt to reconcile global water balance model estimates from different sources with an ensemble of total water storage (TWS) estimates derived from the Gravity Recovery And Climate Experiment (GRACE) satellite mission (Tapley et al., 2004). Various alternative approaches can be conceptualised to achieve this integration and the most appropriate among these is not obvious. Our approach was to use water balance estimates generated by five global hydrological models along with several ancillary data sources to generate an ensemble of prior estimates of monthly water storage changes. Errors in the different model estimates and GRACE products were estimated spatially through triple collocation (Stoffelen, 1998). Subsequently, a DA scheme was designed to sequentially reconcile the model ensemble and GRACE observations. The reanalysis results were evaluated with independent global streamflow records, remote sensing of river water level and snow water equivalent (SWE), and independent glacier mass balance estimates.

## 2. Methods and Data Sources

### 2.1. Overall approach

We conceptualise TWS ($S$, in mm) as the sum of five different water stores ($s$ in mm), *i.e.*, water stored in snow and ice ($s_{snow}$); below the surface in soil and groundwater ($s_{sub}$), and in rivers ($s_{riv}$); lakes ($s_{lake}$), and seas and oceans ($s_{sea}$). We ignore atmospheric water storage changes, which are removed from the signal during the GRACE TWS retrieval process (e.g., Wahr et al., 2006), and vegetation mass changes, which are assumed negligible. The GRACE TWS estimates are denoted by $y$ and have the same units as $S$ but are distinct in their much smoother spatial character.

To date, DA schemes developed for large-scale water cycle analysis typically use Kalman filter approaches (Liu et al., 2012a). This requires calculation of co-variance matrices and, presumably because of complexity and computational burden, has only been applied for single models and limited regions (e.g., Zaitchik et al., 2008). We aimed to develop a DA scheme that made it possible to use water balance estimates derived 'off line' (i.e., in the absence of DA) so we could use an ensemble of already available model outputs. In the DA terminology of Bouttier and Courtier (1999), our scheme could be described as sequential and near-continuous with a spatially variable but temporally stable gain factor. The characteristics of the DA problem to be addressed in this application were as follows:

(1) Alternative GRACE TWS estimates ($y^o$) were available from different processing centres and error estimates were required for each;

(2) Alternative estimates for some of the stores, $s$, were available from different hydrological models with higher definition than $y^o$;

(3) Error estimates were required for each store and data source;

(4) A method was required to spatially transform between $s$ and $y$ as part of the assimilation.

## 2.2. Data sources

The data used include those needed to derive prior estimates for each of the water cycle stores, the GRACE retrievals to be assimilated and independent observations to evaluate the quality of the reanalysis. All are listed in Table 1 and described below.

Monthly water balance components from four global land surface model estimates at 1° resolution were obtained from NASA's Global Data Assimilation System (GLDAS) (Rodell et al., 2004). The four models include CLM, Mosaic, NOAH and VIC which, for the 2003–2012, were forced with "a combination of NOAA/GDAS atmospheric analysis fields, spatially and temporally disaggregated NOAA Climate Prediction Center Merged Analysis of Precipitation (CMAP) fields, and observation-based radiation fields derived using the method of the Air Force Weather Agency's AGRicultural METeorological modelling system" (Rui, 2011). The models are described in Rodell et al. (2004). From the model outputs we used *(i)* snow water equivalent (SWE) depth, *(ii)* total soil moisture storage over a soil depth that varies between models, and *(iii)* generated streamflow, calculated as the sum of surface runoff and sub-surface drainage. In addition to GLDAS, we used global water balance estimates generated by the W3RA model (Van Dijk et al., 2013) in the configuration used in the Asia-Pacific Water Monitor (http://eos.csiro.au/apwm/). For 2003–2008, the model was forced with the 'Princeton' merged precipitation, down-welling short-wave radiation, minimum and maximum daily temperature and air pressure data produced by Sheffield et al. (2006). From 2009 onwards, the model primarily uses 'ERA-Interim' weather forecast model reanalysis data from the European Centre for Medium-Range Weather Forecasts. For low latitudes, these are combined with near-real time TRMM multi-sensor precipitation analysis data (TMPA 3B42 RT) (Huffman et al., 2007) to improve estimates of convective rainfall (Peña-Arancibia et al., 2013). Both were bias-corrected with reference to the Princeton data to ensure homogeneity. W3RA model estimates were conceptually similar to those from GLDAS, except that the model includes deep soil and groundwater stores and sub-grid surface and groundwater routing.

The five hydrological models do not provide estimates of groundwater depletion and storage in rivers, lakes and impoundments and these were therefore derived separately. Groundwater depletion estimates were derived for 1960–2010 by Wada et al. (2012). The time series were calculated as the net difference between estimated groundwater extraction and recharge. National groundwater extraction data compiled by the International Groundwater Resources Assessment Centre (IGRAC) were disaggregated using estimates of water use intensity and surface water availability at 0.5° resolution from a hydrological model (PCR-GLOBWB; see Wada et al., 2012, for details). The model also estimated recharge including return flow from irrigation. Uncertainty information of groundwater depletion was generated by 10,000 Monte Carlo simulations, with 100 realizations of extraction and recharge respectively (Wada et al., 2010). This method tends to overestimate reported depletion in non-arid regions, where groundwater pumping can enhance recharge from surface water. Wada et al. (2012) used a universal multiplicative correction to account for this. Here, the correction was calculated per climate region rather than world-wide, reflecting the dependency of uncertainty on recharge estimates and their errors. Data for 2011–2012 were not available; these were estimated using monthly average depletion and uncertainty values for the preceding 2003–2010 period. Given the regular pattern of depletion in the preceding years this by itself is unlikely to have affected the analysis noticeably.

River water storage was estimated by propagating runoff fields from each of the five models through a global routing scheme. In a previous study, we compared these runoff fields with streamflow records from 6,192 small (<10,000 $km^2$) catchments worldwide and found that observed runoff was 1.28 to 1.77 times greater than predicted by the different models (Van Dijk et al., 2013). The respective values were used to uniformly bias-correct the runoff fields. Next, we used a global 0.5º resolution flow direction grid (Oki et al., 1999; Oki and Sud, 1998) to parameterise a cell-to-cell river routing scheme. We used a linear reservoir kinematic wave approximation (Vörösmarty and Moore III, 1991), similar to that used in several large-scale hydrology models (see recent review by Gong et al., 2011). The monthly 1º runoff fields from each of the five models were oversampled to 0.5º and daily time step before routing, and the river water storage estimates (in mm) were aggregated back to monthly 1º grid cell averages before use in assimilation. The routing function was an inverse linear function of the distance between network nodes and a transfer (or routing) coefficient. For each model, a globally uniform optimal transfer coefficient was found by testing values of 0.3 to 0.9 $day^{-1}$ in 0.1 $day^{-1}$ increments and finding the value that produced best overall agreement with seasonal flow patterns observed in 586 large rivers world-wide. These 586 were a subset of 925 ocean-reaching rivers for which streamflow records were compiled by Dai et al. (2009) from various sources; we excluded locations where streamflow records were available for less than 10 years since 1980 or less than 6 months of the year.

The resulting river flow estimates do not account for the impact of river water use (i.e., the evaporation of water extracted from rivers, mainly for irrigation). We addressed this using global monthly surface water use estimates that were derived in a way similar to that used for groundwater depletion estimates (details in Wada et al., 2013). For each grid cell, mean water use rates for 2002–2010 were subtracted from mean runoff estimates for the same period, and the remaining runoff was routed downstream. The resulting mean net river flow estimates were divided by the original estimates to derive a scaling factor, which was subsequently applied at each time step. Lack of additional global information on river hydrology meant that three simplifications needed to be made: (*i*) our approach implies that for a particular grid cell, monthly river water use is assumed proportional to river flow for that month; (*ii*) the influence of lakes, wetlands and water storages on downstream flows (e.g., through dam operation) is not accounted for, even though their actual storage changes are (see further on); (*iii*) our approach does not account for losses associated with permanent or ephemeral wetlands, channel leakage and net evaporation from the river channel. To some extent, the DA process may correct mass errors resulting from these assumptions.

Variations in lake water storage were not modelled, but water level data for 62 lakes world-wide were obtained from the Crop Explorer web site (Table 1) and include most of the world's largest lakes and reservoirs, including the Caspian Sea. The water level data for these lakes were derived from satellite altimetry and converted to mm water storage. Measurements were typically available every 10 days. The mean and standard deviation for each individual month were used as best estimate and estimation error, respectively. Storage in water bodies without altimetry data was necessarily assumed negligible. This includes many small lakes and dams, but also some larger lakes affected by snow and ice cover (e.g., the Great Bear and Great Slave Lakes in Canada) and ephemeral, distributed or otherwise complex water bodies (e.g., the Okavango delta in Botswana and Lake Eyre in Australia, each of which contains >10 $km^3$ of water when full).

A list of dams was collated by Lehner et al. (2011) and was updated with large dams constructed in more recent years with the ICOLD data base (Table 1). For the period 1998–2012, a total 198 georeferenced dams with a combined storage capacity of 418 $km^3$ were identified. For the Three Gorges Dam (39 $km^3$), reservoir water level time series from *http://www.ctg.com.cn/inc/sqsk.php* were converted to storage volume following Wang et al. (2011). For the remaining dams, we assumed a gradual increase to storage capacity over the first five years after construction with a relative estimation error of 20%.

Delayed time, up-to-date global merged mean sea level anomalies were obtained from the Aviso web site (Table 1). The monthly data were reprojected from the native 1/3° Mercator grid to regular 1° grids. An estimate of uncertainty was derived by calculating the spatial standard deviation in sea level values within a 4° by 4° region around each grid cell during re-projection. When sea level data were missing, because of sea ice, we assumed sea level did not change and assigned an uncertainty of 5 mm. Following the recent global sea level budget study by Chen et al. (2013), we assumed that 75% of the observed sea level change was due to mass increase, and we multiplied altimetry sea level anomalies with this factor.

We did not have spatial global time series of glacier mass changes. The five hydrological models have an oversimplified representation of ice dynamics, and therefore large uncertainties and errors can be expected for glaciated regions. To account for this, we used the 'GGHYDRO' global glacier extent mapping by Cogley (2003) to calculate the percentage glacier area for each grid cell, and assumed a proportional error in monthly glacier mass change estimates corresponding to 300 mm per unit glacier area. This value was chosen somewhat arbitrarily and ensures that a substantial fraction of the analysis increment is assigned to glaciers.

Three alternative GRACE TWS retrieval products were downloaded from the Tellus web site. The three products (coded CSR, JPL and GFZ; release 05) each had 1° (nominal) and monthly resolution. The land and ocean mass retrievals (Chambers and Bonin, 2012) were combined. The land retrievals had been 'de-striped' and smoothed with a 200 km half-width spherical Gaussian filter (Swenson et al., 2008; Swenson and Wahr, 2006), whereas the ocean retrievals had been smoothened with a 500 km filter (Chambers and Bonin, 2012). The DA method we employed is designed to deal with the signal 'leakage' caused by the smoothing process and therefore we did not use the scaling factors provided by the algorithm developers. In addition, gravity fields produced by CNES/GRGS (Bruinsma et al., 2010) at 1° resolution for 10 day periods were used. The three Tellus data sources had been corrected for Glacial Isostatic Adjustment (GIA); we corrected the GRGS data using the same GIA estimates of Geruo et al. (2013). Initial DA experiments produced unexpectedly strong mass trends around the Gulf of Thailand. Inspection demonstrated that all products, to different degrees, contained a mass redistribution signal associated with the December 2004 Sumatera-Andaman earthquake. To account for this, we first calculated a time series of seasonally-adjusted monthly anomalies (i.e., the average seasonal cycle was removed) for the region [5°N–15°, 80–110°E]. Next, we adjusted values after December 2004 by the difference in the mean adjusted anomalies for the year before and after the earthquake, respectively.

## 2.3. Data assimilation scheme

For each update cycle, the DA scheme proceeds through the steps illustrated in Figure 1 and described below.

*1) Deriving the prior estimate for each store*. The way to calculate the prior (or background) estimate of storage $s_t^b$ varied between stores. A systematic and accumulating bias (or 'drift') was considered plausible for the deep soil and groundwater components of model-derived sub-surface storage due to slow groundwater dynamics (including extraction) and ice storage in permanent glaciers and ice sheets, which may be progressively melting or accumulating. In these cases, the model-estimated *change* in storage was assumed more reliable than the actual storage itself, and estimates from the five models were used to calculate storage change, $\Delta s_t^b$ for store *i* (*i*=1,…, *N*) as:

$$\Delta s_t^b(i) = \sum_{l=1}^{L} w_l x_t^l(i) \tag{1}$$

where $x_t^l$ is the estimate of storage change from model *l* (*l*=1,…, *L*) between time *t*–1 and *t*, and $w_l$ the relative weight of model *l* in the ensemble, computed as:

$$w_l = \frac{\sigma_l^{-2}}{\sum_l \sigma_l^{-2}} \tag{2}$$

where $\sigma_{y,l}^2$ is the error for model *l* based on triple collocation (see Section 2.4). Subsequently, $s_t^b$ was calculated as:

$$s_t^b(i) = s_{t-1}^{a*}(i) + \Delta s_t^b(i) \tag{3}$$

where $s_{t-1}^{a*}$ is the posterior (or analysis) estimate from the previous time step. This approach was not suitable for model-estimated seasonal snowpack and river storage, where the ephemeral nature of the storage means that long-term drift is not an issue and Eq. (2) could in fact lead to unrealistic negative storage values. For these cases, $s_t^b$ was computed as:

$$s_t^b(i) = \sum_{l=1}^{L} w_l s_t^l(i) \tag{4}$$

where $s_t^l$ is the storage estimate from model *l*. The glacier extent map was used to identify whether Eq. (3) or (4) should be used for $s_{snow}$. Similarly, no drift was expected in the ocean and lake storage data, and these were used directly as estimates of $s_t^b$.

*2) Deriving the prior estimate of GRACE-like TWS ($y^b$).* This estimate was derived by summing all stores $s_t^b$ as:

$$S_t^b = \sum_{i=1}^{N} s_t^b(i) \tag{5}$$

and subsequently applying a convolution operator Γ to transform $S_t^b$ to a 'GRACE-like' TWS $y^b$. The operator Γ was a Gaussian smoother (cf. Jekeli, 1981) written here as:

$$y_t^b(j_1) = \sum_{j_1} \Gamma(j_1, j_2)\, S_t^b(j_1, j_2) \tag{6}$$

where $j_1$ and $j_2$ in principle should encompass all existing grid cell coordinates. In practice, Γ was applied as a moving Gaussian kernel with a size of 6°×6° and a half-width of 300 km (see further on).

*3) Updating the GRACE-like TWS.* The updated GRACE-like TWS, $y_t^a$, was calculated from the prior (Eq. (6)) and GRACE observations $y_t^o$ for time *t* as (cf. Figure 1 a-d):

$$y_t^a = y_t^b + \delta y_t = y_t^b + k(y_t^o - y_t^b) \tag{7}$$

where $\delta y_t$ is the analysis increment and *k* a temporally static gain factor derived by combining the error variances of modelled and observed *y* as follows:

$$k = \frac{\sum_l w_{y,l}\sigma_{y,l}^2}{\sum_l w_{y,l}\sigma_{y,l}^2 + \sum_m w_{y,m}\sigma_{y,m}^2} \tag{8}$$

where $w_{y,l}$ and $w_{y,m}$ are the weights applied to each of the five GRACE-like TWS estimates and four GRACE data sources, respectively, calculated from their respective error variances $\sigma_{y,l}^2$ and $\sigma_{y,m}^2$ analogous to Eq. (2).*4) Spatially disaggregating the analysis increment to the different stores.* The observation model was inverted and combined with the store error estimates in order to spatially redistribute the analysis increment $\delta y_t$, as follows (cf. Figure 1e-g):

$$\delta s_t(i, j_1) = \sum_{j_2} \Omega(j_1, j_2)\delta y_t(j_2) \tag{9}$$

where the redistribution operator Ω can be written as (cf. Figure 1g):

$$\Omega(j_1, j_2) = \frac{\Gamma(j_1, j_2)\sigma^{-2}(i, j_2)}{\sum_i \sum_{j_1} \Gamma(j_1, j_2)\sigma^{-2}(i, j_2)} \tag{10}$$

To implement this, spatial error estimates are required for each store. For lakes and seas, the errors were estimated from the observations (see Section 2.2). For the model-based estimates, the error was calculated for each time step and store as:

$$\sigma_t^2(i) = \sum_l w_l [x_t^l(i) - \Delta s_t^b(i)]^2 \tag{11}$$

The resulting error estimates are spatially and temporally dynamic and respond to the magnitude of the differences between the different model estimates. For $s_{sub}$ and $s_{snow}$ we combined the error estimates derived by Eq. (11) with the estimated errors in groundwater depletion and glacier mass change, respectively (see Section 2.2), calculating total error as the quadratic sum of the composite errors.

*5) Updating the stores.* In the final step, the state of each store is updated:

$$s_t^a(i) = s_t^b(i) + \delta s_t(i) \tag{12}$$

Subsequently, the procedure is repeated for the next time step.

## 2.4. Error estimation

Spatial error fields are required for all data sets to calculate the gain factor *k* and where necessary these were estimated using the triple collocation technique (Stoffelen, 1998). This technique infers errors in three independent time series by analysing the covariance structure. The approach has been applied widely to estimate errors in, among others, satellite-derived surface soil moisture (Dorigo et al., 2010; Scipal et al., 2009), evapotranspiration (Miralles et al., 2011) and vegetation leaf area (Fang et al., 2012). A useful description of the technique, the assumptions underlying it and an extension of the theory to any number of time series greater than three was provided by Zwieback et al. (2012). Application requires three (or more) estimates of the same quantity. This was achieved by convolving the model-derived storage estimates into large-scale, smoothed TWS estimates equivalent to those derived from GRACE measurements using Eqs. (5) and (6). Inspection of the original Tellus data made clear that the 200 km filter that was already applied as part of the land retrieval had only removed part of the spurious aliasing in the data sets, and propagated these artefacts into the error estimates and reanalysis. Therefore a smoother, 300 km filter was applied to the Tellus TWS data sets. Because conceptual consistency is required for triple collocation, the same filter was applied to the GRGS and model-derived TWS estimates. Several alternative Tellus and model time series were available, and therefore the triple collocation technique could be used to produce alternative error estimates from multiple triplet combinations (i.e., five for Tellus TWS, three for model TWS, and 5×3=15 for GRGS TWS). The agreement between these alternative estimates was calculates as a measure of uncertainty in estimated errors.

Important assumptions of the collocation technique are that: (1) each data set is free of bias relative to each other, (2) errors do not vary over time, (3) there is no temporal autocorrelation in the errors, and (4) there is no correlation between the errors in the respective time series (Zwieback et al., 2012). Each of these assumptions is difficult to ascertain, but some interpretative points can be made. Errors in the GRACE products vary somewhat from month to month depending on data availability, and overall decreased after June 2003. Therefore assumption (2) is a simplification. Assumption (3) is also unlikely to hold fully: there will almost certainly be systematic errors and biases that cause temporal correlation in the errors in the modelled TWS (e.g., due to poorly represented processes causing secular trends such as groundwater extraction or glacier melt). We avoided this assumption by applying the triple collocation to monthly storage changes rather than actual storage, although temporal correlation in storage change errors remains a possibility. However, temporal correlation in the GRACE errors is unlikely. Therefore, the error in individual mass estimates was calculated following conventional error propagation theory, by dividing the estimated error in mass changes by $\sqrt{2}$.

Assumption (4) will not be fully met where estimates are partially based on the same principle or measurement. In this study, arguably the most uncertain assumption is that the GRGS and Tellus errors are to a large extent uncorrelated. The basis for this assumption is that most of the error is likely to derive from the TWS retrieval method rather than the primary measurements (Sakumura et al., 2014). The GRGS time series was selected as the third triple collocation member because the four Tellus products are retrieved by methods that are comparatively more similar than the GRGS method, which uses ancillary observations from the Laser Geodynamics Satellites (Tregoning et al., 2012). Correspondingly, global average correlation among the Tellus TWS time series was stronger (0.61–0.73) than between any of the Tellus and GRGS time series (0.49–0.58). Nonetheless, there may well have been a residual covariance between errors in the GRGS and Tellus products. In triple collocation, this would cause some part of the differences to be wrongly attributed to the prior estimates rather than the observation products. Therefore, we conservatively inflated the calculated value by including an additional error of 5 mm through quadratic summation before calculating the gain factor (Eq. 8).

Uncertainty in the derived error estimates also arises from sample size, i.e. the number of collocated observations (*N*=111). Previous studies have suggested that 100 samples are sufficient to produce a reasonable estimate (Dorigo et al., 2010), although Zwieback et al. (2012) calculate that the relative uncertainty in the estimated errors for *N*=111 can be expected to be in the order of 20%. Such a modest uncertainty in derived errors will not have a strong impact on the reanalysis results.

## 2.5. Evaluation against observations

Evaluation of the reanalysis results for sub-surface storage was a challenge: ground observations are not widely available at global scale, are often conceptually not equivalent to the reanalysis terms, require tenuous scaling assumptions for comparison at 1° grid cell resolution, and many existing data sets contain few or no records during 2003–2012. For example, comparison with in situ soil moisture measurements or groundwater bore data is beset by such problems (Tregoning et al., 2012). Similarly, an initial comparison with near-surface (<5 cm depth) soil moisture estimates from passive and active microwave remote sensing (Liu et al., 2012b; Liu et al., 2011) showed that the conceptual difference between the two quantities was too great for a meaningful comparison.

We were able to evaluate the reanalysis for storage in rivers, seasonal snow pack and glaciers, however. Firstly, a total of 1,264 water level time series for several large rivers worldwide were obtained from the Laboratoire d'Etudes en Geodésie et Océanographie Spatiales (LEGOS) HYDROWEB web site (Table 1). The river levels were retrieved from ENVISAT and JASON-2 satellite altimetry (Crétaux et al., 2011) and included uncertainty information for each data period. From each time series, we removed data points with an estimated error of more than 25% of the temporal standard deviation (SD). Another 165 altimetry time series were obtained from the European Space Agency (ESA) River&Lake web site (Berry, 2009). These were selected to increase measurement period and sample size for the available locations, as well as extending coverage to additional rivers. The ESA time series did not include error estimates; instead data plots were judged visually to assess the likelihood of measurement noise; seemingly affected time series and outlier data points (>3SD) were excluded. The total 1,429 time series were merged for individual 1º grid cells. In each case, the longest time series was chosen as reference. Overlapping time periods were used to remove (typically small) systematic biases in water surface elevation between time series; where there was no overlap the time series were normalised by the median water level. The ESA data were used where or when HYDROWEB data were not available, and merged time series with fewer than 24 data points in total were excluded. The resulting data set contained time series for 442 grid cells with an average 61 (maximum 115) data points during 2003–2012. The relationship between river water level and river discharge (i.e., the discharge rating curve) is usually non-linear but unknown, and therefore a direct comparison could not be made. Instead, we calculated Spearman's rank correlation coefficient ($\rho$) between estimated discharge and observed water level.

Secondly, we used the already mentioned discharge data for 586 ocean-reaching rivers world-wide (Dai et al., 2009). From these, we selected 430 basins for which the reported drainage area was within 20% of the area derived from the 0.5º routing network. The ratio between reported and model-derived drainage area was used to adjust the reanalysis estimates and these were compared with recorded mean streamflow. The recorded mean annual discharge values are not for 2003-2012, but we assume that the differences are not systematic and, therefore, that any large change in agreement may still be a useful indicator of reanalysis quality.

Third, snow storage estimates were evaluated with the European Space Agency GlobSnow product (Luojus et al., 2010). This data set contains monthly 0.25° resolution estimates of snow water equivalent (SWE, in mm) for low relief regions with seasonal snow cover north of 55°N during 2003–2011. The SWE estimates are derived through a combination of AMSR-E passive microwave remote sensing and weather station data (Pulliainen, 2006; Takala et al., 2009). The GlobSnow data were aggregated to 1° resolution. The root mean square error (RMSE) and the coefficient of correlation ($r^2$) were calculated as measures of agreement.

Finally, we compared the estimated trends in storage in different glacier regions to trends for mountain glaciers compiled by Gardner et al. (2013) for 2003–2010 and for Greenland and Antarctica by Jacob et al. (2012) for 2003–2009. In some cases, these mass balance estimates were based on independent glaciological or ICESAT satellite observations and these were the focus of comparison. Other estimates were partially or wholly based on GRACE data, which makes comparison less insightful.

## 3. Results

### 3.1. Error estimation

The mean errors derived by the triple collocation technique were of similar magnitude for the GRACE and model estimates (Table 2; note that the numbers listed are for storage change rather than storage per se and are not adjusted for GRACE error covariance; cf. Section 2.4). The relatively low values for the coefficient of variation suggest that the error estimates are reasonably robust.

The spatial error in merged GRACE and model storage change estimates were calculated analogous to Eq. (8). The resulting GRACE error surface was relatively homogeneous with an estimated error of around 5–20 mm for most regions, but increasing to 20–40 mm over parts of the Amazon and the Arctic (Figure 2a). The combined model error surface suggest that errors are smaller than those in the GRACE data for arid regions (<10 mm) but higher elsewhere, increasing beyond 80 mm in the Amazon region (Figure 2b). The mean errors over non-glaciated land areas were similar, at 18.1 mm for the combined model and 13.5 mm for the combined GRACE data. Assuming no temporal correlation and allowing for error covariance among GRACE products reduces the latter to 10.8 mm (i.e., $\sqrt{13.5^2/2+5^2}$).

### 3.2. Analysis increments

Inspection of the analysis increments and the overall difference between prior and posterior estimates provides insights into the functioning of the assimilation scheme (Figure 3). The spatial pattern in root mean squared (RMS) TWS increments ($\sqrt{\overline{\delta S^2}}$) emphasises the important role of the world's largest rivers in explaining mismatches between expected and observed mass changes, particularly in tropical humid regions (Figure 3a). Large increments also occurred over Greenland (mainly due to updated ice storage changes) and the seasonally-wet regions of Brazil, Angola and south Asia (sub-surface storage). When considering the RMS between prior and posterior estimates of actual TWS as opposed to monthly changes (Figure 3b) a similar pattern emerges, but with more emphasis on the smaller but accumulating difference in estimated storage over Greenland, Alaska and part of Antarctica (due to updated ice mass changes) and northwest India (groundwater depletion).

### 3.3. Mass balance and trends

The trend and monthly fluctuations (expressed in standard deviation, SD) in global mean total water mass provides a test of internal consistency. Among the original GRACE TWS data, the GRG data showed the smallest temporal SD (0.04 mm) and linear trend (0.007 ± 0.001 SD mm $y^{-1}$) in global water mass. The three Tellus retrievals showed larger temporal SD (4.7–6.4 mm) and trends (-0.37 ± 0.21 to -0.23 ± 0.20 mm $y^{-1}$). The merged GRACE TWS data had intermediate SD (3.97 mm) and trend (-0.32 mm $y^{-1}$). Assimilation reduced SD (to 3.1 mm) and removed the residual trend (-0.01±0.10 mm $y^{-1}$). The discrepancies in global water mass trends in the merged GRACE data and in the analysis were mostly located over the oceans, and therefore the achieved mass balance closure can be attributed to the influence of the prior sea mass change estimates (Figure 4).

## 3.4. Regional storage trends

The spatial pattern in linear trends in the merged GRACE TWS ($y_0$) and the synthetic reanalysis signal ($y_b$) agree well (Figure 4bc), suggesting that the assimilation scheme is able to reconcile the prior estimates of storage changes and observed storage as intended. Seasonally adjusted anomalies were calculated for the prior and posterior estimates of the different water cycle components by subtracting the mean seasonal pattern. The 2003–2012 linear trends in these adjusted anomalies (Figure 5) show that the analysis has (*i*) increased spatial variability in sub-surface water storage trends, with amplified increasing and decreasing trends (Figure 5ab); (*ii*) drastically changes trends in snow and ice storage and typically made them more negative (Figure 5cd); (iii) reversed river water storage trends in the lower Amazon and Congo Rivers (Figure 5ef). The reanalysis shows a complex pattern of strongly decreasing and increasing sub-surface water storage trends in northwest India (Figure 5b). This may be an artefact from incorrectly specified errors in the groundwater depletion estimates (see Section 4.2). Less visible is that the analysis often reduced negative storage trends in other regions with groundwater depletion, that is, decreased the magnitude of estimated depletion. Because all sub-surface storage terms were combined, a revised estimate of groundwater depletion cannot calculated directly, but it can be estimated: for all grid cells with significant prior groundwater depletion estimates (>0.5 mm $y^{-1}$, representing 99% of total global groundwater depletion) the 2003–2012 trend in sub-surface storage change was estimated a priori at -168 ± 3 (SD) $km^3$ $y^{-1}$ of which 157 $km^3$ (94%) due to groundwater depletion and the remaining -11 $km^3$ due to climate variability. Analysis reduced the total trend for these grid cells to -103 ± 3 $km^3$ per year, from which a revised groundwater extraction estimate of ca. 92 $km^3$ can be derived.

From the seasonally adjusted anomalies, time series and trends of global storage in different water cycle components were calculated. We calculated snow and ice mass change separately for regions with seasonal snow cover, high (>55º) latitude glaciers, and remaining glaciers (Figure 6). The mean 2003–2012 trends are listed in Table 3; for the posterior estimates also as equivalent sea level rise (SLR, by dividing by the fraction of Earth's surface occupied by oceans, i.e., 0.7116) and volume ($km^3$ $y^{-1}$, equivalent to Gt $y^{-1}$). Some of the effects of the assimilation were to (*i*) remove the decreasing trend in prior global terrestrial sub-surface water storage estimates (Figure 6a), (*ii*) change the poor prior estimates of polar ice cap mass considerably (Figure 6fg), (*iii)* reduce the estimated rate of ocean mass increase from 1.84 ± 0.06 (SD) mm to 1.45±0.05 mm (Table 3), and (*iv*) achieve mass balance closure between net terrestrial and ocean storage changes (cf. Section 3.3).

### 3.5. Evaluation against river level remote sensing

The rank correlation ($\rho$) between river water level and estimated discharge for the 445 grid cells with altimetry time series are shown in Figure 7. Overall there was no significant change in agreement between the prior ($\rho$ = 0.63 ± 0.27 SD) and posterior ($\rho$ = 0.63 ± 0.26) estimates, with an average change of +0.01 ± 0.12. However, $\rho$ did improve for more locations than it deteriorated (286 vs. 159). There are some spatial patterns in the influence of assimilation (Figure 7c): strong improvements in the northern Amazon and Orinoco basins and most African rivers, except for some stations along the Congo and middle Nile Rivers, and reduced agreement for rivers in China (where prior estimates agreed well) and most stations in the Paraná and Uruguay basins (where they did not). In most remaining rivers, agreement did not change much; in some cases because it was already very good (e.g., the Ganges-Brahmaputra and remainder of the Amazon basin). Altimetry and estimated discharge time series are shown in Figure 8 for grid cells with the most data points in three large river systems. In these cases, there is reasonably clear improvement in agreement.

### 3.6. Evaluation against historic river discharge observations

The prior estimate of discharge (i.e., the error-weighted average of the four bias-corrected models) provided estimates that were already considerably better than any of the individual members (Table 4, Figure 9). Assimilation led to small improvements in RMSE, from 47 to

44 km$^3$ y$^{-1}$, and a very slight increase in the median absolute percentage difference, from 40 to 41%. Combined recorded discharge from the 430 selected basins was 20,909 km$^3$ y$^{-1}$, representing 90% of estimated total discharge to the world's oceans according to Dai et al. (2009). Assimilation improved the agreement with this number from -11% to -4%, of which about half (5%) is due to a closer estimate of Amazon River discharge. However, modelled and observed discharge values relate to different time periods and so it is not clear whether this should be considered evidence for improvement or merely reflects multi-annual variability.

### 3.7. Evaluation against snow water equivalent remote sensing

The spatial RMSE and correlation between the prior and posterior snow water equivalent (SWE) estimates and the GlobSnow retrievals are shown in Figure 10. Although RMSE deteriorated in a majority (57%) of grid cells, correlation remained unchanged at $R^2$=0.79 and average RMSE improved slightly from 23.2 to 22.3 mm. Assimilation appeared most successful for grid cells with large prior RMSE in northern Canada (Figure 10a-c).

### 3.8. Evaluation against glacier mass balance estimates

Glacier mass changes reported in literature (Gardner et al., 2013; Jacob et al., 2012) are listed in Table 5 and compared to regional mass trends associated with glaciers and other components of the terrestrial water derived from the analysis. In the polar regions (e.g., Antarctica, Greenland, Iceland, Svalbard, and the Russian Arctic) a large part of the gravity signal is necessarily from glacier mass change. Published trends for most of these regions also heavily rely on GRACE data and hence our estimates are generally in good agreement. Remaining differences can be attributed to the products, product versions and post-processing methods used, without providing insight into the accuracy of our analysis estimates. In the other regions, the glaciated areas are smaller and surrounded by ice-free terrain, which strongly increases the potential for incorrect distribution of analysis increments, as evidenced by the high trend ratios (>47%, last column Table 5). As a consequence, glacier mass trends are not well constrained by GRACE data alone and alternative observations are required. The agreement with independently derived trend estimates varies. For the Canadian Arctic Archipelago, Alaska and adjoining North America, the assimilation scheme assigns only 55% (68 Gt y$^{-1}$) of the total regional negative mass trend (-124 Gt y$^{-1}$) to glacier mass changes, with most of the remainder (40% or 50 Gt y$^{-1}$) assigned to sub-surface water storage changes.

Excluding regions for which independent storage change estimates are not available (Greenland, Antarctica and Patagonia), our estimate of total glacier storage change in the world's glaciers (-114 $km^3$ $y^{-1}$) was 101 $km^3$ $y^{-1}$ less than the estimate of *Gardner et al.* (2013) (-215 $km^3$ $y^{-1}$).

# 4. Discussion

## 4.1. Estimated errors

The triple collocation method produced estimates of errors in month-to-month changes in GRACE TWS estimates of 12.8–14.3 mm over non-glaciated land areas. From these, GRACE TWS errors of 10.4–12.0 mm can be estimated (cf. Section 3.1). By comparison, reported uncertainty estimates based on formal error propagation are larger, usually in the order of 20–25 mm (e.g., Landerer and Swenson, 2012; Tregoning et al., 2012; Wahr et al., 2006). One plausible explanation is that the 5 mm we assumed to correct for potential covariance in errors between the GRACE products is too low, another that the formal uncertainty estimates are too conservative. Inflating the GRACE error estimates by 10 mm instead of 5 mm reduced the gain by 18% on average. The resulting uncertainty in the analysis is modest (see next section). Formal error analyses predict that the retrieval errors decrease towards the poles due to the closer spacing of satellite overpasses (Wahr et al., 2006), but surprisingly we did not find such a latitudinal pattern.

The mean errors in monthly changes in prior TWS for the different models were 16.5–27.9 mm. We do not have independent estimates of errors in modelled large-scale TWS with which to compare, but the estimates would seem plausible and perhaps less than we anticipated. From a theoretical perspective, violation of the assumptions underpinning triple collocation is likely to have produced overestimates of model error, if anything. The calculated error in the prior estimates over oceans and very stable regions such as Mongolia and the Sahara are around 5 mm (Figure 2). This provides some further evidence to suggest that the 5 mm GRACE error inflation we applied may have been reasonable. The largest errors in the merged model estimates (>40 mm) were found for humid tropical regions and high latitudes. The former may be attributed to the combination of large storage variations and often uncertain rainfall estimates. Precipitation measurements are also fewer at high latitudes, and here the poor prediction of snow and ice dynamics and melt water river hydrology are also important factors.

## 4.2. Assimilation scheme performance

The spatial pattern in analysis increments emphasises the importance of water stores other than the soil in explaining discrepancies between model and GRACE TWS estimates (Figure 3). Adjustments to storage changes in large rivers, groundwater depletion, mass changes in high latitude ice caps and glaciers (e.g., Greenland, Alaska and Antarctica) and lake water levels (e.g., the Caspian Sea and the North-American Great Lakes) were all considerable within their region, absorbing monthly analysis increments or long-term trend discrepancies or both.

Uncertainty in error estimates for the different data sources affects the analysis in different ways. Incorrect estimation of GRACE and model-derived TWS errors by the triple collocation method primarily affects (*i*) the weighting of the ensemble members and (*ii*) the gain matrix. Appropriate weighting only requires that the relative magnitude of errors among ensemble members is estimated correctly (cf. Eq. (2)). The average errors for the different GRACE TWS estimates were all within 14% of the ensemble average (Table 2) and did not have strong spatial patterns, and therefore the analysis would likely have been very similar if equal weighting had been applied (cf. Sakumura et al., 2014). Estimated model errors showed greater differences (up to 52% greater than the ensemble mean, Table 2) as well as regional patterns. However, the relative rankings and their spatial pattern were robust to the choice of GRACE TWS members in triple collocation, as evidenced by a low coefficient of variation (Table 2). This suggests that the errors were correctly specified in a relative sense. For the gain matrix, the relative magnitude of errors in GRACE *versus* model TWS ensemble means needed to be estimated correctly (cf. Eq. (8)). The estimated GRACE TWS ensemble errors are reasonably homogeneous in space (Figure 1a) which increases our confidence in their validity. The uncertainty due to the correction for assumed correlation between the GRGS and Tellus TWS (see previous section) is further mitigated by the design of the DA scheme: the gain factor determines how rapidly the analysis converges towards the GRACE observations and therefore is important for month-to-month variations, but long-term trends in TWS will still approach those in the GRACE observations (cf. Figure 4b and c).

The main sources of uncertainty in long-term trends in the individual water balance terms are (*i*) the removal of non-hydrological mass trends in the GRACE TWS time series and (*ii*) accurate specification of relative errors in the individual water balance terms, which is needed for correct redistribution of the integrated TWS analysis increments. For example, the analysis results illustrate the insufficiently constrained problem of separating gravity signals due to mass changes in mountain glaciers from nearby sub-surface water storage changes. This was particularly evident around the Gulf of Alaska and northwest India, where decreases can be expected not only in glacier mass but also in sub-surface storage due to, respectively, a regional drying trend and high groundwater extraction rates (Figure 5a). We suspect that unexpectedly strong increasing storage trends in parts of northwest India are because the prior groundwater depletion estimates were too high and the assigned errors too low, causing the analysis update to distribute increments incorrectly. We could have addressed this by inflating the local groundwater depletion estimation errors, but more research is needed to understand the underlying causes. Plausible causes are that groundwater extraction is overestimated, or that extraction is compensated by induced groundwater recharge (e.g., from connected rivers) (see Wada et al., 2010 for further discussion).

Mass balance closure was not enforced and hence provides a useful diagnostic of reanalysis quality. The GRGS product achieved approximate global mass balance closure at all time scales, but the three Tellus products showed a seasonal cycle and long-term negative trend in global water mass. Accounting for atmospheric water vapour mass changes (from ERA-Interim reanalysis and the NVAP-M satellite product, data not shown) could not explain the trends and in fact increased the seasonal cycle in global water mass. Data assimilation reduced the seasonal cycle and entirely removed the trend in total water mass, thanks to the prior estimates of sea mass increase. For comparison, we calculated average ocean mass increases by an alternative, more conventional method, which involved avoiding areas likely to be affected by nearby land water storage changes. Excluding a 1000 km buffer zone produced a 2003–2012 mass trend of +0.58 to +0.72 mm $y^{-1}$ for the three Tellus retrievals, +1.12 mm $y^{-1}$ for the GRGS retrieval , and +0.75 mm $y^{-1}$ for the merged GRACE data. Data assimilation produced a stronger trend of +1.22 mm $y^{-1}$ due to the influence of the prior estimate of +1.67 mm $y^{-1}$. Our prior estimate followed Chen et al. (2013), who used an iterative modelling approach to attribute 75% of altimetry-observed SLR to mass increase. Chen et al. (2013) argue that the conventional method produces underestimates of ocean mass increase. Indeed, the trends we calculated for the ‘buffered’ ocean regions are lower than for the entire oceans (+1.22 vs. +1.45 mm $y^{-1}$ for the reanalysis, and +1.67 vs. +1.84 mm $y^{-1}$ for the prior estimates; Table 3). However the reduction in sea mass change of 0.39 mm $y^{-1}$ from prior to analysis is likely to reopen the problem of reconciling mass and temperature observations with the altimetry derived mean sea level rise of +2.45 ± 0.08 mm $y^{-1}$ (cf. Chen et al., 2013).

### 4.3. Evaluation against observations

The reanalysis generally did not have much impact on the agreement with river and snow storage observations, with small improvements for some locations and small degradations for others. While a robust increase in the agreement would have been desirable, the fact that agreement was not degraded overall was encouraging. The data assimilation procedure applied has the important benefit of bringing the estimates into agreement with GRACE observations. Moreover, performance improvements with respect to river discharge and level data did occur in the Amazon, where they make an important contribution to TWS changes. Similarly, snow water equivalent estimates were improved in the North-American Arctic, where errors in the prior estimates were largest. This demonstrates that GRACE data can indeed be successfully used to constrain water balance estimates, although further development may be needed to avoid some of the undesired performance degradation for water balance components that do not contribute much to the TWS signal.

The models used for our prior estimates provided poorly constrained estimates of ice mass balance changes, and our reanalysis ice mass loss estimates should not be assumed more accurate than estimates based on more direct methods (Table 5). Our analysis is unique when compared to previous estimates based on GRACE, in that data assimilation allowed some of the observed mass changes to be attributed to other water balance components within the same region, depending on relative uncertainties in the prior estimates. Comparison against independent estimates of glacier mass balance changes also demonstrated the challenge of correct attribution, however. Glacier mass balance estimates were in good agreement for several regions, but estimates for North American glaciers in particular were questionable: their combined mass loss (-68 Gt $y^{-1}$) was much lower than the estimates derived by independent means (-124 Gt $y^{-1}$; Table 5). This can be explained by incorrect specification of errors. Two caveats are made: *(i)* the GIA signal is relatively large for these three regions (+50 Gt $y^{-1}$) and hence GIA estimation errors may have had an impact; and *(ii)* a significant change in sub-surface water storage is plausible in principle; for example, higher summer temperatures could be expected to enhance permafrost melting and runoff, as well as enhance evaporation. More accurate spatiotemporal observation and modelling of glacier dynamics would appear to be necessary to resolve this issue.

### 4.4. Contributions to sea level rise

The reanalysis estimate of net terrestrial water storage change of -495 Gt $y^{-1}$ (Table 3) appears a plausible estimate of ocean mass change, equivalent to ca. +1.4 mm $y^{-1}$ sea level rise. Our results confirmed that mass loss from the polar ice caps is the greatest contributor to net terrestrial water loss, with Antarctica and Greenland together contributing -342 Gt $y^{-1}$. The next largest contribution was from the remaining glaciers. We combine the reanalysis estimate of -129 Gt $y^{-1}$ with another -101 Gt $y^{-1}$ estimated to be misattributed (cf. Section 3.8) and obtain a revised estimate of -230 Gt $y^{-1}$. A small but significant contribution of -18 Gt $y^{-1}$ (Table 3) was estimated to originate from reductions in seasonal snow cover (particularly in Quebec and Siberia; Figure 5cd). Inter-annual changes in river water storage were not significant. Small contributions of -10 Gt $y^{-1}$ and +16 Gt $y^{-1}$ were attributed to storage changes in existing lakes and large new dams, respectively, and compensated each other. The largest change in an individual water body was in the Caspian Sea (-27 Gt $y^{-1}$, cf. Figure 5) which experiences strong multi-annual water storage variations depending on Volga River inflows.

Finally, the analysis suggested at statistically insignificant change of +9 Gt $y^{-1}$ in sub-surface storage globally. Adding back the suspected misattribution of 101 Gt $y^{-1}$ associated with glaciers produces a revised estimate of +110 Gt $y^{-1}$ (cf. Figure 6a). Combining this with the -92 Gt $y^{-1}$ attributed to groundwater depletion suggests that storage over the remaining land areas increased by 202 Gt $y^{-1}$. Calculating sub-surface storage trends by latitude band suggests that most of the terrestrial water ‘sink’ can be found north of 40°N and between 0–30°S and is opposite to the prior estimates (Figure 11). The main tropical regions experiencing increases are in the Okavango and upper Zambezi basins in southern Africa and the Amazon and Orinoco basins in northern South America (Figure 5b). Storage increases for these regions are also evident from the original GRACE data (Figure 4a) and cannot be attributed to storage changes in rivers or large lakes. The affected regions contain low relief, poorly drained areas with (seasonally) high rainfall. In such environments, the storage changes could occur in the soil, groundwater, wetlands, or a combination of these. Further attribution is impossible without additional constraining observations (Tregoning et al., 2012; van Dijk et al., 2011). The ten-year analysis period is short and this cautions against over-interpreting this apparent ‘tropical water sink’. However it is of interest to note that a gradual strengthening of global monsoon rainfall extent and intensity has been observed, and is predicted to continue (Hsu et al., 2012). In any event, the difference between prior and posterior trends in Figure 11 illustrates that the current generation hydrological models, even as an ensemble, should not be assumed a reliable surrogate observation of long-term sub-surface groundwater storage changes. GRACE observations proved valuable in improving these estimates.

## 5. Conclusions

We presented a global water cycle reanalysis that reconciles four total water storage retrieval products derived from GRACE observations with water balance estimates derived from an ensemble of five global hydrological models, water level measurements from satellite altimetry, and ancillary data. We summarise our main findings as follows:

1. The data assimilation scheme generally behaves as desired, but in hydrologically complex regions the analysis can be affected by poorly constrained prior estimates and error specification. The greatest uncertainties occur in regions where glacier mass loss and sub-surface storage declines (may) both occur but are poorly known (e.g., northern India and North-American glaciers).
2. The error in original GRACE TWS data was estimated to be around 11–12 mm over non-glaciated land areas. Errors in the prior estimates of TWS changes are estimated to be 17–28 mm for the five models.
3. Water storage changes in other water cycle components (seasonal snow, ice, lakes and rivers) are often at least as important and uncertain as changes as sub-surface water storage in reconciling the various information sources.
4. The analysis results were compared to independent river water level measurements by satellite altimetry, river discharge records, remotely sensed snow water storage, and independent estimates of glacier mass loss. In all cases the agreement improved or remained stable compared to the prior estimates, although results varied regionally. Better estimates and error specification of groundwater depletion and mountain glacier mass loss are required.
5. Data assimilation achieved mass balance closure over the 2003–2012 period and suggested an ocean mass increase of ca. 1.45 mm $y^{-1}$. This reopens some question about the reasons for an apparently unexplained 0.39 mm $y^{-1}$ (16%) of 2.45 mm $y^{-1}$ satellite observed sea level rise for the analysis period (Chen et al., 2013).

6. For the period 2003–2012, we estimate glaciers and polar ice caps to have lost around 572 Gt $y^{-1}$, with an additional small contribution from seasonal snow (-18 Gt $y^{-1}$). The net change in surface water storage in large lakes and rivers was insignificant, with compensating effects from new reservoir impoundments (+16 Gt $y^{-1}$), lowering water level in the Caspian Sea (-27 Gt $y^{-1}$) and increases in the other lakes combined (+16 Gt $y^{-1}$). The net change in subsurface storage was significant when considering a likely misattribution of glacier mass loss, and may be as high as +202 Gt $y^{-1}$ when excluding groundwater depletion (-92 Gt $y^{-1}$). Increases were mainly in northern temperate regions and in the seasonally wet tropics of South America and southern Africa (+87 Gt $y^{-1}$). Continued observation will help determine if these trends are due to transient climate variability or likely to persist.

## Acknowledgements

GRACE land data were processed by Sean Swenson and the ocean data by Don P. Chambers, both supported by the NASA MEASURES Program, and are available at http://grace.jpl.nasa.gov. The GLDAS and TMPA data used in this study were acquired as part of the mission of NASA's Earth Science Division and archived and distributed by the Goddard Earth Sciences (GES) Data and Information Services Center (DISC). Lake and reservoir surface height variations were from the USDA’s Global Reservoir and Lake (GRLM) web site /www.pecad.fas.usda.gov/cropexplorer/global_reservoir/, funded by USDA/FAS/OGA and NASA Global Agriculture Monitoring (GLAM) Project. Altimetric lake level time-series variations were from the Topex/Poseidon, Jason-1, Jason-2/OSTM, and Geosat Follow-On (GFO) missions.

## References

Boening, C., Willis, J. K., Landerer, F. W., Nerem, R. S., and Fasullo, J.: The 2011 La Niña: So strong, the oceans fell, Geophysical Research Letters, 39, L19602, 10.1029/2012gl053055, 2012.

Bouttier, F., and Courtier, P.: Data assimilation concepts and methods, ECMWF Meteorological Training Course Lecture Series, 14, 1999.

Bruinsma, S., Lemoine, J.-M., Biancale, R., and Valès, N.: CNES/GRGS 10-day gravity field models (release 2) and their evaluation, Advances in Space Research, 45, 587-601, doi: 10.1016/j.asr.2009.10.012, 2010.

Cazenave, A., Dominh, K., Guinehut, S., Berthier, E., Llovel, W., Ramillien, G., Ablain, M., and Larnicol, G.: Sea level budget over 2003-2008: A reevaluation from GRACE space gravimetry, satellite altimetry and Argo, Global and Planetary Change, 65, 83-88, 2009.

Chambers, D. P., and Bonin, J. A.: Evaluation of Release-05 GRACE time-variable gravity coefficients over the ocean, Ocean Sci., 8, 859-868, 10.5194/os-8-859-2012, 2012.

Chen, J. L., Wilson, C. R., and Tapley, B. D.: Contribution of ice sheet and mountain glacier melt to recent sea level rise, Nature Geosci, 6, 549-552, 10.1038/ngeo1829, 2013.

Cogley, J. G.: GGHYDRO-Global Hydrographic Data, release 2.3, Technical Note 2003-1, Dept. of Geographty, Trent University, Peterborough, Ontario, Canada, 2003.

Crétaux, J. F., Jelinski, W., Calmant, S., Kouraev, A., Vuglinski, V., Bergé-Nguyen, M., Gennero, M. C., Nino, F., Abarca Del Rio, R., Cazenave, A., and Maisongrande, P.: SOLS: A lake database to monitor in the Near Real Time water level and storage variations from remote sensing data, Advances in Space Research, 47, 1497-1507, doi: 10.1016/j.asr.2011.01.004, 2011.

Dai, A., Qian, T., Trenberth, K. E., and Milliman, J. D.: Changes in continental freshwater discharge from 1948 to 2004, Journal of Climate, 22, 2773-2792, 2009.

Dorigo, W. A., Scipal, K., Parinussa, R. M., Liu, Y. Y., Wagner, W., De Jeu, R. A. M., and Naeimi, V.: Error characterisation of global active and passive microwave soil moisture datasets, Hydrol. Earth Syst. Sci, 14, 2605-2616, 2010.

Fang, H., Wei, S., Jiang, C., and Scipal, K.: Theoretical uncertainty analysis of global MODIS, CYCLOPES, and GLOBCARBON LAI products using a triple collocation method, Remote Sensing of Environment, 124, 610-621, doi: 10.1016/j.rse.2012.06.013, 2012.

Fasullo, J. T., Boening, C., Landerer, F. W., and Nerem, R. S.: Australia's unique influence on global sea level in 2010–2011, Geophysical Research Letters, 40, 4368-4373, 10.1002/grl.50834, 2013.

Gardner, A. S., Moholdt, G., Cogley, J. G., Wouters, B., Arendt, A. A., Wahr, J., Berthier, E., Hock, R., Pfeffer, W. T., Kaser, G., Ligtenberg, S. R. M., Bolch, T., Sharp, M. J., Hagen, J.

O., van den Broeke, M. R., and Paul, F.: A Reconciled Estimate of Glacier Contributions to Sea Level Rise: 2003 to 2009, Science, 340, 852-857, 10.1126/science.1234532, 2013.

Geruo, A., Wahr, J., and Zhong, S.: Computations of the viscoelastic response of a 3-D compressible Earth to surface loading: an application to Glacial Isostatic Adjustment in Antarctica and Canada, Geophysical Journal International, 192, 557-572, 2013.

Gong, L., Halldin, S., and Xu, C. Y.: Global-scale river routing—an efficient time-delay algorithm based on HydroSHEDS high-resolution hydrography, Hydrological Processes, 25, 1114-1128, 10.1002/hyp.7795, 2011.

Hsu, P.-c., Li, T., Luo, J.-J., Murakami, H., Kitoh, A., and Zhao, M.: Increase of global monsoon area and precipitation under global warming: A robust signal?, Geophysical Research Letters, 39, L06701, 10.1029/2012GL051037, 2012.

Huffman, G. J., Adler, R. F., Bolvin, D. T., Gu, G. J., Nelkin, E. J., Bowman, K. P., Hong, Y., Stocker, E. F., and Wolff, D. B.: The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, Journal of Hydrometeorology, 8, 38-55, 2007.

Jacob, T., Wahr, J., Pfeffer, W. T., and Swenson, S.: Recent contributions of glaciers and ice caps to sea level rise, Nature, 482, 514-518, 2012.

Jekeli, C.: Alternative Methods to Smooth the Earth's Gravity Field. Report 327, Dep. of Geod. Sci. and Surv., Ohio State Univ., Columbus, Ohio, 1981.

Landerer, F. W., and Swenson, S. C.: Accuracy of scaled GRACE terrestrial water storage estimates, Water Resources Research, 48, W04531, 10.1029/2011WR011453, 2012.

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N., and Wisser, D.: High-resolution mapping of the w’rld's reservoirs and dams for sustainable river-flow management, Frontiers in Ecology and the Environment, 9, 494-502, 10.1890/100125, 2011.

Leuliette, E. W., and Miller, L.: Closing the sea level rise budget with altimetry, Argo, and GRACE, Geophys. Res. Lett., 36, L04608, 10.1029/2008gl036010, 2009.

Liu, Weerts, A. H., Clark, M., Hendricks Franssen, H. J., Kumar, S., Moradkhani, H., Seo, D. J., Schwanenberg, D., Smith, P., van Dijk, A. I. J. M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O., and Restrepo, P.: Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, Hydrol. Earth Syst. Sci., 16, 3863-3887, 10.5194/hess-16-3863-2012, 2012a.

Liu, Y. Y., Parinussa, R. M., Dorigo, W. A., De Jeu, R. A. M., Wagner, W., van Dijk, A., McCabe, M. F., and Evans, J. P.: Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals, Hydrol. Earth Syst. Sci, 15, 425-436, 2011.

Liu, Y. Y., Dorigo, W., Parinussa, R., De Jeu, R., Wagner, W., McCabe, M., Evans, J., and Van Dijk, A.: Trend-preserving blending of passive and active microwave soil moisture retrievals, Remote Sensing of Environment, 123, 280-297, 2012b.

Luojus, K., Pulliainen, J., Takala, M., Derksen, C., Rott, H., Nagler, T., Solberg, R., Wiesmann, A., Metsamaki, S., Malnes, E., and Bojkov, B.: Investigating the feasibility of the globsnow snow water equivalent data for climate research purposes, Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International, 2010,

Miralles, D. G., De Jeu, R. A. M., Gash, J. H., Holmes, T. R. H., and Dolman, A. J.: Magnitude and variability of land evaporation and its components at the global scale, Hydrol. Earth Syst. Sci., 15, 967-981, 10.5194/hess-15-967-2011, 2011.

Oki, T., and Sud, Y. C.: Design of Total Runoff Integrating Pathways (TRIP)-A global river channel network, Earth interactions, 2, 1-37, 1998.

Oki, T., Nishimura, T., and Dirmeyer, P. A.: Assessment of Annual Runoff from Land Surface Models Using Total Runoff Integrating Pathways (TRIP), J Meteorol, 77, 235-255, 1999.

Peña-Arancibia, J., Van Dijk, A. I. J. M., Mulligan, M., and Renzullo, L. J.: Evaluation of precipitation estimation accuracy in reanalyses, satellite products and an ensemble method for regions in Australia and in south and east Asia, Journal of Hydrometeorology, accepted 29 January 2013, 2013.

Pulliainen, J.: Mapping of snow water equivalent and snow depth in boreal and sub-arctic zones by assimilating space-borne microwave radiometer data and ground-based observations, Remote Sensing of Environment, 101, 257-269, doi: 10.1016/j.rse.2006.01.002, 2006.

Rodell, M., Houser, P., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C., Arsenault, K., Cosgrove, B., Radakovich, J., and Bosilovich, M.: The global land data assimilation system, Bulletin American Meteorological Society, 85, 381-394, 2004.

Rui, H.: README Document for Global Land Data Assimilation System Version 1 (GLDAS-1) Products, NASA, 2011.

Sakumura, C., Bettadpur, S., and Bruinsma, S.: Ensemble prediction and intercomparison analysis of GRACE time-variable gravity field models, Geophysical Research Letters, 41, 1389-1397, 10.1002/2013GL058632, 2014.

Scipal, K., Holmes, T., de Jeu, R., Naeimi, V., and Wagner, W.: A possible solution for the problem of estimating the error structure of global soil moisture data sets, Geophysical Research Letters, 35, 2009.

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling, Journal of Climate, 19, 3088-3111, 2006.

Sheffield, J., Wood, E. F., and Roderick, M. L.: Little change in global drought over the past 60 years, Nature, 491, 435-438, 2012.

Stoffelen, A.: Toward the true near-surface wind speed: Error modeling and calibration using triple collocation, Journal of Geophysical Research: Oceans, 103, 7755-7766, 10.1029/97jc03180, 1998.

Swenson, S., and Wahr, J.: Post-processing removal of correlated errors in GRACE data, Geophys. Res. Lett., 33, L08402, 10.1029/2005gl025285, 2006.

Swenson, S., Famiglietti, J., Basara, J., and Wahr, J.: Estimating profile soil moisture and groundwater variations using GRACE and Oklahoma Mesonet soil moisture data, Water Resour. Res., 44, W01413, 10.1029/2007wr006057, 2008.

Takala, M., Pulliainen, J., Metsamaki, S. J., and Koskinen, J. T.: Detection of Snowmelt Using Spaceborne Microwave Radiometer Data in Eurasia From 1979 to 2007, Geoscience and Remote Sensing, IEEE Transactions on, 47, 2996-3007, 10.1109/TGRS.2009.2018442, 2009.

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., and Watkins, M. M.: GRACE Measurements of Mass Variability in the Earth System, Science, 305, 503-505, 10.1126/science.1099192, 2004.

Tregoning, P., McClusky, S., van Dijk, A., Crosbie, R. S., and Peña-Arancibia, J. L.: Assessment of GRACE satellites for groundwater estimation in Australia, National Water Commission, Caberra, 82, 2012.

van Dijk, A. I. J. M., and Renzullo, L. J.: Water resource monitoring systems and the role of satellite observations, Hydrology and Earth System Sciences, 15, 39-55, 10.5194/hess-15-39-2011, 2011.

van Dijk, A. I. J. M., Renzullo, L. J., and Rodell, M.: Use of Gravity Recovery and Climate Experiment terrestrial water storage retrievals to evaluate model estimates by the Australian water resources assessment system, Water Resources Research, 47, W11524., 10.1029/2011WR010714, 2011.

Van Dijk, A. I. J. M., Peña-Arancibia, J. L., Wood, E. F., Sheffield, J., and Beck, H. E.: Global analysis of seasonal streamflow predictability using an ensemble prediction system and observations from 6192 small catchments worldwide, Water Resources Research, DOI: 10.1002/wrcr.20251, 10.1002/wrcr.20251, 2013.

Vörösmarty, C. J., and Moore III, B. I.: Modeling basin-scale hydrology in support of physical climate and global biogeochemical studies: An example using the Zambezi River, Surveys in Geophysics, 12, 271-311, 10.1007/bf01903422, 1991.

Wada, Y., van Beek, L. P. H., van Kempen, C. M., Reckman, J. W. T. M., Vasak, S., and Bierkens, M. F. P.: Global depletion of groundwater resources, Geophysical Research Letters, 37, L20402, 10.1029/2010gl044571, 2010.

Wada, Y., van Beek, L. P. H., Sperna Weiland, F. C., Chao, B. F., Wu, Y.-H., and Bierkens, M. F. P.: Past and future contribution of global groundwater depletion to sea-level rise, Geophysical Research Letters, 39, L09402, 10.1029/2012GL051230, 2012.

Wada, Y., Van Beek, R., Wanders, N., and Bierkens, M. F. P.: Human water consumption intensifies hydrological drought worldwide, Environmental Research Letters, 8, 034036, 2013.

Wahr, J., Swenson, S., and Velicogna, I.: Accuracy of GRACE mass estimates, Geophysical Research Letters, 33, L06401, 10.1029/2005GL025305, 2006.

Wang, X., de Linage, C., Famiglietti, J., and Zender, C. S.: Gravity Recovery and Climate Experiment (GRACE) detection of water storage changes in the Three Gorges Reservoir of China and comparison with in situ measurements, Water Resources Research, 47, 2011.

Zaitchik, B. F., Rodell, M., and Reichle, R. H.: Assimilation of GRACE Terrestrial Water Storage Data into a Land Surface Model: Results for the Mississippi River Basin, Journal of Hydrometeorology, 9, 535-548, doi:10.1175/2007JHM951.1, 2008.

Zwieback, S., Scipal, K., Dorigo, W., and Wagner, W.: Structural and statistical properties of the collocation technique for error characterization, Nonlinear Processes in Geophysics, 19, 69-80, 2012.

Table 1. Description and sources of data used in this analysis. Acronyms are explained in the text.

| Description | Source | Data access |
|---|---|---|
| ***Prior estimates*** | | |
| model estimates (CLM, MOS, NOAH, VIC) | GLDAS | ftp://hydro1.sci.gsfc.nasa.gov/data/s4pa/GLDAS_V1/ (data accessed 17 April 2013). |
| Model estimates (W3RA) | | available from author Van Dijk |
| groundwater depletion | | available from author Wada |
| river flow direction | TRIP | http://hydro.iis.u-tokyo.ac.jp/~taikan/TRIPDATA/Data/trip05.asc (downloaded 10 May 2013) |
| discharge from small catchments | | available from author Van Dijk |
| discharge from large basins | | http://www.cgd.ucar.edu/cas/catalog/surface/dai-runoff/index.html |
| surface water extraction | | available from author Wada |
| lake water level | Crop Explorer | http://www.pecad.fas.usda.gov/cropexplorer/global_reservoir/ (downloaded 9 May 2013) |
| new dam impoundments | GranD | http://atlas.gwsp.org/ (accessed 14 May 2014) |
| new dam impoundments | ICOLD | http://www.icold-cigb.org/ (accessed 14 May 2014) |
| sea level | AVISO | http://www.aviso.oceanobs.com/en/data/products/sea-surface-height-products/global/ (downloaded 7 November 2013) |
| glacier extent | GGHYDRO | http://people.trentu.ca/~gcogley/glaciology/ (downloaded 12 June 2013) |
| ***Assimilated data*** | | |
| TWS: CSR, GFZ, JPL | Tellus | ftp://podaac-ftp.jpl.nasa.gov/allData/tellus/L3/land_mass/RL05/netcdf/ (downloaded 16 April 2013) |
| TWS: GRGS | CNES | http://grgs.obs-mip.fr/grace/variable-models-grace-lageos/grace-solutions-release-02 (downloaded 16 April 2013) |
| glacial isostatic adjustment | Tellus | ftp://podaac-ftp.jpl.nasa.gov/allData/tellus/L3/land_mass/RL05/netcdf/ (downloaded 16 April 2013) |
| ***Evaluation data*** | | |
| water level in large rivers | LEGOS HYDROWEB | http://www.legos.obs-mip.fr/en/soa/hydrologie/hydroweb/ (downloaded 13 October 2013) |
| *idem* | ESA River&Lake | http://tethys.eaprs.cse.dmu.ac.uk/RiverLake/shared/main (downloaded 25 October 2012) |
| snow depth | GLOBSNOW | http://www.globsnow.info/swe/archive_v1.3/ (downloaded 9 October 2013) |

Table 2. Spatial mean values (non-glaciated land areas only) of the error in monthly mass change estimates for different GRACE and model sources as derived through triple collocation. Also listed is the number of triple collocation estimates derived (*N*) and the spatial mean of the coefficient of variation (C.V.) in these *N* estimates.

| | Mean error mm | Mean C.V. % | N |
|---|---|---|---|
| *GRACE* | | | |
| GRG | 14.3 | 15 | 15 |
| CSR | 12.8 | 15 | 5 |
| GFZ | 15.5 | 11 | 5 |
| JPL | 15.2 | 12 | 5 |
| Merged | 13.5 | – | – |
| *Models* | | | |
| CLM | 26.7 | 6 | 3 |
| MOS | 21.9 | 7 | 3 |
| NOAH | 16.6 | 9 | 3 |
| VIC | 27.7 | 6 | 3 |
| W3RA | 17.9 | 7 | 3 |
| Merged | 18.1 | – | – |

Table 3. Calculated linear trends in global mean seasonally-adjusted anomalies associated with different water cycle components for 2003–2012. The posterior trend estimates are also expressed in equivalent sea level rise (SLR) and volume. Second number is standard deviation.

| Store | Prior global mean mm $y^{-1}$ | Posterior global mean mm $y^{-1}$ | SLR mm $y^{-1}$ | Volume $km^3$ $y^{-1}$ |
|---|---|---|---|---|
| Sub-surface | -0.572 ± 0.029 | 0.017 ± 0.023 | 0.024 ± 0.032 | 9 ± 12 |
| Rivers | 0.012 ± 0.009 | 0.003 ± 0.01 | 0.004 ± 0.014 | 1 ± 5 |
| Lakes | -0.012 ± 0.005 | -0.021 ± 0.005 | -0.029 ± 0.006 | -11 ± 2 |
| New dams | 0.043 ± 0.001 | 0.032 ± 0.002 | 0.045 ± 0.003 | 16 ± 1 |
| Seasonal snow | -0.022 ± 0.007 | -0.035 ± 0.007 | -0.049 ± 0.01 | -18 ± 4 |
| Arctic glaciers (>55°N) | 0.265 ± 0.004 | -0.604 ± 0.009 | -0.849 ± 0.013 | -308 ± 5 |
| Antarctic glaciers (>55°S) | - | -0.301 ± 0.007 | -0.423 ± 0.01 | -154 ± 4 |
| Remaining glaciers | -0.029 ± 0.004 | -0.061 ± 0.003 | -0.086 ± 0.004 | -31 ± 2 |
| Total terrestrial | - | -0.97 ± 0.035 | -1.364 ± 0.049 | -495 ± 18 |
| Oceans | 1.309 ± 0.044 | 1.029 ± 0.039 | 1.446 ± 0.054 | 525 ± 20 |

Table 4. Evaluation of alternative estimates of mean basin discharge using observations collated by Dai et al. (2009). Listed is the agreement for the ensemble models (without bias correction), the merged prior estimate and the posterior estimates resulting from reanalysis.

| | CLM | MOS | NOAH | VIC | W3RA | prior | posterior |
|---|---|---|---|---|---|---|---|
| Combined discharge ($km^3$ $y^{-1}$) | 21,874 | 9,003 | 11,474 | 13,666 | 16,518 | 18,663 | 20,149 |
| Diff. total (%) | 5 | -57 | -45 | -35 | -21 | -11 | -4 |
| RMSE ($km^3$ $y^{-1}$) | 114 | 184 | 126 | 147 | 63 | 47 | 44 |
| Median \|%\| diff. | 60 | 63 | 57 | 48 | 61 | 40 | 41 |

Table 5. Published trends in glacier water storage (Gardner et al., 2013; Jacob et al., 2012) compared to estimates from reanalysis. Uncertainties are given at the 95% (2 standard deviation) interval, superscripts refer to estimates derived from GRACE (g) or independent methods (i). Also listed are regional trends attributed to other parts of the hydrological cycle, and the ratio of the relative magnitude of that residual trends over estimated glacier mass change.

| Region | Reported trend (Gt y$^{-1}$) | | This study glacier trend (Gt y$^{-1}$) | other components (Gt y$^{-1}$) | ratio (%) |
|---|---|---|---|---|---|
| Greenland ice sheet + PGICs | -222 ± 9 | g | -203 ± 10 | -5 ± 1 | 3 |
| Canadian Arctic Archipelago | -60 ± 6 | i,g | -48 ± 3 | -19 ± 2 | 39 |
| Alaska | -50 ± 17 | i,g | -23 ± 6 | -23 ± 6 | 101 |
| Northwest America excl. Alaska | -14 ± 3 | i | 3 ± 3 | -8 ± 9 | 275 |
| Iceland | -10 ± 2 | i,g | -6 ± 1 | -0.6 ± 0.2 | 10 |
| Svalbard | -5 ± 2 | i,g | -2 ± 1 | 0.1 ± 0.1 | 3 |
| Scandinavia | -2 ± 0 | i | 0.4 ± 1.0 | 5 ± 2 | >500 |
| Russian Arctic | -11 ± 4 | i,g | -4 ± 1 | 2 ± 2 | 47 |
| High Mountain Asia | -26 ± 12 | i,g | -29 ± 4 | -15 ± 11 | 51 |
| South America excl. Patagonia | -4 ± 1 | i | -2 ± 1 | -21 ± 33 | >500 |
| Patagonia | -29 ± 10 | g | -15 ± 1 | 1 ± 2 | 4 |
| Antarctica ice sheet + PGICs | -165 ± 72 | g | -139 ± 8 | 0 | 0 |
| Rest of world | -4 ± 0 | | -3 ± 1 | 82 ± 107 | >500 |
| Total | -549 ± 57 | | -471 ± 25 | | |

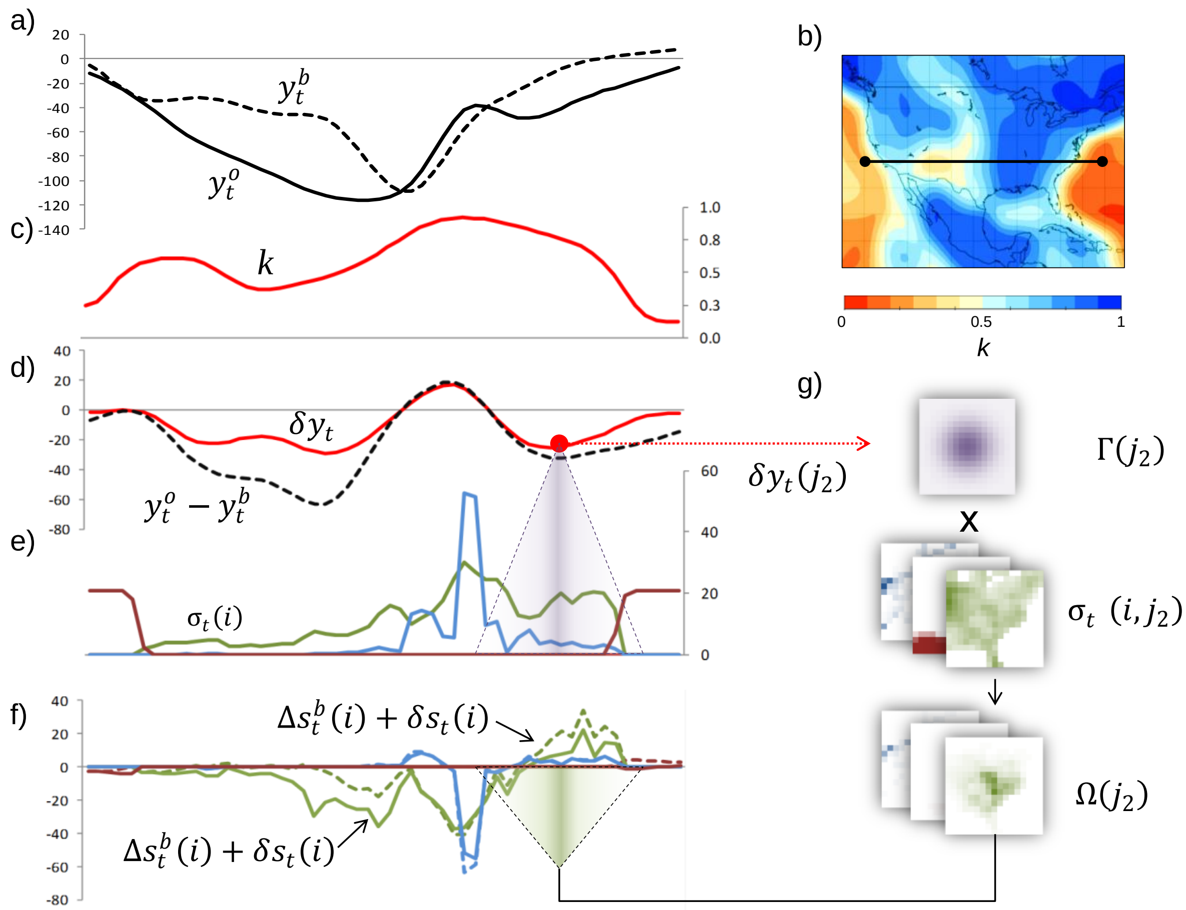



Figure 1. Illustration of the data assimilation approach followed using data along a transect through the USA for August 2003. Shown are: a) monthly satellite-derived TWS, $y_t^o$, and the equivalent prior estimate, $y_t^b$; b) location of the West-East transect on a map of the gain matrix, *k*; c) profile of *k* along the transect (cf. Figure 2c); d) calculation of the TWS analysis increment, $\delta y_t$ , from *k* and innovation, $(y_t^o - y_t^b)$; e) the prior error in the change of each of the stores, $\sigma_t(i)$; f) the prior and posterior estimate of change in each store, $\Delta s_t^b(i)$ and $\Delta s_t^b(i) + \delta s_t(i)$, resp.; and g) visual illustration of the disaggregation of the TWS analysis increments to the different stores. All units are in mm unless indicated otherwise; see text for full explanation of symbols; stores shown include the sub-surface (green), rivers (blue) and sea (dark red; remaining stores not shown for clarity).

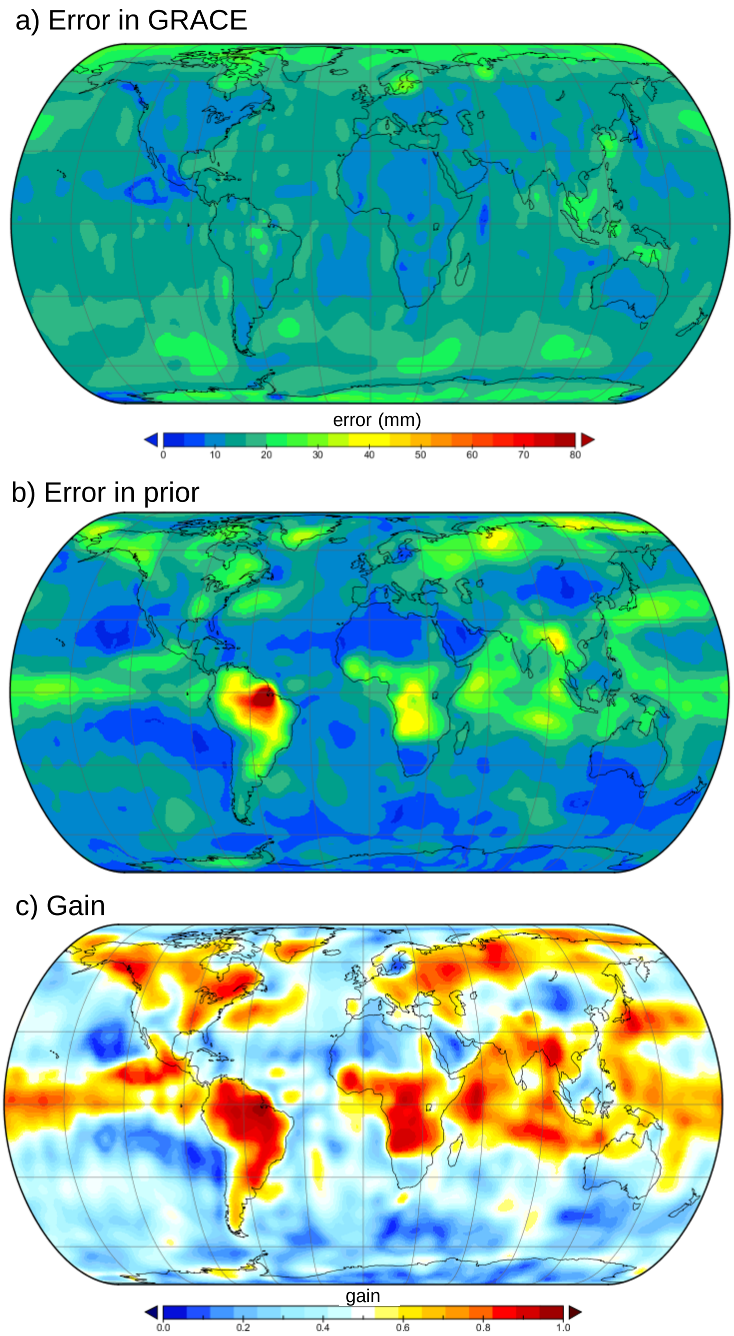



Figure 2. Triple collocation estimated error in storage change from the merged (a) GRACE and (b) prior estimates, and (c) resulting gain matrix.

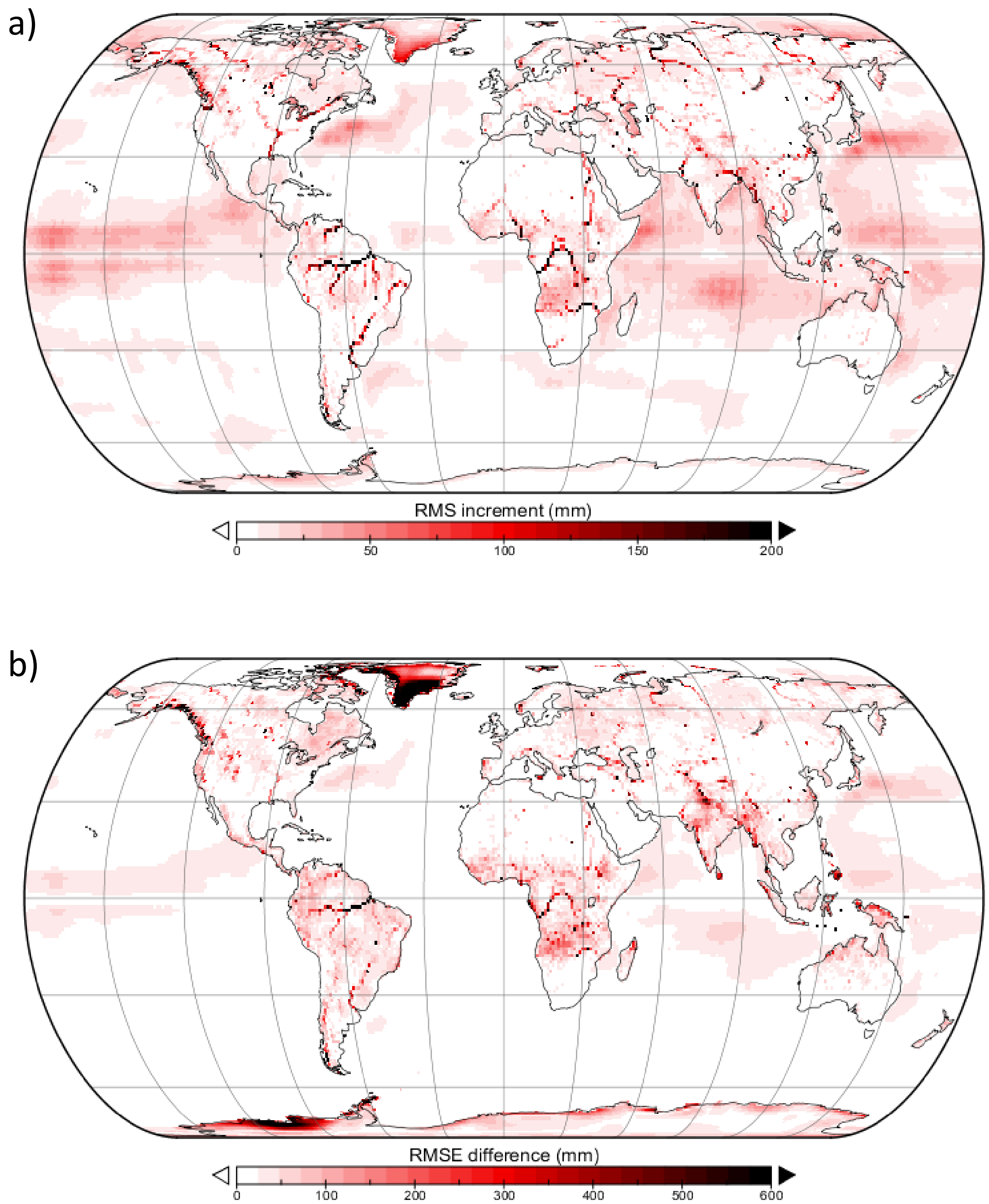
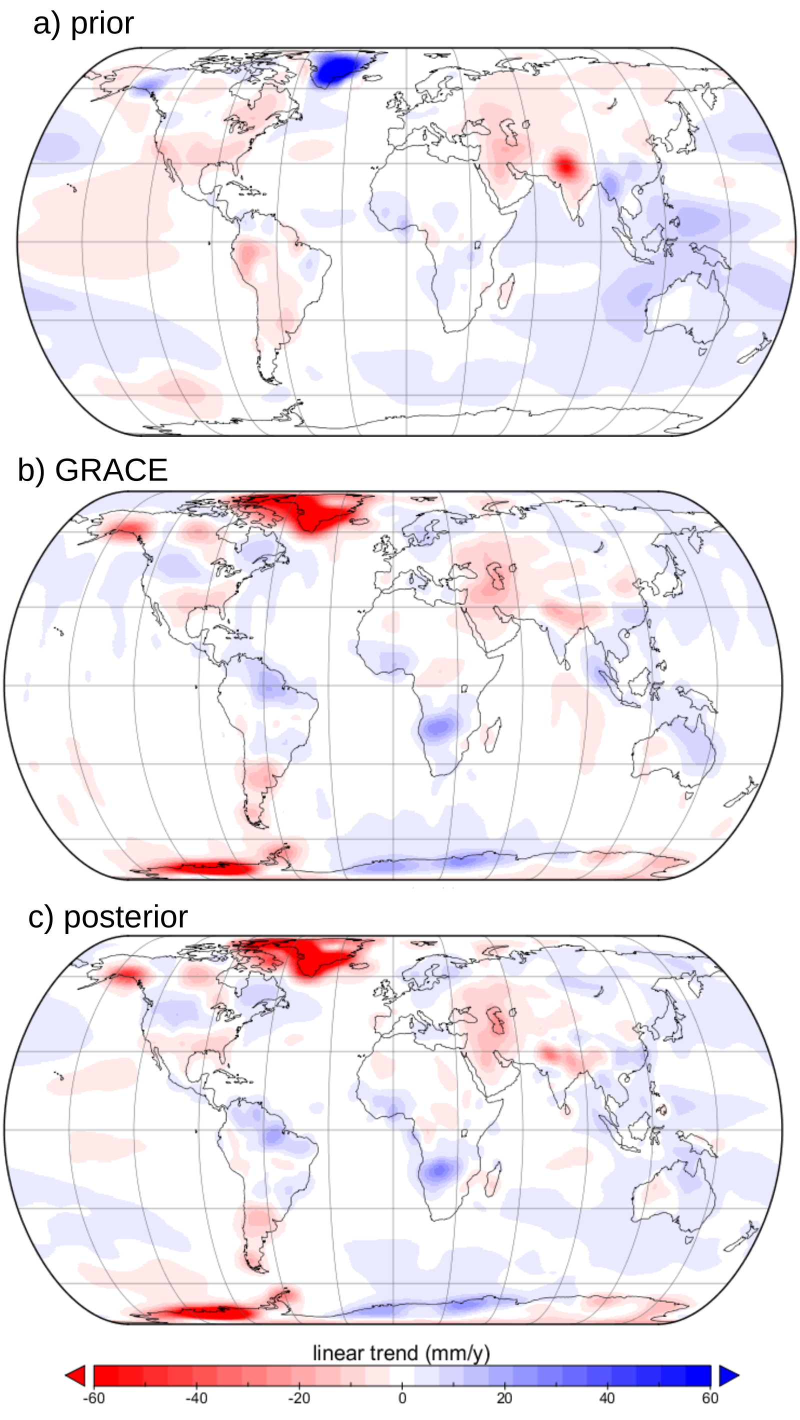



Figure 3. The impact of GRACE data assimilation on total water storage expressed as (a) the root mean square (RMS) analysis increment and (b) the RMS difference between prior and posterior storage time series.




952

953 Figure 4. Trends in GRACE total water storage as derived from (a) prior storage estimates;
954 (b) merged satellite retrievals; and (c) posterior estimates.

955

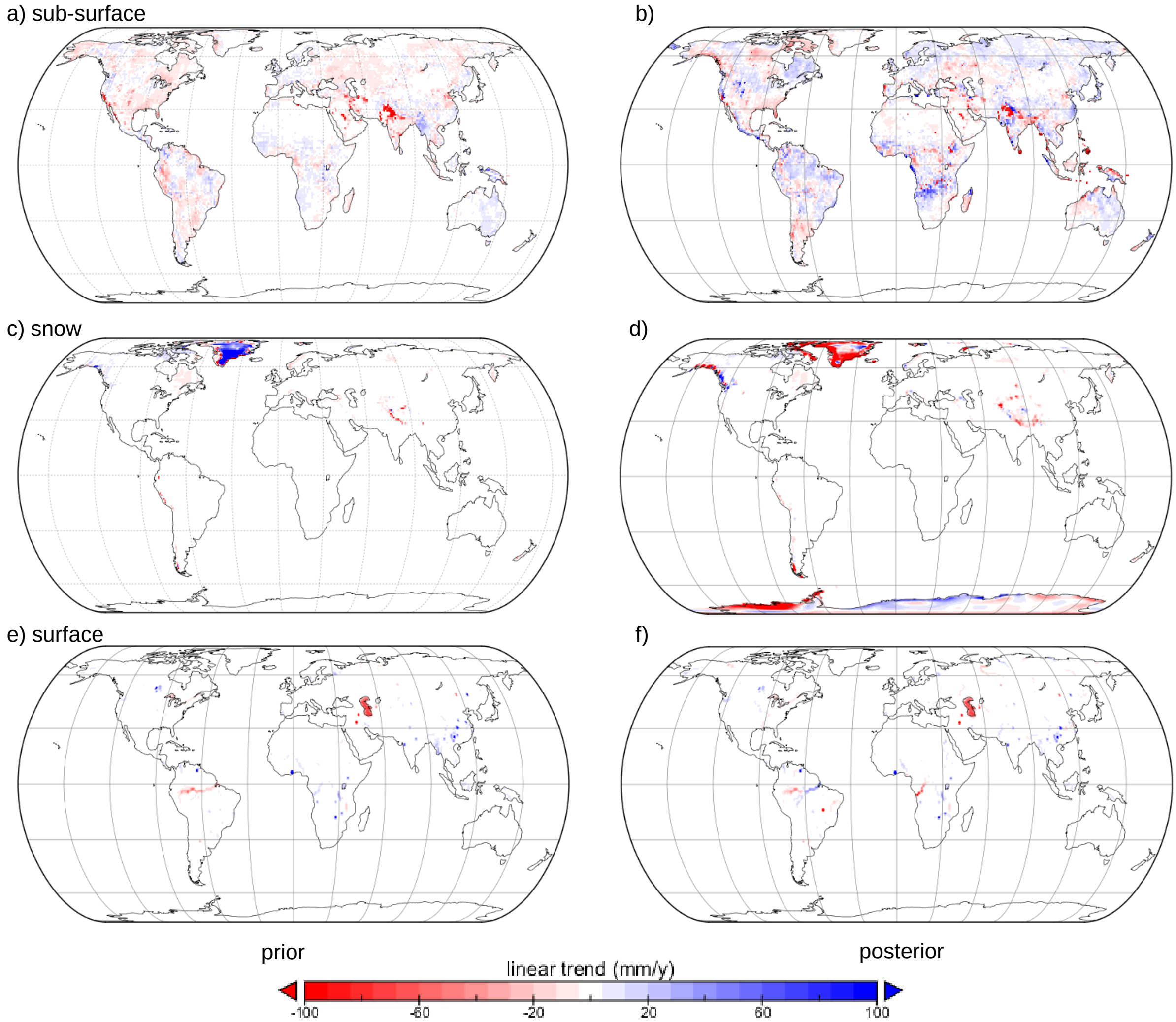



Figure 5. Trends in seasonal anomalies of prior (left column) and posterior (right column) estimates of (a-b) sub-surface, (c-d) snow and (e-f) surface water (i.e., lake and river) water storage.

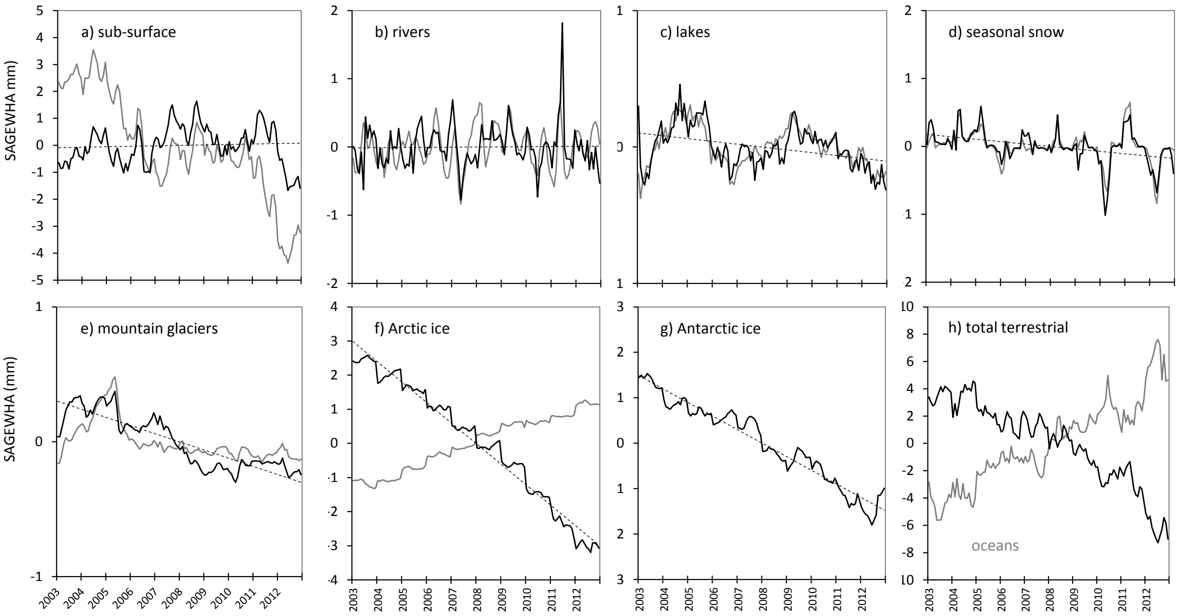



Figure 6. Time series of the prior (grey lines) and posterior (black lines) estimates of global average seasonally-adjusted storage anomalies in different water cycle components. Dashed lines show linear trends for 2003–2012 as listed in Table 3.

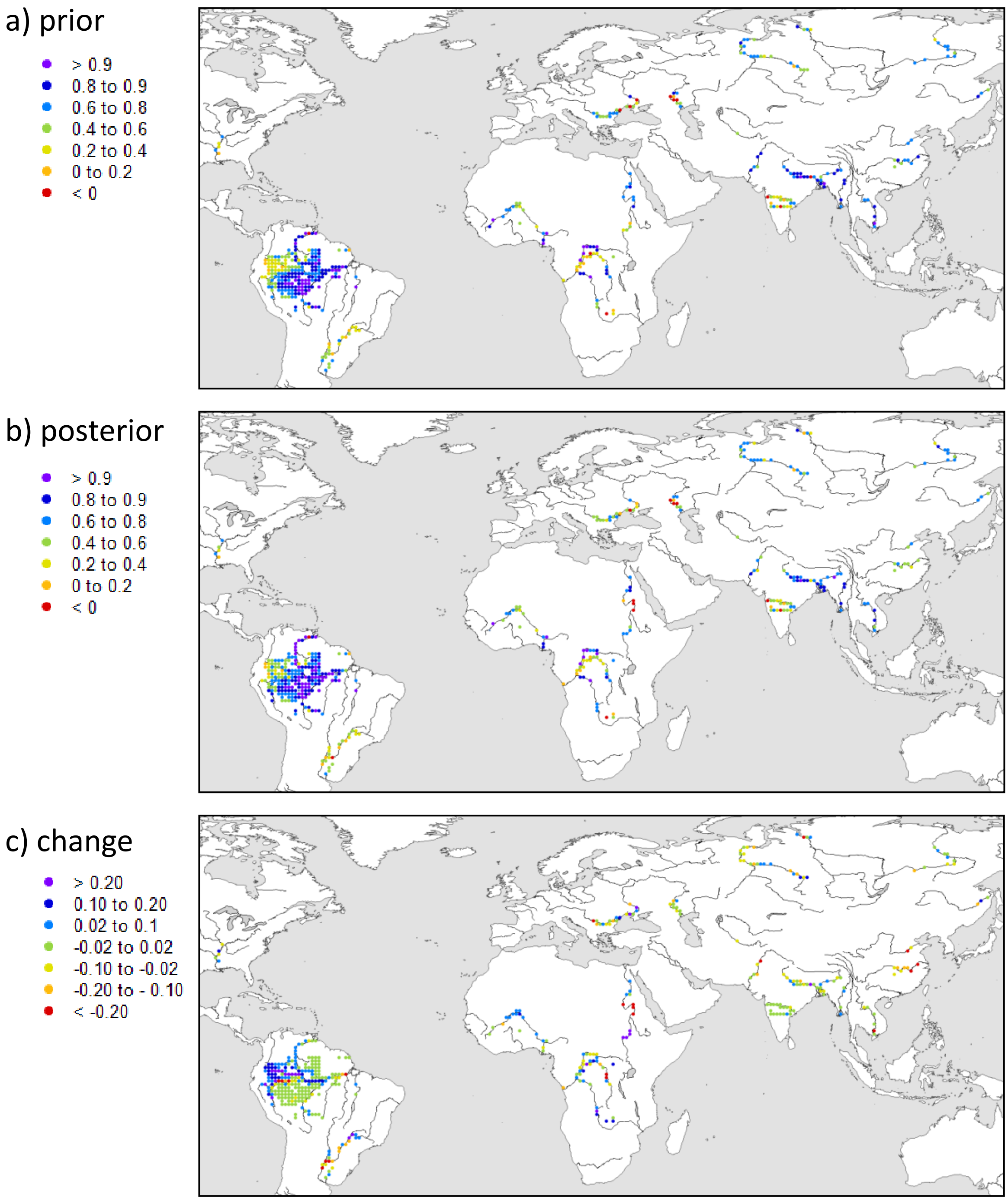



Figure 7. Effect of assimilation agreement with satellite altimetry river water levels: Spearman's rank correlation coefficient ($\rho$) for (a) prior and (b) posterior estimates and (c) difference between the two.

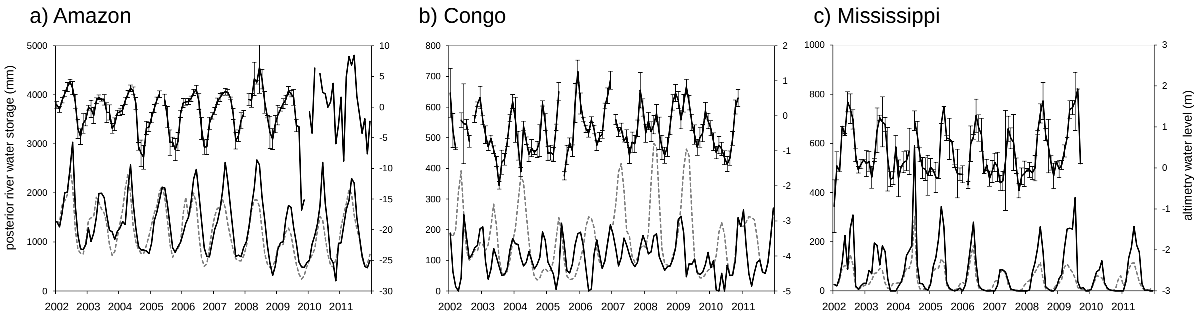



Figure 8. Effect of assimilation agreement with satellite altimetry river water levels for grid cells including the a) Amazon River (~2.5°S, 65.5°W; $\rho$ changed from 0.71 for prior to 0.80 for posterior estimates); b) Congo River (~2.5°N, 21.5°E; $\rho$ from 0.28 to 0.47) and Mississippi River (35.5°, 90.5°W; $\rho$ from 0.37 to 0.56).

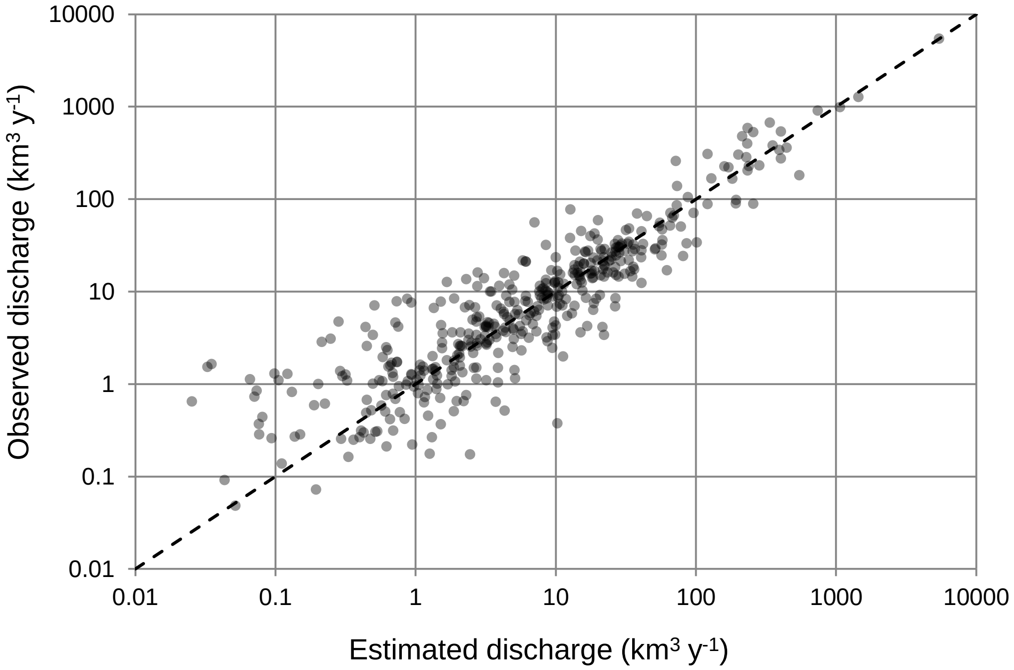



Figure 9. Comparison of mean basin discharge resulting from the analysis ($Q_a$) and values based on observations (Dai et al., 2009) (darker areas indicate overlapping data points).

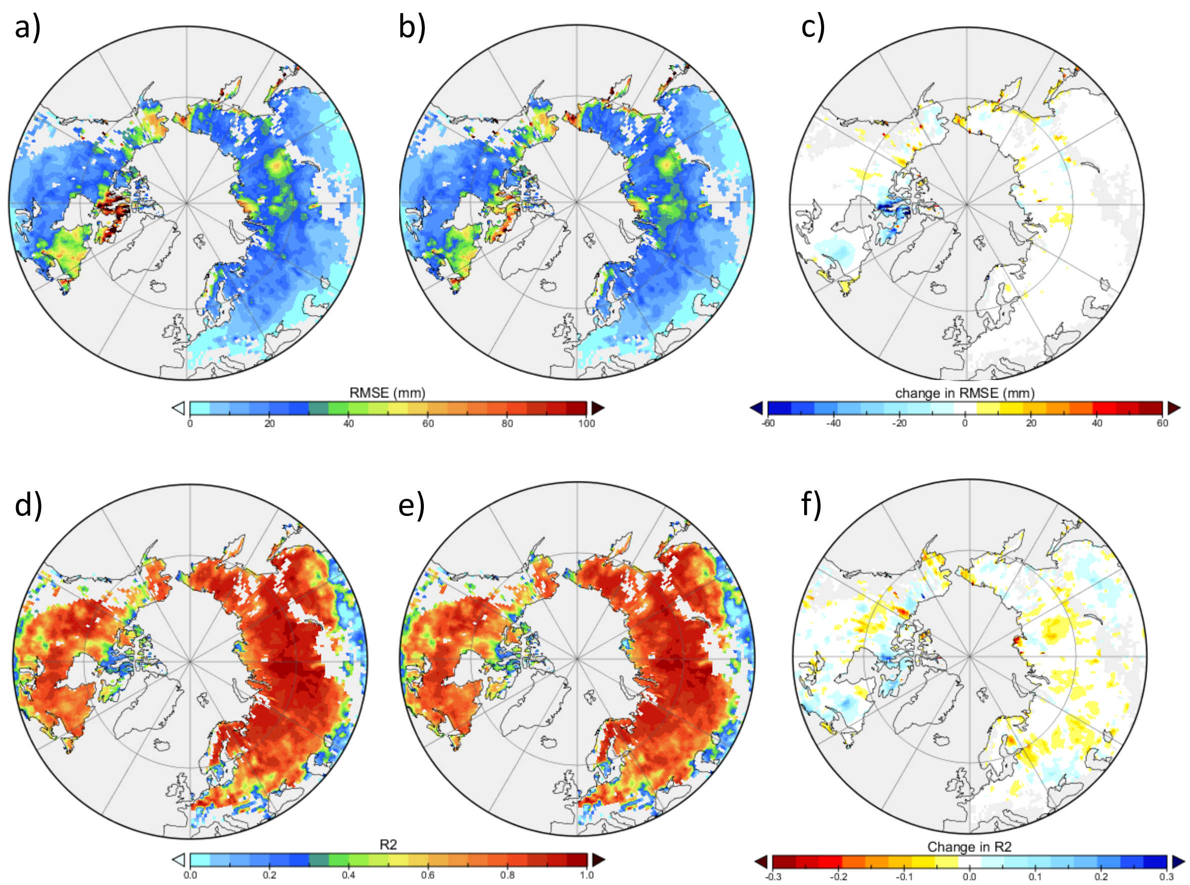



Figure 10. Effect of assimilation on agreement with GlobSnow snow water equivalent (SWE) estimates, showing (a-c) root mean square error (RMSE) and (d-f) the coefficient of correlation ($R^2$). From left to right, agreement for (a,d) prior and (b, e) posterior estimates as well as (c, f) the change in agreement.

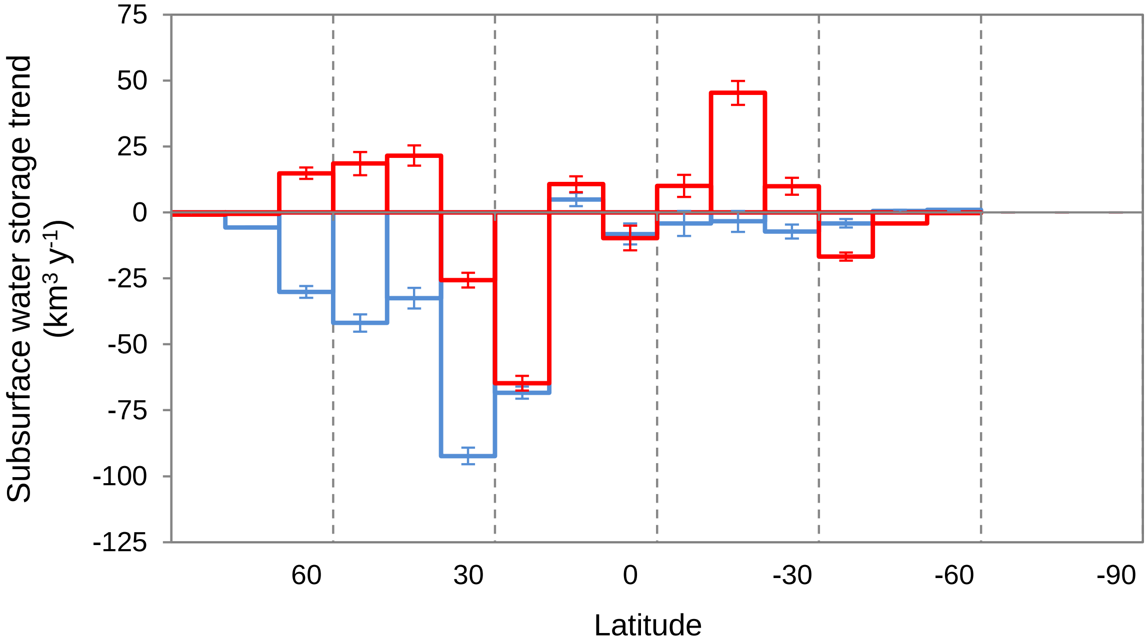



Figure 11. Linear 2003–2012 trends in sub-surface water storage by 10° latitude band, showing prior (blue) and posterior (red) estimates.