Comparing Global Hydrological Models and Combining them with GRACE by Dynamic Model Data Averaging (DMDA)

Nooshin Mehrnegar[a], Owen Jones[b], Michael Bliss Singer[a,c], Maike Schumacher[d], Paul Bates[e], Ehsan Forootan[a,d]

[a]*School of Earth and Ocean Sciences, Cardiff University, Cardiff CF103AT, UK*
[b]*School of Mathematics, Cardiff University, Cardiff CF244AG, UK*
[c]*Earth Research Institute, University of California Santa Barbara, Santa Barbara, 91306, USA*
[d]*Institute of Physics and Meteorology (IPM), University of Hohenheim, Stuttgart D70593, Germany*
[e]*School of Geographical Sciences, University of Bristol, University Road, Clifton, Bristol BS81SS, UK*

Abstract

Historically, hydrological models have been developed to represent land-atmosphere interactions by simulating water storage and water fluxes. These models, however, have their own unique characteristics (strength and weakness) in capturing different aspects of the water cycle, and their results are typically compared to or calibrated against in-situ observations such as river runoff measurements. As a result, there may be gross inaccuracies in the estimation of water storage states produced by these models. In this study, we present the novel approach of Dynamic Model Data Averaging (DMDA), which can be used to compare and merge multi-model water storage simulations with monthly Terrestrial Water Storage (TWS, a vertical summation of surface and sub-surface water storage) estimates from the Gravity Recovery And Climate Experiment (GRACE) satellite mission. Here, the main hypothesis is that merging GRACE data with multi-model outputs likely provides more skillful hydrological estimations compared to a single model or data set. Theoretically, the proposed DMDA combines the benefits of the Kalman Filter (KF) and Bayesian Model Averaging (BMA) techniques and has the capability to deal with various observations and models with different error structures. Based on the Bayes theory, DMDA provides time-variable weights for hydrological models to compute an average of their outputs that are best fited to GRACE TWS estimates. Numerically, the DMDA method is implemented by integrating the output of six hydrological and land surface models (PCR-GLOBWB, SURFEX-TRIP, LISFLOOD, HBV-SIMREG, W3RA, and ORCHIDEE) and monthly GRACE TWS estimates (2002–2012) within the world's 33 largest river basins, while considering the inherent uncertainties of all inputs. Our results indicate that DMDA correctly separates GRACE TWS estimates into surface water, soil moisture and groundwater compartments. Linear trends fitted to the DMDA-derived groundwater compartment are found to be different from those of original models. This means that anthropogenic influences within the GRACE data, which are not well reflected by models, are introduced by DMDA. We also find that temporal correlation coefficients between the DMDA-derived individual water storage estimations (surface water, soil moisture, and groundwater) and the El Niño Southern Oscillation (ENSO) index are considerably increased compared to those derived between individual model simulations and ENSO (e.g., an increase from -0.2 to 0.6 in the Murray River Basin). For the Nile River Basin, they changed from 0.1 to 0.4 for the soil moisture, and from 0.3 to 0.7 for the surface water compartment. Comparisons between the DMDA-derived surface water and those from independent satellite altimetry observations indicate that after implementing DMDA, temporal correlation coefficients within major lakes are increased. Based on these results, we have gained confidence in the DMDA water storage estimates to be used for improving the characterization of water storage over broad regions of the globe.

Keywords: GRACE, Terrestrial Water Storage (TWS), Dynamic Model Data Averaging (DMDA), Kalman Filter (KF), Bayesian Model Averaging (BMA), Multi-Hydrological Models, Satellite Altimetry

1. Introduction

Studying global water storage changes and their relationships with climate variability and exploring their trends are important to understand the interactions between the Earth's water, energy, and carbon cycles. It is also essential for managing water resources and understanding floods and food risks in a changing climate. In-situ and/or remote sensing observations provide estimates of different aspects of the Earth system, but they do not provide water cycle closure due to sampling and retrieval errors. In practice, hydrological models are used to quantify hydro-meteorological processes such as interactions between the global climate system and the water cycle (*Sheffield et al.*, 2012), the contribution of land hydrology to global sea level rise (*Boening et al.*, 2012), as well as to support applications related to water resources planning and management (*Hanington et al.*, 2017). However, model simulations are prone to errors due to imperfect model structure, as well as errors in inputs and forcing data that are used to run model simulations. As a result, available models operating at regional to global scales have limited skills to reflect human impacts on water storage and runoff changes (*Wada et al.*, 2012; *Scanlon et al.*, 2018; *Singer et al.*, 2018).

Among available remote sensing techniques, the Gravity Recovery And Climate Experiment (GRACE, 2002–2017) satellite mission (*Tapley et al.*, 2004) and its Follow-On mission (GRACE-FO, 2018–onward) provide an opportunity to assess the global water cycle by monitoring time-variable gravity fields. Global GRACE-derived time-variable gravity field data can be used to estimate changes in Terrestrial Water Storage (TWS), which is a vertical summation of canopy, surface water (lakes, rivers, and wetlands), as well as soil moisture and groundwater storage. Changes in TWS provide a critical measure of regional and global water balances, which cannot be measured by any other satellite mission. A review of GRACE applications in hydrology, and particularly for groundwater monitoring, can be found in *Frappart and Ramillien* (2018).

GRACE data can be used in conjunction with hydrological models to maximize information gained from modelling with rationalisation and separation of GRACE TWS. Thus, the gravimetric data from GRACE can inject realism into regional hydrological predictions, which are often poorly constrained in terms of TWS. Generally speaking, integrating GRACE data with hydrological models is important from two perspectives: (1) it can update (modify) water storage simulation within hydrological models and (2) it vertically separates GRACE TWS into storage compartments. The first point is of interest for hydrologists since most global models are not usually combined with water storage observations (*Bai et al.*, 2018). Therefore, such updates may lead to more realistic water storage simulations, which makes these models more useful for water resource applications (see e.g., *Werth et al.*, 2009; *Mostafaie et al.*, 2018). Regarding the second point, it is important to state that any attempt to vertically separate GRACE-derived TWS into its individual components requires a priori information from other sources, such as, hydrological models, satellite altimetry observations to estimate surface water storage, and soil moisture remote sensing data to estimate shallow depth soil moisture storage changes (*Forootan et al.*, 2014).

Various studies have developed techniques to merge multi-resources and achieve vertical separation of surface and sub-surface water storage compartments by several methods outlined below.

(a) Forward modeling techniques are used to evaluate different compartments of mass variations through a simple reduction process, relying on model and/or observation data for other compartments, e.g., surface water and soil moisture, if groundwater should be estimated (e.g., *Tiwari et al.*, 2009; *Rodell et al.*, 2009; *Strassberg et al.*, 2009; *Feng et al.*, 2013; *Khandu et al.*, 2016). This method is relatively straightforward, but it is not necessarily the most accurate way to separate GRACE signals, due to the reflection of modeling error and/or observation errors on the final estimation of mass changes. Also, the spatial and temporal resolution of the observations (from satellites or in-situ) and model outputs, as well as their signal content are not necessarily consistent (see the discussions in, e.g., *Forootan et al.*, 2014). Most of these limitations are taken into account by the methods described in what follows.

(b) Statistical inversion techniques, which are formulated based on statistical signal decomposition techniques, such as Principal Component Analysis (PCA, *Lorenz*, 1956) and its alternatives, e.g., Independent Component Analysis (ICA, *Forootan and Kusche*, 2012, 2013), have been used in previous studies to separate GRACE TWS into individual water storage estimates. For example, *Schmeer et al.* (2012) used PCA to generate a priori information about mass changes from global ocean, atmosphere, and land hydrology models. Then, they applied a least squares technique to use GRACE TWS to modify their priori estimates. A statistical inversion, which works based on both PCA and ICA, was proposed in *Forootan et al.* (2014, 2017) and *Awange et al.* (2014) to separate GRACE TWS using auxiliary data of surface water from satellite altimetry and individual sub-surface water storage estimate from a land surface model (Global Land Data Assimilation System (GLDAS, *Rodell et al.*, 2004)). This inversion harmonizes the use of all available data sets within a single least squares framework. As a result, a more consistent mass estimate (than that of the forward modeling in (a)) for individual water storage components can be achieved.

(c) Data Assimilation (DA) as well as simultaneous Calibration/Data Assimilation (C/DA) have been used in recent years to merge GRACE data with hydrological model outputs or other types of observations. These techniques rely on the model equations to relate water and energy fluxes to water storage changes. Therefore, unlike the inversion approach (b), combining information from observations (e.g., GRACE TWS estimates) and a model is performed in a physically justifiable way. DA or C/DA can potentially increase physical understanding of the model and improve the model states by decreasing the simulation errors. For example, DA is used in *Zaitchik et al.* (2008); *Girotto et al.* (2016, 2017); *Tian et al.* (2017); *Khaki et al.* (2018d,e), while C/DA is applied in *Schumacher et al.* (2016, 2018) to improve global models such as GLDAS (*Rodell et al.*, 2004), World-Wide Water Resources Assessment (W3RA, *Van Dijk*, 2010), WaterGap Global Hydrological Model (WGHM, *Döll et al.*, 2003), and NOAH Multi Parameterization Land Surface Model (NOAH-MP LSM, *Niu et al.*, 2011). Most of the previous DA and C/DA are implemented regionally (except *Van Dijk et al.* (2014); *Khaki et al.* (2017a, 2018a)) for example over the Mississippi River Basin (*Zaitchik et al.*, 2008; *Schumacher et al.*, 2016), Bangladesh (*Khaki et al.*, 2018d), the Middle East (*Khaki et al.*, 2018e), and the Murray-Darling

River Basin (*Tian et al.*, 2017; *Schumacher et al.*, 2018). In addition, these studies rely on simulation from (only) one selected hydrological model, which could contain errors in the model structure such as biases in the model's internal parameters and boundary conditions. In each of these studies, multiple realisations of the model-derived water storage simulations were generated by perturbing the input forcing data and/or model parameters. A sequential integration techniques such as the Ensamble Kalman Filtering (EnKF, *Evensen*, 1994) or its extensions was then used to merge GRACE data with the (ensemble) outputs of a single model (e.g., *Schumacher et al.*, 2016, 2018; *Khaki et al.*, 2017b). *Van Dijk et al.* (2014) used EnKF to merge GRACE data with a priori data from models and other remote sensing techniques. Their study covered the period of 2003-2012 and focused on updating the individual water storage estimates rather than interpreting the water storage estimates in terms of trends or addressing the suitability of models used to perform the analyses.

(d) In recent years, Bayesian-based techniques have been used to combine different observations with models and update their outputs. For example, *Long et al.* (2017) applied the Bayesian Model Averaging (BMA, *Hsu et al.*, 2009) technique to average multiple GRACE TWS products and global hydrological models to analyse spatial and temporal variability of global TWS. However, their study did not assess the update of individual surface and sub-surface water storage estimates. *Sha et al.* (2018) used a model-data synthesis framework based on Bayesian Hierarchical Modelling (BHM, see e.g., *Banerjee et al.*, 2004) to use GRACE TWS estimates to update land surface deformations derived from Glacial Isostatic Adjustment (GIA) models. Their study did not, however, address global hydrological mass changes.

It is worth mentioning here that the Ensamble Kalman Filter used for DA and C/DA can also be classified as a Bayesian-based technique because the cost function for updating unknown state parameters condition on the measurement data, is formulated based on the Bayes theory (see e.g., *Evensen*, 2003; *Schumacher*, 2016; *Fang et al.*, 2018). Methods, such as Particle Filter (PF) and Particle Smoother (PS) are also Bayesian (*Särkkä*, 2013), and have already been applied in a wide range of geophysical and hydrological applications. For example, *Weerts and El Serafy* (2006) compared the capability of EnKF and PF to update a conceptual rainfall-runoff model using discharge and rainfall data. *Plaza Guingla et al.* (2013) also used the standard PF to assimilate a densely sampled discharge records into a conceptual rainfall-runoff model. However, *Bain and Crisan* (2008) and *Del Moral and Miclo* (2000) show that the rate of convergence of the approximate probability distribution until attainment of the true posterior is inversely proportional to the number of particles used in the filter. This means that the filter perfectly approximates the posterior distribution when the number of particles tends to infinity. However, since the computational cost of PF grows with the number of particles, choosing a specific number of particles in the design of filters is a key parameter for these methods. The rationale for introducing a new Bayesian data-model merging algorithm in this study is described in (e).

(e) In this study, we present the Dynamic Model Data Averaging method (DMDA, i.e., a modified version of Dynamic Model Averaging (DMA) approach presented by *Raftery et al.*, 2010) to merge multi-model derived water storage simulations with GRACE TWS estimates, as an alternative technique to that described in (d). Our main goal is to evaluate available model outputs against GRACE TWS and merge them in a sensible way to gain more realistic insights about global surface and sub-surface water storage changes. The main hypothesis behind the presented approach is that each global hydrological model has its own unique characteristics and strengths in capturing different aspects of the water cycle. Therefore, relying on a single model often leads to predictions that represent some phenomena or events well at the expenses of others. *Scanlon et al.* (2018) recently compared GRACE TWS with the outputs of global models, whose results indicated inconsistencies in long-term trends and cyclic (e.g., seasonal) components. Besides, many studies have concluded that effective combination of multiple models may provide more skillful hydrological simulations compared to a single model (*Duan et al.*, 2007). Therefore, a multi-model choice is considered in this study.

Our motivation to formulate the DMDA is based on its capability to deal with various observations and models with different structures. In summary, DMDA is based on the Bayes theory and provides time-variable weights to compute an average of hydrological model outputs, yielding the best fit to GRACE TWS estimates, while considering their errors (see section 3). These time-variable weights indicate which of the available models at a given point in time fits better to GRACE TWS estimates. These weights can then be used to separate the components of TWS and modify the estimation of water storage in these individual components. Therefore, the DMDA-derived ensemble is expected to yield more skillful (realistic) hydrological simulations compared to any individual model (see similar arguments in *Duan et al.*, 2007). Here, we promote the use of DMDA over the previously introduced EnKF, PF, and PS methods because it is computationally more efficient in handling large dimensional problems such as the global integration implemented in this study. In addition, the DMDA's time-variable weights can be used to assess the performance of hydrological models, whereas this aspect is missing in other merging techniques. More details about the computational aspects of DMDA are provided in section 3.

To implement the DMDA method, surface and sub-surface water storage simulations of the six published global hydrological and land surface models (*Schellekens* *et al.*, 2017) are used. These models are structurally different but they are all forced by the same reanalysis data set (WATCH-Forcing-Data-ERA-Interim, WFDEI *Weedon et al.*, 2014) as inputs. GRACE-derived TWS estimates are then used in the DMDA method to compare their outputs and merge them. A challenging problem in merging GRACE TWS with the outputs from multiple hydrological models is related to their different spatial and temporal resolutions. To overcome the computational problem caused by the spatial and temporal mismatch, *Schumacher et al.* (2016) introduced spatial and temporal matching functions, which are able to avoid computational problems. In this study, we did not implement the spatial/temporal operator because both model outputs and GRACE data were set at monthly (temporal) and basin-averaged (spatial). Handling the differences in spectral domain is described in section 2.2. A realistic synthetic example is presented in section 4.1 to test the performance of the DMDA method, where the true merged values are known and the method can be evaluated to provide the confidence that it can be applied to a real case study. Our numerical results cover the world's 33 largest river basins (see Figure ESM.1 in Electronic Supporting Material, ESM) for the period of 2002–2012, during which both GRACE data and model simulations are available. Global hydrological model outputs are compared against GRACE TWS, using DMDA-derived temporal weights, within the largest river basins for the period of this study (see section 4.2). The DMDA-derived updates, which are assigned to the long-term trend of surface and sub-surface water storage components, are explored and interpreted (see section 4.3).

Among many climatic factors that influence inter-annual to decadal TWS changes, the El Niño Southern Oscillation (ENSO, *Barnston and Livezey*, 1987) events represent a dominant impact on global precipitation and TWS changes (see, e.g., *Hurkmans et al.*, 2009; *Chen et al.*, 2010; *Zhang et al.*, 2015; *Forootan et al.*, 2016; *Ni et al.*, 2018; *Anyah et al.*, 2018; *Forootan et al.*, 2019). In this study, temporal correlation coefficients between model-derived storage outputs and the ENSO index are used as a measure to determine whether implementing the DMDA helps to derive realistic storage simulations (see section 4.3.1). In addition, independent surface water level observations from satellite altimetry within14 major lakes, located in different river basins around the world, are used to validate our results (see section 4.4). This paper contains an Electronic Supporting Material (ESM) document that provide auxiliary information to improve understanding of the performed investigations.

2. Data sources

The data used in this paper include the monthly GRACE data to compute Terrestrial Water Storage (TWS) and individual water storage estimates from global models to provide a priori estimates to perform a Bayesian signal separation. GRACE TWS estimates are used in the DMDA to modify the multi-model water storage outputs.

2.1. GRACE Data

The latest release of the monthly GRACE level-2 (L2) product (RL06), expressed as dimensionless spherical harmonic coefficients up to degree and order 90, are downloaded for the period of April 2002 to December 2012 from the Center for Space Research (CSR, `http://www2.csr.utexas.edu/grace/RL06.html`). A limited length of the GRACE data is used here since the global hydrological model outputs of *Schellekens et al.* (2017) were available until 2012.

Recommended corrections are applied to generate monthly TWS fields from the GRACE product, i.e., degree 1 coefficients are replaced by those from *Swenson et al.* (2008) to account for the movement of the Earth's center of mass. The zonal degree 2 spherical harmonic coefficients (C20) are replaced by more stable ones derived from Satellite Laser Ranging (SLR) data (*Chen et al.*, 2007). Surface deformations known as the Glacial Isostatic Adjustment (GIA) are reduced using the output of the model provided by *Wahr and Zhong* (2012). GRACE level-2's correlated errors are reduced by applying the DDK2 an-isotropic de-correlation filter (*Kusche et al.*, 2009). The application of smoothing filters causes a spatial leakage problem, which is evaluated in terms of TWS errors following the approach in *Wahr et al.* (1998); *Khaki et al.* (2018c) over the world's 33 largest river basins as shown in Fig. ESM.1. An overview of the TWS's strength and our error estimates is shown in ESM-section 2 (see Figure ESM.2).

2.2. Global Hydrological Model (GHM) Outputs

Monthly water balance components from six large-scale Global Hydrological Models (GHMs) including PCR-GLOBWB (*Van Beek et al.*, 2011; *Wada et al.*, 2014), SURFEX-TRIP (*Decharme et al.*, 2013), LISFLOOD (*Van Der Knijff et al.*, 2010), HBV-SIMREG (*Lindström et al.*, 1997), W3RA (*Van Dijk*, 2010), and ORCHIDEE (*Polcher et al.*, 2011) are used in this study to provide a priori information about groundwater, soil moisture, surface water, canopy, and snow water storage components. The output of these models are published by the eartH2Observe Tier-1 (*Schellekens et al.*, 2017), and are available at 0.5° spatial resolution covering the period of 1979–2012 which can be downloaded from `http://earth2observe.github.io/water-resource-reanalysis-v1`.

Although, these models are structurally different, i.e., they use different methodology to simulate water changes, they are driven by the same reanalysis-based forcing data set, WFDEI (WATCH Forcing Data methodology applied to ERA-Interim reanalysis *Weedon et al.*, 2014). In other words, all hydrological models that are used in this study may represent the TWS, but their respective approaches for simulating TWS and its corresponding storage compartments are not identical. For example, *Schellekens et al.* (2017) state that PCR-GLOBWB and SURFEX-TRIP contain all surface and sub-surface water storage components in their TWS estimation. In contrast, TWS derived from LISFLOOD, HBV-SIMREG, and W3RA are equal to the summation of groundwater, soil moisture, and snow, while that of ORCHIDEE is the summation of soil moisture, surface water, and snow storage components.

An overview of the model outputs used in this study is provided in Table 1, and the linear trend (as a representative of monotonic long-term storage changes) fitted to the model outputs are shown in ESM-section 3.

TABLE 1

To ensure that the TWS estimates from GRACE L2 data and model outputs have the same spectral content, 0.5° resolution hydrological model outputs are transformed into the spectral domain and truncated to the maximum degree and order 90. The conversion follows an ordinary integration while considering the Gibbs effect along the coast lines (for more details please see, e.g., *Wang et al.*, 2006; *Forootan et al.*, 2013). Basin averages of each model components and their errors in terms of water storage are obtained from the same procedure used to process GRACE L2 data, i.e., implemented here following *Wahr et al.* (1998); *Khaki et al.* (2018c).

2.3. El Niño Southern Oscillation (ENSO) Index

The El Niño Southern Oscillation (ENSO, *Barnston and Livezey*, 1987) is a large-scale inter-annual climate variability phenomenon in the Tropical Pacific Ocean, which affects the climate of many regions of the Earth due to its ability to change the global atmospheric circulation, which influences temperature and precipitation across the globe (*Trenberth*, 1990; *Forootan et al.*, 2016). The positive phase on ENSO is known as El Niño, and its opposite phase is known as La Nina. The ENSO index used in this study is derived from sea surface temperature in the Niño 3.4 region $(5°N - 5°S, 170°E - 120°W)$. Monthly ENSO index (Niño 3.4 index), which is provided by the NOAA National Center for Environmental Information (NCEI) covering 1948 onward, is downloaded from `https://www.esrl.noaa.gov/psd/data/correlation/nina34.data`. This index will be used later in this study to demonstrate whether the DMDA-derived surface and sub-surface water storage estimates are closer to the reality than those from individual models.

2.4. Satellite Altimetry of Major Lakes

Water level measurement by satellite altimetry has been developed and optimised for open oceans, yet improved post-processing techniques can be used to obtain reliable satellite altimetry-derived height measurements within inland water bodies such as lakes, rivers, floodplains and wetlands (e.g., *Moore and Williams*, 2014; *Uebbing et al.*, 2015). In this study, satellite altimetry-derived surface water observations are used to validate TWS changes of GRACE and models as well as surface water derived from GHMs and the DMDA method. Satellite altimetry time series of major global lakes are available from the U.S. Department of Agriculture (USDA) (https://ipad.fas.usda.gov/). Repeated observations of the TOPEX/Poseidon (T/P), Jason-1, and Jason2/OSTM altimetry missions are included in this database. USDA provides time series of lake water level variations from 1992 to the present-day within 81 lakes, and from 2008 to present-day within more than 280 lakes around the world. An assessment over 14 lakes located within 8 river basins of this study is presented in section 4.4 for the period of 2002–2012. Details of these lakes are reported in Table 2.

TABLE 2

3. Dynamic Model Data Averaging (DMDA) Method

In this section, we present the mathematical formulation of Dynamic Model Data Averaging (DMDA), which follows the method of Dynamic Model Averaging (DMA, *Raftery et al.*, 2010) but with some modifications to achieve a recursive update of hydrological model outputs using GRACE TWS data (Fig. 1 summarises the DMDA method). It will also be shown that the implementation of DMDA combines the benefits of state-space merging techniques, such as Kalman Filtering (KF, *Evensen*, 1994) or Particle Filtering (PF, *Gordon et al.*, 1993), Markov Chain (MC, *Metropolis et al.*, 1953; *Chan and Geyer*, 1994; *Kuczera and Parent*, 1998), and Bayesian Model Averaging (BMA, *Hsu et al.*, 2009). DMDA can be applied in data assimilation applications that work with only one model, e.g., (*Girotto et al.*, 2016; *Khaki et al.*, 2017c,b; *Schumacher et al.*, 2018), as well as in handling multi-model outputs as in *Van Dijk et al.* (2014).

DMDA is formulated based on the representation of a state-space equation, which dynamically relates the GRACE TWS estimates and hydrological model outputs as:

$$y_t = z_t\theta_t + \epsilon_t, \tag{1}$$

$$\theta_t = \theta_{t-1} + \delta_t, \tag{2}$$

Equation (1) is known as 'observation equation' and represents a linear regression between the observation y_t (GRACE TWS estimates) and the vector of predictors

z_t (model-derived water storage simulations). The unknown regression parameter θ_t, commonly known as the 'state vector' (*Bernstein*, 2005), is allowed to evolve in time, according to equation (2), and is known as the 'state equation'. In equations (1) and (2), ϵ_t and δ_t can be interpreted as the residual of output vector and state parameters, respectively. They are usually defined using a normal distribution with the mean value of zero and a standard deviation, which will be computed during the DMDA procedure.

It is worth mentioning here that the EnKF (*Evensen*, 1994) and PF are among popular algorithms that can be used to recursively update an estimate of the model states and produce corresponding innovation values given a sequence of observations in the state-space equation (similar to what introduced above). In theory, EnKF accomplishes this goal by linear projections, and the estimations in PF are performed through a Sequential Monte Carlo sampling. Comparing EnKF and PF, the latter includes a random element so it converges to the true posterior probability function if the number of samples is very large. While the strength of PF is in its ability to account for both Gaussian and non-Gaussian error distributions, it suffers from the curse of dimensionality, which means that the sample size increases exponentially with the dimension of the state-space in order to achieve a certain performance. This fact precludes the use of PF in high-dimensional data-model fusion problems (*Bengtsson et al.*, 2008; *Daum and Huang*, 2003; *Snyder et al.*, 2008). For linear and Gaussian-type state-space models, as presented in this study, the PF method will yield the same likelihood as EnKF when the number of simulations is large enough (this has been tested but the results are not shown to keep the focus of this study on presenting the DMDA). Therefore, the DMDA, which combines the benefits of the EnKF and it is mathematically rigorous like PF, is adopted for the global data-model integration of this study.

Equations (1) and (2) are formulated with the main assumption that there is little physical knowledge about how the defined regression model and its parameters are likely to evolve in time. However, we will show that, by introducing two parameters of λ and α, which are referred to as 'forgetting factors', one can control the temporal dependency of the DMDA solutions. These two parameters provide the opportunity to treat model simulations and observations of each step temporally dependent on, or independent from, previous steps. Since changes in water storage depend on the history of hydrological processes, accounting for temporal dependency between water states sounds logical.

Formulating DMDA to Update Multi-Model Outputs using GRACE TWS

Here the DMDA method is formulated to update the outputs of multi-hydrological models, M_k, (for six models: $k = 1, ..., 6$). It is worth mentioning that since available models have different storage definitions, the length of the state vector can change from one model to another. Additionally, the structure of each individual storage components can also be defined differently in different models (e.g., the number of soil layers does not remain constant in different hydrological models). These differences can be handled by DMDA.

In the following, $Y_t = [y_1, ..., y_t]$ represents the vector of observations (i.e., GRACE TWS estimates in our study) up to the time step t. To use this vector to update the water storage simulation of a single-model, one can estimate the unknown (linear) regression parameters (θ_t) as

$$\theta_{t-1}|Y_{t-1} \sim N(\hat{\theta}_{t-1}, \hat{\Sigma}_{t-1}). \tag{3}$$

The distribution of each parameter can be assumed to be normal with unknown mean $\hat{\theta}_{t-1}$ and the variance $\hat{\Sigma}_{t-1}$. The regression coefficients at time t (θ_t) can then be obtained using θ_{t-1} from equation (3) and by introducing $\delta_t \sim \mathcal{N}(0, W_t)$ to the state equation (equation (2)). Therefore, the desired parameters at time t are defined by

$$\theta_t|Y_{t-1} \sim N(\hat{\theta}_{t-1}, R_t), \tag{4}$$

where

$$R_t = \hat{\Sigma}_{t-1} + W_t. \tag{5}$$

In equation (5), W_t is the covariance matrix of the state innovation vector (δ_t in equation (2)) and it shows the dependency of the regression parameters at each time point to the previous time. However, in practice, there is no information about the temporal relationship between GRACE TWS estimates and hydrological model outputs to be used to define W_t. Therefore, to mathematically define a temporal dependency, R_t in equation (4) can be replaced by

$$R_t = \lambda^{-1}\hat{\Sigma}_{t-1}, \tag{6}$$

where λ ($0 < \lambda \leq 1$) controls the influence of previous observations on the regression value at time t, and is known as 'forgetting factor' in the DMDA method (see, e.g., *Fagin*, 1964; *Jazwinski*, 2007).

Hannan et al. (1989) indicated that in the recursive estimation of auto-regressive models, the covariance of previous steps is derived as a weighted product of the current step (i.e., weighted by λ^{-1} in equation (6)). By this assumption, the effective window size of temporal dependency is estimated by $1/(1-\lambda)$. In our case, we choose λ to be 0.95, which means that for monthly data, the effective window size is equivalent to 18 months. This value is chosen experimentally because it minimized the Root Mean Square (RMS) of differences between TWS derived from DMDA and GRACE.

To apply DMDA and update water storage simulated by K different models, the parameter prediction of equation (4) is extended as

$$\theta_t^{(k)}|M_t = k, Y_{t-1} \sim N(\hat{\theta}_{t-1}^{(k)}, \lambda^{-1}\hat{\Sigma}_{t-1}^{(k)}),\ k = 1, ..., K, \tag{7}$$

where $M_t = k$ denotes which model (from the $k = 1, 2, ..., K$ available models) applies at time t, and the solution $\theta_t^{(k)}$ and $\hat{\Sigma}_{t-1}^{(k)}$ can be obtained using a Kalman Filter (KF)-type update conditional on $M_t = k$ for each sample. This (KF-type) update at time t is derived as

$$\theta_t^{(k)}|Y_t \sim N(\hat{\theta}_t^{(k)}, \hat{\Sigma_t}^{(k)}). \tag{8}$$

Regression parameters to update multi-model storage simulations can be estimated as

$$\hat{\theta}_t^{(k)} = \hat{\theta}_{t-1}^{(k)} + R_t^{(k)} z_t^{(k)} (V_t + z_t^{(k)}(R_t^{(k)} + Q_t^{(k)}) z_t^{(k)T})^T (y_t^{(k)} - z_t^{(k)}\hat{\theta}_{t-1}^{(k)}), \tag{9}$$

where V_t is the covariance matrix of GRACE TWS estimates (our observation), and Q_t is the covariance matrix of predictor z_t (see equation (1)). In this study, the leakage errors of model-derived TWS are estimated for the world's 33 river basins (similar to those of GRACE). These errors are used to generate Q_t, which is therefore a diagonal matrix in the DMDA implementation of this study. For a grid based implementation of DMDA, one can use the full covariance matrix of GRACE TWS similar to *Schumacher et al.* (2016). The covariance matrix $\hat{\Sigma}_t$ in equation (8) can be estimated from

$$\hat{\Sigma_t}^{(k)} = R_t^{(k)} - R_t^{(k)} z_t^{(k)T} (V_t + z_t^{(k)}(R_t^{(k)} + Q_t^{(k)}) z_t^{(k)T})^{-1} z_t^{(k)} R_t^{(k)}. \tag{10}$$

It is evident from equations (9) and (10) that the estimation of regression parameter $\hat{\theta}_t$ is conditional on a particular model. Therefore, the DMDA solution to obtain unconditional results and update multi-model simulations involves calculating the posterior model probability $P(M_t = k|Y_t)$ as a weight for each model, which changes at each time step. In the following, we show that time-variable weights need to be computed for each model k by choosing a forgetting factor α in a recursive method, where $k = 1, ..., K$. These weights are then used to average the models, which leads to the best fit to the GRACE TWS estimates. This justifies the term 'Dynamic' in the DMDA and makes the method different from other averaging techniques such as the Bayesian Model Averaging (BMA).

Let us assume that $P(M_t = k|Y_t) = \pi_{t|t,k}$, then the posterior model probability for each model k at time t can be estimated as

$$\pi_{t|t,k} = \frac{\pi_{t|t-1,k} P(y_t|M_t = k, Y_{t-1})}{\sum_{l=1}^{K} \pi_{t|t-1,l} P(y_t|M_t = l, Y_{t-1})}, \tag{11}$$

where, $P(y_t|M_t = k, Y_{t-1})$ is the density of the observation at time t, conditional on model k, as well as $Y_{t-1} = [y_1, y_2, ..., y_{t-1}]$, which is estimated by a normal distribution as

$$y_t|M_t = k, Y_{t-1} \sim N(z_t^{(k)} \hat{\theta}_{t-1}^{(k)}, V_t + z_t^{(k)} (R_t^{(k)} + Q_t^{(k)}) z_t^{(k)T}), \tag{12}$$

and, $\pi_{t|t-1,k}$ is the model prediction equation, which is defined by

$$\pi_{t|t-1,k} = \Sigma_{l=1}^{K} \pi_{t-1|t-1,k} a_{kl}. \tag{13}$$

In equation (12), $\hat{\theta}_{t-1}^{(k)}$ is estimated using the KF-type update as formulated in equations (9) and (10), while $R_t^{(k)}$ is obtained from equation (6) by choosing a forgetting factor λ, i.e., between 0 and 1.

In equation (13) $a_{kl} = P(M_t = l|M_{t-1} = k)$ is the element of the $K \times K$ transition matrix $A(a_{kl})$ between models, which can be onerous when the number of models is large, e.g., for K models and τ time steps, the number of combinations of models will be $K^{2\tau}$. In our study, we have 6 hydrological models, and 122 time steps over the entire period of the study (2002–2012), which leads to 6^{244} combinations of models. To specify the transition matrix A, one way is to use the Markov Chain Monte Carlo method (MCMC, *Geyer*, 2011), which will typically be computationally expensive. Therefore, in this study, we avoid the implicit specification of the transition matrix using the forgetting factor of $0 < \alpha < 1$, which has the same role as λ in equation (6). As a result, the model prediction equation (13) can be rewritten as

$$\pi_{t|t-1,k} = \frac{\pi_{t-1|t-1,k}^{\alpha}}{\sum_{l=1}^{K} \pi_{t-1|t-1,l}^{\alpha}}. \tag{14}$$

The posterior model probability, or weights, for each model at time t is estimated in a recursive solution between equations (11), (12), and (14). This process is initialized by setting $\pi_{0|0,k} = \frac{1}{K}$ for $k = 1, ..., K$, and assigning a prior values to the initial condition of the states $\theta_0^{(k)} \sim N(0, \Sigma_0^{(k)})$ and $\Sigma_0^{(k)}$ =Variance $(y_t^{(k)})$/Variance $(z_t^{(k)})$. The reason of choosing this prior value is that in a linear regression, a regression coefficient for a predictor z_t is likely to be less than the standard deviation of the observations y_t divided by the standard deviation of predictors z_t (for more information see e.g., *Raftery*, 1993). In our numerical evaluation of DMDA with six hydrological models, the optimum regression estimates are found when $0.85 < \alpha < 0.9$, because the RMS of differences between the DMDA-derived TWS and those of GRACE were at a minimum here. By choosing a forgetting factor $\alpha = 0.9$, we assume a temporal smoothing window with 36 month time steps between 6 hydrological model ensembles to predict posterior probability values of each model k at time t. It means that the contribution of hydrological models at time $t-37$ in to the posterior model probability of each model k at time t is negligible. The length of this smoothing window is reduced e.g., to 8 months if we choose $\alpha = 0.2$.

The multi-model predictions of y_t is a weighted average of model specific prediction $\hat{y}_t$, using the posterior model probabilities, $\pi_{t|t,k} = Pr(M_t = k|Y_t)$, as its weights, i.e.,

$$\hat{y}_t^{DMDA} = \sum_{l=1}^{K} \pi_{t|t,l} \hat{y}_t^{(l)}, \tag{15}$$

where $\hat{y}_t^{(k)} = z_t^{(k)} \hat{\theta}_t^{(k)}$.

The posterior model probability for each model at time t, along with the estimated time-variable regression parameter $\theta_t^{(k)}$ from KF-type updating equation (9) are used to estimate the multi-model prediction of water storage components as

$$\hat{z}_{j,t}^{DMDA} = \sum_{l=1}^{K} \pi_{t|t,l} z_{j,t}^{(l)} \hat{\theta}_{j,t}^{(l)}, \tag{16}$$

where j represents each of the water storage components, i.e. groundwater, soil moisture, surface water, canopy, and snow. To update the water storage simulations of a single-model using the GRACE TWS estimates and the DMDA approach, K needs to be set to 1, and the prediction step is limited to the conditional estimation of the parameter $\theta_t^{(k)}|M_t^{(k)}$ using equation (9).

The posterior model probability can also be used to estimate unconditional probability distribution of regression parameters $\Theta_t = (\theta_t^{(1)}, ..., \theta_t^{(K)})$ given by observation Y_t following

$$p(\Theta_t|Y_t) = \sum_{l=1}^{K} p(\theta_t^{(l)}|M_t = k, Y_t) P(M_t = k|Y_t), \tag{17}$$

where $p(\theta_t^{(k)}|M_t^{(k)}, Y_t)$ shows the conditional distribution of $\theta_t^{(k)}$ which is approximated by a normal distribution as:

$$\theta_t^{(k)}|M_t^{(k)}, Y_t \sim N(\hat{\theta}_t^{(k)}, \hat{\Sigma}_t^{(k)}). \tag{18}$$

The DMDA approach can be recovered to a standard Bayesian Model Averaging (BMA, *Hoeting et al.* (1999)) when $\alpha = \lambda = 1$. Then the posterior model probability of model k is given by

$$P(M_t = k|Y_t) = \frac{p(Y_t|M_t = k)}{\sum_{l=1}^{K} p(Y_t|M_t = l)}, \tag{19}$$

where $p(Y_t|M_t = k)$ is the marginal likelihood, obtained by integrating the product of the likelihood, $P(Y_t|\theta^{(k)}, M_t = k)$, and the prior, $P(\theta^{(k)}|M_t = k)$, over the parameter space (see also *Hsu et al.*, 2009). Figure 1 summarises the work-flow of the DMDA approach.

FIGURE 1

4. Results

4.1. Setup a Simulation to Test the Performance of DMDA

Before applying the DMDA method on real data, its performance is tested in a controlled synthetic simulation, where the results of the Bayesian update are known by definition. In the first step of our simulation, we aim to compare DMDA and BMA in terms of updating hydrological model outputs with respect to the observations (i.e., GRACE TWS estimates in this study). In the second step, it will be shown that the DMDA-derived time-variable weights are the same as the expected values.

To make the synthetic study simple, we assumed that TWS is defined as the summation of just groundwater and soil moisture components. By this definition, the time series of groundwater and soil moisture of two hydrological models, i.e., here selected as LISFLOOD (M_1) and SURFEX-TRIP (M_2), are introduced as predictors to the DMDA, and TWS derived from a third model, here selected to be PCR-GLOBWB, is considered as the observation (here standing in for GRACE derived TWS). By this choice, after applying DMDA to merge M_1 and M_2 with simulated observed TWS, we expect that the updated (DMDA-derived) groundwater and soil moisture storage estimates will be fitted to those of simulated observation. Here, we selected results within the Niger River Basin (id:20 in Fig. ESM.1), covering the period of 2002–2012. Figure 2 (A) shows the PCR-GLOBWB TWS as our observation,

Fig. 2 (B) represents the time series of groundwater and soil moisture derived from M_1 (B1, B3, blue curves) and M_2 (B2, B4, green curves), while the expected value of DMDA-derived groundwater and soil moisture (simulated observation) are shown with the red color curves in these figures.

The magnitude of minimum (Min), maximum (Max) and the Root Mean Square (RMS) of the signal for all simulated data sets can be found in Table 3. The uncertainty of these data sets are computed following a least squares error propagation, while considering the leakage error of GRACE TWS in the Niger River Basin. It is worth mentioning that the final results of the simulation do not depend on the selection of models and the adopted simplification. The RMS of differences between the simulated TWS and two selected models (reported in Table 3) indicates that M_2 (RMS of $\Delta_{TWS} = 14.1$ mm) had a better agreement with the observations compared to M_1 (RMS of $\Delta_{TWS} = 18.6$ mm). Figure 2 (C1) shows the estimated weights for the first model (W_1, Mean= 0.47) and second model (W_2, Mean= 0.53) obtained using DMDA (equation (11)). These results show that the model which had a better agreement with observations gained higher weights.

To compare DMDA and BMA methods to average hydrological components, we apply both of these methods on simulated data sets. The final results are shown in Fig. 2 (D1: groundwater) and (D2: soil moisture). Groundwater, soil moisture, and consequently TWS derived from DMDA shows better agreement with the expected values in comparison to the BMA results. The RMS of errors for both methods are reported in Table 3, which indicates that although TWS derived from BMA follow the expected value (RMS of error= 8.4 mm), the obtained individual components from this method are not close to the simulated values (RMS of errors of 20.4 mm and 18.6 mm are found for groundwater and soil moisture, respectively). A considerable decrease in the differences between hydrological components and the expected values of DMDA shows that the method is suitable to update multi-model water storage estimates. Details of the numerical comparisons can be found in Table 3.

In the second step of our simulation, we use the weights of the first step (W_1, W_2, Fig. 2 (C1)) plus a temporal white noise with standard deviation of 0.02 m (equal to the standard deviation of GRACE TWS error within the Niger River Basin) to simulate GRACE like TWS estimates. Reconstructed weights after applying the DMDA for the second time, using the new synthetic TWS observations, are shown in Fig. 2 (C2). The correlation coefficient between W_1 and W_2 with their reconstructed values is found to be 0.73 and the RMS of the reconstruction's errors is found to be 0.18. This indicates that the DMDA-derived weights are close to reality and further motivates us to apply it on real data sets.

FIGURE 2

TABLE 3

4.2. DMDA Weights to Compare Global Hydrological Models

TWS derived from DMDA is a weighted average of selected models by estimating time varying weights based on the Bayes rule as in equation (15). Figure 3 shows the estimated weights for ten basins with the largest RMS of differences between TWS derived from individual models and GRACE TWS. Time-variable weights derived from DMDA allow us (1) to quantify the quality and compare individual water storage simulations derived from each global hydrological model against GRACE TWS for different periods of time, and (2) to separate GRACE TWS in a Bayesian framework, while considering different model structures and errors within and between model simulations and GRACE data. The average of weights during 2002–2012 is considered as the basis to select the best model in DMDA results over 33 river basins which is shown in the middle of Fig. 3. From our numerical results, PCR-GLOBWB is found to gain the largest weights during this period, thus, it contributed the most in the DMDA-derived TWS in North Asia, Central Africa, and North America. The weights computed for SURFEX-TRIP are found to be larger than other models within the snow-dominated regions, such as, the Yukon and Mackenzie in the north part of America and the Lena in the Northeast Asia. Our results confirm the investigations by *Schellekens et al.* (2017), who compared the mentioned models against the Interactive Multi-sensor snow and Ice Mapping System (IMS, *Ramsay*, 1998). Apparently, multiple snow layers of SURFEX-TRIP helps it to better simulate snow dynamics during the cold seasons.

We also find that SURFEX-TRIP received the highest averaged weights (compared to other models) within the Amazon and Brahmaputra River Basins during 2002–2012. The explanation is that SURFEX-TRIP likely better accounts for (1) the snow coverage of the Brahmaputra River Basin, (2) the considerable contribution of surface water storage components in the TWS changes within the Amazon River Basin, and (3) the overall dry period within both basins (*Chen et al.*, 2009; *Khandu et al.*, 2016), specially the extreme hydrological droughts of 2005 and 2010 (*Forootan et al.*, 2019). In the Amazon River Basin, we also find the highest performance for SURFEX-TRIP between 2009-2011. *Chen et al.* (2009) reported that in 2009 the Amazon River Basin experienced an extreme flood, which increased the magnitude of inter-annual TWS in this basin. TWS changes within the Amazon are also closely connected to the ENSO events in the tropical Pacific (*Kousky et al.*, 1984; *Ropelewski* *and Halpert*, 1987). Later we will show that surface water derived from SURFEX-TRIP shows the highest correlation with ENSO index in comparison with the other models of this study. This could be another reason that we derive the highest weights for SURFEX-TRIP between 2009-2011 within the Amazon River Basin.

Our results (Fig. 3) indicate that within the river basins with considerable irrigation (such as the Indus, Euphrates, and Orange River Basins), the relatively highest weights are assigned to the LISFLOOD and ORCHIDEE, where both account for human water-use (*Schellekens et al.*, 2017). ORCHIDEE is also found to perform well within the Brahmaputra, Ganges, and Murray River Basins, each of which experienced a strong decline in rainfall over the entire period of our study (e.g., 9.0 $\pm$ 4.0 mm/decade between 1994–2014 over Ganges and Brahmaputra *Khandu et al.*, 2016). Specifically, ORCHIDEE contains 14 soil layers (see Table 1) that help it to better resolve vertical water exchange within the irrigated regions. In ESM-section 2, it is shown that GRACE TWS changes within the Murray River Basin are considerably influenced by ENSO events (see also *Forootan et al.*, 2012, 2016), and the simulated outputs of ORCHIDEE reflects these changes better than the other tested models justifying the higher weights that are assigned to this model within the DMDA procedure. In ESM-section 5, we show that after applying the DMDA, model-derived TWS simulations are tuned to GRACE TWS.

FIGURE 3

4.3. DMDA-Derived Individual Water Storage Estimates

The estimated weights for the six models of section 4.2 along with the computed regression coefficients $\hat{\theta}_t$ (see the flowchart of Fig. 1), are used to compute the DMDA-derived groundwater, soil moisture, and surface water. In order to interpret the monotonic changes of water storage changes within the river basins, a long-term linear trend is fitted to the DMDA results that are shown in Figure 4, and the numerical values are reported in Table 4.

FIGURE 4

TABLE 4

Figure 4 (a1) and (a2) show the linear trend fitted to the DMDA-derived groundwater and its uncertainty. The results indicate a decrease in groundwater in 42% of the assessed river basis (i.e., 14 of 33). The largest decreasing trends are found in basins with large-scale irrigation such as the Ganges (-14.77 $\pm$ 0.25 mm/yr), Indus (-8.26 $\pm$ 0.16 mm/yr) and Euphrates (-5.36 $\pm$ 0.23 mm/yr). The results confirm the findings by *Khandu et al.* (2016), *Forootan et al.* (2019), and *Voss et al.* (2013), respectively. The strongest increasing trends in groundwater are seen in the Tocantins basin (South America) at the rate of 2.41 $\pm$ 0.47 mm/yr, the Okavango (South Africa) with a rate of 1.74 $\pm$ 1.31 mm/yr, and the Lena (Northeast Asia) with 1.74 $\pm$ 0.11 mm/yr. However, all of these trends are not statistically significant. The positive trends in groundwater storage in these last two basins are associated to the heavy rainfalls, seasonal floods and the geographical location of the Okavango Delta (*McCarthy et al.*, 1998), and underground ice melting caused by global warming (*Dzhamalov et al.*, 2012), respectively. Comparisons between the DMDA-derived groundwater and those of hydrological models indicate that after merging GRACE TWS with output from multiple hydrological models, the linear trend has changed considerably. This means that introducing GRACE data can successfully modify the anthropogenic effects, which are not well simulated by models (linear trends of the modelled groundwater are shown in ESM-section 3).

The linear trend fitted to the DMDA-derived soil moisture and its uncertainty are shown in Fig. 4 (b1) and (b2). We find strongest increasing trends in soil moisture estimates within the Murray (Australia), Okavango, and Orinoco (South America) River Basins with rates of 6.66 $\pm$ 0.15, 3.92 $\pm$ 0.55, and 3.45 $\pm$ 0.26 mm/yr respectively, and largest decreasing trends in the Brahmaputra and Euphrates with rates of -7.00 $\pm$ 0.69 and -5.75 $\pm$ 0.39 mm/yr.

Figure 4 (c1) and (c2) show the linear trends and their uncertainty fitted to the surface water storage estimated through the DMDA method. Linear trends of surface water within the 28 out of the 33 river basins are found to be statistically insignificant (values between -1 and +1 mm/yr). The strongest negative trends are found in the Euphrates, Murray, and Okavango River Basins with rates of -2.09 $\pm$ 0.09, -1.47 $\pm$ 0.04, and -1.42 $\pm$ 0.37 mm/yr respectively. In contrast, the largest positive trends are found within the Amazon and Colorado, at the rate of 1.43 $\pm$ 0.06 and 1.04 $\pm$ 0.04 mm/yr, respectively. The heavy flood during the summer of 2008–2009 (*Marengo et al.*, 2011; *Chen et al.*, 2010), which was considerably bigger than the temporal mean, likely caused these positive trend in the Amazon River Basin. Negative trends in all three water storage compartments of the Euphrates River Basin (groundwater -5.36 $\pm$ 0.23 mm/yr, soil moisture -5.75 $\pm$ 0.39 mm/yr, and surface water -2.09 $\pm$ 0.09 mm/yr) can be associated to both irrigation and long-term drought as shown by *Forootan et al.* (2017).

4.3.1. Contribution of ENSO in DMDA-Derived Water Storage Components

To demonstrate that the DMDA-derived surface and sub-surface water storage estimates are closer to the reality than those from any individual model, we extract the dominant ENSO mode from the DMDA estimates and compare them with climate indices (see e.g., *Anyah et al.*, 2018) in terms of temporal correlation coefficients with the ENSO index (-Niño 3.4 index, Fig. 5, 6, and 7). The reason for this comparison is that GRACE captures considerable variability due to the ENSO events (*Phillips et al.*, 2012; *Forootan et al.*, 2018). Therefore, by merging multi-model outputs with GRACE data, their skill in representing water storage changes due to large-scale teleconnections would be improved.

In order to extract the ENSO modes from the DMDA-derived water storage estimates and the original outputs of the six models (PCRGLOB-WB, SURFEX-TRIP, LISFLOOD, HBV-SIMREG, W3RA, and ORCHIDEE) Principal Component Analysis (PCA, *Lorenz*, 1956) method is applied after removing the long-term linear trend and seasonality from hydrological components. More details about PCA results and extracting ENSO modes from DMDA water storage components are reported in ESM-section 6.

Figure 5 shows temporal correlations between the ENSO mode of groundwater (from DMDA and original models) and the ENSO index. Maximum and minimum correlation of 0.75 and 0.53 corresponding to a maximum lag of up to 2 months are found globally between the DMDA groundwater and the ENSO index, respectively. Smaller correlations are found between the original models and the ENSO index. Among these models, W3RA and HBV-SIMREG indicate stronger correlations ($\sim$ 0.6 and $\sim$ 0.4 respectively) with the ENSO index with a maximum lag of 2 months. Other models such as LISFLOOD and SURFEX-TRIP indicate notably different correlations (compared to HBV-SIMREG and W3RA as well as that of DMDA) with ENSO in various basins. We find small positive correlations with a maximum value of 0.3 between original PCR-GLOBWB's groundwater and the ENSO index. Although the maximum lag of 3 month is estimated in most of the 33 basins, a lag of 15 months is estimated for the Nile, Okavango, and Zambezi (Africa), Colorado and Nelson (North America), Ob, Lena, and Yellow (Asia) River Basins, which are likely not realistic (see, e.g., *Awange et al.*, 2014; *Anyah et al.*, 2018).

FIGURE 5

Similar assessments are performed between the soil moisture and surface water storage changes with the ENSO index and the results are shown in Figs. 6 and 7. Correlation coefficients of up to 0.8 are computed from the DMDA estimates with a maximum lag of up to 2 months. Among the six models, correlation in soil moisture of the SURFEX-TRIP and LISFLOOD models is found to be the highest, i.e., correlations of 0.6 to 0.8 within the 33 river basins examined here. PCR-GLOBWB and W3RA show a correlation of $\sim$ 0.5, while those from HBV-

SIMREG and ORCHIDEE are different from our other estimations, for example, less than 0.1 in the Niger and Nile River Basins, and greater than 0.75 in North Asia. *Khaki et al.* (2018b) indicate that over the Nile River Basin, all the three hydrological components, (i.e., groundwater, surface water, and soil moisture) are strongly influenced by ENSO. Therefore, the obtained correlation of 0.1 in the Nile River Basin from HBV-SIMREG is likely not realistic.

FIGURE 6

The DMDA-derived surface water storage is compared with those of PCR-GLOBWB, SURFEX-TRIP, and ORCHIDEE, which contain the surface water storage compartment. The correlation coefficients are found to be generally smaller than those of soil moisture and groundwater components (with a maximum of 0.5), which likely shows that the modelling of surface water needs improvement because in reality surface water in lakes and rivers within regions like East Africa shows an immediate response to ENSO (e.g., *Becker et al.*, 2010; *Khaki et al.*, 2018b). Figure 7 shows that the surface water storage output of SURFEX-TRIP had the highest correlations with the ENSO index in all basins of America (values between 0.33 and 0.51) and Africa (values between 0.23 and 0.48), while ORCHIDEE shows the highest correlations (values between 0.32 and 0.58) in most parts of Asia. The correlations for PCR-GLOBWB are found to be relatively smaller, i.e., between 0.1 and 0.2 with lags of between 5-12 months. Comparisons between the DMDA and original model outputs indicate that combining models with GRACE data improve the correlations with the ENSO index and the correlation lags are considerably reduced globally. It is worth mentioning that the DMDA results that are presented here are derived by setting the α value in equation (14) to 0.9. This means that we assume a 36 month temporal correlations between water storage simulations of the six models. This value guarantee an extraction of the ENSO modes within two PCA modes after merging GRACE and model outputs.

FIGURE 7

4.4. Evaluating the DMDA Results with satellite altimetry observation

To validate our results, TWS and surface water derived from DMDA and six hydrological models are compared with independent surface water observations from satellite altimetry. The results are shown for various regions with reliable satellite altimetry measurements such as the Nile, Niger, and Zambezi River Basins in Africa, Ob and Euphrates in Asia, St' Lawrence and Nelson in North America, and Orinoco in South Africa. Here, we assessed 14 lakes located in the 8 mentioned river basins.

Comparisons are performed in terms of correlation coefficients between TWS and surface water estimates (within the river basins), and water mass variations within the lakes (i.e., lake level heights from satellite altimetry data are converted to mass variations following *Moore and Williams* (2014)). The numerical results are summarized in Table 5, which indicates that after implementing the DMDA method, correlation coefficients are increased in most of the lakes. High values are found in the Nile River Basin, e.g., Tana Lake (0.718), Euphrates (Tharthar Lake, 0.569), and Niger (Chad Lake, 0.558), while low values are found in the Kainiji Lake of the Niger River Basin (0.102) and Winnipegosis of the Nelson River Basins (0.249). It should be noted here that although low correlations are found for some lakes, the values are increased when compared with the original model simulations. More details can be found in ESM-section 7.

TABLE 5

5. Summary and Conclusion

In this study, the method of Dynamic Model Data Averaging (DMDA) is introduced, which can be used (1) to compare multi-model (individual) water storage simulations with GRACE-derived Terrestrial Water Storage (TWS) estimates; and (2) to separate GRACE TWS into horological water storage compartments. DMDA combines the property of Kalman Filter (equations (9), (10)) and a Bayesian weighting (equation (11)) to fit multi-model water storage changes to GRACE TWS estimates. The method is flexible in accounting for errors in observations and a priori information (equation 9 and equation 10), and can deal with state vectors of different length.

The benefit of the DMDA method over the commonly used PF or PS methods are twofold: 1) these methods might not be efficient for high-dimensional fusion tasks (e.g., *Snyder et al.*, 2008; *Van Leeuwen*, 2009) such as the global hydrological application presented here, but the DMDA's computational load is lower than these techniques; 2) DMDA provides time-variable weights that can be used to understand the behavior of a priori information (here the output of hydrological models) against GRACE TWS estimates, while considering their errors. The advantage of the DMDA over the Ensemble Kalman Filter-based of techniques is that the posterior distributions are computed through a Bayesian rule that result in more reliable estimations of states and their errors, while avoiding the high computational loads of the PF techniques.

A realistic synthetic example was defined to evaluate the performance of DMDA (Fig. 2), which showed that the method is able to correctly separate GRACE TWS estimates into its individual hydrological components. We also showed that the DMDA's estimation of temporal weights (for each model) was close to the reality, and can be used to assess the performance of available models. Based on the real data, we showed that the representation of linear trends and seasonality within global hydrological models, as well as their water storage changes due to the El Niño Southern Oscillation (ENSO) can be improved using DMDA, while considering the uncertainties of models and observations (see Fig. 1). Our results also showed that how the DMDA method is able to deal with models with different structures, and how it updates their water storage simulations while considering their errors. Considering these arguments, we believe that the new water storage estimates, i.e., models combined with GRACE, are of great values and can be used for further hydrological and climate research investigations compared to model or GRACE only estimates. Therefore, the presented results can be considered as one step forward to improve model deficiencies following the insights of *Scanlon et al.* (2018). In what follows, the main conclusions and remarks of this study are summarized.

- Estimated weights (Fig. 3) showed that the PCR-GLOBWB model gained the largest weights, thus, it contributed the most in the DMDA-derived TWS in North Asia, North America, and the center of Africa. SURFEX-TRIP performed best within basins with dominant surface water storage changes, as well as in snow-dominant regions. The LISFLOOD and ORCHIDEE models were found to perform well within irrigated basins, and those affected by ENSO events.

- DMDA results in Fig. 4 (a1) showed that considerable trends exist in groundwater storage changes within the Ganges, Indus, and Euphrates basins during 2002–2012. These changes are dominantly influenced by anthropogenic modifications. Trends in soil moisture (Fig. 4 (b1)) were found to be mostly related to meteorological prolonged drought events such as those in the Brahmaputra and Euphrates River Basins.

- DMDA was able to modify the ENSO mode of water storage variability in most of the world's 33 largest river basins (see Fig. 5, Fig. 6, and Fig. 7). DMDA assigned the biggest corrections of ENSO mode in groundwater to the Nile, Murray, Tocantins, Ob, Okavango and Orange River Basins. The highest corrections of the ENSO mode in soil moisture were found for the Nile, Niger, Zambezi, and Amur River Basins, and in surface water to Nile, Niger, Congo, Tocantins, and Murray River Basin. For example, the correlation coefficient between groundwater storage and ENSO in the Murray River Basin changed from -0.2 to 0.6. For the Nile River Basin, they changed from 0.1 to 0.4 for soil moisture, and from 0.3 to 0.7 for the surface water compartment.

- Comparison between TWS and surface water derived from DMDA with independent surface water observations from satellite altimetry (Fig. ESM.15 and Fig. ESM.16 in ESM-section 7) showed that, DMDA was able to correctly detect the best performing model and maximize its contribution in the dynamic averaging process which enhanced the reality of water storage estimates.

- To implement the DMDA in this study a forgetting factor of 0.95 was considered in equation (6), which is equivalent to the temporal dependency in estimating time variable regression parameters in equation (2). In section 3, it was shown that this selection is equivalent to 18 months temporal dependency between GRACE TWS observations and model simulations. This value is selected because the DMDA results were closest to that of GRACE. After selecting this value, we also obtained a distinguishable ENSO mode from the DMDA-derived TWS and individual water storage estimates. Therefore, we conclude that this temporal lag might be considered in other works that attempt to apply sequential mergers or smoothers to assimilate observed water storage data into models.

- In order to reduce the computational load of this work, instead of implementing a Markov Chain Monte Carlo (MCMC) technique to estimate the transition matrix between models in equation (13), a forgetting factor of 0.9 was considered in equations (14). This might be replaced with an efficient MCMC implementation in future.

The DMDA method, introduced in this study, has the potential to be used in different climate and hydrological applications to compare available models (which can be of various types of hydrological or climate models) against reliable observations. It can also be used to generate ensembles from multi-model outputs such as climate projections. The application of this study can also be extended by incorporating other types of remote sensing observations such as satellite based soil moisture or water level data beside those of GRACE. A secondary application of the DMDA can also be devoted to its application for predicting (or extrapolating) water storage estimates. To achieve this purpose, however, the DMDA's formulation needs to be extended. For example, one approach can be to use the DMDA weights, which are computed for the period of study, to identify best models in different river basins covering different seasons. By analysing this information and knowing the TWS in the future, one can use a combination of different model runs (weighted by the DMDA outputs) and extrapolate the surface and sub-surface water storage estimates.

References

Anyah, R. O., E. Forootan, J. L. Awange, and M. Khaki (2018), Understanding linkages between global climate indices and terrestrial water storage changes over africa using grace products, *Science of The Total Environment*, *635*, 1405–1416, doi:10.1016/j.scitotenv.2018.04.159.

Awange, J. L., E. Forootan, M. Kuhn, J. Kusche, and B. Heck (2014), Water storage changes and climate variability within the nile basin between 2002 and 2011, *Advances in Water Resources*, *73*, 1–15, doi:10.1016/j.advwatres.2014.06.010.

Bai, P., X. Liu, and C. Liu (2018), Improving hydrological simulations by incorporating grace data for model calibration, *Journal of Hydrology*, *557*, 291–304, doi:10.1016/j.jhydrol.2017.12.025.

Bain, A., and D. Crisan (2008), *Fundamentals of stochastic filtering*, vol. 60, Springer Science & Business Media, doi:10.1007/978-0-387-76896-0.

Banerjee, S., B. P. Carlin, and A. E. Gelfand (2004), *Hierarchical modeling and analysis for spatial data*, Chapman and Hall/CRC.

Barnston, A. G., and R. E. Livezey (1987), Classification, seasonality and persistence of low-frequency atmospheric circulation patterns, *Monthly weather review*, *115*(6), 1083–1126, doi:10.1175/1520-0493(1987)115⟨1083:CSAPOL⟩2.0.CO;2.

Becker, M., W. LLovel, A. Cazenave, A. Güntner, and J.-F. Crétaux (2010), Recent hydrological behavior of the east african great lakes region inferred from grace, satellite altimetry and rainfall observations, *Comptes Rendus Geoscience*, *342*(3), 223–233, doi:10.1016/j.crte.2009.12.010.

Bengtsson, T., P. Bickel, B. Li, et al. (2008), Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems, in *Probability and statistics: Essays in honor of David A. Freedman*, pp. 316–334, Institute of Mathematical Statistics, doi:10.1214/193940307000000518.

Bernstein, D. S. (2005), *Matrix mathematics: Theory, facts, and formulas with application to linear systems theory*, vol. 41, Princeton university press Princeton.

Boening, C., J. K. Willis, F. W. Landerer, R. S. Nerem, and J. Fasullo (2012), The 2011 la niña: So strong, the oceans fell, *Geophysical Research Letters*, *39*(19), doi:10.1029/2012GL053055.

Chan, K. S., and C. J. Geyer (1994), Discussion: Markov chains for exploring posterior distributions, *The Annals of Statistics*, *22*(4), 1747–1758, doi:10.1214/aos/1176325754.

Chen, J. L., C. R. Wilson, J. S. Famiglietti, and M. Rodell (2007), Attenuation effect on seasonal basin-scale water storage changes from grace time-variable gravity, *Journal of Geodesy*, *81*(4), 237–245, doi:10.1007/s00190-006-0104-2.

Chen, J. L., C. R. Wilson, B. D. Tapley, Z. L. Yang, and G. Y. Niu (2009), 2005 drought event in the amazon river basin as measured by grace and estimated by climate models, *Journal of Geophysical Research: Solid Earth*, *114*(B5), doi:10.1029/2008JB006056.

Chen, J. L., C. R. Wilson, and B. D. Tapley (2010), The 2009 exceptional amazon flood and interannual terrestrial water storage change observed by grace, *Water Resources Research*, *46*(12), doi:10.1029/2010WR009383.

Daum, F., and J. Huang (2003), Curse of dimensionality and particle filters, in *2003 IEEE Aerospace Conference Proceedings (Cat. No. 03TH8652)*, vol. 4, pp. 4_1979–4_1993, IEEE, doi:10.1109/AERO.2003.1235126.

Decharme, B., E. Martin, and S. Faroux (2013), Reconciling soil thermal and hydrological lower boundary conditions in land surface models, *Journal of Geophysical Research: Atmospheres*, *118*(14), 7819–7834, doi:10.1002/jgrd.50631.

Del Moral, P., and L. Miclo (2000), Branching and interacting particle systems approximations of feynman-kac formulae with applications to non-linear filtering, in *Seminaire de probabilites XXXIV*, pp. 1–145, Springer, doi:10.1007/BFb0103798.

Döll, P., F. Kaspar, and B. Lehner (2003), A global hydrological model for deriving water availability indicators: model tuning and validation, *Journal of Hydrology*, *270*(1-2), 105–134, doi:10.1016/S0022-1694(02)00283-4.

Duan, Q., N. K. Ajami, X. Gao, and S. Sorooshian (2007), Multi-model ensemble hydrologic prediction using bayesian model averaging, *Advances in Water Resources*, *30*(5), 1371–1386, doi:10.1016/j.advwatres.2006.11.014.

Dzhamalov, R. G., G. N. Krichevets, and T. I. Safronova (2012), Current changes in water resources in lena river basin, *Water Resources*, *39*(2), 147–160, doi:10.1134/S0097807812020042.

Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics, *Journal of Geophysical Research: Oceans*, *99*(C5), 10,143–10,162, doi:10.1029/94JC00572.

Evensen, G. (2003), The ensemble kalman filter: Theoretical formulation and practical implementation, *Ocean dynamics*, *53*(4), 343–367, doi:10.1007/s10236-003-0036-9.

Fagin, S. L. (1964), Recursive linear regression theory, optimal filter theory, and error analysis of optimal systems, in *IEEE international convention record*, vol. 12, pp. 216–245.

Fang, H., N. Tian, Y. Wang, M. Zhou, and M. A. Haile (2018), Nonlinear bayesian estimation: from kalman filtering to a broader horizon, *IEEE/CAA Journal of Automatica Sinica*, *5*(2), 401–417, doi:10.1109/JAS.2017.7510808.

Feng, W., M. Zhong, J. M. Lemoine, R. Biancale, H.-T. Hsu, and J. Xia (2013), Evaluation of groundwater depletion in north china using the gravity recovery and climate experiment (grace) data and ground-based measurements, *Water Resources Research*, *49*(4), 2110–2118, doi:10.1002/wrcr.20192.

Forootan, E., and J. Kusche (2012), Separation of global time-variable gravity signals into maximally independent components, *Journal of Geodesy*, *86*(7), 477–497, doi:0.1007/s00190-011-0532-5.

Forootan, E., and J. Kusche (2013), Separation of deterministic signals using independent component analysis (ica), *Studia Geophysica et Geodaetica*, *57*(1), 17–26, doi:10.1007/s11200-012-0718-1.

Forootan, E., J. Awange, J. Kusche, B. Heck, and A. Eicker (2012), Independent patterns of water mass anomalies over australia from satellite data and models, *Remote sensing of environment*, *124*, 427–443, doi:10.1016/j.rse.2012.05.023.

Forootan, E., O. Didova, J. Kusche, and A. Löcher (2013), Comparisons of atmospheric data and reduction methods for the analysis of satellite gravimetry observations, *Journal of Geophysical Research: Solid Earth*, *118*(5), 2382–2396, doi:10.1002/jgrb.50160.

Forootan, E., R. Rietbroek, J. Kusche, M. A. Sharifi, J. L. Awange, M. Schmidt, P. Omondi, and J. S. Famiglietti (2014), Separation of large scale water storage patterns over iran using grace, altimetry and hydrological data, *Remote Sensing of Environment*, *140*, 580–595, doi:10.1016/j.rse.2013.09.025.

Forootan, E., J. L. Awange, M. Schumacher, R. O. Anyah, A. I. J. M. van Dijk, J. Kusche, et al. (2016), Quantifying the impacts of enso and iod on rain gauge and remotely sensed precipitation products over australia, *Remote sensing of Environment*, *172*, 50–66, doi:10.1016/j.rse.2015.10.027.

Forootan, E., A. Safari, A. Mostafaie, M. Schumacher, M. Delavar, and J. L. Awange (2017), Large-scale total water storage and water flux changes over the arid and semiarid parts of the middle east from grace and reanalysis products, *Surveys in Geophysics*, *38*(3), 591–615, doi:10.1007/s10712-016-9403-1.

Forootan, E., J. Kusche, M. Talpe, C. K. Shum, and M. Schmidt (2018), Developing a complex independent component analysis (cica) technique to extract non-stationary patterns from geophysical time series, *Surveys in Geophysics*, *39*(3), 435–465, doi:10.1007/s10712-017-9451-1.

Forootan, E., M. Khaki, M. Schumacher, V. Wulfmeyer, N. Mehrnegar, A. I. J. M. van Dijk, L. Brocca, S. Farzaneh, F. Akinluyi, G. Ramillien, C. Shum, J. L. Awange, and A. Mostafaie (2019), Understanding the global hydrological droughts of 2003–2016 and their relationships with teleconnections, *Science of The Total Environment*, *650*, 2587–2604, doi:10.1016/j.scitotenv.2018.09.231.

Frappart, F., and G. Ramillien (2018), Monitoring groundwater storage changes using the gravity recovery and climate experiment (grace) satellite mission: A review, *Remote Sensing*, *10*(6), 829, doi:10.3390/rs10060829.

Geyer, C. J. (2011), Introduction to markov chain monte carlo, *Handbook of Markov Chain Monte Carlo*, p. 46.

Girotto, M., G. J. De Lannoy, R. H. Reichle, and M. Rodell (2016), Assimilation of gridded terrestrial water storage observations from grace into a land surface model, *Water Resources Research*, *52*(5), 4164–4183, doi:10.1002/2015WR018417.

Girotto, M., G. J. De Lannoy, R. H. Reichle, M. Rodell, C. Draper, S. N. Bhanja, and A. Mukherjee (2017), Benefits and pitfalls of grace data assimilation: A case study of terrestrial water storage depletion in india, *Geophysical research letters*, *44*(9), 4107–4115, doi:10.1002/2017GL072994.

Gordon, N. J., D. J. Salmond, and A. F. Smith (1993), Novel approach to nonlinear/non-gaussian bayesian state estimation, in *IEE proceedings F (radar and signal processing)*, vol. 140, pp. 107–113, IET, doi:10.1049/ip-f-2.1993.0015.

Hanington, P., Q. T. To, P. D. T. Van, N. A. V. Doan, and A. S. Kiem (2017), A hydrological model for interprovincial water resource planning and management: a case study in the long xuyen quadrangle, mekong delta, vietnam, *Journal of Hydrology*, *547*, 1–9, doi:10.1016/j.jhydrol.2017.01.030.

Hannan, E. J., A. McDougall, and D. Poskitt (1989), Recursive estimation of autoregressions, *Journal of the Royal Statistical Society. Series B (Methodological)*, *51*(2), 217–233, doi:10.1111/j.2517-6161.1989.tb01759.x.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999), Bayesian model averaging: A tutorial, *Statistical Science*, *14*(4), 382–417.

Hsu, K.-l., H. Moradkhani, and S. Sorooshian (2009), A sequential bayesian approach for hydrologic model selection and prediction, *Water Resources Research*, *45*(12), doi:10.1029/2008WR006824.

Hurkmans, R., P. A. Troch, R. Uijlenhoet, P. Torfs, and M. Durcik (2009), Effects of climate variability on water storage in the colorado river basin, *Journal of hydrometeorology*, *10*(5), 1257–1270, doi:10.1175/2009JHM1133.1.

Jazwinski, A. H. (2007), *Stochastic processes and filtering theory*, Courier Corporation.

Khaki, M., B. Ait-El-Fquih, I. Hoteit, E. Forootan, J.L., and M. Kuhn (2017a), A two-update ensemble kalman filter for land hydrological data assimilation with an uncertain constraint, *Journal of Hydrology*, *555*, 447–462, doi:10.1016/j.jhydrol.2017.10.032.

Khaki, M., I. Hoteit, M. Kuhn, J. L. Awange, E. Forootan, A. I. J. M. Van Dijk, M. Schumacher, and C. Pattiaratchi (2017b), Assessing sequential data assimilation techniques for integrating grace data into a hydrological model, *Advances in Water Resources*, *107*, 301–316, doi:10.1016/j.advwatres.2017.07.001.

Khaki, M., M. Schumacher, E. Forootan, M. Kuhn, J. L. Awange, and A. I. J. M. van Dijk (2017c), Accounting for spatial correlation errors in the assimilation of grace into hydrological models through localization, *Advances in Water Resources*, *108*, 99–112, doi:10.1016/j.advwatres.2017.07.024.

Khaki, M., B. Ait-El-Fquih, I. Hoteit, E. Forootan, J. L. Awange, and M. Kuhn (2018a), Unsupervised ensemble kalman filtering with an uncertain constraint for land hydrological data assimilation, *Journal of Hydrology*, *564*, 175–190, doi:10.1016/j.jhydrol.2018.06.080.

Khaki, M., J. L. Awange, E. Forootan, and M. Kuhn (2018b), Understanding the association between climate variability and the nile's water level fluctuations and water storage changes during 1992–2016, *Science of The Total Environment*, *645*, 1509–1521, doi:10.1016/j.scitotenv.2018.07.212.

Khaki, M., E. Forootan, M. Kuhn, J. L. Awange, L. Longuevergne, and Y. Wada (2018c), Efficient basin scale filtering of grace satellite products, *Remote Sensing of Environment*, *204*, 76–93, doi:10.1016/j.rse.2017.10.040.

Khaki, M., E. Forootan, M. Kuhn, J. L. Awange, F. Papa, and C. K. Shum (2018d), A study of bangladesh's sub-surface water storages using satellite products and data assimilation scheme, *Science of the Total Environment*, *625*, 963–977, doi:10.1016/j.scitotenv.2017.12.289.

Khaki, M., E. Forootan, M. Kuhn, J. L. Awange, A. I. J. M. van Dijk, M. Schumacher, and M. A. Sharifi (2018e), Determining water storage depletion within iran by assimilating grace data into the w3ra hydrological model, *Advances in Water Resources*, *114*, 1–18, doi:10.1016/j.advwatres.2018.02.008.

Khandu, K., E. Forootan, M. Schumacher, J. L. Awange, and H. Müller-Schmied (2016), Exploring the influence of precipitation extremes and human water use on total water storage (tws) changes in the ganges-brahmaputra-meghna river basin, *Water Resources Research*, *52*(3), 2240–2258, doi:10.1002/2015WR018113.

Kousky, V. E., M. T. Kagano, and I. F. Cavalcanti (1984), A review of the southern oscillation: oceanic-atmospheric circulation changes and related rainfall anomalies, *Tellus A*, *36*(5), 490–504, doi:10.1111/j.1600-0870.1984.tb00264.x.

Kuczera, G., and E. Parent (1998), Monte carlo assessment of parameter uncertainty in conceptual catchment models: the metropolis algorithm, *Journal of Hydrology*, *211*(1-4), 69–85, doi:10.1016/S0022-1694(98)00198-X.

Kusche, J., R. Schmidt, S. Petrovic, and R. Rietbroek (2009), Decorrelated grace time-variable gravity solutions by gfz, and their validation using a hydrological model, *Journal of geodesy*, *83*(10), 903–913, doi:10.1007/s00190-009-0308-3.

Lindström, G., B. Johansson, M. Persson, M. Gardelin, and S. Bergström (1997), Development and test of the distributed hbv-96 hydrological model, *Journal of hydrology*, *201*(1-4), 272–288, doi:10.1016/S0022-1694(97)00041-3.

Long, D., Y. Pan, J. Zhou, Y. Chen, X. Hou, Y. Hong, B. R. Scanlon, and L. Longuevergne (2017), Global analysis of spatiotemporal variability in merged total water storage changes using multiple grace products and global hydrological models, *Remote sensing of environment*, *192*, 198–216, doi:10.1016/j.rse.2017.02.011.

Lorenz, E. N. (1956), Empirical orthogonal functions and statistical weather prediction.

Marengo, J. A., C. A. Nobre, G. Sampaio, L. F. Salazar, and L. S. Borma (2011), Climate change in the amazon basin: Tipping points, changes in extremes, and impacts on natural and human systems, in *Tropical rainforest responses to climatic change*, pp. 259–283, Springer, doi:10.1007/978-3-642-05383-2_9.

McCarthy, T. S., A. Bloem, and P. A. Larkin (1998), Observations on the hydrology and geohydrology of the okavango delta, *South African Journal of Geology/Suid-Afrikaanse Tydskrif vir Geologie*, *101*(2), 101–117.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953), Equation of state calculations by fast computing machines, *The journal of chemical physics*, *21*(6), 1087–1092, doi:10.1063/1.1699114.

Moore, P., and S. D. P. Williams (2014), Integration of altimetric lake levels and grace gravimetry over africa: Inferences for terrestrial water storage change 2003–2011, *Water Resources Research*, *50*(12), 9696–9720, doi:10.1002/2014WR015506.

Mostafaie, A., E. Forootan, A. Safari, and M. Schumacher (2018), Comparing multi-objective optimization techniques to calibrate a conceptual hydrological model using in situ runoff and daily grace data, *Computational Geosciences*, *22*(3), 789–814, doi:10.1007/s10596-018-9726-8.

Ni, S., J. Chen, C. R. Wilson, J. Li, X. Hu, and R. Fu (2018), Global terrestrial water storage changes and connections to enso events, *Surveys in Geophysics*, *39*(1), 1–22, doi:10.1007/s10712-017-9421-7.

Niu, G.-Y., Z.-L. Yang, K. E. Mitchell, F. Chen, M. B. Ek, M. Barlage, A. Kumar, K. Manning, D. Niyogi, E. Rosero, et al. (2011), The community noah land surface model with multiparameterization options (noah-mp): 1. model description and evaluation with local-scale measurements, *Journal of Geophysical Research: Atmospheres*, *116*(D12), doi:10.1029/2010JD015139.

Phillips, T., R. Nerem, B. Fox-Kemper, J. Famiglietti, and B. Rajagopalan (2012), The influence of enso on global terrestrial water storage using grace, *Geophysical Research Letters*, *39*(16), doi:10.1029/2012GL052495.

Plaza Guingla, D. A., R. De Keyser, G. J. De Lannoy, L. Giustarini, P. Matgen, and V. R. Pauwels (2013), Improving particle filters in rainfall-runoff models: Application of the resample-move step and the ensemble gaussian particle filter, *Water Resources Research*, *49*(7), 4005–4021, doi:10.1002/wrcr.20291.

Polcher, J., N. Bertrand, H. Biemans, D. B. Clark, M. Floerke, N. Gedney, D. Gerten, T. Stacke, M. Van Vliet, and F. Voss (2011), Improvements in hydrological processes in general hydrological models and land surface models within watch.

Raftery, A. (1993), *Model selection and accounting for model uncertainty in linear regression models*, Citeseer.

Raftery, A. E., M. Kárnỳ, and P. Ettler (2010), Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill, *Technometrics*, *52*(1), 52–66, doi:10.1198/TECH.2009.08104.

Ramsay, B. H. (1998), The interactive multisensor snow and ice mapping system, *Hydrological Processes*, *12*(10-11), 1537–1546, doi:10.1002/(SICI)1099-1085(199808/09)12:10/11⟨1537::AID-HYP679⟩3.0.CO;2-A.

Rodell, M., P. R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C.-J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, et al. (2004), The global land data assimilation system, *Bulletin of the American Meteorological Society*, *85*(3), 381–394, doi:10.1175/BAMS-85-3-381.

Rodell, M., I. Velicogna, and J. S. Famiglietti (2009), Satellite-based estimates of groundwater depletion in india, *Nature*, *460*(7258), 999, doi:10.1038/nature08238.

Ropelewski, C. F., and M. S. Halpert (1987), Global and regional scale precipitation patterns associated with the el niño/southern oscillation, *Monthly weather review*, *115*(8), 1606–1626, doi:10.1175/1520-0493(1987)115⟨1606:GARSPP⟩2.0.CO;2.

Särkkä, S. (2013), *Bayesian filtering and smoothing*, vol. 3, Cambridge University Press, doi:10.1201/b16018.

Scanlon, B. R., Z. Zhang, H. Save, A. Y. Sun, H. M. Schmied, L. P. van Beek, D. N. Wiese, Y. Wada, D. Long, R. C. Reedy, et al. (2018), Global models underestimate large decadal declining and rising water storage trends relative to grace satellite data, *Proceedings of the National Academy of Sciences*, *115*(6), E1080–E1089, doi:10.1073/pnas.1704665115.

Schellekens, J., E. Dutra, A. Martínez-de la Torre, G. Balsamo, A. van Dijk, F. S. Weiland, M. Minvielle, J.-C. Calvet, B. Decharme, S. Eisner, et al. (2017), A global water resources ensemble of hydrological models: the earth2observe tier-1 dataset, *Earth System Science Data*, *9*(2), 389–413, doi:10.5194/essd-9-389-2017.

Schmeer, M., M. Schmidt, W. Bosch, and F. Seitz (2012), Separation of mass signals within grace monthly gravity field models by means of empirical orthogonal functions, *Journal of Geodynamics*, *59*, 124–132, doi:10.1016/j.jog.2012.03.001.

Schumacher, M. (2016), Methods for assimilating remotely-sensed water storage changes into hydrological models, Ph.D. thesis, Universitäts-und Landesbibliothek Bonn.

Schumacher, M., J. Kusche, and P. Döll (2016), A systematic impact assessment of grace error correlation on data assimilation in hydrological models, *Journal of Geodesy*, *90*(6), 537–559, doi:10.1007/s00190-016-0892-y.

Schumacher, M., E. Forootan, A. I. J. M. van Dijk, H. M. Schmied, R. S. Crosbie, J. Kusche, and P. Döll (2018), Improving drought simulations within the murray-darling basin by combined calibration/assimilation of grace data into the watergap global hydrology model, *Remote Sensing of Environment*, *204*, 212–228, doi:10.1016/j.rse.2017.10.029.

Sha, Z., J. Rougier, M. Schumacher, and J. Bamber (2018), Bayesian model-data synthesis with an application to global glacio-isostatic adjustment, *Environmetrics*, doi:10.1002/env.2530.

Sheffield, J., E. F. Wood, and M. L. Roderick (2012), Little change in global drought over the past 60 years, *Nature*, *491*(7424), 435, doi:10.1038/nature11575.

Singer, M. B., K. Michaelides, and D. E. Hobley (2018), Storm 1.0: a simple, flexible, and parsimonious stochastic rainfall generator for simulating climate and climate change, *Geoscientific Model Development*, *11*(9), 3713–3726, doi:10.5194/gmd-11-3713-2018.

Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson (2008), Obstacles to high-dimensional particle filtering, *Monthly Weather Review*, *136*(12), 4629–4640, doi:10.1175/2008MWR2529.1.

Strassberg, G., B. R. Scanlon, and D. Chambers (2009), Evaluation of groundwater storage monitoring with the grace satellite: Case study of the high plains aquifer, central united states, *Water Resources Research*, *45*(5), doi:10.1029/2008WR006892.

Swenson, S., D. Chambers, and J. Wahr (2008), Estimating geocenter variations from a combination of grace and ocean model output, *Journal of Geophysical Research: Solid Earth*, *113*(B8), doi:10.1029/2007JB005338.

Tapley, B. D., S. Bettadpur, J. C. Ries, P. F. Thompson, and M. M. Watkins (2004), Grace measurements of mass variability in the earth system, *Science*, *305*(5683), 503–505, doi:10.1126/science.1099192.

Tian, S., P. Tregoning, L. J. Renzullo, A. I. J. M. van Dijk, J. P. Walker, V. R. Pauwels, and S. Allgeyer (2017), Improved water balance component estimates through joint assimilation of grace water storage and smos soil moisture retrievals, *Water Resources Research*, *53*(3), 1820–1840, doi:10.1002/2016WR019641.

Tiwari, V., J. Wahr, and S. Swenson (2009), Dwindling groundwater resources in northern india, from satellite gravity observations, *Geophysical Research Letters*, *36*(18), doi:10.1029/2009GL039401.

Trenberth, K. E. (1990), Recent observed interdecadal climate changes in the northern hemisphere, *Bulletin of the American Meteorological Society*, *71*(7), 988–993, doi:10.1175/1520-0477(1990)071⟨0988:ROICCI⟩2.0.CO;2.

Uebbing, B., J. Kusche, and E. Forootan (2015), Waveform retracking for improving level estimations from topex/poseidon, jason-1, and jason-2 altimetry observations over african lakes, *IEEE Transactions on Geoscience and Remote Sensing*, *53*(4), 2211–2224, doi:10.1109/TGRS.2014.2357893.

Van Beek, L. P. H., Y. Wada, and M. F. P. Bierkens (2011), Global monthly water stress: 1. water balance and water availability, *Water Resources Research*, *47*(7), doi:10.1029/2010WR009791.

Van Der Knijff, J. M., J. Younis, and A. P. J. De Roo (2010), Lisflood: a gis-based distributed model for river basin scale water balance and flood simulation, *International Journal of Geographical Information Science*, *24*(2), 189–212, doi:10.1080/13658810802549154.

Van Dijk, A. I. J. M. (2010), The australian water resources assessment system: Technical 901 report 3, landscape model (version 0.5) technical description, csiro, *Water for a Healthy Country National Research Flagship*.

Van Dijk, A. I. J. M., L. J. Renzullo, Y. Wada, and P. Tregoning (2014), A global water cycle reanalysis (2003–2012) merging satellite gravimetry and altimetry observations with a hydrological multi-model ensemble, *Hydrology and Earth System Sciences*, *18*(8), 2955–2973, doi:10.5194/hess-18-2955-2014.

Van Leeuwen, P. J. (2009), Particle filtering in geophysical systems, *Monthly Weather Review*, *137*(12), 4089–4114, doi:10.1175/2009MWR2835.1.

Voss, K. A., J. S. Famiglietti, M. Lo, C. De Linage, M. Rodell, and S. C. Swenson (2013), Groundwater depletion in the middle east from grace with implications for transboundary water management in the tigris-euphrates-western iran region, *Water resources research*, *49*(2), 904–914, doi:10.1002/wrcr.20078.

Wada, Y., L. P. H. van Beek, F. C. S. Weiland, B. F. Chao, Y.-H. Wu, and M. F. P. Bierkens (2012), Past and future contribution of global groundwater depletion to sea-level rise, *Geophysical Research Letters*, *39*(9), doi:10.1029/2012GL051230.

Wada, Y., D. Wisser, and M. Bierkens (2014), Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources, *Earth System Dynamics Discussions*, *5*(1), 15–40, doi:10.5194/esdd-4-355-2013.

Wahr, J., and S. Zhong (2012), 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*(2), 557–572, doi:10.1093/gji/ggs030.

Wahr, J., M. Molenaar, and F. Bryan (1998), Time variability of the earth's gravity field: Hydrological and oceanic effects and their possible detection using grace, *Journal of Geophysical Research: Solid Earth*, *103*(B12), 30,205–30,229, doi:10.1029/98JB02844.

Wang, H., P. Wu, and Z. Wang (2006), An approach for spherical harmonic analysis of non-smooth data, *Computers and geosciences*, *32*(10), 1654–1668, doi:10.1016/j.cageo.2006.03.004.

Weedon, G. P., G. Balsamo, N. Bellouin, S. Gomes, M. J. Best, and P. Viterbo (2014), The wfdei meteorological forcing data set: Watch forcing data methodology applied to era-interim reanalysis data, *Water Resources Research*, *50*(9), 7505–7514, doi:10.1002/2014WR015638.

Weerts, A. H., and G. Y. El Serafy (2006), Particle filtering and ensemble kalman filtering for state updating with hydrological conceptual rainfall-runoff models, *Water resources research*, *42*(9), doi:10.1029/2005WR004093.

Werth, S., A. Güntner, S. Petrovic, and R. Schmidt (2009), Integration of grace mass variations into a global hydrological model, *Earth Planet. Sci. Lett.*, *277(1)*, 166–173, doi:10.1016/j.epsl.2008.10.021.

Zaitchik, B. F., M. Rodell, and R. H. Reichle (2008), Assimilation of grace terrestrial water storage data into a land surface model: Results for the mississippi river basin, *Journal of Hydrometeorology*, *9*(3), 535–548, doi:10.1175/2007JHM951.1.

Zhang, Z., B. Chao, J. Chen, and C. Wilson (2015), Terrestrial water storage anomalies of yangtze river basin droughts observed by grace and connections with enso, *Global and Planetary Change*, *126*, 35–45, doi:10.1016/j.gloplacha.2015.01.002.

Figures and Tables

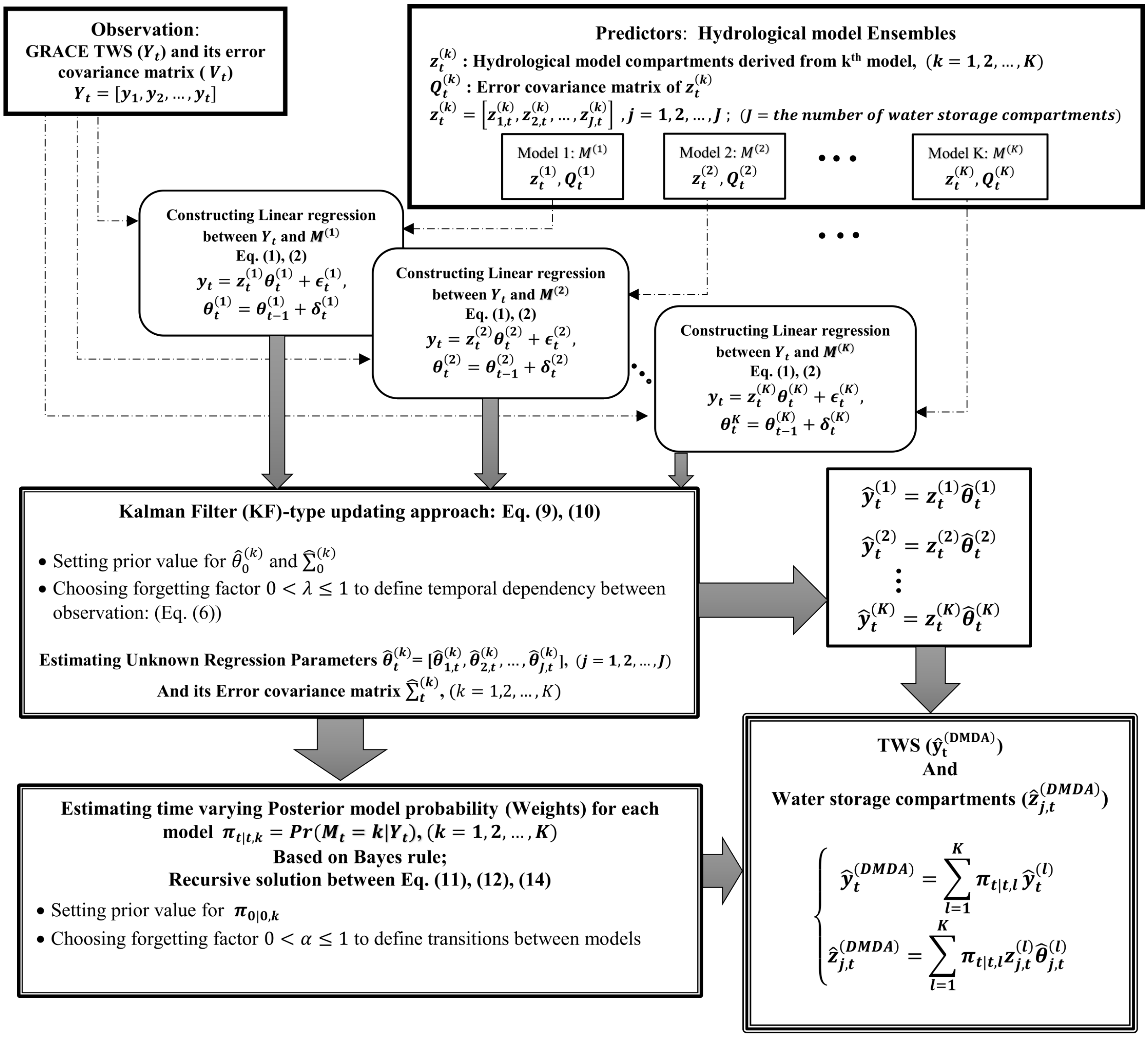

Figure 1: Flowchart of the Dynamic Model Data Averaging (DMDA) method. The framework can accept an arbitrary number of models and it can be extended to accept various type of observations.

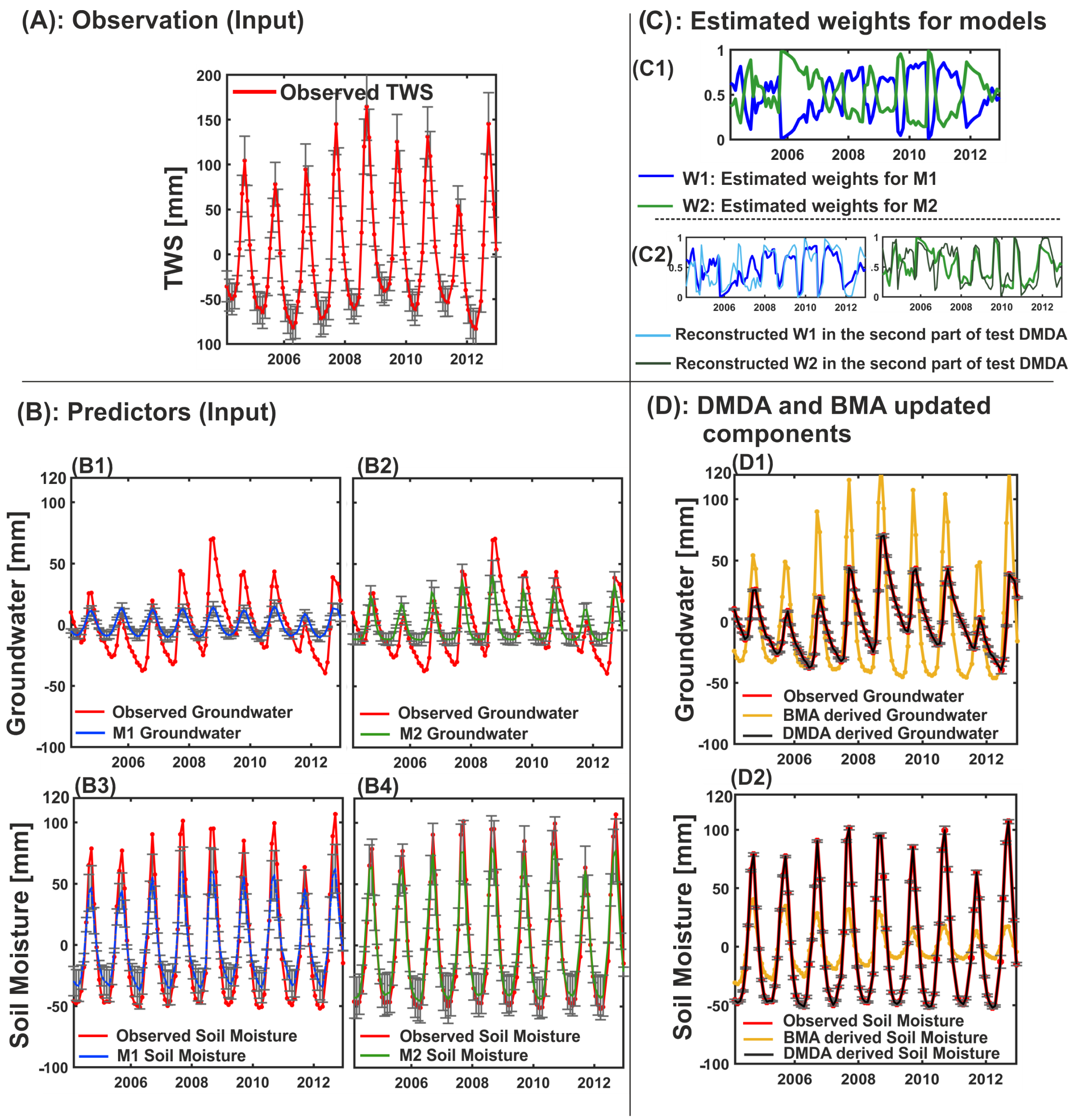

Figure 2: A synthetic example, where DMDA is applied in a controlled set up, to integrate 2 hydrological models (here selected as SURFEX-TRIP and LISFLOOD) with simulated observed TWS to separate its compartments (i.e., groundwater and soil moisture). All data sets in this simulation is related to the Niger River Basin and covering the period between 2002–2012; Figure 2 (A) shows TWS simulated from PCR-GLOBWB (here standing in for observed TWS); Figure 2 (B) shows the time series of groundwater and soil moisture derived from model 1 (B1, B3) and model 2 (B2, B4), which are considered as the input predictors in DMDA; Figure 2 (C1) presents the time varying weights estimated for two selected model,and Figure 2 (C2) shows the reconstructed of weights in the second step of our simulation. Figure 2 (D1) and (D2) show the updated hydrological components obtained from the DMDA and BMA method and comparison between the obtained results and the expected values derived form simulated observation data.

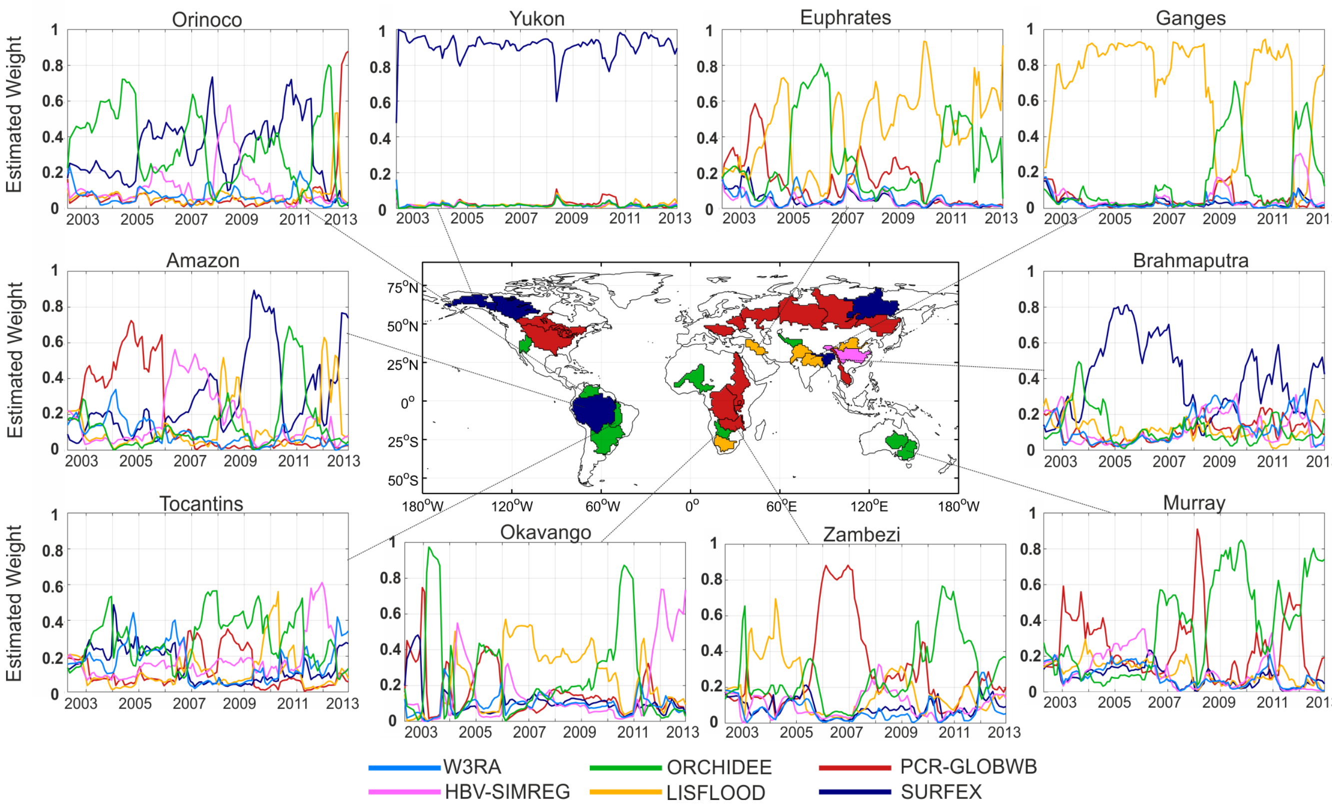

Figure 3: Posterior model probabilities for the six initially considered models, over 10 selected river basins with the biggest RMSEs computed using GRACE and models-derived TWS. In the middle of Fig 3 the most contributed models in the DMDA-derived TWS are shown over the world's 33 largest river basins, covering the period of 2002–2012.

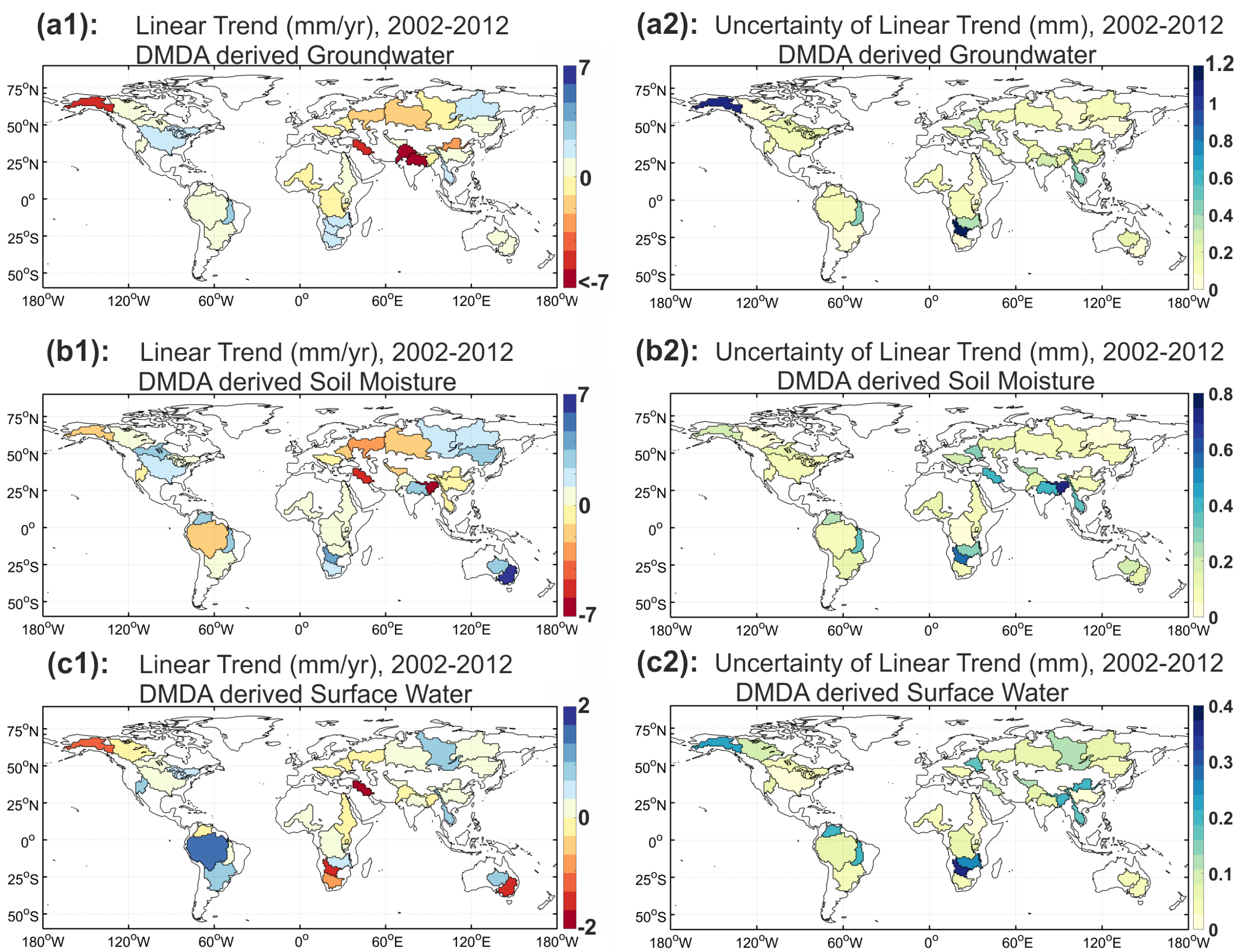

Figure 4: Long-term (2002–2012) linear trend in the DMDA-derived groundwater (a1), soil moisture (b1), and surface water (c1) components, expressed in mm/yr. The uncertainty of these fitted linear trends are shown in (a2), (b2), (c2) respectively.

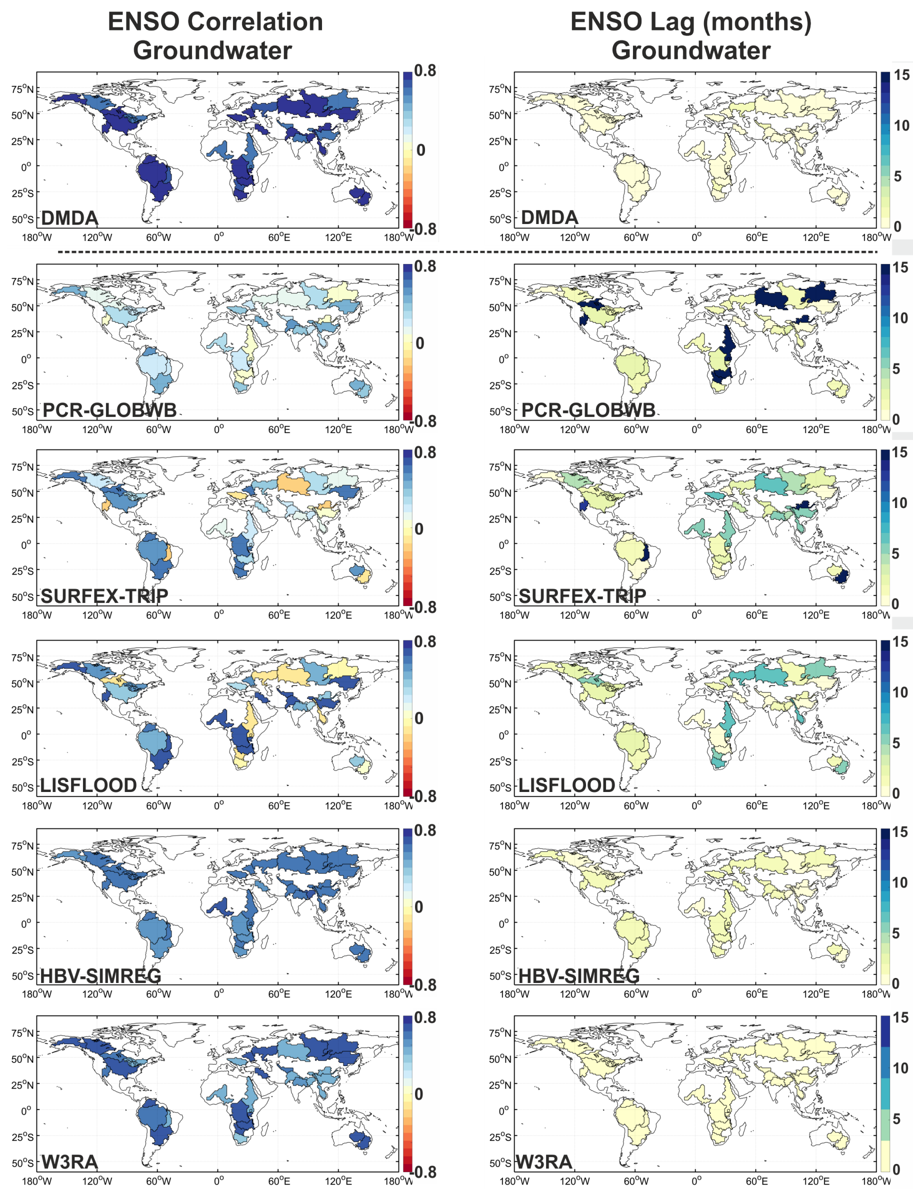

Figure 5: Correlation coefficients and their lags between the ENSO (-Niño 3.4 index) and groundwater estimates derived from the DMDA method and hydrological models used in this study for the period of 2002–2012.

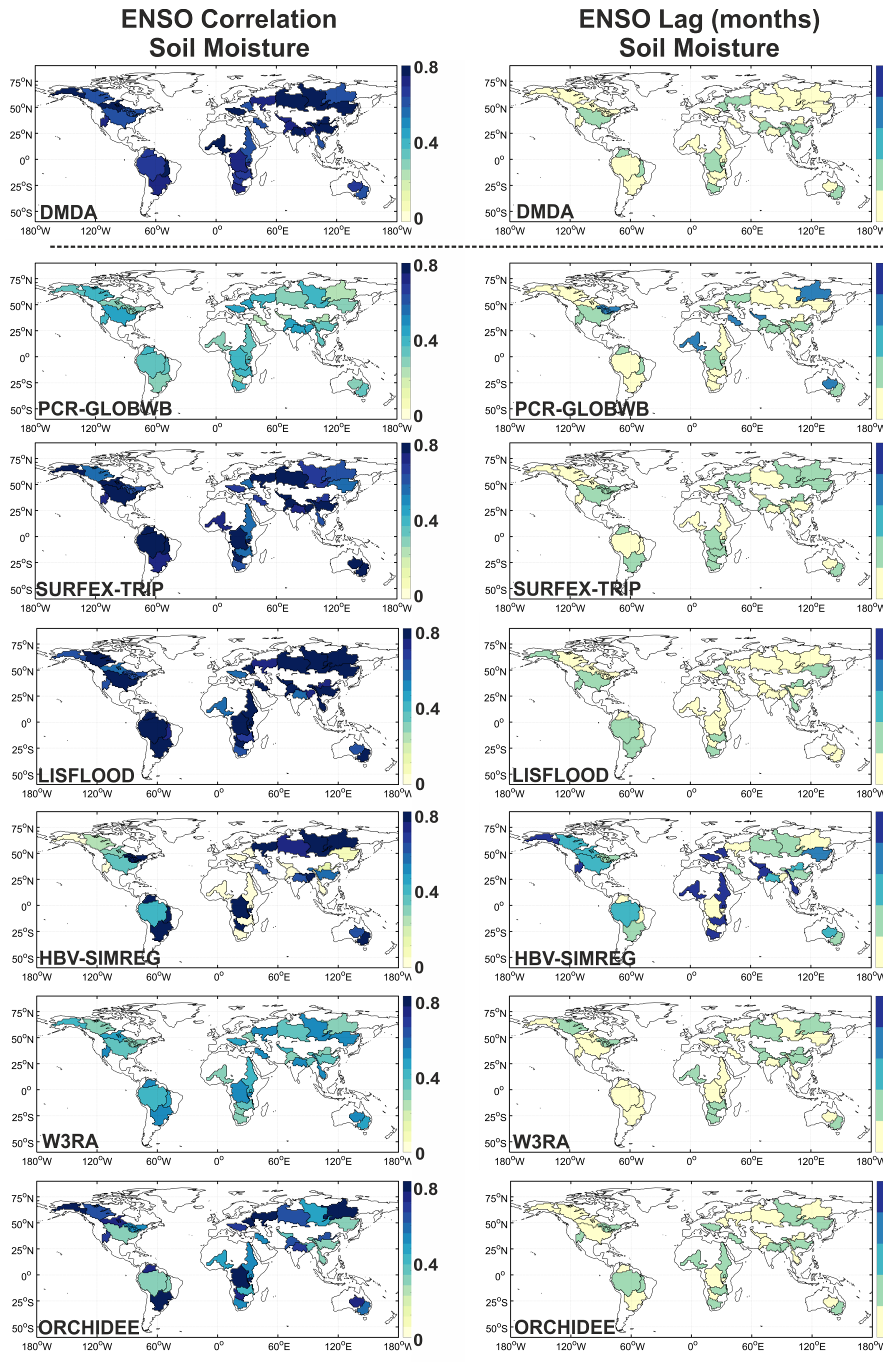

Figure 6: Correlation coefficients and their lags between the ENSO (-Niño 3.4 index) and soil moisture estimates derived from the DMDA method and hydrological models used in this study for the period of 2002–2012.

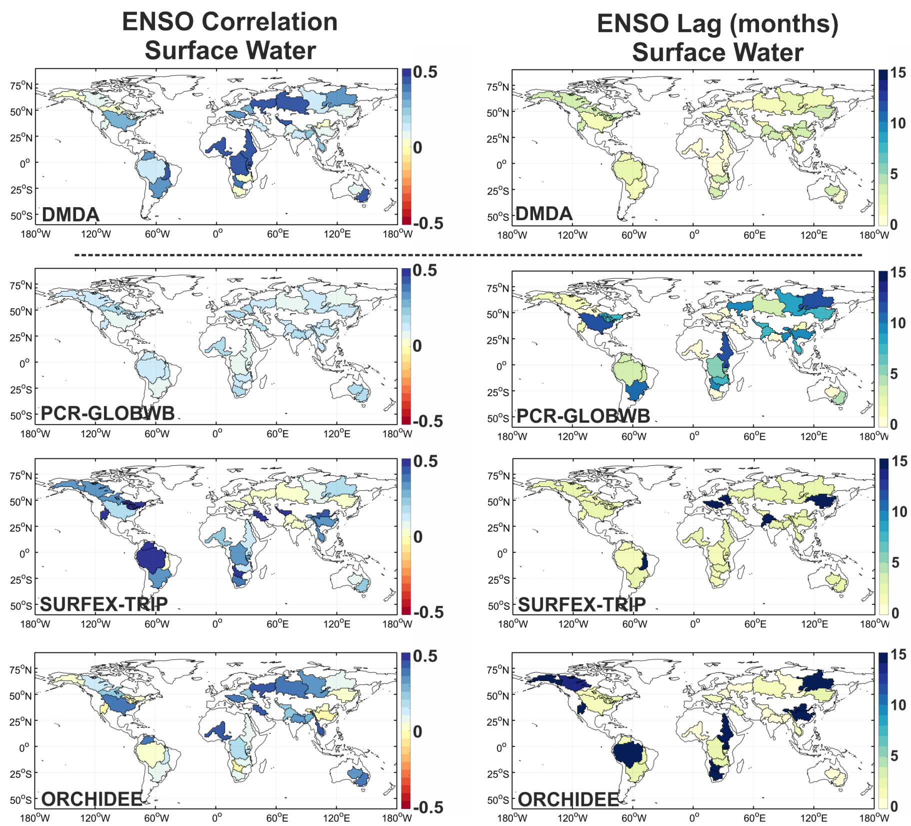

Figure 7: Correlation coefficients and their lags between the ENSO (-Niño 3.4 index) and surface water estimates derived from the DMDA method and hydrological models used in this study for the period of 2002–2012.

Table 1: Overview of models used in this study and their water storage components.

Model	Water Storage Compartments						
	GroundWater	Soil layer	Surface Water	Canopy	Snow	Snow layer	Water Use
PCR-GLOBWB	Yes	2	Yes	Yes	Yes	1	No
W3RA	Yes	3	No	No	Yes	1	No
HBV-SIMREG	Yes	1	No	No	Yes	1	No
SURFEX-TRIP	Yes	14	Yes	Yes	Yes	12	No
LISFLOOD	Yes	2	No	No	Yes	1	Yes
ORCHIDEE	No	11	Yes	No	Yes	6	irrigation

Table 2: An overview of satellite altimetry observation used to validate DMDA results.

Lake	River Basin	Lake mid point	Latitude range of pass	Satellite pass	Cycle
Nasser	Nile	23.31°N 32.83°E	[22.91°N 23.66°N]	94	48
Tana	Nile	12.11°N 37.40°E	[11.95°N 12.19°N]	94	38
chad	Niger	13.01°N 14.38°E	[12.94°N 13.05°N]	248	25
Kainiji	Niger	10.49°N 4.50°E	[10.40°N 10.50°N]	135	21
Malawi	Zambezi	10.84°S 34.40°E	[12.042°S 9.70°S]	44	4
Tanganyika	Zambezi	6.41°S 29.23°E	[8.44°S 4.461°S]	222	11
Guri	Orinoco	7.37°N 117.12°W	[7.06°N 7.67°N	152	69
Winnipeg	Nelson	53.18°N 98.21°W	[52.82°N 53.55°N]	195	9
Winnipegosis	Nelson	51.91°N 100.01°W	[51.85°N 52.05°N	195	17
Erie	St. Lawrence	42.11°N 81.48°W	[41.60°N 42.54°N]	193	45
Ontario	St. Lawrence	43.56°N 77.47°W	[43.35°N 43.83°N]	15	36
Tharthar	Euphrates	33.87°N 43.37°E	[33.75°N 34.00°N]	133	70
Urmia	Euphrates	37.25°N 45.45°E	[37.25°N 37.31°N]	133	4
Chany	Ob	54.96°N 77.33°E	[54.94°N 55.02°N]	5	28

Table 3: Magnitude of simulated predictors, observations, and DMDA results in a controlled synthetic simulation.

Hydrological Compartment	Model name	Min [mm]	Max [mm]	RMS [mm]
Groundwater (First model)	LISFLOOD	-10.5	16.1	7.9
Groundwater (Second model)	SURFEX-TRIP	-12.1	39.8	14.2
Groundwater (Expected value of DMDA)	PCR-GLOBWB	-39.5	70.4	24.2
Groundwater (DMDA result)	DMDA Output	-35.3	92.3	19.9
Groundwater (BMA result)	BMA Output	-46.0	130.2	43.8
Soil Moisture (First model)	LISFLOOD	-37.4	62.2	30.8
Soil Moisture (Second model)	SURFEX-TRIP	-45.7	79.9	41.5
Soil Moisture (Expected value of DMDA)	PCR-GLOBWB	-52.0	107.9	48.7
Soil Moisture (DMDA result)	DMDA Output	-58.5	113.8	51.2
Soil Moisture (BMA result)	BMA Output	-40.8	49.6	21.0
TWS (First model)	LISFLOOD	-46.8	75.5	37.2
TWS (Second model)	SURFEX-TRIP	-57.6	115.2	54.6
TWS (Expected value of DMDA results)	PCR-GLOBWB	-83.3	164.5	64.2
TWS (DMDA result)	DMDA Output	-77.8	153.8	63.2
TWS (BMA result)	BMA Output	-77.8	153.8	63.2
$\|\Delta\|_{\text{Groundwater}}$	|LISFLOOD − Expected value|	0	58.1	11.2
$\|\Delta\|_{\text{Groundwater}}$	|SURFEX − Expected value|	0	45.8	10.3
$\|\Delta\|_{\text{Groundwater}}$	|DMDA − Expected value|	0	31.2	5.3
$\|\Delta\|_{\text{Groundwater}}$	|BMA − Expected value|	0	87.6	20.4
$\|\Delta\|_{\text{Soil Moisture}}$	|LISFLOOD − Expected value|	0	46.8	9.6
$\|\Delta\|_{\text{Soil Moisture}}$	|SURFEX − Expected value|	0	29.3	5.7
$\|\Delta\|_{\text{Soil Moisture}}$	|DMDA − Expected value|	0	29.2	5.2
$\|\Delta\|_{\text{Soil Moisture}}$	|BMA − Expected value|	0	89.5	18.6
$\|\Delta\|_{\text{TWS}}$	|LISFLOOD − Expected value|	0	94.7	18.6
$\|\Delta\|_{\text{TWS}}$	|SURFEX − Expected value|	0	60.9	14.1
$\|\Delta\|_{\text{TWS}}$	|DMDA − Expected value|	0	24.2	6.2
$\|\Delta\|_{\text{TWS}}$	|BMA − Expected value|	0	31.4	8.4

Table 4: The amplitude of linear trend [mm/yr] and its uncertainty, fitted to the DMDA-derived groundwater, soil Moisture, and surface water, during 2002–2012.

Basin		**DMDA**	**DMDA**	**DMDA**
ID	**Name**	**GroundWater**	**Soil Moisture**	**Surface Water**
1	Amazon	0.17 ± 0.12	-1.92 ± 0.09	1.43 ± 0.06
2	Amur	0.46 ± 0.06	2.61 ± 0.09	0.25 ± 0.03
3	Aral	0.02 ± 0.08	-1.43 ± 0.22	0.21 ± 0.12
4	Brahmaputra	-0.44 ± 0.16	-7.00 ± 0.69	-0.13 ± 0.21
5	Caspian-Volga	-2.06 ± 0.15	-2.98 ± 0.16	-0.02 ± 0.07
6	Colorado	0.80 ± 0.11	-0.75 ± 0.09	0.82 ± 0.08
7	Congo	-0.72 ± 0.08	0.59 ± 0.03	0.06 ± 0.06
8	Danube	-0.47 ± 0.18	-0.75 ± 0.21	-0.08 ± 0.04
9	Dnieper	-0.5 ± 0.29	-2.27 ± 0.28	-0.03 ± 0.18
10	Euphrates	-5.36 ± 0.23	-5.75 ± 0.39	-2.09 ± 0.09
11	Lake Eyre	0.55 ± 0.16	2.42 ± 0.19	0.77 ± 0.04
12	Ganges	-14.77 ± 0.25	2.69 ± 0.40	0.29 ± 0.05
13	Indus	-8.26 ± 0.16	1.10 ± 0.13	-0.06 ± 0.07
14	Lena	1.74 ± 0.11	1.94 ± 0.05	0.20 ± 0.08
15	Mackenzie	0.51 ± 0.06	0.12 ± 0.05	-0.05 ± 0.10
16	Mekong	1.58 ± 0.43	-0.79 ± 0.33	0.83 ± 0.17
17	Mississippi	1.25 ± 0.09	1.36 ± 0.09	0.33 ± 0.02
18	Murray	0.06 ± 0.06	6.66 ± 0.15	-1.47 ± 0.04
19	Nelson	0.70 ± 0.18	2.45 ± 0.15	0.11 ± 0.03
20	Niger	-1.14 ± 0.15	0.75 ± 0.15	0.32 ± 0.05
21	Nile	0.45 ± 0.06	0.77 ± 0.06	-0.05 ± 0.02
22	Ob	-1.42 ± 0.08	-1.54 ± 0.06	0.05 ± 0.07
23	Okavango	1.74 ± 1.31	3.92 ± 0.55	-1.42 ± 0.37
24	Orange	1.32 ± 0.05	1.28 ± 0.06	-0.85 ± 0.05
25	Orinoco	0.87 ± 0.11	3.45 ± 0.26	-0.22 ± 0.19
26	Parana	0.68 ± 0.08	0.03 ± 0.13	1.04 ± 0.04
27	St. Lawrence	1.49 ± 0.18	1.07 ± 0.07	0.48 ± 0.05
28	Tocantins	2.41 ± 0.47	2.37 ± 0.35	0.08 ± 0.21
29	Yangtze	0.55 ± 0.23	-0.30 ± 0.09	0.20 ± 0.02
30	Yellow	-3.50 ± 0.14	-0.27 ± 0.05	0.08 ± 0.21
31	Yenisei	-0.26 ± 0.07	1.79 ± 0.06	0.75 ± 0.11
32	Yukon	-4.73 ± 1.08	-1.52 ± 0.20	-1.11 ± 0.23
33	Zambezi	1.19 ± 0.38	0.65 ± 0.31	0.35 ± 0.25

Table 5: Correlation between satellite altimetery observation and: I) TWS , II) Surface Water (SW) derived from GRACE, DMDA, and individual models, during 2002–2012.

Basin	Water storage	GRACE	DMDA	PCR-GLOBWB	SURFEX-TRIP	LISFLOOD	HBV-SIMREG	W3RA	ORCHIDEE
		Correlation between Altimetry Obs. and:							
Nile	TWS	0.358	**0.381**	0.326	0.239	0.095	-0.082	0.001	0.180
(Nasser Lake)	SW	-	**0.462**	0.363	0.441	-	-	-	-0.046
Nile	TWS	0.682	**0.718**	0.602	0.569	0.517	0.302	0.231	0.635
(Tana Lake)	SW	-	0.492	0.340	**0.603**	-	-	-	0.455
St. Lawrence	TWS	**0.353**	**0.261**	0.271	0.010	-0.121	-0.114	-0.087	-0.010
(Erie Lake)	SW	-	0.432	**0.483**	0.126	-	-	-	0.227
St. Lawrence	TWS	**0.410**	**0.364**	0.353	0.110	-0.063	-0.064	-0.023	0.037
(Ontario Lake)	SW	-	**0.582**	0.572	0.273	-	-	-	0.239
Euphrates	TWS	**0.698**	**0.569**	0.225	0.021	0.103	-0.057	0.043	0.182
(Tharthar Lake)	SW	-	**0.236**	0.127	0.093	-	-	-	-0.282
Euphrates	TWS	**0.737**	**0.628**	0.223	0.080	0.148	0.021	0.095	0.185
(Urmia Lake)	SW	-	**0.172**	0.170	0.131	-	-	-	-0.325
Ob	TWS	0.393	**0.482**	0.371	0.303	0.336	0.338	0.348	0.328
(Chany Lake)	SW	-	**0.296**	0.278	0.177	-	-	-	-0.333
Zambezi	TWS	0.552	**0.632**	0.362	0.277	0.346	0.225	0.246	0.391
(Malawi Lake)	SW	-	0.382	0.247	**0.410**	-	-	-	0.394
Zambezi	TWS	**0.414**	**0.365**	0.231	0.192	0.121	0.117	0.128	0.160
(Tanganyika Lake)	SW	-	**0.243**	0.096	0.241	-	-	-	-0.093
Niger	TWS	**0.576**	**0.558**	0.436	0.318	0.308	0.065	0.188	0.519
(Chad Lake)	SW	-	0.657	0.511	0.616	-	-	-	**0.689**
Niger	TWS	**0.132**	**0.102**	-0.002	-0.149	-0.174	-0.383	-0.278	0.079
(Kainiji Lake)	SW	-	**0.282**	0.126	0.200	-	-	-	0.214
Orinoco	TWS	**0.585**	**0.539**	0.332	0.427	0.431	0.321	0.301	0.434
(Guri Lake)	SW	-	**0.421**	0.314	0.390	-	-	-	0.318
Nelson	TWS	**0.285**	**0.270**	0.139	-0.185	-0.444	-0.440	-0.389	-0.279
(Winnipeg Lake)	SW	-	**0.104**	-0.290	0.072	-	-	-	0.012
Nelson	TWS	0.216	**0.249**	0.238	0.135	-0.09	-0.164	-0.088	-0.065
(Winnipegosis Lake)	SW	-	**0.098**	-0.321	-0.015	-	-	-	-0.480