Improving drought simulations within the Murray-Darling Basin by combined calibration/assimilation of GRACE data into the WaterGAP Global Hydrology Model

M. Schumacher[a,b,c], E. Forootan[d], A.I.J.M. van Dijk[c], H. Müller Schmied[e,f], R.S. Crosbie[g], J. Kusche[b], P. Döll[e,f]

[a]*School of Geographical Sciences, University of Bristol, Bristol, UK*
[b]*Institute of Geodesy and Geoinformation, Bonn University, Bonn, Germany*
[c]*Fenner School of Environment and Society, The Australian National University, Canberra, Australia*
[d]*School of Earth and Ocean Sciences, Cardiff University, Cardiff, UK*
[e]*Institute of Physical Geography, University of Frankfurt, Frankfurt am Main, Germany*
[f]*Senckenberg Biodiversity and Climate Research Centre (BiK-F), Frankfurt am Main, Germany*
[g]*CSIRO Land and Water, Adelaide, Australia*

Abstract

Simulating hydrological processes within the (semi-)arid region of the Murray-Darling Basin (MDB), Australia, is very challenging specially during droughts. In this study, we investigate whether integrating remotely sensed terrestrial water storage changes (TWSC) from the Gravity Recovery And Climate Experiment (GRACE) mission into a global water resources and use model enables a more realistic representation of the basin hydrology during droughts. For our study, the WaterGAP Global Hydrology Model (WGHM), which simulates the impact of human water abstractions on surface water and groundwater storage, has been chosen for simulating compartmental water storages and river discharge during the so-called 'Millennium Drought' (2001-2009). In particular, we test the ability of a parameter calibration and data assimilation (C/DA) approach to introduce long-term trends into WGHM, which are poorly represented due to errors in forcing, model structure and calibration. For the first time, the impact of the parameter equifinality problem on the C/DA results is evaluated. We also investigate the influence of selecting a specific GRACE data product

Email address: `maike.schumacher@bristol.ac.uk` (M. Schumacher)

and filtering method on the final C/DA results. Integrating GRACE data into WGHM does not only improve simulation of seasonality and trend of TWSC, but also it improves the simulation of individual water storage components. For example, after the C/DA, correlations between simulated groundwater storage changes and independent in-situ well data increase (up to 0.82) in three out of four sub-basins. Declining groundwater storage trends - found mainly in the south, i.e. Murray Basin, at in-situ wells - have been introduced while simulated soil water and surface water storage do not show trends, which is in agreement with existing literature. Although GRACE C/DA in MDB does not improve river discharge simulations, the correlation between river storage simulations and gauge-based river levels increases significantly from 0.15 to 0.52. By adapting the C/DA settings to the basin-specific characteristics and reducing the number of calibration parameters, their convergence is improved and their and uncertainty is reduced. The time-variable parameter values resulting from C/DA allow WGHM to better react to the very wet Australian summer 2009/10. Using solutions from different GRACE data providers produces slightly different C/DA results. We conclude that a rigorous evaluation of GRACE errors is required to realistically account for the spread of the differences in the results.

Keywords: GRACE, WGHM, Data Assimilation, Calibration, Murray Darling Basin, Drought

1. Introduction

The Murray-Darling Basin (MDB) in south-eastern Australia is one of the driest river basins over the world. Long-term hydro-meteorological records indicate that the MDB is prone to extreme hydrological events (Verdon-Kidd et al., 2009; Gallant et al., 2011; Gergis et al., 2012). Particularly, a long drought period, the so-called 'Millennium Drought' (Ummenhofer et al., 2009; Leblanc et al., 2012; van Dijk et al., 2013), occurred during 2001-2009 and affected environment, agriculture, and therefore economic activities within the basin. Subsequently, during 2010-2012, the MDB received above average precipitation, mainly driven by the El Niño Southern Oscillation (ENSO, see e.g., Boening et al., 2012) and to a smaller extent the Indian Ocean Dipole (IOD, see e.g., Forootan et al., 2016). Although this helped refilling its terrestrial water storage, studies indicate an overall water availability decline that is likely due to climate change (e.g., Grafton et al., 2014) noting that the sensitivity of stream-flow generation to changes in climate drivers varies spatially (Donohue et al., 2011).

Various remote sensing data and hydrological models have been applied to monitor water variability of the MDB. For example, terrestrial water storage changes (TWSC) can be derived from the Gravity Recovery And Climate Experiment (GRACE) satellite mission (Tapley et al., 2004). The measurements represent the vertical integration of above- and below-surface water storage compartments, and have been used to study the distribution of water and the impact of climate variability within the MDB (e.g., Brown and Tregoning, 2010; Awange et al., 2011; Garćia-Garćia et al., 2011; Forootan et al., 2012). In addition, remotely sensed surface soil moisture and vegetation water content variations have been analyzed to quantify the influence of large-scale climate variability, such as ENSO and IOD, on the basin hydrology (Liu et al., 2009; Bauer-Marschallinger et al., 2013). Hydrological models have also been applied over the MDB, such as the WaterGAP Global Hydrology Model (WGHM, Döll et al., 2003), the Global Land Data Assimilation System (GLDAS, Rodell et al., 2009), and the high resolution continental model of AWRA (Australian Water Resources Assessment, van Dijk and Renzullo, 2011; van Dijk et al, 2011; Vaze et al., 2013).

WGHM simulates daily water storage changes in several individual compartments, including canopy, snow, soil, lake, wetland, man-made reservoirs, river and groundwater. The groundwater compartment is often not explicitly realized in other hydrological models (such as GLDAS). In addition, WGHM considers anthropogenic water abstraction, which makes the model distinct from most others. Accurate estimation of water storage variability, including variability of the surface and sub-surface (soil moisture and groundwater) storage compartments, as well as river discharge within the MDB is difficult due to its complex geomorphology, the definition of water connection within the basin (Lamontagne et al., 2014), and the strong dependence of hydrology on antecedent rainfall (Beaumont, 2012). In general, the simulation skill of hydrological models is limited by uncertainties in: climate forcing (particularly precipitation), model parameters, and deficiencies in the model structure (Müller Schmied et al., 2014, 2016). Abelen and Seitz (2013) reported inconsistencies between WGHM and remotely sensed soil moisture variations, which might be due to neglected physical processes. For example, the soil water compartment is defined by a single layer in WGHM with its depths depending on the plants' root zone. GLDAS simulations also do not perfectly represent the hydrological property of the MDB due to the missing groundwater compartment, as well as ignoring the influence of human water use (e.g., Tregoning et al., 2012). Similarly, the AWRA model does not account for extensive pumping, which occurs during drought periods. During flood events also, less accurate discharge/recharge estimations are reported (e.g., in Crosbie et al., 2011). van Dijk and Renzullo (2011) and Forootan et al. (2012) showed inconsistencies in the linear trend (2003-2011) between GRACE TWSC and that of AWRA.

To understand the hydrological behavior of the MDB, in most of previous studies, GRACE TWSC estimates were compared directly to the storage variability or surface loading estimations simulated by hydrological models or observed by other techniques e.g., GPS, satellite altimetry, soil moisture remote sensing, and in-situ observation wells (e.g., Leblanc et al., 2009; Chen et al., 2016). Variability of a particular storage compartment, e.g., groundwater, is usually computed by reducing other storage compartments (e.g., surface, canopy and soil storage compartments) derived from complimentary sources (see an extensive review in Tregoning et al., 2012, chapter 2). Leblanc et al. (2009), for instance, conducted a multi-sensor analysis over the MDB, and found a rapid decline in soil moisture and surface water of about 80 km^3 and 12 km^3, respectively, during 2001-2003 and low storage levels in the following years. They also reported that the in-situ groundwater measurements are highly correlated with GRACE TWSC (correlation coefficients of 0.94) and found a groundwater loss of about 104 km^3 during 2003-2007. Chen et al. (2016) focused on Victoria, southern Australia, and estimated changes in groundwater by subtracting simulations of the other storage compartments from GRACE TWSC. The authors found a good agreement between their estimations and in-situ observation wells, i.e. a declining trend of about 8.0-8.3 km^3/year during 2005-2009.

The validity of hydrological assessments in previous works might be limited due to the inconsistencies between GRACE TWSC and model simulations or other observation techniques. Therefore, inversion (e.g., Forootan et al., 2014, 2017; Al-Zyoud et al., 2015) and data assimilation techniques (e.g., Zaitchik et al., 2008; Eicker et al., 2014; Van Dijk et al., 2014) should be applied to consistently merge observations with hydrological model simulations.

In this study, we pursue the recently improved calibration and data assimilation (C/DA) framework based on ensemble Kalman filtering (EnKF, Schumacher et al., 2016) to merge GRACE TWSC estimation with WGHM simulations for the MDB. Unlike other hydrological measurements GRACE TWSC constrains the sum of changes within all individual water storage compartments including groundwater, which cannot be measured by any other remote sensing techniques. Using GRACE data, it is not possible to distinguish changes in individual storage components, i.e. whether these changes occur in canopy, soil water, surface water or groundwater. To vertically disaggregate the GRACE-derived TWSC into its individual components, one needs a priori information from other sources, for example, hydrological models, i.e. WGHM in our study. In addition, GRACE observations only provide a coarse horizontal resolution. Data assimilation provides a realistic way to downscale GRACE observations based on the equations implemented in hydrological models. Recently, Khaki et al. (2017a,b) applied GRACE data and Tian et al. (2017) used GRACE and soil moisture data simultaneously in an ensemble-based assimilation framework to update storage estimation of a hydrological model in Australia and the MDB. Although their studies indicate improvements in soil and groundwater storage estimations, no attempts have been made to calibrate model parameters. In this study, we show to what extent adding water storage information from GRACE, through a C/DA procedure, is able to improve WGHM's TWSC, individual water storage simulations and its parameters. Hereby, the main focus of our paper is on the effect of the Millennium Drought on the groundwater storage. It is also investigated whether a C/DA of GRACE data affects WGHM's river discharge simulations. This study is the first attempt to assess the impact of GRACE data assimilation on hydrological simulations during a long-term drought period, i.e. here the Millennium Drought.

WGHM has 22 parameters that ensure its realistic simulations. However, several parameter combinations may be able to restore observed TWSC and thus GRACE-based calibration alone would be plagued by the equifinality problem. We will show here that, by reducing the number of calibrated parameters, deficiencies in model outputs reduces, and subsequently hydrological estimations within the MDB are improved. The implemented C/DA framework has already been successfully applied to improve simulations of total and individual water storage compartments in the Mississippi River Basin (Eicker et al., 2014). Their study was however limited to one year, and the results were not validated with independent data sets. The novelty of the presented framework compared to previous approaches is the extension to model parameter calibration, as well as the implementation of spatial GRACE TWSC error correlations in the ensemble filter update.

The objectives of this paper are: (1) to transfer and assess the C/DA approach (Schumacher et al., 2016) to a (semi-)arid region experiencing a severe long-term drought without tuning the approach; (2) to investigate the impact of GRACE data products and its post-processing on the C/DA results; (3) to address the equifinality problem that occurs in the parameter calibration stage; (4) to identify changes in hydrological behavior of the basin within and after the Millennium Drought; and (5) validating the C/DA results using independent in-situ data, i.e. here river level and river discharge from gauge stations, as well as groundwater well data. The designed objectives will address important technical issues related to the combination of GRACE and hydrological models:

Objective (1) will show whether by applying the C/DA and using GRACE data it is possible to restore long-term trends (water decline in our case) in a particular water storage compartment. This is important since models usually do not realistically represent long-term decline or rising of water levels in the MDB that have been found in GRACE data (Döll et al., 2014). To our knowledge, this is the first application of GRACE-based model parameter calibration via ensemble-based data assimilation for this purpose. An independent validation against in-situ groundwater measurements is also performed.

Objective (2) helps assessing the robustness of the C/DA approach with respect to the choice of data products. This investigation is also important for other studies since there is currently no clear guidance on the "best" selection of a GRACE product and of its post-processing for assimilation studies.

Objective (3) has not yet been tackled in the context of parameter calibration against GRACE data. Therefore, we will discuss how selecting a sub-set of model parameters improves the C/DA.

Objective (4) provides insights about spatial and temporal variations of soil water and groundwater storage changes within the MDB after implementing a C/DA. The combined results are likely more reliable than interpreting WGHM simulations or GRACE data individually.

Objective (5) shows to what extent C/DA can improve water storage simulations and its impact on river discharge simulations can be identified.

2. Study Area and Data

The MDB, with an area of $\sim 1,060,000$ km^2, is home of two major rivers; the Murray River and the Darling River, which joins the Murray River around 500 km upstream from the basin outlet. It extends from the subtropics of central Queensland to the southern alps of Victoria and the Southern Ocean, therefore, it has been under influence of both humid and arid climates and their variabilities (Connell and Grafton, 2011). Most of the basin is flat, low-lying and far inland, and receives 477 mm area-averaged annual rainfall (Fu et al., 2010). Its tributary rivers tend to be long and slow-flowing, and carry a volume of water that is large only by Australian standards. The sedimentary rocks have a maximum depth of 600 m; thus, groundwater storage is relatively small. The MDB is essentially a closed groundwater basin, where groundwater drainage is directed internally towards the central subsidence and thicker sediments, rather than towards the side where the Murray connects to the sea (Grafton et al., 2014).

We consider four sub-basins within the MDB: the arid north-western Darling area (NW), which contains the Darling and Warrego Rivers, and the north-eastern Darling area (NE) in which the Balonne River and several other northern rivers flow. The other two consist of the south-eastern Murray area (SE) with the first half of the Murray River, and the whole Lachlan and Murrumbidgee Rivers, as well as the south-western Murray area (SW) with the second half of the Murray River. These regions are defined (i) based on the hydrological sub-basins and underlying river routing system considered in WGHM, as well as (ii) the spatial area detectable by GRACE. The shapes of the sub-basins and their areas are reported in Fig. 1.

2.1. Hydrological Model: WGHM

The WaterGAP Global Hydrology Model (WGHM) and five water use models together form the global water availability and use model Water - Global Assessment and Prognosis (WaterGAP). WGHM uses a number of water storage equations that describe the daily vertical water balance and horizontal routing, with a spatial resolution of $0.5° \times 0.5°$ for the global land area excluding Antarctica. Detailed descriptions of the model equations are given in Döll et al. (2003) and Müller Schmied et al. (2014). In this study, we use the model version WaterGAP 2.2 for calibration and data assimilation (C/DA) of GRACE TWSC. The model has already been calibrated against mean annual river discharge at 1319 Global Runoff Data Centre (GRDC) stations, of which 11 are located in the MDB (Müller Schmied et al., 2014). The monthly forcing fields of tempera-

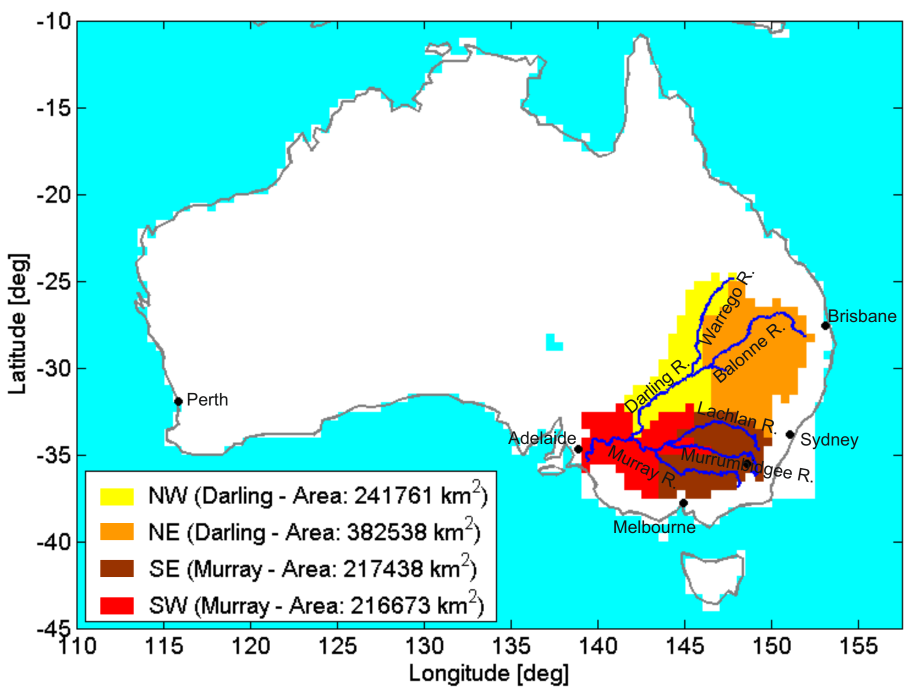

Figure 1: The Murray-Darling Basin (MDB) and its four sub-basins considered here to integrate GRACE TWSC with the WGHM model simulations.

ture, cloud cover, and the number of wet days were obtained from the Climate Research Unit's Time Series (CRU TS 3.2; Harris et al., 2013) and precipitation provided by the Global Precipitation Climatology Center (GPCC v6; Schneider et al., 2014), which at the date of our study were available until end of 2010.

2.2. GRACE TWSC

Monthly GRACE level 2 products, expressed as dimensionless spherical harmonics of the geopotential up to degree and order 90, are available from different sources. Here, the RL05 of GFZ and JPL (`ftp://podaac-ftp.jpl.nasa.gov/allData/grace/L2/`) are considered, as well as those of ITSG-Grace2014 (`http://portal.tugraz.at/portal/page/portal/TU_Graz/Einrichtungen/Institute/Homepages/i5210/research/ITSG-Grace2014`). Degree 1 coefficients are replaced by those from Swenson et al. (2008). The zonal degree 2 spherical harmonic coefficients (C_{20}) are replaced by Satellite Laser Ranging (SLR) data (Cheng et al., 2013, see also `grace.jpl.nasa.gov`).

GRACE level 2 products contain correlated errors, visible as striping patterns in the spatial domain (Kusche, 2007). Therefore, before computing monthly TWS fields, the DDK3 anisotropic decorrelation filter (Kusche et al., 2009) is applied to suppress such errors. Monthly residual gravity field solutions are computed by subtracting the temporal average of 2003-2010 from each month. The residual coefficients are then converted to gridded TWSC fields (on the $0.5^\circ \times 0.5^\circ$ grid used in WGHM) following Wahr et al. (1998). The same steps are repeated for the ITSG-Grace2014 product, while applying a Gaussian filter with 300 km and 500 km radii to investigate the influence of smoothing of GRACE TWSC on the C/DA results. A formal variance-covariance error propagation is carried out to obtain the observation error covariance matrices (Schumacher et al., 2016). It is worth mentioning that the TWSC estimations from CSR data lie within the GRACE ensemble (ITSG-GRACE2014, GFZ, JPL). Thus, here, we do not explicitly report the results based on CSR data. In total, five different GRACE TWSC variants are considered in this study. For all variants, the full error covariance matrix of the ITSG-Grace2014 product smoothed by a 300 km Gaussian filter is used.

For the C/DA, Schumacher et al. (2016) suggest to integrate GRACE TWSC and model simulations either on coarse grids, e.g., $5.0^\circ \times 5.0^\circ$ or as (sub-) basin averages. In this study, we select GRACE TWSC averaged over the four subbasins of Fig. 1 for assimilation into WGHM. To account for the signal damping and spatial leakage due to the application of filtering, constant and time-variable scaling factors are estimated (see Sect. 6 of the Supplementary Data for details). The scaling values are found to be close to 1. The main C/DA results are presented with respect to the ITSG-Grace2014 product, which is filtered by DDK3, and called ITSG-DDK3 in the following.

2.3. Groundwater Observations

Groundwater changes from around 15800 observation wells within the MDB are applied to validate the C/DA results. The measurements were spatially averaged over $1° \times 1°$ grid cells, including between one to around 2680 wells per grid cell. The locations of the individual observation wells are provided in (Tregoning et al., 2012). It was reported that these wells might be influenced by local effects such as pumping that might cause draw-down or recharge due to irrigation. The observations are expressed as groundwater levels, and converted to equivalent water heights (EWH) by considering aquifer specific yield, which is usually unknown and cannot be measured at this scale. Here, we use an estimate of 0.1 as a typical value for water aquifers as proposed by Tregoning et al. (2012). To demonstrate the effect of the choice of the specific yield, additionally specific yield maps based on surface geology are considered (Viney et al., 2015, , Sect. 4.3.2).

3. Calibration and Data Assimilation (C/DA) Framework

An overview of the calibration and data assimilation (C/DA) study set-up is given in Fig. 2. To run the hydrological simulation, WGHM is initialized during 1995-2000. Then, an ensemble of N_e=30 runs is generated to represent uncertainties in forcing data, model parameters (see Tab. 1), initial water states and errors in the model structure. For this, a priori Probability Density Functions (PDF) are considered for the model parameters based on literature (Döll et al., 2003; Kaspar, 2004; Schumacher et al., 2015). A multiplicative error model is assumed for precipitation fields centered around 1 and with limits of 0.7 and 1.3, and an additive error model for temperature fields centered at 0 and limits of $\pm 2°$C; both are added as white noise. The generated ensembles are used in a two years model spin-up phase during 2001-2002 to generate an ensemble of initial water states. Our experiments with the initialization and spin-up length indicate that these have negligible influence on the model runs (details in Sect. 7, Supplementary Data.

First, an open loop (OL) run during 2003-2010, i.e. WGHM runs are performed with each of the 30 ensemble members (first column in Fig. 2, and Tab. 2). Within WGHM, parameter values are set globally, i.e. the same values are used in all river basins world-wide. Moreover, the parameters are temporally constant. Subsequently, WGHM is run in C/DA mode, i.e. GRACE TWSC observations along with their full error covariance information are assimilated monthly into WGHM (second column in Fig. 2, and Tab. 2) using the EnKF (Evensen, 1994; Burgers et al., 1998). In the EnKF updates, the water mass balance is not conserved, i.e. water mass can be introduced to or removed from WGHM. By applying the C/DA, model parameters are calibrated sequentially each time that GRACE observations are available within the MDB. Therefore, the calibrated parameters are the most appropriate for the MDB but not necessarily for other river basins. The adjusted parameter values are then used to start the WGHM runs for the next months. This is done for the entire 2003-2010. In summary, parameter values after the C/DA vary in time and are not identical to the parameters used in the OL run. Since the updated water states and parameters are adjusted to the GRACE observations within each EnKF update step, the model uncertainties decrease successively. Thus, an inflation factor of 10%, based on findings in Schumacher et al. (2016), is used to ensure a contribution of GRACE TWSC to the updated water states and parameters during the entire study period (addressing Objective 1).

We also carry out five experiments with a range of configurations (Tab. 2): (i) different GRACE products (ITSG, GFZ, JPL) are used for introducing the observed TWSC, and (ii) various spatial filters applied to the ITSG-Grace2014 data product (300 and 500 km Gaussian filter, as well as DDK3), to account for the impact of GRACE post-processing (addressing Objective 2).

Another experiment is designed, in which only the three parameters of the root depth multiplier, net radiation multiplier and groundwater outflow coefficient are calibrated instead of the 22 model parameters (C/DA (v2) in Tabs. 1 and 2). These three parameters are selected since they are relatively independent and have considerable influence on simulating relevant water compartments in the MDB, i.e. soil water and groundwater. By this reduction and comparing to the C/DA version, in which all 22 parameters are calibrated, we can investigate the equifinality problem using GRACE TWSC for model calibration (addressing Objective 3).

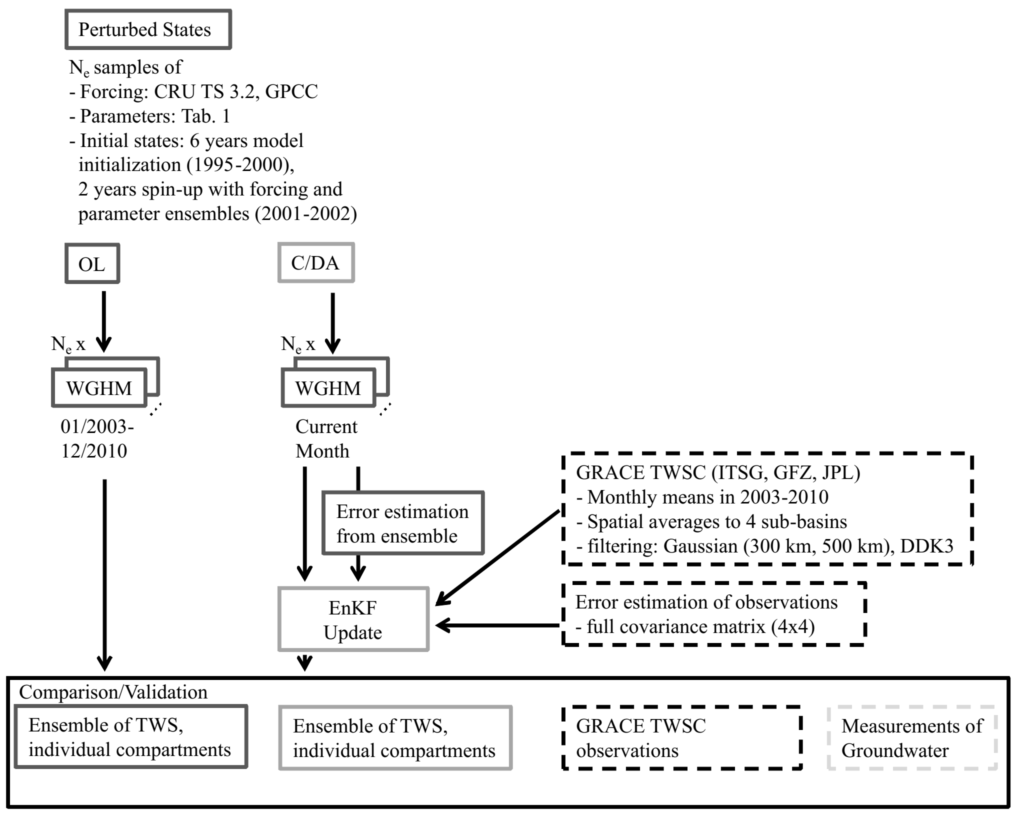

Figure 2: Set-up of study for the Murray-Darling Basin (MDB). First, open loop (OL) model runs are performed over 2003-2010 (left column). Subsequently, GRACE TWSC averaged over the 4 major sub-basins of the MDB are assimilated into WGHM testing different configurations (center and right column) and simultaneously the WGHM's parameters are calibrated (see Tab 1). To assess the C/DA results, simulated TWSC and groundwater changes are compared to GRACE TWSC and independent groundwater well measurements.

Table 1: Model parameters that are calibrated within the EnKF, where "IN" indicates the identification number, "mode" represents the value used in the original WGHM run, and under "limits" the spread of parameter values used for ensemble generation are summarized. The last two columns indicate whether a parameter is calibrated against GRACE. For the C/DA version 2 (v2) run, the mode and limits of parameters 3, 4 and 19 are modified. These values are provided in brackets.

IN	Calibration Parameter	Mode	Limits	C/DA	C/DA (v2)
1	root depth multiplier	1	[0.5 2.0]	yes	yes
2	river roughness coefficient multiplier	1	[0.5 2.0]	yes	-
3	lake depth (m)	5	[1 20]	yes	-
		(4)	([1 10])		
4	wetland depth (m)	2	[0.5 5]	yes	-
		(1)	([0.5 2])		
5	surface water outflow coefficient (day^{-1})	0.01	[0.001 0.1]	yes	-
6	net radiation multiplier	1	[0.5 2.0]	yes	yes
7	Priestley-Taylor coefficient (humid)	1.26	[0.885 1.65]	yes	-
8	Priestley-Taylor coefficient (arid)	1.74	[1.365 2.115]	yes	-
9	maximum daily potential evapotranspiration (mm/day)	15	[7.25 22.5]	yes	-
10	maximum canopy water height per leaf area (mm)	0.3	[0.1 1.4]	yes	-
11	specific leaf area multiplier	1	[0.5 2.0]	yes	-
12	snow freeze temperature (°C)	0	[-1.0 3.0]	yes	-
13	snow melt temperature (°C)	0	[-3.75 3.75]	yes	-
14	degree day factor multiplier	1	[0.5 2.0]	yes	-
15	temperature gradient (°C/m)	0.006	[0.004 0.01]	yes	-
16	groundwater recharge factor multiplier	1	[0.5 2.0]	yes	-
17	maximum groundwater recharge multiplier	1	[0.5 2.0]	yes	-
18	critical precipitation for groundwater recharge (mm/day)	10	[2.5 20.0]	yes	-
19	groundwater outflow coefficient (day^{-1})	0.006	[0.006 0.018]	yes	yes
		(0.01)	([0.004 0.016])		
20	net abstraction surface water multiplier	1	[0.5 2.0]	yes	-
21	net abstraction groundwater multiplier	1	[0.5 2.0]	yes	-
22	precipitation multiplier	1	[0.8 1.2]	yes	-

Table 2: Overview of model simulations and assimilation runs that are analyzed in this study. The main results are presented with respect to the C/DA variant ITSG-DDK3 and the C/DA version 2 (v2), in which only three model parameters are calibrated (see Tab. 1). The remaining C/DA variants are discussed in the Supplementary Data.

Run	Method	GRACE Product	GRACE Filtering
OL	Open Loop	-	-
ITSG-DDK3	EnKF	ITSG-Grace2014	DDK3
ITSG-300km	EnKF	ITSG-Grace2014	300 km Gaussian
ITSG-500km	EnKF	ITSG-Grace2014	500 km Gaussian
GFZ-DDK3	EnKF	GFZ RL05	DDK3
JPL-DDK3	EnKF	JPL RL05	DDK3
C/DA (v2)	EnKF	ITSG-Grace2014	DDK3

4. Results

4.1. Meteorological and Hydrological Conditions

During the Millennium Drought (2001-2009), the MDB has received below average precipitation (see e.g., Leblanc et al., 2012; van Dijk et al., 2013). Basin-averaged annual precipitation from the Australian Bureau of Meteorology (BoM) during 1981-2013 shows that 2001-2009 was the longest period with below the mean precipitation of 477 mm (Fig. 3 (A), see also Forootan et al., 2016). Compared to the previous three decades, particularly, 2002 and 2006 were the driest years with up to 41% below average precipitation, followed by the wettest year in 2010 with 66% higher annual precipitation. The distribution of precipitation is however not homogeneous over the basin. In Fig. 3 (B), the differences between the mean annual precipitation over the Millennium Drought, and during 1981-2013 are shown on a $0.5° \times 0.5°$ grid. In the Darling Basin (northern part), precipitation is found to be overall higher during 2001-2009 compared to the three decade mean with a maximum value of +38 mm/year. In contrast, precipitation in the Murray Basin (southern part) is found smaller with a maximum of -40 mm/year. Therefore, we expect strong impact from the meteorological drought predominantly in the south.

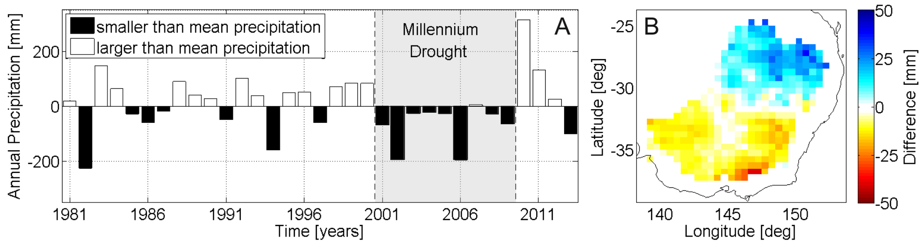

Figure 3: (A) Divergence of annual precipitation in mm (from the long-term temporal mean of 477 mm) averaged over the entire Murray-Darling Basin (MDB). (B) Difference in mean annual precipitation during 2001-2009 and 1981-2013 on a 0.5°×0.5° grid.

In Fig. 4, monthly TWSC derived from the open loop (OL) run during 1995-2010 and from GRACE during 2003-2013 over the entire MDB are shown. The WGHM simulation shows a strong decline in TWSC during 2001-2002, as well as a strong increase in 2010, which are clearly related to the extreme meteorological conditions. However, no further water decline is visible in the very dry year 2006. In contrast, during 2003-2007, the GRACE-derived TWSC decreased and is found mostly below the temporal mean until 2009. The strong rainfall events in 2010 and 2011 resulted in an increase of the total water mass (Forootan et al., 2012). Afterwards, TWSC values are found to be mostly above the temporal mean.

No significant linear trend is visible in TWSC from the WGHM OL run during 2003-2009. On the contrary, the estimation from the ITSG-DDK3 GRACE solution (see Tab. 2) shows a decrease of -7.6 mm/year over the entire MDB, ranging from -2.9 mm/year in the north-eastern Darling Basin (NE) to -14.0 mm/year in the south-eastern Murray Basin (SE, Tab. 3). Although precipitation is above the three decadal average (see Fig. 3 (B)), the linear trends in the Darling Basins are found to be negative. The application of different filtering to smooth GRACE TWSC represents a small impact on the linear trend estimation in the Darling sub-basins (differences of around 0.3 mm/year, see column

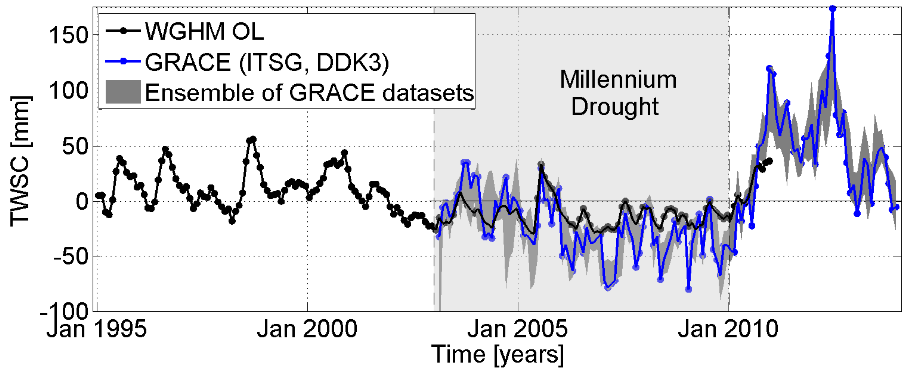

Figure 4: TWSC (in mm) derived from the WGHM open loop (OL) run and from GRACE averaged over the entire Murray-Darling Basin (MDB). The black line shows the WGHM OL, the blue line indicates GRACE (using ITSG-Grace2014), which is smoothed by the DKK3 filter, while the dark gray area represents the range of all investigated GRACE datasets (see Tab. 2).

"GRACE Filtering" in Tab. 3), and a higher influence in the Murray sub-basins (differences of up to 3.0 mm/year, see Tab. 3). Using different GRACE products for the trend estimation has a similar impact on the results (see column "GRACE Products" in Tab. 3). However, all analyzed GRACE data sets indicate negative trends in TWSC for the entire MDB. Therefore, an improved representation of the TWSC decline between 2003-2009 is expected by merging GRACE and WGHM in the C/DA framework.

4.2. TWSC Simulations from WGHM

4.2.1. Improving the Representation of TWSC

TWSC time series from the open loop (OL) simulations, GRACE and the calibration and data assimilation (C/DA) results after assimilating ITSG-DDK3, are shown in Fig. 5. A much better agreement is found between C/DA results (and the ensemble of all C/DA variants) with GRACE TWSC compared to the OL variant of WGHM. In terms of root mean square errors (RMSE), the fit for the entire basin is improved by 50% (from 21.4 to 10.7 mm), ranging from 45% in the north-western Darling Basin (NW) to 53% in both Murray sub-basins

Table 3: Linear trend (in mm/year) during 2003-2009 and its error derived by ITSG-Grace2014 (filtered by DDK3) for the averages over the entire MDB and its four major sub-basins (see the basins in Fig. 1). Averaged linear trends and their uncertainties estimated from different GRACE products, as well as after applying different filtering techniques are presented.

Basin	ITSG-DDK3	GRACE Products	GRACE Filtering
MDB	-7.6 ± 0.6	-5.9 ± 1.5	-6.8 ± 1.0
NW	-3.8 ± 0.8	-2.7 ± 1.0	-4.2 ± 0.3
NE	-2.9 ± 0.8	-0.8 ± 2.1	-3.2 ± 0.3
SE	-14.0 ± 0.7	-11.7 ± 2.1	-11.1 ± 3.0
SW	-13.5 ± 0.7	-12.8 ± 0.6	-11.4 ± 2.4

(Tab. 4). Applying different filtering techniques or using different GRACE products indicate improvements for the entire basin of up to 51% in terms of RMSE with respect to the OL variant. Furthermore, the correlation coefficient of WGHM simulated TWSC after C/DA with GRACE TWSC improves by 37% (from 0.58 to 0.92) for the entire MDB compared to OL. For the sub-basins, the improvements range between 28% in the south-eastern Murray Basin (SE) and 72% in the north-western Darling Basin (NW). Assessing the different C/DA variants in Tab. 2 indicates improvements for the entire MDB in terms of correlation coefficients of up to 36% compared to OL. After calibrating only three model parameters in C/DA (v2), the correlation coefficients are still high and the RMSE has been reduced compared to the OL. The individual RMSE and correlation coefficient values of all C/DA variants can be found in Tabs. S1 and S2 of the Supplementary Data.

The influence of assimilation on WGHM in simulating TWSC on the $0.5^\circ \times 0.5^\circ$ grid is assessed in Fig. 6, which shows correlation coefficients and RMSE between model simulations (from OL and C/DA) and GRACE TWSC after applying DDK3 filtering for both. Low to moderate improvements in correlations are found after C/DA all over the basin. The RMSE values between the WGHM simulated TWSC after C/DA and GRACE TWSC are found also to be smaller compared to the OL variant.

Table 4: Agreement between model predicted and observed TWSC in terms of correlation coefficients (CC) and root mean square errors (RMSE) in mm. Improvements are reported in the brackets.

Basin	CC OL	CC ITSG-DDK3	CC C/DA (v2)	RMSE OL	RMSE ITSG-DDK3	RMSE C/DA (v2)
MDB	0.61	0.92 (+0.31)	0.87 (+0.26)	21.7	10.7 (-11.0)	13.3 (-8.3)
NW	0.23	0.75 (+0.52)	0.58 (+0.36)	23.3	15.7 (-7.6)	19.0 (-4.2)
NE	0.45	0.89 (+0.44)	0.79 (+0.34)	27.8	14.7 (-13.1)	19.4 (-8.4)
SE	0.73	0.95 (+0.22)	0.93 (+0.20)	30.2	13.7 (-16.5)	16.3 (-14.0)
SW	0.52	0.91 (+0.39)	0.83 (+0.30)	33.8	16.1 (-17.7)	22.1 (-11.8)

4.2.2. Linear Trends and Seasonality in TWSC

The estimated linear trends in TWSC from the OL and C/DA variants of WGHM are summarized in Tab. 5. The standard deviations of the WGHM variant ITSG-DDK3 and C/DA (v2) are determined by formal error propagation based on the error covariance matrices of the EnKF updates. A comparison of the trends after C/DA with the trends from OL, and different GRACE products shows that the negative trends in the WGHM TWSC are reasonably intensified. The mean difference of the trends from the C/DA variants compared to GRACE is 1.5 mm/year, while the mean difference to the TWSC outputs of the OL simulations is 5 mm/year. The trends of the C/DA (v2) variant are somewhat smaller in the western parts of the MDB.

In order to assess whether the contribution of GRACE TWSC in the updated WGHM simulations (after C/DA) is realistically distributed, in Fig. 7, we show those statistically significant linear rates in TWSC that are found in the MDB during 2003-2009. A t-test with a significance level of 97.5 % is applied for this

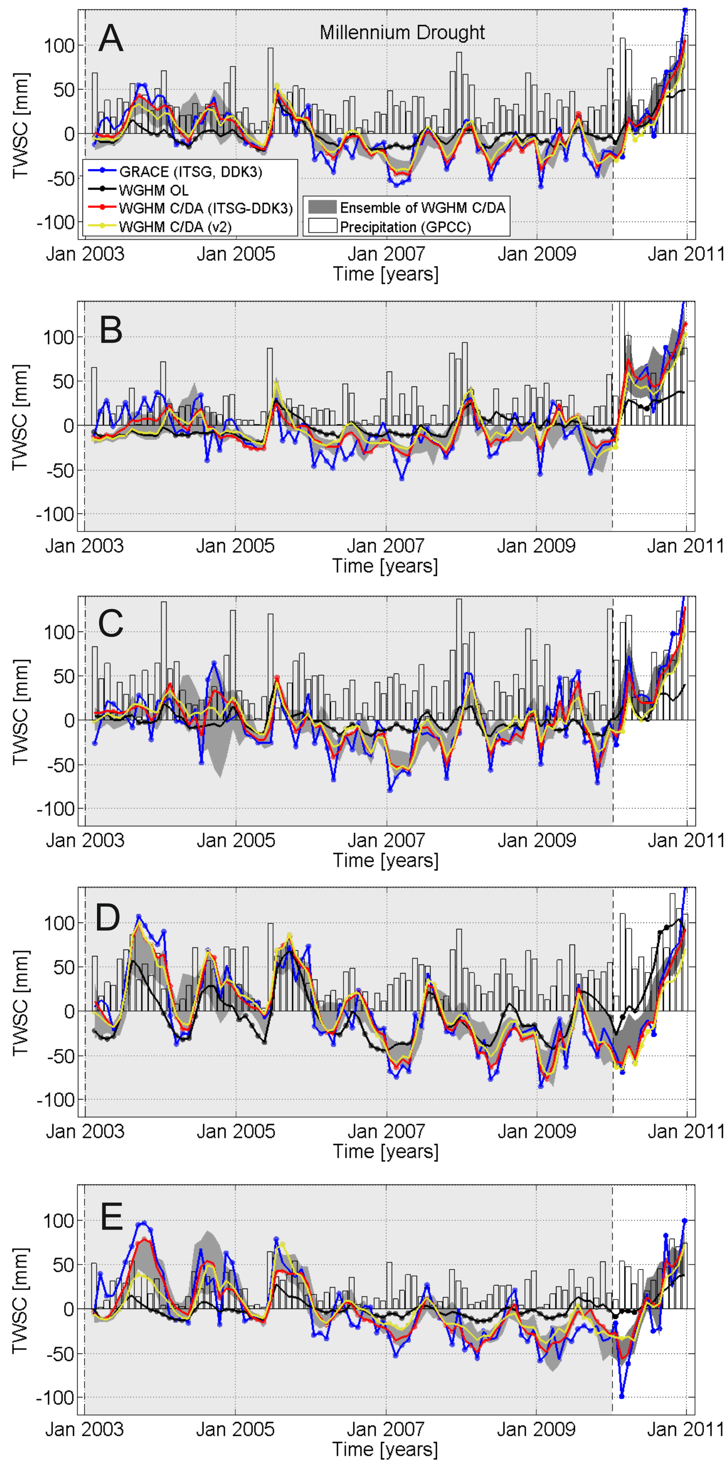

Figure 5: Monthly TWSC in mm averaged (A) over the entire MDB, (B) over NW, (C) over NE, (D) over SE, and (E) over SW. The blue line indicates the TWSC from GRACE (ITSG, DDK3); the black line indicates the WGHM OL simulation; the red line indicates the WGHM simulation after C/DA of GRACE (ITSG, DDK3), and the yellow line the WGHM simulation after C/DA (v2) of GRACE (ITSG, DDK3). The dark gray area represents the range of all C/DA results (see Tab. 2 for C/DA configurations).

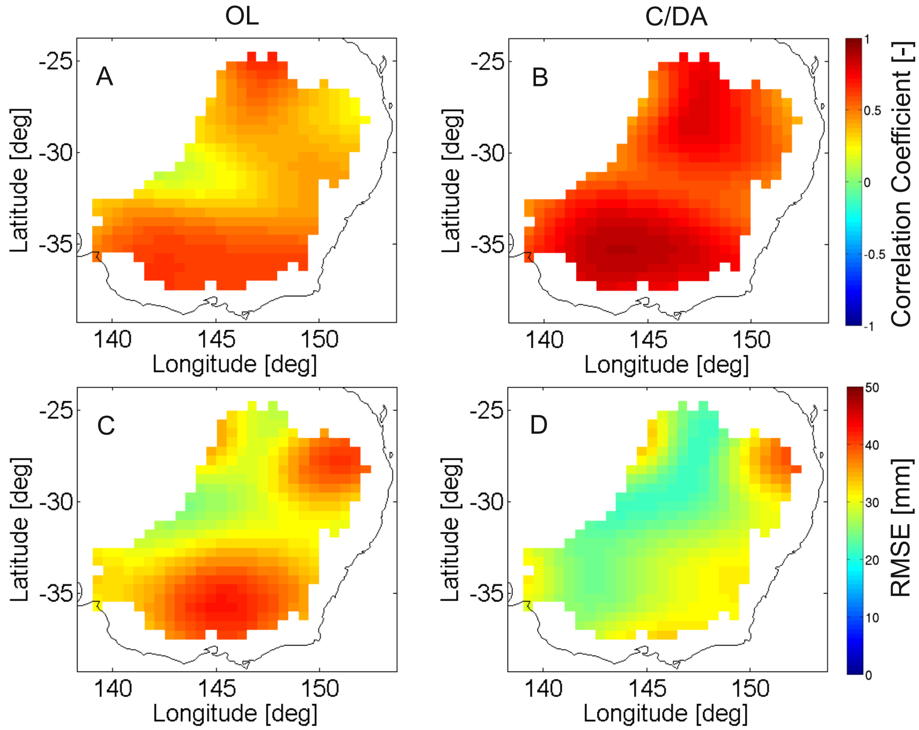

Figure 6: Gridded correlation coefficients between WGHM TWSC simulation and ITSG-Grace2014 TWSC after applying DDK3 filtering for both; (A) for the OL, (B) after applying the C/DA. Gridded root mean square error (RMSE) in mm estimated (C) from the differences between OL TWSC and those of GRACE, and (D) from the C/DA TWSC and GRACE TWSC.

assessment. As it was expected from the basin averaged results (Fig. 4), the DDK3-filtered OL TWSC does not contain significant linear trends (see Fig. 7 (A)), while in the non-smoothed simulations, moderate negative trends can be found over parts of the north and south-west of the MDB (see Fig. 7 (D)). After applying the C/DA based on ITSG-DDK3, a negative trend in TWSC is introduced mainly to the south, which can be seen in Fig. 7 (B) and (E). The restored linear trends (Fig. 7 (B)) are in better agreement with those of GRACE compared to the OL simulation (Fig. 7 (C)).

Our results indicate that the CD/A also influences the seasonal skill of WGHM. In Fig. 8, the annual amplitude of TWSC for 2003-2009 is shown. The DDK3-filtered values, estimated from the OL, C/DA, and ITSG-Grace2014, are shown in Fig. 8 (A), (B), and (C), respectively. Comparing the spatial distributions and magnitude of the annual cycle, one can easily see that the C/DA results (in B) are tuned towards GRACE estimation (in (C)) compared to those of the OL (in A). In Fig. 8 (D) and (E), the annual amplitudes of TWSC, without applying a filter, are shown, which indicate that the OL simulation underestimates the annual cycle mainly over the south and north-east (Fig. 8 (D)). This is however improved after applying C/DA (see Fig. 8 (E)).

Table 5: Linear trends (in mm/year) of TWSC and their uncertainty during 2003-2009 computed for the entire MDB and the four sub-basins (basins are shown in Fig. 1). The OL results and those after the C/DA of WGHM using ITSG-Grace2014-DDK3 are shown in the second and third columns, respectively. The averages of linear trends and their errors from different GRACE products, and after applying different filtering techniques are reported in the fourth and fifth columns, respectively. Results of the C/DA (v2) is reported in the last column.

Basin	OL	ITSG-DDK3	GRACE Products	GRACE Filtering	C/DA (v2)
MDB	-0.9 ± 0.05	-6.5 ± 0.3	-5.3 ± 1.6	-5.7 ± 1.1	-5.5 ± 0.1
NW	2.1 ± 0.09	-1.0 ± 0.2	-0.8 ± 1.0	-2.0 ± 1.0	-0.3 ± 0.2
NE	-1.6 ± 0.04	-4.2 ± 0.5	-2.3 ± 2.1	-3.9 ± 0.4	-3.8 ± 0.1
SE	-3.7 ± 0.13	-13.0 ± 0.7	-10.9 ± 2.5	-9.7 ± 3.4	-12.2 ± 0.2
SW	-0.4 ± 0.11	-10.0 ± 0.3	-9.7 ± 0.6	-9.0 ± 1.8	-7.3 ± 0.1

4.3. Details of Groundwater Storage Changes

4.3.1. Improvements of the Representation of Groundwater Changes

Among various water storage compartments simulated by WGHM, our results indicate that the negative linear trends, restored in WGHM by assimilating GRACE TWSC, are predominantly associated with the groundwater compartment, and much less with the surface water and soil water storage compartments (see the results of the surface and soil compartments in the Supplementary Data, Figs. S1 and S2). While in van Dijk et al. (2013) a decrease in public reservoirs is reported for 2006-2007, our analysis agrees well with the findings in Leblanc et al. (2009), who did not find considerable trend in surface water and soil moisture in MDB since 2003. This comparison does not allow to distinguish whether OL or the C/DA results are better. However, it clearly shows that C/DA did not erroneously introduce decreasing trends to the soil and surface water components (as could have happened given the decreasing trend in TWSC). This was, however, correctly translated by C/DA to a water decline in the groundwater storage only.

In Fig. 9, WGHM's groundwater time series (derived by OL runs and after C/DA) and the observed groundwater well time series are shown. Results are averaged over the entire MDB and its four sub-basins of Fig. 1. All graphs in Fig. 9 (A) to (E) indicate nearly constant values in the OL simulations (black lines), which are not consistent with the well measurements (blue lines) that show strong annual variability and linear trends within most sub-basins. After C/DA, the agreement of simulated and observed groundwater is clearly improved for the entire MDB and all four sub-basins: Seasonal variability and

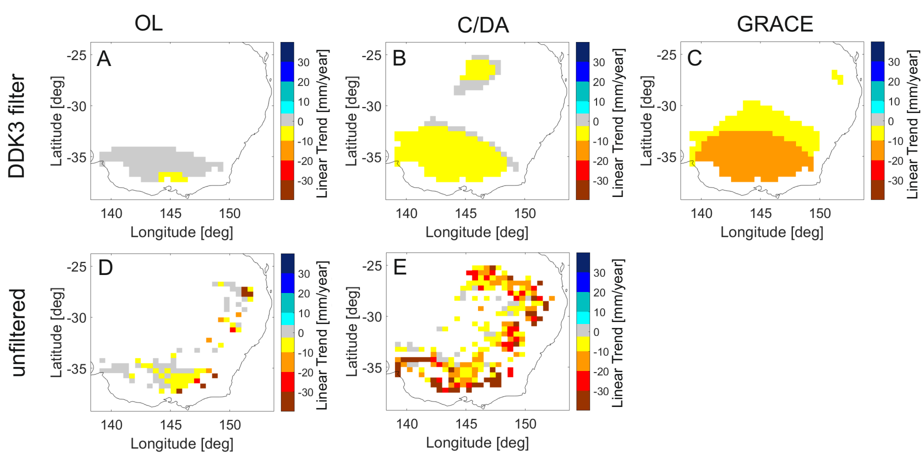

Figure 7: An overview of statistically significant linear trend in TWSC (in mm/year) within the MDB during 2003-2009. The results in (A), (B), and (C) are respectively derived after applying the DDK3 filter to the WGHM OL runs, improved WGHM after C/DA, and from ITSG-Grace2014. In (D) and (E), the linear trend from the original OL TWSC simulations of WGHM and after applying C/DA without any spatial filtering are shown, respectively.

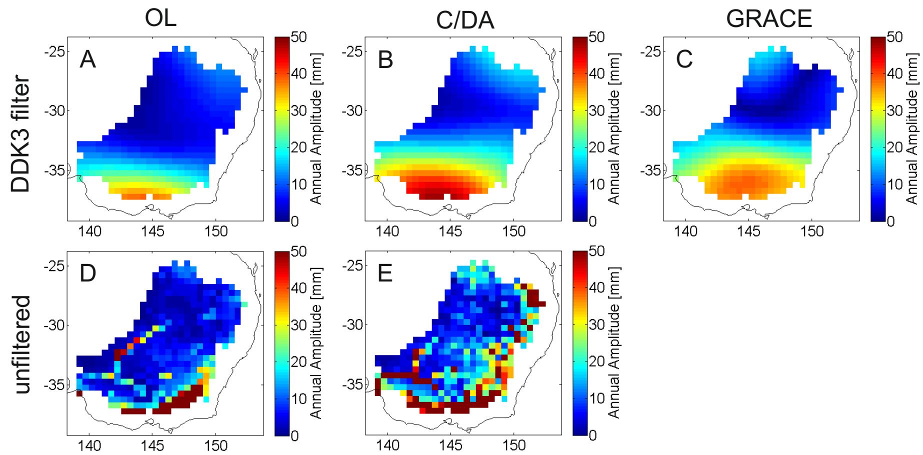

Figure 8: Annual amplitude of TWSC (in mm) from WGHM and GRACE. The DDK3-filtered results are shown in (A) using WGHM OL, (B) using improved WGHM after C/DA, and (C) using ITSG-Grace2014. In (D) and (E), the annual amplitudes from the original OL TWSC simulations of WGHM and after applying C/DA without any spatial filtering are shown, respectively.

negative linear trends are merged towards groundwater observations. The correlation coefficients of the OL and C/DA time series with respect to the groundwater observation time series are shown in Tab. 6.

The correlation coefficients are found to be even higher for the C/DA (v2) variant except for the south-western Murray region. The groundwater changes from the OL are found to be phase shifted compared to the wells observations, especially over the Murray sub-basins. As a result, small correlation coefficients are found between them. After C/DA, the phase shift is reduced over all regions except for the north-eastern Darling Basin (NE). The improvements occur mainly during 2006-2009, which are reflected in the higher correlation coefficients (Tab. 6). However, the inter-annual variability during 2003-2005 seems to be clearly underestimated in all regions. In 2010, the increase in groundwater is not yet captured by the C/DA variants that calibrate all 22 WGHM parameters. In contrast, the C/DA (v2) is able to reflect this increase in the groundwater compartment since the adjusted parameters are more efficient.

Groundwater observations are provided to us on $1^\circ \times 1^\circ$ grid cells. Thus, the OL and C/DA groundwater simulations are averaged on the same grid and the correlation coefficients before and after C/DA are shown in Fig. 10. Correlation coefficients are found to be increased in some grid points, while for others no changes are observed. C/DA (v2) further improves the correlation coefficients over the Darling and Murray regions.

Table 6: Correlation coefficients between WGHM simulated groundwater changes (OL and after C/DA) and well measurements covering 2003-2009. MDB and its sub-basins are defined according to Fig. 1.

Basin	OL	ITSG-DDK3	C/DA (v2)
MDB	0.53	0.66 (+0.13)	0.72 (+0.19)
NW	-0.01	0.74 (+0.75)	0.82 (+0.83)
NE	0.32	0.16 (-0.17)	0.28 (-0.04)
SE	0.01	0.36 (+0.34)	0.41 (+0.39)
SW	-0.05	0.77 (+0.82)	0.69 (+0.75)

4.3.2. Spatial Distribution of the Groundwater Depletion

In Fig. 11 (A), (B) and (C), statistically significant linear trends in groundwater changes from the OL and C/DA variants of WGHM and the well measurements are shown. The OL simulation shows no trend in the majority of the grid cells. Assimilating ITSG-DDK3 TWSC observations into WGHM, restores negative trends to more than half of the grid cells. These trends correspond well to the linear trends derived from groundwater well measurements, which show strong linear trends (up to more than 40 mm/year) predominantly in the north and the south-east of the MDB. Also for the original WGHM groundwater time series on the $0.5° \times 0.5°$, OL shows no linear trend nearly all over the MDB (Fig. 11 (D)). The more highly resolved grid values show that assimilating GRACE TWSC restores a negative trend predominantly in the north, east and south-east of the MDB (Fig. 11 (E)). Several grid cells especially in the south-east exhibit water decline of more than 40 mm/year. In case of C/DA (v2), the linear trends restored to the groundwater compartment are smaller for various grid cells compared to Fig. 11 (E) but considerably improved compared to the

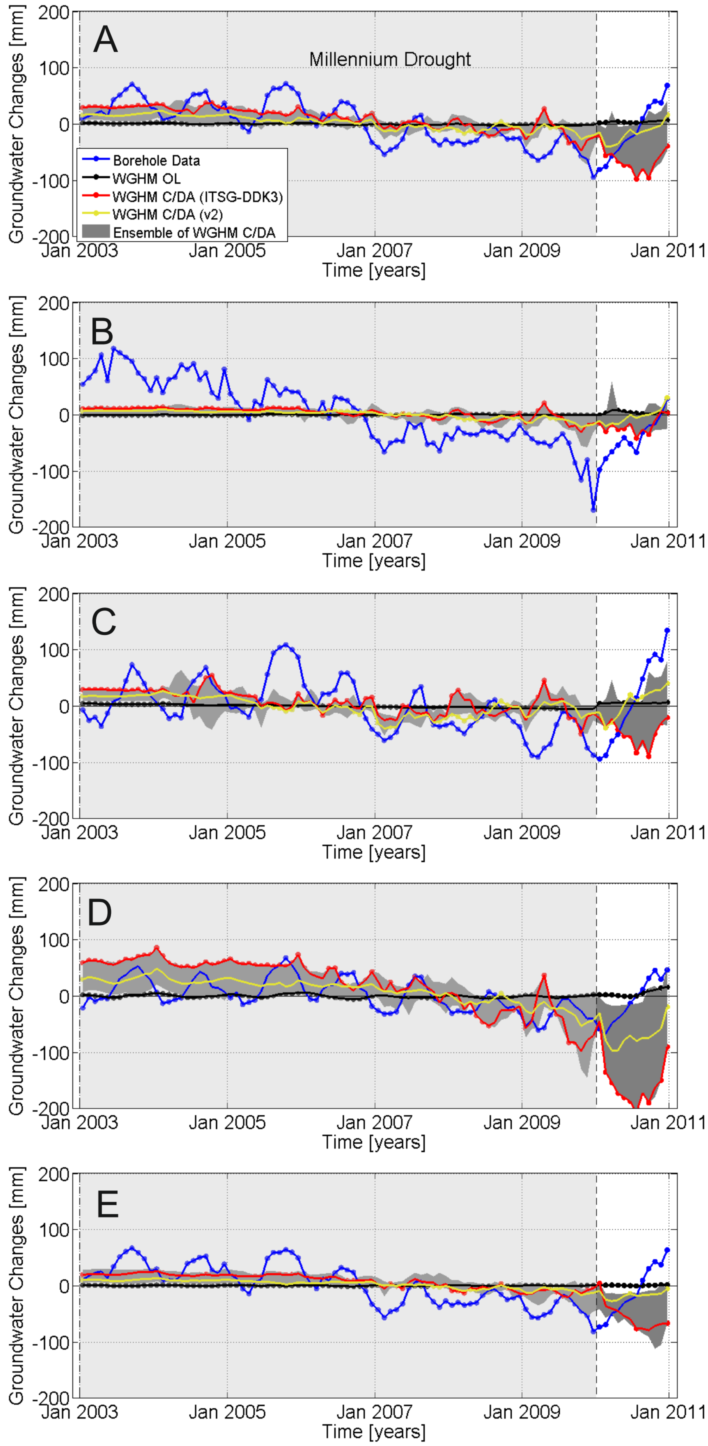

Figure 9: Monthly time series of groundwater changes (in mm) averaged (A) over the entire MDB, (B) over NW, (C) over NE, (D) over SE, and (E) over SW. The blue line indicates the groundwater observations; the black line indicates the WGHM OL simulation; the red line indicates the WGHM simulation after C/DA of GRACE (ITSG, DDK3), and the yellow line the WGHM simulation after C/DA (v2) of GRACE (ITSG, DDK3). The gray area represents the range of all C/DA results (see Tab. 2 for C/DA configurations).

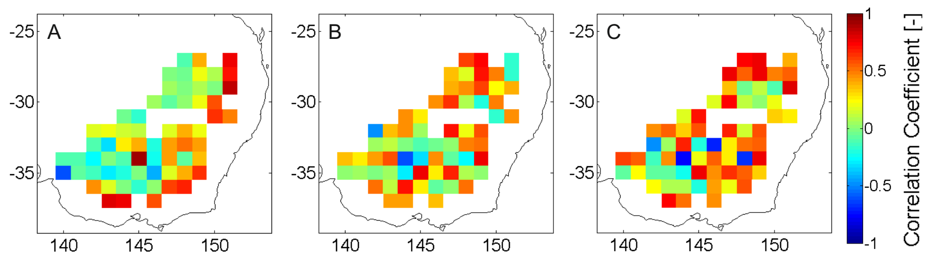

Figure 10: Correlation coefficients between wells data and: (A) the OL groundwater simulations, (B) the C/DA simulations (case ITSG-DDK3 while calibrating all 22 model parameters), and (C) the C/DA (v2) simulations (calibrating only 3 parameters).

OL variant.

The spatially averaged linear trends for the MDB and its four sub-basins are reported in Tab. 7. We have good confidence in the spatial averages of GRACE-derived TWSC over large areas such as the sub-basins of the MDB and their spatial distributions. These are accordingly integrated into the WGHM after C/DA. In contrast, the spatial averages over large areas from in-situ groundwater measurements are strongly influenced by interpolation errors, especially if well observations are obtained close to irrigation wells. More generally, groundwater observation wells tend to be positioned in reliable and productive aquifers. These may occupy only a small part of the landscape, and thus are not representative for the entire MDB (Tregoning et al., 2012, chapters 5 and 6). The ranking based on GRACE and the C/DA variants of WGHM also fits well to the spatial distribution of the difference in mean annual precipitation. Thus, it seems justified to trust the GRACE observations more than the groundwater well interpolation at large scales.

As for the estimation of linear trends in TWSC after C/DA, the choice of GRACE products and filtering clearly affects the linear trends in groundwater, which reaches up to 2 mm/year averaged over the entire MDB. The smallest impact of up to 1 mm/year occurred in the north-western Darling Basin (NW), which also exhibits the smallest linear trend among the sub-basins. In contrast, the linear trend in the south-eastern Murray Basin (NE) is affected by more than 6 mm/year.

In order to demonstrate the impact of post-processing of groundwater measurements on the validation of results, we modify the post-processing in two ways: First, instead of using an average specific yield value of 0.1, values based on a geology map are applied to convert groundwater levels to equivalent water heights (Viney et al., 2015), i.e. values between 0.06 and 0.30; Second, we identify those (gridded) groundwater time series that exhibit the highest RMSE compared to the sub-basin averaged time series. It is assumed that these time series might be representative for the $1° \times 1°$ grid cell but not for the sub-basin average. Therefore, these grids are neglected and the sub-basin averages are recomputed. From the different post-processing strategies an average water storage decline of -11.6 mm/year is determined with a standard deviation of $\pm$ 6.5 mm/year within the south-eastern Murray Basin (SE) and an average decline of -33.3 mm/year with a standard deviation of $\pm$ 14.5 mm/year within the north-western Darling Basin (NW; see last column in Tab. 7). These large differences indicate the high dependency of the groundwater estimations on the choice of specific yield and on the errors for computing (sub-)basin averages from point measurements. The effect is found to be considerably higher than the effect of the chosen GRACE product and the choice of the TWSC filtering approach.

4.4. Model Parameter Calibration

An extensive section is provided in the Supplementary Data to discuss the calibration of all the 22 WGHM parameters within the C/DA against calibrating only the 3 parameters of the root depth multiplier, the net radiation multiplier, and the groundwater outflow coefficient, which the implementation is called C/DA (v2) from now on. We also modify a priori PDFs of the wetland and lake depth and the groundwater outflow coefficient based on the investigation of the update increments (see Tab. 1). The calibrated parameter values are shown in Sect. 8 of the Supplementary Data. In general, our results indicate that by calibrating all 22 parameters in some instances one can find few of them that

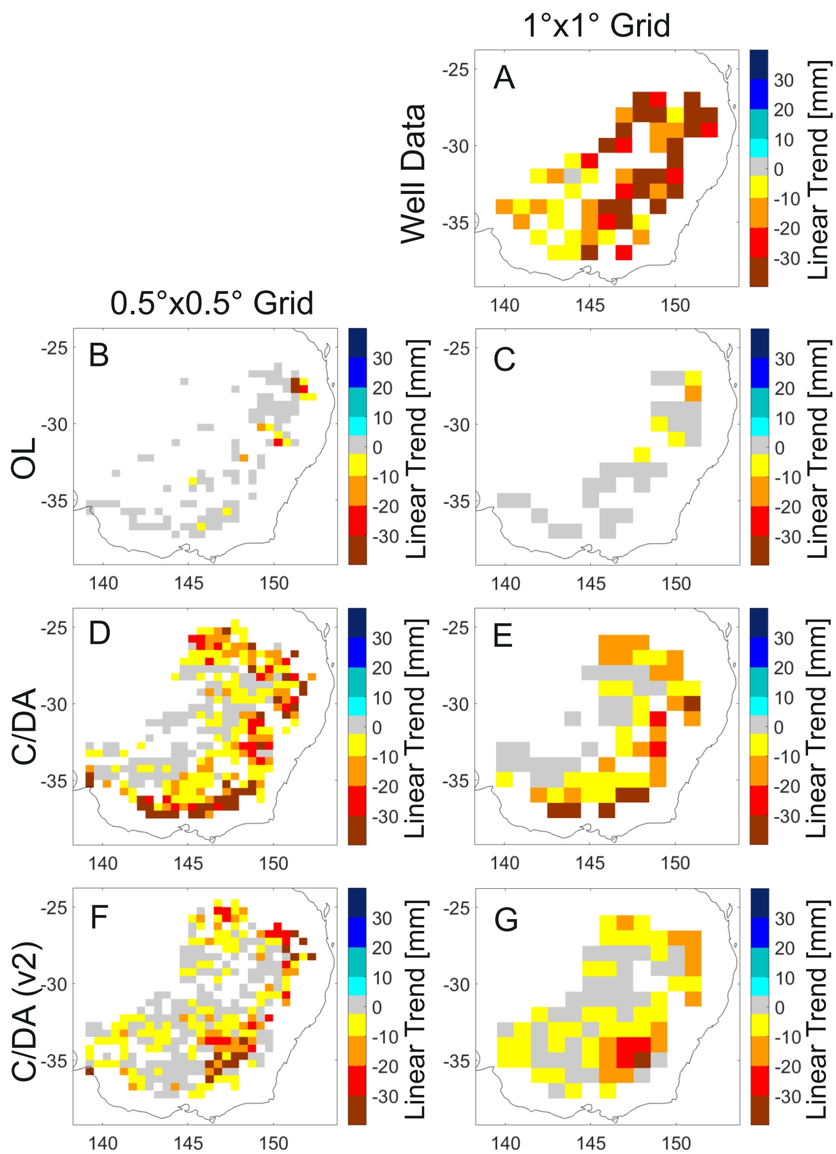

Figure 11: Significant linear trend in groundwater changes (in mm/year) within the MDB during 2003-2009. The results in (A), (C), (E) and (G) are respectively derived from the groundwater measurements, the WGHM OL, WGHM after C/DA while calibrating all 22 parameters, and from WGHM after C/DA (v2) while calibrating only 3 parameters. Results are spatially averaged over 1°×1° grid cells. In (B), (D), and (F), the linear trend from the original OL groundwater simulations of WGHM and after applying C/DA and C/DA (v2) are shown, respectively.

Table 7: Linear trends (in mm/year) in groundwater changes and their uncertainties during 2003-2009 computed for the entire MDB and the four sub-basins. The linear trends estimated from groundwater measurements (specific yield = 0.1) are provided in the second column. The results of WGHM OL and after C/DA of ITSG-DDK3 are shown in the third and fourth columns, respectively. The averages of linear trends and standard deviations from different GRACE products, and after applying different filtering techniques are reported in the fifth and sixth columns, respectively. The results of C/DA (v2) are provided in the seventh column. In the last column, the averages of linear trends and standard deviations from different post-processing strategies (specific yield modification, removing outliers) for the groundwater measurements are shown.

Basin	Data	OL	ITSG-DDK3	GRACE Product	GRACE Filtering	C/DA (v2)	Groundwater Variant
MDB	-16.1	-0.6 ± 0.01	-8.3 ± 0.2	-7.7 ± 2.4	-5.9 ± 2.2	-5.4 ± 0.1	-20.5 ± 4.0
NW	-28.7	0.1 ± 0.00	-3.6 ± 0.2	-4.5 ± 1.0	-2.9 ± 0.8	-3.1 ± 0.1	-33.3 ± 14.5
NE	-12.6	-1.4 ± 0.02	-6.4 ± 0.5	-5.2 ± 2.5	-5.0 ± 1.4	-5.6 ± 0.1	-22.5 ± 15.6
SE	-8.4	-0.4 ± 0.02	-19.2 ± 0.6	-16.3 ± 6.3	-12.1 ± 6.3	-9.6 ± 0.1	-11.6 ± 6.5
SW	-14.9	-0.1 ± 0.01	-5.8 ± 0.3	-7.2 ± 1.6	-4.8 ± 1.0	-3.4 ± 0.1	-14.7 ± 9.5

are not converged to a value within a priori range, while in C/DA (v2), all three parameters converge and their uncertainties are considerably reduced. This does not however necessary imply that one version is better suited to achieve more accurate water storage simulations. Therefore, in the following, we mainly focus on interpreting the C/DA results derived from both versions.

The C/DA update increments, i.e. the difference between model prediction and model update, of the total and individual water storage compartments are presented in Fig. 12. Since mass is not conserved in the EnKF updates, these increments indicate how the water mass balance is violated by data assimilation (see also Sect. 5 of the Supplementary Data). The updates of soil water are higher in the east and south-east of the MDB, and decrease in western direction (Fig. 12 (B)). For groundwater, the same spatial pattern is visible but the amount of water mass associated with the groundwater compartment is considerable larger (Fig. 12 (C)). In Sect. 4.3, it is already shown that the updates for the groundwater compartment lead to improved agreements with in-situ observations. In addition, the updates for the soil water compartments improve the seasonal representation of simulated TWSC after C/DA compared to the OL results (see Fig. S1 in the Supplementary Data). We find only small update increments for lakes, which seems to be reasonable, since only a few small surface water bodies are located in the MDB (Fig. 12 (D)).

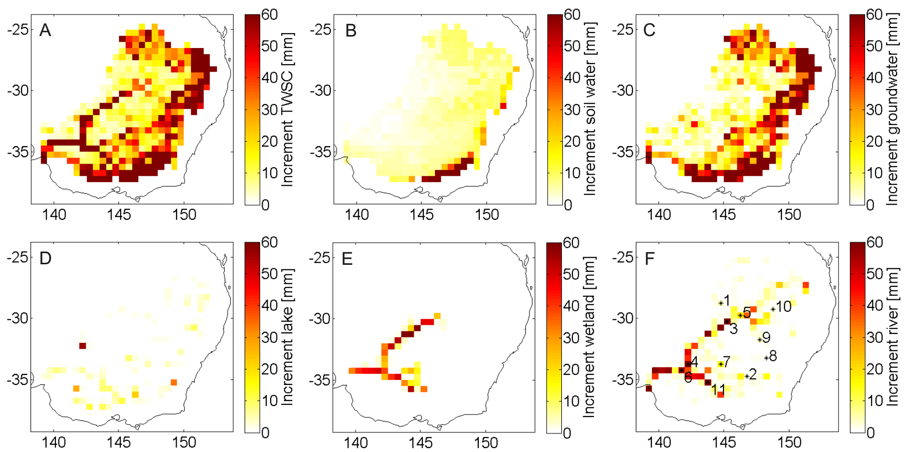

Figure 12: Root mean square (RMS) of monthly update increments after applying the C/DA to integrate WGHM with TWSC from ITSG-DDK3 (calibrating all 22 parameters) for (A) TWSC, (B) soil water, (C) groundwater, (D) lakes, (E) wetlands, and (F) rivers. In (F), the locations of the river discharge stations that have been used to calibrate the WaterGAP 2.2 model version are shown by the black dots.

4.5. River Discharge and River Level

To answer the Objective (5) of this paper, in sections 4.2 and 4.3, we showed how the C/DA improves total and individual water storage simulations of WGHM. Further insights will be provided in section 5. In this section, the impact of C/DA on WGHM's river discharge and river level (storage) simulations is provided. Since GRACE data have a direct influence on water storage simulations and indirectly change simulated fluxes (e.g., river discharge, see Schumacher et al., 2015), one only needs to show the latter has not been worsen by the C/DA.

We use river discharge observations provided by the Bureau of Meteorology (BoM, `http://www.bom.gov.au/waterdata/`) to validate the updated river compartment. In Fig. 13, the time series of river discharge are shown for three selected stations while calibrating 22 parameters in (A), (C) and (E), as well as for the C/DA (v2) in (B), (D) and (F). At the Paroo River at Caiwarro (BoM station number 424201A; number 1 in Fig. 12 (F)), the WGHM OL simulated river discharge fits quite well to the observations but the high flows in 2004, 2008 and 2010 are underestimated (Fig. 13 (A)). After performing the C/DA run with 22 parameters, the discharge values represent the high flows better than OL.

For other stations, the river compartment is found to be overestimated e.g., during 2003-2004, 2008-2009, and during the wet year 2010. In Fig. 13 (B) and (C), we show the time series at Darling River at Burtundy (BoM station number 425007; number 4 in Fig. 12 (F)) and Lachlan River at Booligal (BoM station number 412005, number 7 in Fig. 12 (F)) as examples. After reducing the number of calibration parameters, i.e. within the C/DA (v2) run, the river discharge simulation is found to be improved. At Caiwarro (Fig. 13 (B)), the high flows in 2004 and 2008 are better represented compared to the OL and the previous C/DA run. However, in spring 2008 still two peaks are simulated although only one of them is observed. At the other river discharge station, the simulations are also improved. The high flows in 2010 are found to be much closer to the observations for the C/DA (v2) run, especially at Burtundy (Fig. 13 (F)) but during the drought period they are still found to be overestimated.

We also compare simulated river storage with a number of stations provided by the Murray-Darling Basin Authority (`https://riverdata.mdba.gov.au/system-view`). For example, in Fig. 14, river storage outputs from WGHM are compared with the time series of level changes derived from Murray's upstream, which is close to station 4 in Fig. 12(F). The comparison is limited to 2007.5-2011 during which the gauge data is available. Our results indicate that the open-loop river storage is not well compared with observations (RMSE of 1.42), for example, high peaks are detected in 2008 and 2010, which are not found in the measured levels. After applying the C/DA (both versions, however, the mentioned peaks are vanished and the general evolution of estimated

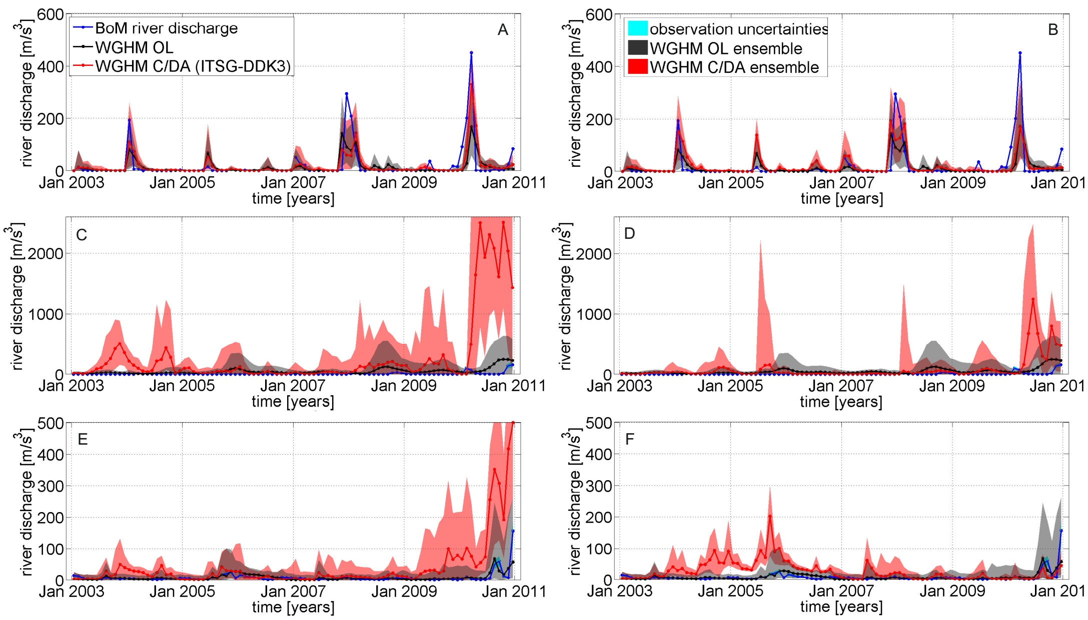

Figure 13: Time series of river discharge (in m/s^3) at three selected river discharge stations: (A, B) Paroo River at Caiwarro (BoM station number 424201A; number 1 in Fig. 12 (F)), (C, D) Darling River at Burtundy (BoM station number 425007; number 4 in Fig. 12 F); and (E, F) Lachlan River at Booligal (BoM station number 412005, number 7 in Fig. 12 (F)). The left column presents C/DA results from the ITSG-DDK3 case for which all 22 parameters have been calibrated, and the right column presents the C/DA (v2) while calibrating only 3 parameters.

river storage fairly well follows that of the gauge data, i.e., RMSE reduces to 0.6. Correlation coefficients between the OL river level simulations and gauge observations indicate a weak correspondence of 0.15 (p-value showed that this correlation is not significant). This is increased to the statistically significant value of 0.52 (significant according to p-values) after implementing the C/DA. Impact of the 2010's La Niña is fairly well reflected in the C/DA derived river storage (compare the red and yellow curves in Fig. 14 with the observation curve in blue). Comparable results are found for the downstream station, which is not shown here.

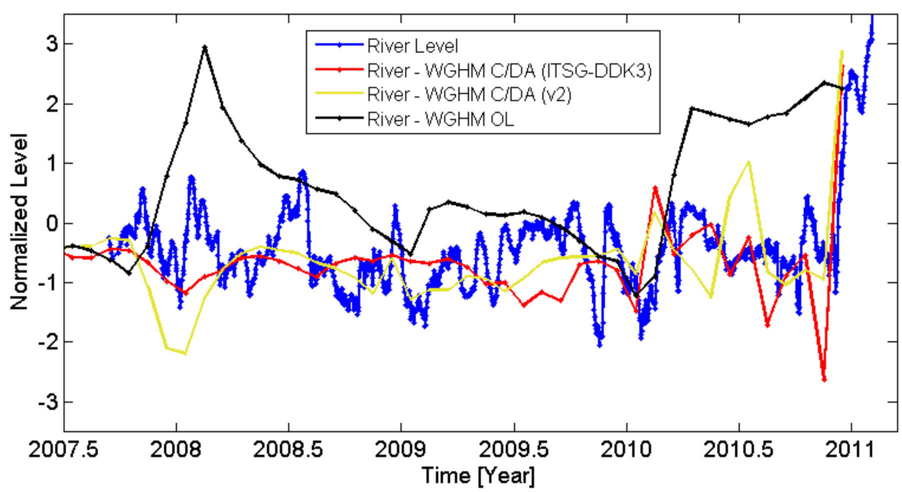

Figure 14: Time series of river level at the station 4 in Fig. 12 (F). The time series are temporally normalized, thus, they are unit-less.

5. Discussion

5.1. Choice of GRACE Product and Post-Processing

Several GRACE products (ITSG-Grace2016, GFZ, and JPL) with different spatial filters (the isotropic Gaussian and the anisotropic DDK filter) are assessed within the proposed C/DA in the MDB. Our analysis of the updated TWSC and groundwater changes is not able to suggest a single product or spatial filtering strategy that exhibits always superior metrics (here in terms of

RMSE and correlation coefficients). The magnitude of the differences among the EnKF variants is similar to the magnitude of the differences between the considered GRACE variants itself. The uncertainty information obtained for the ITSG-DDK3 results represents these differences among the EnKF variants fairly well. Thus, a careful incorporation of the GRACE TWSC uncertainty information provides reliable information of the spread of the EnKF updates that might have been obtained when selecting another data product.

5.2. Effect of Equifinality of Calibration Parameters on C/DA Results

We test calibrating only three parameters within the C/DA in order to mitigate the equifinality problem. We find that the three selected parameters converge to a constant value during the drought period and their uncertainty is clearly reduced. Although, improvements are already found for groundwater simulations during the drought period when calibrating 22 model parameters, it is not possible to constrain these many parameters using GRACE to improve the simulation of individual water storages when climate conditions rapidly and strongly change, i.e. the occurrence of strong rainfall events in 2010 after a long drought. This is, however, achieved by reducing the number of calibrated parameters. As a result, we find a strong positive impact on the EnKF updated of groundwater changes, especially in 2010.

In summary, parameter updating using GRACE observations is very challenging. Due to its current coarse spatial resolution and highly correlated errors, it might have limitations and might result in poorly constrained WGHM parameters that actually steer the simulation of individual water storage compartments or fluxes. An improved spatial resolution, which is expected from the GRACE follow on (GRACE-FO) mission (scheduled launch at the end of 2017), and a combination with other remote sensing observations might lead to better constrained parameter values.

5.3. Application of the C/DA Framework within a (semi-)arid River Basin

We find that all the EnKF variants improve the WGHM simulations and outperform the original simulations in terms of RMSE and correlation for the (semi-)arid basin of the Murray and Darling rivers and its four sub-basins, and even on the $0.5^\circ \times 0.5^\circ$ grid. The WGHM grid is much finer resolved than the spatial resolution of GRACE data and therefore this result is not self-evident. We would like to recall that we integrated GRACE data averaged over the four major sub-basins of the MDB and not at each individual WGHM grid point. Thus, the results give confidence that GRACE data can be horizontally downscaled by the C/DA within (semi-)arid regions.

The water decline is primarily associated with the groundwater compartment, which is confirmed through validation with independent well measurements. However, in three out of the four MDB sub-basins, the restored trends are much smaller than the observed ones. For a realistic assessment of the C/DA performance, it is important to be aware that uncertainties exist also for the ground-based validation data and these should not be treated as truth. Thus, a perfect agreement between groundwater simulations after C/DA and groundwater measurements cannot be expected. Using groundwater simulations improved by C/DA of GRACE data has therefore the advantage that no specific yield estimate and no spatial interpolation are required. The results indicate that the groundwater simulations in the Darling Basin (NE) are less improved compared to other regions in terms of correlation coefficients. The hydrological reason for this is a different behavior in terms of annual cycles between GRACE TWSC and groundwater well observations in this region. In fact, seasonality of GRACE TWSC is less pronounced in the Darling Basin (NE), but it is visible in the in-situ well measurements. Thus, C/DA is not able to correct the seasonality of WGHM's groundwater simulations in this sub-basin.

No significant trends are found in the surface water and soil water storage compartments after 2003, which is in agreement with the analysis performed in Leblanc et al. (2009). If the water decline was solely climate related, we would expect more or less similar rates of decline in the surface, soil and groundwater compartments. Our investigations however suggest that anthropogenic influence on the hydrological cycle, in form of groundwater abstraction, is the reason for the significant water decline within a wide area of the MDB (see, e.g., Fig. S8

(C), in which the net abstraction multiplier for groundwater is mostly larger than 1), which is supported by local reports (e.g., from the Australian Bureau of Meteorology).

The impact of C/DA on TWSC in the northern and southern regions of the MDB is found to be different. Stronger seasonal amplitudes in the south result in higher correlation coefficients but also higher RMSE values. The response of the hydrological resources within the four sub-basins to the meteorological drought also differs for the northern and southern sub-basins. The spatial distribution of the BoM precipitation data shows that more rainfall occurred in the northern MDB, especially in the Darling Basin (NW), compared to the other sub-basins. Thus, the impact of the Millennium Drought is found to be predominant in the southern MDB, which is in agreement with the pronounced hydrological drought in the south observed by GRACE. The negative linear trends of TWSC, as well as groundwater are less strong in the north compared to the south. The reason might not only be related to the climatological conditions but also to the human influence on the water resources in the MDB. Due to surface water subtractions, e.g., from the Darling River in the north, less water enters the Murray sub-basins in the south. In order to ensure irrigation and therefore continue agricultural activities, groundwater is even more heavily pumped resulting in the observed decline of TWSC and groundwater resources. This statement is supported by the engagement of the Murray Darling Basin Authority (see https://www.mdba.gov.au/) that established a Basin Plan to manage the entire basin as one system beyond political boarders in order to balance the water use and to ensure a sustainable use of the water resources. The hydrological drought is therefore a consequence of the mixture of dry meteorological conditions and human impact on the water cycle, which is especially pronounced in the southern MDB.

According to the results we show above, we are confident to state that the C/DA approach can be applied to use GRACE and improve a model (here WGHM) in a (semi-)arid region without tuning its setting. However, few problems remain for the simulation of river discharge. It is important to keep in mind that assimilating GRACE data into a model does not directly affect the river discharge simulation but rather through the calibration of several model parameters. Therefore, a perfect agreement with river discharge observations for the entire basin cannot be expected at least by the current resolution of GRACE products. However, after applying the C/DA we find a good agreement between river storage simulation of WGHM and gauge observations at the Murray's upstream and downstream. Therefore, our conclusion is that the C/DA successfully improves storage simulation of WGHM. To achieve better discharge simulations, one likely needs to assimilate observations in the form of water fluxes (e.g., river flow and/or multiple altimetry observations), which will be addressed in future.

5.4. Groundwater and Soil Storage Response to Climate Variability and Water Abstraction

In this section, we explore the spatial and temporal variability of soil water storage and groundwater changes within the entire Murray Darling Basin by applying a principal component analysis (PCA, Forootan, 2014, chapter 3) on the outputs of WGHM before and after implementing C/DA. This analysis helps us to understand how these storages evolve after a dry season and how they response to climate variability.

In Figs. 15 and 16, PCA results of soil water and groundwater storage changes are shown, respectively. In both figures, the spatial patterns are empirical orthogonal function (EOF) in mm that can be interpreted as anomaly maps and their corresponding temporal evolutions are unit-less (normalized) evolutions shown on right and labeled as principal component (PC). By multiplying EOF and PC, one can reconstruct spatio-temporal variability of soil and groundwater storage changes in the region, while representing their maximum variance. Our computations indicate that the first mode of soil (EOF1 and PC1 of soil in Fig. 15) is equivalent with 62% of the total variance and the one of groundwater (EOF1 and PC1 in Fig. 16) represents 78% of the total variance. For brevity, in both Figs. 15 and 16, we only show the EOF that corresponds to the open loop output but PCs are estimated separately by applying PCA on the soil water and groundwater storage outputs of open loop, C/DA with all parameters, and C/DA with 3 parameters. The presentation of PCs is limited to the period of 2007.5-2011, within which the PCs are better distinguishable. In both figures, we also show a measure of ENSO events, reflected in the southern oscillation index (SOI), which is downloaded from the website of BoM (`http://www.bom.gov.au/climate/current/soi2.shtml`). Sustained positive values of the SOI used here represent La Niña episodes and its negative values represent El Niño, which respectively correspond to higher and lower than normal precipitation in Australia.

PCA results of soil storage from the open loop output indicate stronger anomalies on the east and north parts of the basin (see EOF1 in Fig. 15), as well as a temporal delay of $\sim$6 months between peaks of ENSO and soil moisture in 2008 and 2009. The strong La Niña in 2010 is found to change the open loop's soil storage outputs quite immediately. We find no obvious trend in the open loop results, which apparently indicate that the history of water storage does not play a major role in simulating the maximum peaks derived from WGHM (see the black curve in Fig. 15). PCs derived from the C/DA outputs reflect the ENSO activity on the basin's soil water storage more realistically. Particularly, we find the dry period of 2008.8-2010.2 causes a decline in soil storage (covering 2009.2-2010.6), which is recovered by the La Niña in the middle of 2010 (see the red and yellow curves in Fig. 15).

Application of C/DA is found very beneficial for improving the representation of groundwater in the basin. The PCA results derived from groundwater output of the open loop run (see the black curve in Fig. 16) indicate a moderate decline until 2010, which is followed by a sudden groundwater recharge that is likely caused by the extensive rainfall in 2010-2011. Groundwater anomalies are found stronger along the river (see EOF in Fig. 16). The computed groundwater PCs, derived after implementing the C/DA (both versions), evolve more naturally than that of the open loop. For example, it is clear that within the La Niña years of 2007.5-2009.5, the rate of groundwater storage decline is quite moderate (see the red and yellow curves in Fig. 16), which likely reflects the impact of water use. An accelerated groundwater depletion is found during 2009-2010.2, which reflects both a strong El Niño and extensive irrigations. Then, the water decline has been gradually recovered by the 2010's La Niña.

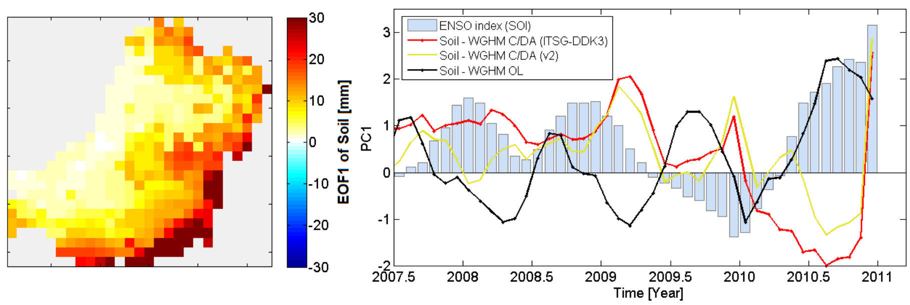

Figure 15: First dominant orthogonal mode, including EOF and its corresponding PC, derived from soil moisture outputs of WGHM. Here EOF1 is derived from the open loop run, but PC1 is derived by applying PCA on the open loop, and two versions of the C/DA outputs and compared to the ENSO index (SOI). This dominant mode represents 62% of variance in soil moisture variability in the region.

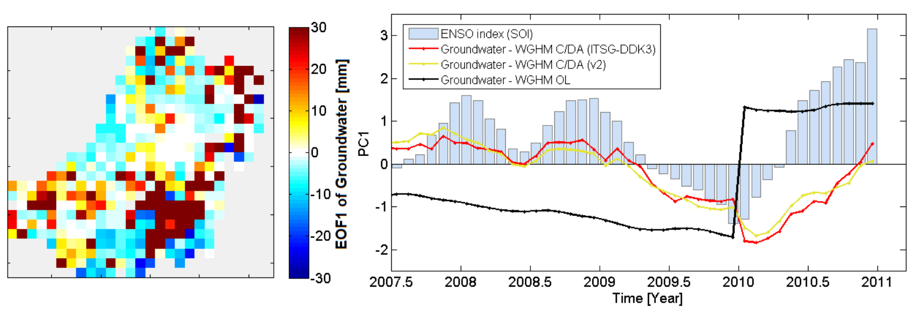

Figure 16: First dominant orthogonal mode, including EOF1 and its corresponding temporal pattern PC1, derived from groundwater outputs of WGHM is shown. Here EOF1 is derived from the open loop run, but PC1 is derived by applying PCA on the open loop, and two versions of the C/DA outputs and compared to the ENSO index (SOI). This dominant mode (EOF1 and PC1 together) represents 78% of variance in groundwater variability in the region.

6. Conclusions and Outlook

A novel calibration and data assimilation (C/DA) framework (Schumacher et al., 2016) is applied here to integrate terrestrial water storage changes (TWSC) observed by GRACE satellites into WGHM within the Murray-Darling Basin (MDB) during 2003-2010. Several technical insights are revealed from this assessment that are summarized in the following:

1. By applying the C/DA approach to the (semi-)arid region of the MDB, it is possible to restore linear trends into WGHM, and also improve the seasonality. As droughts in the MDB are well studied, they can act as a reference for impact models like WGHM. The association of the water decline with the correct water storage compartment, i.e. groundwater in our study, is achieved and validated against ground-based well measurements. Our results show that by implementing C/DA the response of soil water and groundwater storage to climate variability within the MDB has been improved. Our results indicate that although river discharge simulation WGHM in the MDB cannot be improved by assimilating limited resolution GRACE data, its river storage simulations can be considerably (positively) influenced by the C/DA.
2. Difficulties exist when combining information from different sources, i.e. model simulations, remote sensing and ground-based measurements, and of different spatial resolution and accuracy. Uncertainties of ground-based data have to be considered for independent validation of the C/DA performance and a perfect agreement might not be expected.
3. Adapting the C/DA settings to basin-specific characteristics (in this study by modifying a priori PDFs of parameters) and reducing the number of calibration parameters to avoid equifinality has several positive impacts on the C/DA results: (i) the uncertainties of calibration parameters are clearly reduced and their values converge; (ii) the influence of climate condition on the groundwater compartments is captured; and (iii) the

representation of river discharge is clearly improved, especially within the wet year 2010.

4. The calibration of a smaller parameter sub-set clearly suggests that parameter values vary with changes of climatic conditions within the river basin. Therefore, allowing the model parameters to change over time results in a better representation of water storage variability and water fluxes within MDB.

5. Parameter updating using GRACE observations is very challenging, even if the number of calibration parameters is reduced. Combined C/DA using GRACE data is a highly under-determined system that might be limited in constraining individual model parameters, while an optimal parameter set with respect to TWSC simulations is always achieved.

6. Comparing WGHM outputs with in-situ observations indicates that C/DA of GRACE data does not improve river discharge simulations in the MDB, but river storage simulations are significantly improved. This is likely caused by limitation in model equations that transfer storage information to water fluxes (Müller Schmied et al., 2014). This limitation is not only an issue for WGHM but also most of existing hydrological or land surface models.

7. Comparing GRACE data from different providers and using different filtering techniques, it seems that their impact on the final C/DA results is smaller than GRACE data errors.

The assessment of our C/DA approach for assimilating GRACE TWSC into a hydrological model has clearly shown the strengths and limitations of the current implementation. For future work, the application of a multi-criteria C/DA approach in which data on river discharge and possibly surface water level variations are taken into account might further help to improve the C/DA results.

Acknowledgments

The Authors are grateful to Dr Tim R. McVicar (Associate Editor) and five anonymous reviewers, whose constructive comments are used to improve the quality of this study. The support of the German Research Foundation (DFG) within the framework of the Special Priority Program "Mass transport and mass distribution in the system Earth" (SPP1257) under the project REGHYDRO and BAYES-G is gratefully acknowledged. M. Schumacher is thankful for the exchange grant (2015/16 57044996) awarded by the German Academic Exchange Service (DAAD) to visit the Australian National University (ANU). We are grateful to the data providers of GRACE, in-situ water wells, river discharge and level, as well as climate indices used in this study.

References

Abelen S, Seitz F (2013) Relating satellite gravimetry data to global soil moisture products via data harmonization and correlation analysis. Remote Sens Environ 136:89-98. doi:10.1016/j.rse.2013.04.012

Al-Zyoud S, Rühaak W, Forootan E, Sass I (2015) Over exploitation of groundwater in the centre of Amman Zarqa basin-Jordan: evaluation of well data and GRACE satellite observations. Resources 4:819-830. doi:10.3390/resources4040819

Awange J, Fleming KM, Kuhn M, Featherstone WE, Heck B, Anjasmara I (2011) On the suitability of the $4° \times 4°$ GRACE mascon solutions for remote sensing Australian hydrology. Remote Sens Environ 115(3):864-875. doi:10.1016/j.rse.2010.11.014

Bauer-Marschallinger B, Dorigo WA, Wagner W, van Dijk AIJM (2013) How oceanic oscillation drives soil moisture variations over Mainland Australia: an analysis of 32 years of satellite observations. J Climate 26:10159-10173. doi:10.1175/JCLI-D-13-00149.1

Beaumont DA (2012) Floodplain wetland biota in the Murray-Darling Basin: Water and Habitat requirement. Edited by Rogers K and Ralph TJ. CSIRO Publishing: Collingwood, VIC., Australia, 348. doi: 10.1002/rra.1512

Boening C, Willis JK, Landerer FW, Nerem RS, Fasullo J (2012) The 2011 La Niña: So strong, the oceans fell. Geophys Res Lett 39:L19602. doi:10.1029/2012GL053055

Brown NJ, Tregoning P (2010) Quantifying GRACE data contamination effects on hydrological analysis in the Murray-Darling Basin, southeast Australia. Aust J Earth Sci 57(3):329-335. doi:10.1080/08120091003619241.

Burgers G, Van Leeuwen PJ, Evensen G (1998) Analysis scheme in the ensemble Kalman filter. Mon Weather Rev 126:1719-1724. doi:10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2

Chen JL, Wilson CR, Tapley BD, Scanlon B, Güntner A (2016) Long-term groundwater storage change in Victoria, Australia from satellite gravity and in situ observations. Global Planet Change 139:56-65. doi:10.1016/j.gloplacha.2016.01.002

Cheng M, Tapley BD, Ries J C (2013) Deceleration in the Earth's oblateness. J Geophys Res Solid Earth 118:740-747. doi:10.1002/jgrb.50058

Connell D, Grafton Q (Eds) (2011) Basin futures: water reform in the Murray Darling Basin. ANU E Press, Canberra. http://press.anu.edu.au?p=115431

Crosbie RS, Peeters L, Doble RC, Joehnk K, Carrara E, Daamen C, Frost A (2011). AWRA-G: A groundwater component for a continental scale land surface model. 19th International Congress on Modelling and Simulation, Perth, Australia, 12-16 December 2011. http://mssanz.org.au/modsim2011

Döll P, Kaspar F, Lehner B (2003) A global hydrological model for deriving water availability indicators: model tuning and validation. J Hydrol 207:105-134. doi:10.1016/S0022-1694(02)00283-4

Döll P, Müller Schmied H, Schuh C, Portmann FT, Eicker A (2014) Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites. Water Resour Res 50:56985720. doi:10.1002/2014WR015595

Donohue RJ, Roderick ML, McVicar TR (2011) Assessing the differences in sensitivities of runoff to changes in climatic conditions across a large basin. J Hydrol 406:234244. doi:10.1016/j.jhydrol.2011.07.003

Eicker A, Schumacher M, Kusche J, Döll P, Müller Schmied H (2014) Calibration/Data Assimilation Approach for Integrating GRACE Data into the WaterGAP Global Hydrology Model (WGHM) Using an Ensemble Kalman Filter: First Results. Surve Geophys 35(6):1285-1309. doi:10.1007/s10712-014-9309-8

Evensen G (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J Geophys Res 99(C5):10143-10162. doi:10.1029/94JC00572

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

Forootan E (2014) Statistical signal decomposition techniques for analyzing time-variable satellite gravimetry data. PhD thesis, University of Bonn, pp131. http://hss.ulb.uni-bonn.de/2014/3766/3766.htm

Forootan E, Rietbroek R, Kusche J, Sharifi MA, Awange JL, Schmidt M, Famiglietti J (2014) Separation of large scale water storage patterns over Iran usingGRACE, altimetry and hydrological data. Remote Sens Environ 140:580-595. doi:10.1016/j.rse.2013.09.025

Forootan E, Khandu K, Awange J, Schumacher M, Anyah R, van Dijk AIJM, Kusche J (2016) Quantifying the impacts of ENSO and IOD on rain gauge and remotely sensed precipitation products over Australia. Remote Sens Environ 172:50-66. doi:10.1016/j.rse.2015.10.027

Forootan E, Safari A, Mostafaei A, Schumacher M, Delavar M, Awange JL (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 38:591-615. doi:10.1007/s10712-016-9403-1

Fu G, Viney N R, Charles S P, Liu J (2010) Long-termtemporal variation of extreme rainfall events in Australia: 19102006. Journal of Hydrometeorology 11:950-965. http://dx.doi.org/10.1175/2010JHM1204.1

Gallant AJE, Gergis J (2011) An experimental streamflow reconstruction for the River Murray, Australia, 1783-1988. Water Resour Res 47(12):W00G04. doi:10.1029/2010WR009832

Garćia-Garćia D, Ummenhofer CC, Zlotnicki V (2011) Australian water mass variations from GRACE data linked to Indo-Pacific climate variability. Remote Sens Environ 115:2175-2183. doi:10.1016/j.rse.2011.04.007

Gergis J, Gallant AJE, Braganza K, Karoly DJ, Allen K, Cullen L, D'Arrigo R, Goodwin I, Grierson P, McGregor S (2012) On the long-term context of the 1997-2009 'Big Dry'in South-Eastern Australia: insights from a 206-year multi-proxy rainfall reconstruction. Climatic Change 111(3-4):923-944. doi:10.1007/s10584-011-0263-x

Grafton RQ, Pittock J, Williams J, Jiang Q, Possingham H, Quiggin J (2014) Water planning and hydro-climatic change in the Murray-Darling Basin. Australia Ambio 43(8):1082-1092. doi:10.1007/s13280-014-0495-x

Harris I, Jones P, Osborn T, Lister D (2013) Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. Int J Climatol 34(3):623-642. doi:10.1002/joc.3711

Kaspar F (2004) Entwicklung und Unsicherheitsanalyse eines globalen hydrologischen Modells (in German). Dissertation, University of Kassel

Khaki M, Hoteit I, Kuhn M, Awange JL, Forootan E, van Dijk AIJM, Schumacher M, Pattiaratchi C (2017a) 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, Schumacher M, Forootan E, Kuhn M, Awange JL, van Dijk AIJM (2017b) 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

Kusche J (2007) Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models. J Geodesy 81:733-749. doi: 10.1007/s00190-007-0143-3

Kusche J, Schmidt R, Petrovic S, Rietbroek R (2009) Decorrelated GRACE time-variable gravity solutions by GFZ, and their validation using a hydrological model. J Geodesy 83(10):903-913. doi:10.1007/s00190-009-0308-3

Lamontagne S, Taylor A.R, Cook PG, Crosbie RS, Brownbill R, Williams RM, Brunner P (2014) Field assessment of surface water-groundwater connectivity in a semi-arid river basin (Murray-Darling, Australia). Hydrol Process 28:1561-1572. doi: 10.1002/hyp.9691

Leblanc MJ, Tregoning P, Ramillien G, Tweed SO, Fakes A (2009) Basin-scale, integrated observations of the early 21st century multiyear drought in southeast Australia. Water Resour Res 45(4). doi:10.1029/2008WR007333

Leblanc MJ, Tweed SO, van Dijk AIJM, Timbal B (2012) A review of historic and future hydrological changes in the Murray-Darling Basin. Global Planet Change 80-81:226-246. doi:10.1016/j.gloplacha.2011.10.012

Liu Y, van Dijk AIJM, de Jeu RAM, Holmes TRH (2009) An analysis of spatiotemporal variations of soil and vegetation moisture from a 29-year satellite-derived data set over mainland Australia. Water Resour Res 45:W07405. doi:10.1029/2008WR007187.

Müller Schmied H, Eisner S, Franz D, Wattenbach M, Portmann FT, Flörke M, Döll P (2014) Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. Hydrol Earth Syst Sci 18:3511-3538. doi:10.5194/hess-18-3511-2014

Müller Schmied H, Adam L, Eisner S, Fink G, Flrke M, Kim H, Oki T, Portmann F T, Reinecke R, Riedel C, Song Q, Zhang J, Döll P (2016) Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use. Hydrol Earth Syst Sci Discussion. doi:10.5194/hess-2015-527

Rodell M, Houser P, Jambor U, Gottschalck J, Mitchell K, Meng CJ, Arsenault K, Cosgrove B, Radakovich J, Bosilovich M, Entin JK, Walker JP, Lohmann D, Toll D (2004) The global land data assimilation system. B Am Meteorol Soc 85(3):381-394. doi:10.1175/BAMS-85-3-381.

Schneider U, Becker A, Finger P, Meyer-Christoffer A, Ziese M, Rudolf B (2014) GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. Theor Appl Climatol 115:15-40. doi:10.1007/s00704-013-0860-x

Schumacher M, Eicker A, Kusche J, Müller Schmied H, Döll P (2015) Covariance analysis and sensitivity studies for GRACE assimilation into WGHM. In: Rizos C., Willis P. (eds) IAG 150 Years. International Association of Geodesy Symposia 143:241-247. Springer, Cham. doi:10.1007/1345_2015_119

Schumacher M, Kusche J, Döll P (2016) A systematic impact assessment of GRACE error correlation on data assimilation in hydrological models. J Geodesy. doi:10.1007/s00190-016-0892-y

South Australian Murray-Darling Basin Natural Resources Management Board (2008) Potential impacts from extended closure of six temporarily disconnected River Murray wetlands. Government of South Australia Report. http://sabirding.com/refill%20sog%20sites%20report.pdf

Swenson S, Chambers D, Wahr J (2008) Estimating geocenter variations from a combination of GRACE and ocean model output. J Geophys Res 113:B08410. doi:10.1029/2007JB005338

Tapley BD, Bettadpur S, Watkins M, Reigber C (2004) The gravity recovery and climate experiment: Mission overview and early results. Geophys Res Lett 31:L09607. doi:10.1029/2004GL019920

Tian S, Tregoning P, Renzullo LJ, van Dijk AIJM, Walker JP, Pauwels VRN, Allgeyer S (2017) Improved water balance component estimates through joint assimilation of GRACE water storage and SMOS soil moisture retrievals. Water Resour Res, 53, 1820-1840. doi:10.1002/2016WR019641.

Tregoning P, McClusky S, van Dijk AIJM, Crosbie RS, Peña-Arancibia JL (2012) Assessment of GRACE satellites for groundwater estimation in Australia. National Water Commission, Waterlines report No 71, Canberra. doi:http://archive.nwc.gov.au/library/waterlines/71

van Dijk AIJM, Renzullo LJ (2011) Water resource monitoring systems and the role of satellite observations. Hydrol Earth Syst Sci 15:3955. doi:10.5194/hess-15-39-2011

van Dijk AIJM, Renzullo LJ, Rodell M (2011) Use of gravity recovery and climate experiment terrestrial water storage retrievals to evaluate model estimates by the Australian water resources assessment system. Water Resour Res 47:W11524. doi:10.1029/2011WR010714

van Dijk AIJM, Beck HE, Crosbie RS, de Jeu RAM, Liu YY, Podger GM, Timbal B, Viney NR (2013) The Millennium Drought in southeast Australia (2001-2009): Natural and human causes and implications for water resources, ecosystems, economy, and society. Water Resour Res 49(2):1040-1057. doi:10.1002/wrcr.20123

van Dijk AIJM, Renzullo LJ, Wada Y, Tregoning P (2014) A global water cycle reanalysis (2003-2012) merging satellite gravimetry and altimetry observations with a hydrological multi-model ensemble. Hydrol Earth Syst Sci 18:2955-2973. doi:10.5194/hess-18-2955-2014

Vaze J, Viney N, Stenson M, Renzullo L, Van Dijk AIJM, Dutta D, Crosbie R, Lerat J, Penton D, Vleeshouwer J, Peeters L, Teng J, Kim S, Hughes J, Dawes W, Zhang Y, Leighton B, Perraud J-M, Joehnk K, Yang A, Wang B, Frost A, Elmahdi A, Smith A, Daamen C (2013) The Australian water resource assessment modelling system (AWRA). 20th International Congress on Modelling and Simulation, Adelaide, Australia, 1-6 December 2013, www.mssanz.org.au/modsim2013 http://www.mssanz.org.au/modsim2013/L17/vaze.pdf

Ummenhofer CC, England MH, McIntosh PC, Meyers GA, Pook MJ, Risbey JS, Sen Gupta A, Taschetto AS (2009) What causes southeast Australia's worst droughts? Geophys Res Lett 36:L04706. doi:10.1029/2008GL036801

Verdon-Kidd DC, Kiem AS (2009) Nature and causes of protracted droughts in southeast Australia: Comparison between the Federation, WWII, and Big Dry droughts. Geophys Res Lett 36(22):L22707. doi: 10.1029/2009GL041067

Viney N, Vaze J, Crosbie R, Wang B, Dawes W, Frost A (2015) AWRA-L v5.0: technical description of model algorithms and inputs. CSIRO, Australia.

Wahr JM, Molenaar M, Bryan F (1998) Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE. J Geophys Res 108(B12):30205-30229. doi:10.1029/98JB02844

Zaitchik BF, Rodell M, Reichle RH (2008) Assimilation of GRACE Terrestrial Water Storage Data into a Land Surface Model: Results for the Mississippi River Basin. J Hydrometeorol 9(3):535-548. doi:10.1175/2007JHM951.1

List of Figure Captions

1 The Murray-Darling Basin (MDB) and its four sub-basins considered here to integrate GRACE TWSC with the WGHM model simulations. 9

2 Set-up of study for the Murray-Darling Basin (MDB). First, open loop (OL) model runs are performed over 2003-2010 (left column). Subsequently, GRACE TWSC averaged over the 4 major sub-basins of the MDB are assimilated into WGHM testing different configurations (center and right column) and simultaneously the WGHM's parameters are calibrated (see Tab 1). To assess the C/DA results, simulated TWSC and groundwater changes are compared to GRACE TWSC and independent groundwater well measurements. 13

3 (A) Divergence of annual precipitation in mm (from the long-term temporal mean of 477 mm) averaged over the entire Murray-Darling Basin (MDB). (B) Difference in mean annual precipitation during 2001-2009 and 1981-2013 on a $0.5° \times 0.5°$ grid. 16

4 TWSC (in mm) derived from the WGHM open loop (OL) run and from GRACE averaged over the entire Murray-Darling Basin (MDB). The black line shows the WGHM OL, the blue line indicates GRACE (using ITSG-Grace2014), which is smoothed by the DKK3 filter, while the dark gray area represents the range of all investigated GRACE datasets (see Tab. 2). 17

5 Monthly TWSC in mm averaged (A) over the entire MDB, (B) over NW, (C) over NE, (D) over SE, and (E) over SW. The blue line indicates the TWSC from GRACE (ITSG, DDK3); the black line indicates the WGHM OL simulation; the red line indicates the WGHM simulation after C/DA of GRACE (ITSG, DDK3), and the yellow line the WGHM simulation after C/DA (v2) of GRACE (ITSG, DDK3). The dark gray area represents the range of all C/DA results (see Tab. 2 for C/DA configurations). 20

6 Gridded correlation coefficients between WGHM TWSC simulation and ITSG-Grace2014 TWSC after applying DDK3 filtering for both; (A) for the OL, (B) after applying the C/DA. Gridded root mean square error (RMSE) in mm estimated (C) from the differences between OL TWSC and those of GRACE, and (D) from the C/DA TWSC and GRACE TWSC. 21

7 An overview of statistically significant linear trend in TWSC (in mm/year) within the MDB during 2003-2009. The results in (A), (B), and (C) are respectively derived after applying the DDK3 filter to the WGHM OL runs, improved WGHM after C/DA, and from ITSG-Grace2014. In (D) and (E), the linear trend from the original OL TWSC simulations of WGHM and after applying C/DA without any spatial filtering are shown, respectively. . . . 23

8 Annual amplitude of TWSC (in mm) from WGHM and GRACE. The DDK3-filtered results are shown in (A) using WGHM OL, (B) using improved WGHM after C/DA, and (C) using ITSG-Grace2014. In (D) and (E), the annual amplitudes from the original OL TWSC simulations of WGHM and after applying C/DA without any spatial filtering are shown, respectively. 24

9 Monthly time series of groundwater changes (in mm) averaged (A) over the entire MDB, (B) over NW, (C) over NE, (D) over SE, and (E) over SW. The blue line indicates the groundwater observations; the black line indicates the WGHM OL simulation; the red line indicates the WGHM simulation after C/DA of GRACE (ITSG, DDK3), and the yellow line the WGHM simulation after C/DA (v2) of GRACE (ITSG, DDK3). The gray area represents the range of all C/DA results (see Tab. 2 for C/DA configurations). 26

10 Correlation coefficients between wells data and: (A) the OL groundwater simulations, (B) the C/DA simulations (case ITSG-DDK3 while calibrating all 22 model parameters), and (C) the C/DA (v2) simulations (calibrating only 3 parameters). 27

11 Significant linear trend in groundwater changes (in mm/year) within the MDB during 2003-2009. The results in (A), (C), (E) and (G) are respectively derived from the groundwater measurements, the WGHM OL, WGHM after C/DA while calibrating all 22 parameters, and from WGHM after C/DA (v2) while calibrating only 3 parameters. Results are spatially averaged over $1^\circ \times 1^\circ$ grid cells. In (B), (D), and (F), the linear trend from the original OL groundwater simulations of WGHM and after applying C/DA and C/DA (v2) are shown, respectively. 29

12 Root mean square (RMS) of monthly update increments after applying the C/DA to integrate WGHM with TWSC from ITSG-DDK3 (calibrating all 22 parameters) for (A) TWSC, (B) soil water, (C) groundwater, (D) lakes, (E) wetlands, and (F) rivers. In (F), the locations of the river discharge stations that have been used to calibrate the WaterGAP 2.2 model version are shown by the black dots. 31

13 Time series of river discharge (in m/s^3) at three selected river discharge stations: (A, B) Paroo River at Caiwarro (BoM station number 424201A; number 1 in Fig. 12 (F)), (C, D) Darling River at Burtundy (BoM station number 425007; number 4 in Fig. 12 F); and (E, F) Lachlan River at Booligal (BoM station number 412005, number 7 in Fig. 12 (F)). The left column presents C/DA results from the ITSG-DDK3 case for which all 22 parameters have been calibrated, and the right column presents the C/DA (v2) while calibrating only 3 parameters. 33

14 Time series of river level at the station 4 in Fig. 12 (F). The time series are temporally normalized, thus, they are unit-less. 34

15 First dominant orthogonal mode, including EOF and its corresponding PC, derived from soil moisture outputs of WGHM. Here EOF1 is derived from the open loop run, but PC1 is derived by applying PCA on the open loop, and two versions of the C/DA outputs and compared to the ENSO index (SOI). This dominant mode represents 62% of variance in soil moisture variability in the region. 40

16 First dominant orthogonal mode, including EOF1 and its corresponding temporal pattern PC1, derived from groundwater outputs of WGHM is shown. Here EOF1 is derived from the open loop run, but PC1 is derived by applying PCA on the open loop, and two versions of the C/DA outputs and compared to the ENSO index (SOI). This dominant mode (EOF1 and PC1 together) represents 78% of variance in groundwater variability in the region. 40