Water Resources Research

RESEARCH ARTICLE
10.1029/2018WR024670

Special Section:
Advancing process representation in hydrologic models: Integrating new concepts, knowledge, and data

On the Use of Adaptive Ensemble Kalman Filtering to Mitigate Error Misspecifications in GRACE Data Assimilation

Ashkan Shokri[1], Jeffrey P. Walker[1], Albert I. J. M. van Dijk[2], and Valentijn R. N. Pauwels[1]

[1]Department of Civil Engineering, Monash University, Clayton, Victoria, Australia, [2]Fenner School of Environment and Society, Australian National University, Canberra, ACT, Australia

Key Points:
- There is a strong potential to improve GRACE data assimilation systems through optimization of the error estimates
- An adaptive EnKF can detect and correct the model error misspecifications in a GRACE data assimilation system

Correspondence to:
A. Shokri,
ashkan.shokri@monash.edu

Citation:
Shokri, A., Walker, J. P., van Dijk, A. I. J. M., & Pauwels, V. R. N. (2019). On the use of adaptive ensemble Kalman filtering to mitigate error misspecifications in GRACE data assimilation. *Water Resources Research*, *55*. https://doi.org/10.1029/2018WR024670

Received 27 DEC 2018
Accepted 19 AUG 2019
Accepted article online 21 AUG 2019

Abstract The ensemble Kalman filter (EnKF) has been proved as a useful algorithm to merge coarse-resolution Gravity Recovery and Climate Experiment (GRACE) data with hydrologic model results. However, in order for the EnKF to perform optimally, a correct forecast error covariance is needed. The EnKF estimates this error covariance through an ensemble of model simulations with perturbed forcing data. Consequently, a correct specification of perturbation magnitude is essential for the EnKF to work optimally. To this end, an adaptive EnKF (AEnKF), a variant of the EnKF with an additional component that dynamically detects and corrects error misspecifications during the filtering process, has been applied. Due to the low spatial and temporal resolutions of GRACE data, the efficiency of this method could be different than for other hydrologic applications. Therefore, instead of spatially or temporally averaging the internal diagnostic (normalized innovations) to detect the misspecifications, spatiotemporal averaging was used. First, sensitivity of the estimation accuracy to the degree of error in forcing perturbations was investigated. Second, efficiency of the AEnKF for GRACE assimilation was explored using two synthetic and one real data experiment. Results show that there is considerable benefit in using this method to estimate the forcing error magnitude and that the AEnKF can efficiently estimate this magnitude.

1. Introduction

It has been proven that the ensemble Kalman filter (EnKF) can improve the accuracy of hydrologic models by merging observations with model predictions. Such observations include soil moisture (e.g., Aubert et al., 2003; Crow & Ryu, 2009; Houser et al., 1998; Pauwels et al., 2001; Reichle et al., 2004, 2008; Walker & Houser, 2001), snow water equivalent (e.g., Barrett, 2003; Slater & Clark, 2006; Sun et al., 2004), streamflow (e.g., Lee et al., 2011; Clark et al., 2008), groundwater levels (e.g., Hendricks Franssen et al., 2017), turbulent heat fluxes (e.g., Bateni & Entekhabi, 2012; Pipunic et al., 2013; Xu et al., 2018) , microwave radiances (e.g., Dechant & Moradkhani, 2011), and terrestrial water storage (TWS; e.g., Ellett et al., 2006; Forman & Reichle, 2013; Forman et al., 2012; Girotto et al., 2016; 2017; Houborg et al., 2012; Khaki, Ait-El-Fquih, et al., 2017; Khaki, Hoteit, et al.,2017; Khaki, Schumacher, et al., 2017; Kumar et al., 2016; Li et al., 2012; Li & Rodell, 2015; Smith, 2013; Tian et al., 2017; van Dijk et al., 2014; Zaitchik et al., 2008). The idea behind the EnKF is to combine observations and model estimates of state variables considering their relative error covariances.

Although the effect of accurately determining the observation and model error specifications is significant, it is difficult in real-world problems to accurately determine their correct values. One possibility is to find the optimal error specification off-line using a calibration approach (Reichle et al., 2008; Reichle et al., 2002a), requiring a considerable number of computationally expensive assimilation runs. Moreover, the error characteristics may vary during the simulation, and thus identifying the spatiotemporal variations of error characteristics could be challenging.

Adaptive filtering provides a tool to analyze the EnKF internal metrics, such as innovations and increments, to understand if there is any misspecification in error characteristics, for example, identifying a correction to improve the accuracy of the error specification for the next assimilation time step. Different variations of these approaches have been developed in previous studies (e.g., Berry & Sauer, 2013; Hong et al., 2009; Mehra, 1970; Moghaddamjoo & Kirlin, 1993). These tools have also been successfully applied to a number of land data assimilation applications (e.g., Crow & Van Loon, 2006; Daley, 1992; De Lannoy et al., 2009;

Reichle et al., 2008; Van Geer et al., 1991; Zhang et al., 2004). However, such approaches have not yet been applied to a Gravity Recovery And Climate Experiment (GRACE) data assimilation framework.

The idea behind the Adaptive EnKF (AEnKF) is that the expected value of a number of metrics, calculated based on internal diagnostics (e.g., innovations), should be consistent with the values that are expected from the assumed error specification. Any difference indicates a misspecification in error characteristics and therefore provides an opportunity to suggest corrections. Since the internal diagnostics are very noisy, different approaches have been adopted to obtain a usable expected value depending on circumstances. Some of these assume that the error characteristics are spatially homogeneous. Therefore, using a spatial averaging approach, a single expected value of the metrics can be calculated. Other studies assume that the error characteristics are temporally stable, and thus they have averaged the metrics across time.

The methodology of the current paper is adopted from Crow and Van Loon (2006). Their study addressed a synthetic soil moisture data assimilation problem with high temporal and spatial resolution. They showed that the methodology performs well in such an information-rich situation, and after data assimilation there was enough information in the internal diagnostics to correct the assumed error magnitudes. One of the challenges they identified was choosing an approach to filter the noise in the adaptive measure (normalized innovations) to provide an appropriate modification coefficient. Crow and Van Loon (2006) used a temporal averaging approach. However, GRACE TWS assimilation has to address the situation, with far less information in the observations, due to their coarse temporal and spatial resolution, and vertical integration. Moreover, GRACE retrievals represent TWS anomalies from a temporal average rather than absolute values. It is not clear that the methodology developed by Crow and Van Loon (2006) can correctly extract the assumed error magnitudes under these conditions. Our objective was to determine if their methodology can improve estimation skill of the model system. The averaging approach of the original methodology was altered as well, by aggregating the information across the domain and over time. Moreover, the method here was tested in both synthetic and real-world experiments. The objective of this paper was therefore to assess if there is any possibility to remove the effect of error misspecification in GRACE assimilation using this approach.

2. Study Area, Model, and Data Set

2.1. The Study Area

The Murray-Darling basin (MDB) has been selected as the study domain. This basin, located in southeastern Australia, covers approximately 1×10^6 km^2 including parts of New South Wales, Victoria, the Australian Capital Territory, South Australia, and Queensland (Figure 1). The basin contains 23 major rivers providing water for approximately 4 million people. This catchment receives 6% of total Australian precipitation producing 24×10^9 m^3 of runoff per year on average (Pigram, 2007).

2.2. The Model

Following Tian et al. (2017) and Shokri, Walker, van Dijk, and Pauwels (2018), the hydrological model used here was the AWRA-L version 0.5 (van Dijk, 2010). This model simulates a variety of state and flux variables on a daily basis. The variables include groundwater, streamflow, evapotranspiration, and soil moisture content in three different layers; specifically, in the surface (0–10 cm), shallow (10–100 cm), and deep (100–500 cm) layers. AWRA-L is a parsimonious model that has been applied successfully in Australia.

This model is a grid-based model (with a grid size of approximately 0.05°). Each grid consists of two hydrological response units (HRU), representing the deep and shallow-rooted vegetation landscape parts, respectively. Each HRU contains three soil layers, and the groundwater and runoff storage is shared between each pixel. The soil water balance and heat flux are simulated separately for each HRU, and the groundwater is simulated for each pixel. In this model, shallow-rooted vegetation has access to just shallow soil moisture and deep-rooted vegetation has access to both shallow and deep soil moisture. The groundwater storage is also simulated as a linear storage without any lateral flow between pixels. Surface and groundwater storage, soil moisture of different layers, evapotranspiration, and the flux between different storages are the model outputs (for more details refer to van Dijk et al., 2012, 2011; Renzullo et al., 2014; Shokri, Walker, van Dijk & Pauwels ; Shokri et al., 2018; Van Dijk & Warren, 2010).

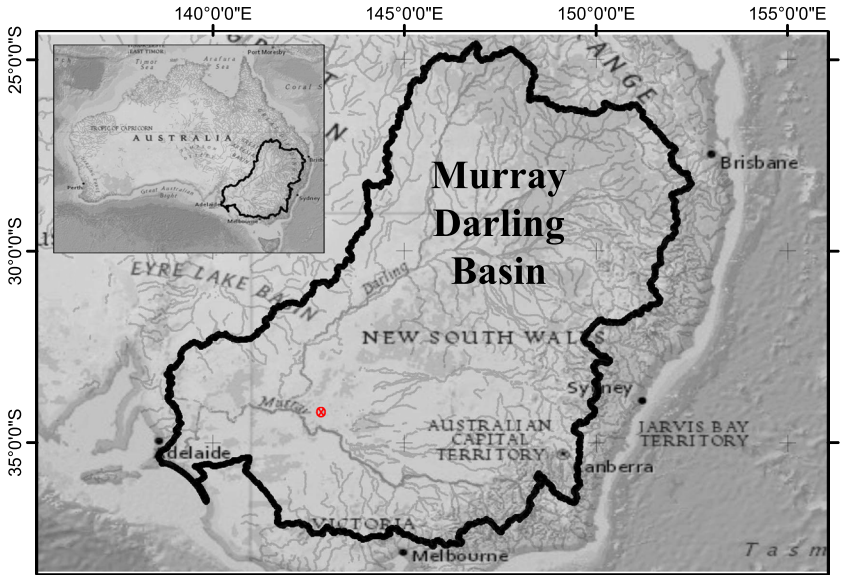

Figure 1. The Murray-Darling basin study area and its location in southeast Australia. The circle shows the location of the example time series of Figure 7.

2.3. The Data

The resolution of the model is 0.05°. The model was forced with meteorological data from the Australian Water Availability Project (AWAP; http://www.bom.gov.au/jsp/awap/index.jsp), including daily precipitation, minimum and maximum temperature, and radiation. The remaining input data, including wind and land use, are identical with those of Shokri, Walker, van Dijk, and Pauwels (2018). The model was calibrated using the patient rule induction method parameter estimation (Shokri, Walker, van Dijk, Wright & Pauwels, 2018) using streamflow observations from 190 unregulated catchments within the MDB (Zhang et al., 2013) from 1990 to 2000. Similar to Shokri, Walker, van Dijk, and Pauwels (2018), two independent "good enough" parameter sets were found. One of them was used to generate the truth data (in the synthetic experiments), while the other was used for the simulation (for more details refer to Shokri, Walker, van Dijk, Wright & Pauwels 2018; Shokri, Walker, van Dijk & Pauwels 2018).

2.4. The Data Assimilation Experiments

A variety of different synthetic observations were assimilated into the AWRA-L model. To generate synthetic observations, a set of simulations were performed with perturbed forcing data, and a good enough parameter set to provide what was then considered as the synthetic "truth." We refer to this parameter set as the "true" parameter set. The synthetic TWS observations were generated by spatiotemporally upscaling and adding synthetic observation noise to the TWS estimates from a model with the true parameter set. The target synthetic observations match with the GRACE pixel sizes (3×3 degree) and temporal resolution (monthly). To validate and compare the results of different approaches, the root-mean-square error (RMSE), correlation coefficient (R^2), and exceedance ratio (ER95; Moradkhani et al., 2006) were used. The ER95 is the fraction of the observations which fall outside the estimated 95% confidence bounds. For a perfect ensemble and noiseless observations, the expected value of ER95 is 5%. Therefore, in this study, the ER95 was calculated using noiseless observations (the aggregated TWS before the perturbation). The main reason for using synthetic data was that it is impossible to determine accurate in situ TWS values for an area similar in size to the very large GRACE footprints. However, the different filtering approaches were also evaluated using real observations including observed groundwater levels, streamflow, and soil moisture values.

The GRACE observations used here were the NASA Jet Propulsion Laboratory's (JPL) RL05M (Watkins et al., 2015; Wiese et al., 2016 https://podaac.jpl.nasa.gov/dataset/TELLUS_GRACE_MASCON_CRI_GRID_RL05_V2). This product is a monthly $3° \times 3°$ TWS anomaly (TWSA) relative to a long-term TWS average of 2004–2009. So the TWSA were converted to TWS by adding the average TWS from a model simulation over the same period. The observation error is also provided with a monthly interval.

The streamflow observations were from the same gauges used in the calibration process but for a different period (from 2000 to 2018). The groundwater observations were level measurements at 14,903 bores across the MDB, selected from 260,186 monitoring points (http://www.bom.gov.au/water/groundwater/explorer/) based on their data availability (bores with less than 24 observations were ignored) during the simulation period. The soil moisture observations were obtained from OZnet (http://www.oznet.org.au/ A. B. Smith et al., 2012). OZnet measures the soil moisture for a 0- to 90-cm profile so the surface and root zone soil moisture can be validated using these observations. The model provides the water states in millimeters of extractable water, while the observations of the groundwater are the standing water level, and the soil moisture observations are volumetric wetness in $\frac{m^3}{m^3}$. Therefore, to evaluate the accuracy of different in real-world experiments, the correlation coefficient (R) was used. Eight state variables per pixel per day were updated in the data assimilation experiments, which were the surface (s0), shallow (ss), and deep (sd) soil moisture for both HRUs of each pixel, and the surface (sr) and groundwater storages (sg).

3. EnKF

The EnKF (Evensen, 2003) is a Monte Carlo variant of the Kalman Filter (KF) which was developed to deal with nonlinear assimilation problems while bypassing the computational challenge of the extended Kalman filter. The computational feasibility and simplicity of the EnKF make it one of the more popular Data Assimilation (DA) approaches. In the EnKF, a relatively small ensemble of model integrations is used to represent the error of the (potentially nonlinear) forward model. These ensembles are formed by adding randomly generated errors to the model inputs and model error covariance estimated from the spread of the resulting state variables. These matrices are then used to determine the model correction. The error characteristics have significant effects on the model error estimation, and consequently the amount of update is highly sensitive to them. Therefore, it is crucial to accurately specify the error characteristics to have an optimal filter.

An EnKF consists of two main steps which are performed sequentially: (1) a forward model which forecasts an ensemble of the state variables by using an ensemble of perturbed forcing data and (2) an update step in which the error covariances inferred from the first step and the estimates of observation errors are used to modify the forecasted state variables. For GRACE assimilation, different variants of the EnKF are developed. Shokri, Walker, van Dijk, and Pauwels (2018) compared different variants of the EnKF for a synthetic GRACE TWS data assimilation and concluded that an EnKF structure which distributes the observation innovation into all state variables during the month had the most reliable performance in terms of reducing the error. The adaptive error tuning strategy was combined with this approach: After finding the predicted observation of each member, all state variables inside each GRACE pixel and during the days of the month were used to calculate the covariance matrix. Therefore, the Kalman gain will have a length equal to the number of model pixels inside a GRACE pixel times the number of time steps (days) of a month. The forward model (first step) can be written as

$$\mathbf{x}_{\tau}^{i-} = F(\mathbf{x}_{\tau-1}^{i+}, \mathbf{q}_{\tau}^{i}), \tag{1}$$

where $\mathbf{x}_{\tau}^{i-}$ is the forecast state vector of member i at time τ derived by a nonlinear forecast model $F(.)$. The model propagates the updated state vector from the previous time step, $\mathbf{x}_{\tau-1}^{i+}$, and imposes an ensemble of model errors $\mathbf{q}_{\tau}^{i}$, with covariance of $\mathbf{Q}_{\tau}$. Once an observation becomes available (one a month in the case of the GRACE retrievals) the observation operator, $\mathbf{M}(.)$, predicts the observation based on the forecast state vectors, $\mathbf{x}_{\tau}^{i-}$. Such that

$$\mathbf{M}(\mathbf{x}_{\tau}^{i}) = \frac{1}{N_{\text{pixel}}} \sum_{k=1}^{N_{\text{pixel}}} \left(\frac{1}{3} \sum_{d \in \{5,15,25\}} TWS_{\tau,d,k}^{i} \right), \tag{2}$$

where N_{pixel} is the number of model pixels in the observation grid, $TWS_{T,d,k}^{i}$ is the TWS of the kth model pixel at the dth day of the Tth monthly time window of the ith member. The reason that the TWS is averaged for 3 days (5, 15, 25) is to mimic the three overpasses of the GRACE satellites used to measure the TWS change in each month. The increments to update the forecast state vector can then be calculated as

$$\Delta\mathbf{x}_{\tau}^{i} = \mathbf{K}_{\tau}[\mathbf{y}_{\tau}^{i} - \mathbf{M}(\mathbf{x}_{\tau}^{i})], \tag{3}$$

where $\mathbf{y}_{\tau}^{i}$ is a vector that contains the perturbed observations, and $\mathbf{K}_{\tau}$ is the Kalman gain at τ observation time step. The Kalman gain, $\mathbf{K}_{\tau}$, converts the innovations $(\mathbf{y}_{\tau}^{i} - \mathbf{M}(\mathbf{x}_{\tau}^{i}))$ into increments $\Delta\mathbf{x}_{\tau}^{i}$ that are then

used to correct the state vector forecast. The relative weight of the observation and model are determined during the update step by adjusting the magnitude of increments based on the error covariance between state variables and model observation predictions ($\mathbf{C}_{\mathbf{xM},\tau}$), the error covariance of the model observation prediction ($\mathbf{C}_{\mathbf{MM},\tau}$) and the observation error covariance ($\mathbf{C}_{\mathbf{yy},\tau}$). The Kalman gain is calculated as

$$\mathbf{K}_\tau = \mathbf{C}_{\mathbf{xM},\tau}.(\mathbf{C}_{\mathbf{MM},\tau} + \mathbf{C}_{\mathbf{yy},\tau})^{-1}, \tag{4}$$

where $\mathbf{C}_{\mathbf{xM}}$ and $\mathbf{C}_{\mathbf{MM}}$ are calculated based on the ensemble simulation results and $\mathbf{C}_{\mathbf{yy}}$ reflects the uncertainty (or error) of the observations (Reichle et al., 2002b).

4. Adaptive EnKF

The optimality of the Kalman gain is subject to assumptions, including that the error characteristics are accurately quantified. It is difficult to satisfy this condition in a hydrological model data assimilation framework with limited observations. Adaptive EnKF strategies can help to mitigate the effect of the potential error misspecification on the performance of the filter.

Adaptive filtering strategy aim to detect the misspecification in error characteristics and modify them using internal diagnostics. Usually, the innovations $[\mathbf{y}_\tau^i - \mathbf{M}(\mathbf{x}_\tau^i)]$ are used for correction. Crow and Van Loon (2006) modified the adaptive EnKF approach from Dee (1995) for land data assimilation systems. This method is adapted to a GRACE retrieval data assimilation system. This approach is based on the mean of the innovations across the ensemble, $v_{\tau,m}$, which can be calculated as

$$\boldsymbol{v}_{\tau,m} = \frac{1}{n_r}\sum_{r=1}^{n_r} v_{\tau,r,m}, \tag{5}$$

where n_r is the size of the ensemble and $v_{\tau,r,m}$ is the innovation of ensemble member r at time step τ for the mth observation grid. In an optimal filter,

$$\mathrm{E}[\boldsymbol{v}_{\tau,m}\boldsymbol{v}_{\tau,m}^T] = \mathbf{C}_{\mathbf{MM},\tau} + \mathbf{C}_{\mathbf{yy},\tau}, \tag{6}$$

should be satisfied, where $\boldsymbol{v}_{\tau,m}^T$ is the transpose of $\boldsymbol{v}_{\tau,m}$. Based on this equation, Dee (1995) defined the normalized innovation mean, $\chi_{\tau,m}$, as a measure to assess the error characteristics correctness by rearranging equation (6) as

$$\chi_{\tau,m} = \frac{\mathrm{E}[\boldsymbol{v}_{\tau,m}\boldsymbol{v}_{\tau,m}^T]}{\mathbf{C}_{\mathbf{MM}} + \mathbf{C}_{\mathbf{yy}}}. \tag{7}$$

In an EnKF which is running optimally, the expected value of $\chi_{\tau,m}$ is one and its temporal autocorrelation is 0. If the expected value of $\chi_{\tau,m}$ is more than unity, it can be inferred that the model error magnitudes are underestimated relative to observation errors, and if it is less than unity, the error magnitudes are overestimated. There are two challenges in using $\chi_{\tau,m}$ to modify the error magnitudes. First, $\chi_{\tau,m}$ is noisy and finding its expected value needs numerous realizations. To address this challenge, $\mathbf{Q}$ and consequently $\boldsymbol{v}_{\tau,m}$ are usually assumed to be temporally or spatially fixed, so that the expected value can be calculated by averaging $(\boldsymbol{v}_{\tau,m}\boldsymbol{v}_{\tau,m}^T)$ over the domain or period. The temporal and spatial resolution of the GRACE retrievals are coarse, so the results of either spatial or temporal averaging is expected to be noisy. Therefore, the $(\boldsymbol{v}_{\tau,m}\boldsymbol{v}_{\tau,m}^T)$ values were first averaged spatially across the catchment. Next, the averaged values smoothed using a temporally exponential moving average. To find the $\chi_{\tau,m}$, they were then divided by the similarly averaged $(\mathbf{C}_{\mathbf{MM}} + \mathbf{C}_{\mathbf{yy}})$ over the same domain and time period.

The second challenge is, although $\chi_{\tau,m}$ shows the direction of the misspecification appropriately, it is difficult to specify the exact amount of required modification in a nonlinear system. To overcome this challenge, a dynamic modification process has been applied in previous studies. Here, instead of trying to correct all errors in one attempt, the error magnitudes were fixed by a small amount of modification toward the direction suggested by $\chi_{\tau,m}$ at each time step. We used this approach for applying the temporal exponential moving average which can be formulated as

$$\mathbf{Q}_\tau = \alpha_\tau \times \mathbf{Q}_0, \tag{8}$$

where $\mathbf{Q}_0$ is the assumed initially model error covariance and α_τ is a scale coefficient calculated dynamically for each time step as

$$\alpha_\tau = \alpha_{\tau-1} \times f_\tau, \tag{9}$$

where α_τ evolves during the time by being multiplied with f_τ which is calculated as

$$f_\tau = \begin{cases} (1+\delta)^{-1} & , \quad \overline{\chi_{\tau,\cdot}} < (1+\delta)^{-1} \\ \overline{\chi_{\tau,\cdot}} & , \ (1+\delta)^{-1} \leqslant \overline{\chi_{\tau,\cdot}} \leqslant (1+\delta) \\ (1+\delta) & , \quad (1+\delta) < \overline{\chi_{\tau,\cdot}} \end{cases}, \tag{10}$$

where f_τ considered to be equal to the averaged normalized innovation mean, $\overline{\chi_{\tau,\cdot}}$, which is constrained by a narrow range. This range is defined with δ which should be a very small number. The dot shows that the variable is averaged over ensemble members. $\overline{\chi_{\tau,\cdot}}$ can be calculated according to

$$\overline{\chi_{\tau,\cdot}} = \beta \frac{MA(\overline{(\boldsymbol{\nu}^2_{\tau,\cdot})})}{MA(C_\tau)} = \beta \frac{\left(\overline{(\boldsymbol{\nu}^2_{\tau,\cdot})}\right)\gamma + \left(\overline{(\boldsymbol{\nu}^2_{\tau-1,\cdot})}\right)(1-\gamma)}{(C_\tau)\gamma + (C_\tau - 1)(1-\gamma)}, \tag{11}$$

$$\mathbf{C}_\tau = \mathbf{C}_{\mathbf{MM},\tau} + \mathbf{C}_{\mathbf{yy},\tau}, \tag{12}$$

where $MA(.)$ is the exponential moving average operator, and γ is the parameter which determines how smooth the moving average is and should be chosen small enough to ensure that noise is removed properly. $\overline{(\boldsymbol{\nu}^2_{\tau,\cdot})}$ is the spatial average of $(\boldsymbol{\nu}_{\tau,m}\boldsymbol{\nu}^T_{\tau,m})$ which is calculated as

$$\boldsymbol{\nu}_{\tau,\cdot}\boldsymbol{\nu}^T_{\tau,\cdot} = \overline{(\boldsymbol{\nu}^2_{\tau,\cdot})} = \frac{1}{n_m}\sum_{m=1}^{n_m} \nu^2_{\tau,m}, \tag{13}$$

where $\nu^2_{\tau,m}$ is the mean of innovations across the ensemble which are calculated using equation (5). The other parts of the AEnKF structure used in the current paper are identical to the EnKF structure explained in the previous section.

5. Results

Two synthetic and one real experiment sets were conducted to explore the feasibility of using adaptive filtering approaches for GRACE TWS assimilation to mitigate potentially misspecified error characteristics. The first and second experiments investigated the effect of error magnitude and accuracy, respectively.

The feasibility of using the AEnKF to mitigate the impact of error misspecification was investigated. Finally, the performance of the AEnKF was evaluated in an experiment with the real GRACE observations. These experiments were conducted on the MDB at a resolution of 0.05° over a 19-year time frame (from 1 January 2000 to 31 December 2018).

An error model similar to Shokri, Walker, van Dijk, and Pauwels (2018) was used here to generate cross-correlated noise fields for the forcing data which are temporally and spatially autocorrelated. Error model parameters are listed in Table 1. Following Shokri, Walker, van Dijk, and Pauwels (2018), an ensemble size of 20 was selected for all DA experiments. To test this number was sufficiently large enough, one of the experiments (EnKF with correct error magnitude and $Q_\alpha = 1$) was tested for different ensemble sizes (between 5 to 50 members). The conclusions did not change when the ensemble members increased from 20 to 50.

5.1. Impact of Error Magnitude

For the first set of experiments, nine open loop (OL) ensembles (with perturbed meteorological data and without assimilation) were generated. Each of these OL ensembles had a specific error magnitude which was controlled by a scale factor α_Q. A set of nominal error characteristics was considered as the center point (Table 1), and by multiplying them with α_Q different spreads of the OL ensemble were obtained. Figure 2 shows three samples of the simulated TWS using $\alpha_Q = \{0.5, 1, 2\}$, averaged across the catchment. For each α_Q, the TWS time-series of a random member was selected as the reference (true) value. The synthetic observations were generated by adding observation white noise to the reference coarse TWS. These observations

Table 1
Perturbation Parameters Used in Generating Spatiotemporal Correlated Meteorological Forcing Ensembles (Adopted From Shokri, Walker, van Dijk & Pauwels, 2018)

Parameter	Type[a]	Standard deviation	Spatial correlation	Temporal correlation	Cross correlation			
					Precipitation	Minimum temperature	Maximum temperature	Radiation
Precipitation	M	0.50	2°	3 days	n/a	−0.8	−0.2	−0.1
Radiation	M	0.30	2°	3 days	−0.8	n/a	0.6	0.5
Minimum temperature	A	0.30 (° C)	2°	3 days	−0.2	0.6	n/a	0.7
Maximum temperature	A	0.25 (° C)	2°	3 days	−0.1	0.5	0.7	n/a

[a]The symbols A and M indicate additive and multiplicative perturbations.

were then used to the EnKF with correct error magnitude specifications. The nine synthetic problems (α_Q = {0.10, 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00}) were generated and solved.

As an example, Figure 3 shows the improvement in the TWS estimate accuracy for the problem with α_Q = 1.00. According to this figure, the mean absolute error (MAE) and root-mean-square error (RMSE) between the true and estimated TWS were reduced considerably when the EnKF with an accurate error magnitude was applied. A similar conclusion was obtained for all other scaling factors. A summary of these results is shown in Figure 4 as using spatiotemporally averaged RMSE for the OL and EnKF. The temporal average lines in Figure 3a provide a single point in Figure 4.

As can be expected, their RMSE increases monotonically by increasing the error magnitude. The RMSE TWS estimates of the OL ensemble with α_Q = 0.1 was 1.51, and by increasing the error magnitude, it reached to 22.75 (for α_Q = 2.0). The rate of the increase was not constant and for a smaller value of α_Q a slightly higher sensitivity of the RMSE was observed. Meanwhile, the EnKF reduces the RMSE by approximately 20% for all tests with different values of α_Q. These results indicate that the skill of an optimal EnKF (with correct error specification) in removing the error for the designed experiments. The tests were repeated with the same conditions except with incorrect error magnitudes introduced to the EnKF, which were doubled. More specifically, for each run in this new experiment set, the observations were generated based on the resulting TWS of a single ensemble member, perturbed with a specific α_Q, but using twice this value in the EnKF. Deliberately feeding wrong information to the EnKF move the filter to a suboptimal operation. Due to the lack of knowledge about specific characteristics of uncertainties associated with different parts of the system, in a real-world problem this suboptimality can easily occur. The results indicate that the RMSE of the EnKF with deliberately wrong error characteristics was considerably degraded and with accuracy of the result becoming even worse than the OL runs (Figure 4). This illustrates the importance of using correct error characteristics in the EnKF applications. The fourth set of experiments were conducted with the same (doubled) error magnitudes, but this time with the AEnKF method applied. The AEnKF reduces the sensitivity of the filtering procedure to the accuracy of the error specification. Results confirmed the capability of the AEnKF in recovering the correct error magnitudes from deliberately misspecified errors. As depicted in Figure 4, the RMSE of the results obtained from the AEnKF, even with doubled error magnitudes,

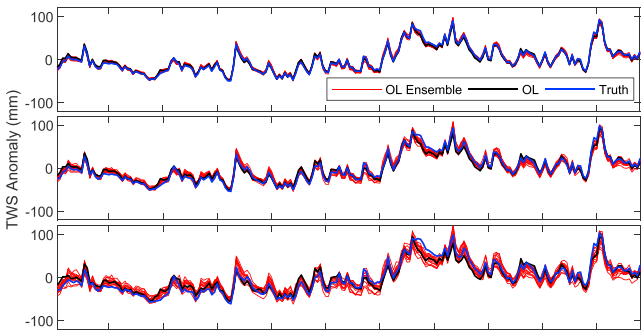

Figure 2. Mean (OL) and spread (OL Ensemble) of the ensemble open loop (with perturbation but without assimilation). Prediction of TWS anomaly for α_Q = {0.5, 1, 2}. Simulation starts from 1 January 2000. OL = open loop; TWS = terrestrial water storage.

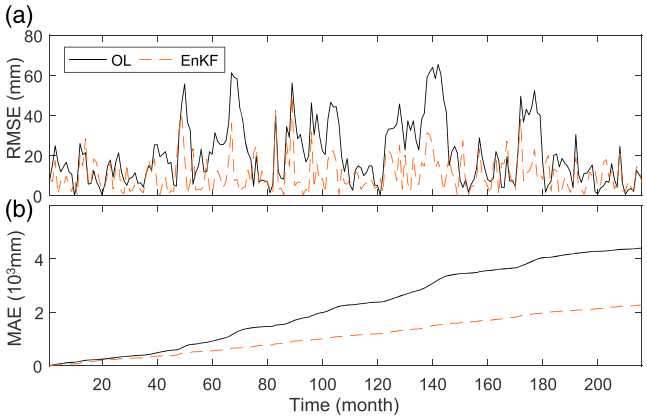

Figure 3. Spatially averaged (a) root-mean-square error (RMSE) and (b) cumulative mean absolute error (MAE) for monthly and coarse scale terrestrial water storage (TWS) estimates with $\alpha_Q = 1$. Simulation starts from from 1 January 2000. OL = open loop; EnKF = ensemble Kalman filter.

were approximately similar to apply the EnKF with correct error specifications. This is also reflected in the spatial RMSE maps (Figure 4c).

5.2. Impact of Error Accuracy

The previous section investigated, the impact of introducing double the optimal error magnitude to the EnKF. However, the impact of a differing levels of inaccuracy was not tested. To do so, a single scaling factor ($\alpha_Q = 1$) was used to generate one set of synthetic observations. Nine different deliberately wrong error magnitudes then introduced to the OL, EnKF, and AEnKF runs. Figure 5a shows the TWS RMSE of the ensemble OL and EnKF with different degrees of misspecification in error magnitudes are compared with the truth. Figure 5 also shows the impact of different β values (equation (11)). As expected the RMSE of both the EnKF and OL were increased when the error magnitudes were over or underestimated. In operational systems, this sensitivity is problematic as these parameters are difficult to measure and will likely be misspecified. The results of the experiments using the AEnKF suggest that this approach can correct the results of using an erroneous value of α_Q. As is illustrated in Figure 5a, the RMSE of the TWS estimation is almost independent of the assigned error magnitude.

A notable feature in Figure 5 is the shift of the minimum RMSE in the regular EnKF toward the left. Ideally, the highest efficiency is expected to correspond to the correct error specification, in this case $\alpha_Q = 1$. However, it is not uncommon in hydrological DA for similar patterns with reported in Crow and Van Loon (2006) and Durand and Margulis (2008). The reason for this shift is not simply due to the selection of random numbers. Durand and Margulis (2008) have prove this by investigating the problem using three different random seed numbers. The results of the current study confirm this claim by reproducing a similar pattern (not shown here) when using several random number seeds. Since the EnKF was originally designed

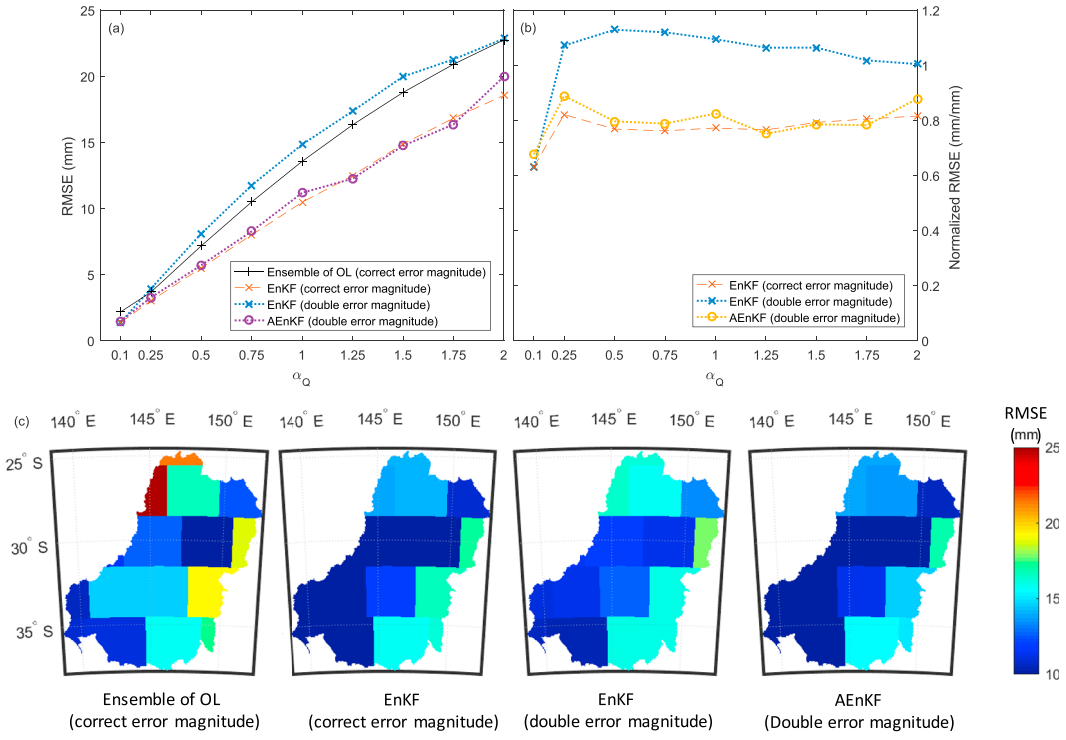

Figure 4. (a) Absolute, (b) normalized spatiotemporally averaged relative to the open loop simulations (for different levels of α_Q), and (c) temporally averaged maps (for α_Q=1.0) of terrestrial water storage RMSE. Plot compares different filtering approaches. EnKF = ensemble Kalman filter; OL = open loop; AEnKF = adaptive EnKF; RMSE = root-mean-square error.

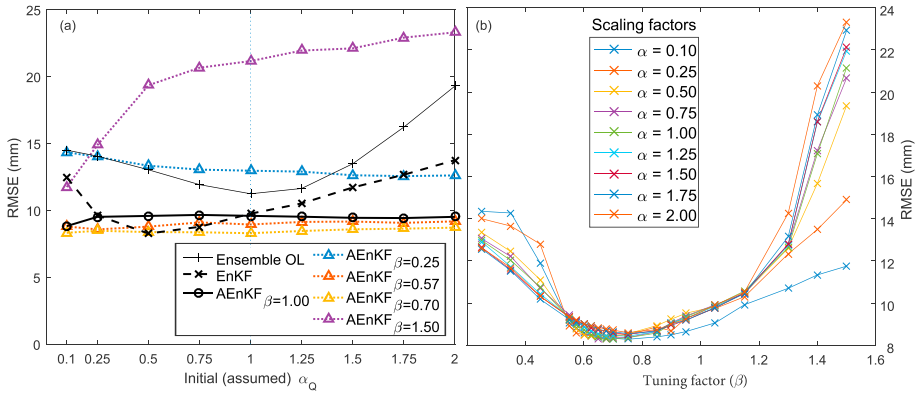

Figure 5. (a) Terrestrial water storage root-mean-square error values of the open loop, ensemble Kalman filter, and adaptive ensemble Kalman filter with deliberately wrong initial α_Q values. (b) Root-mean-square error of the model using different α and β. OL = open loop; EnKF = ensemble Kalman filter; AEnKF = adaptive EnKF; RMSE = root-mean-square error.

for linear systems, the nonlinearity of the hydrological model may be the reason for this shift. The effects of nonlinearity on the effect of the filter have been discussed in several studies (e.g., Evensen, 1994, 1997; Verlaan & Heemink, 2001; Yang et al., 2012).

In the current AEnKF approach a tuning factor (β) was embedded in equation (12) to compensate for nonlinearity. A value less than one intentionally forces the adaptive procedure to underestimate the magnitude of the errors. A range of different β values were tested with the results shown in Figure 5a.

These results indicate that both large and small values of β lead to a suboptimal performance of the AEnKF, while for values closer to the optimal value (between 0.6 and 0.9) the AEnKF performs similar to the optimal EnKF. It should also be noted that the values of α_Q and β can compensate for each other. For a value of 0.1 for α_Q the use of a value of 1.5 for β improved the performance of the AEnKF, while for a value of 2 for α_Q a value of 0.25 for β improved the AEnKF filter performance.

Equation (11) shows that β forces the value of α_Q toward its optimal value. For an initially low α_Q a high value for β assists the adaption procedure achieve a quicker retrieval of its correct value. Similarly, if α_Q is initially too high, a low value for β will again help the adaption procedure to more quickly retrieve its correct value.

These experiments demonstrate the importance of an adequate estimation of β. However, it should be emphasized that without this factor (or $\beta = 1$) the results marginally degraded compared to the optimal EnKF but still the filter is working efficiently.

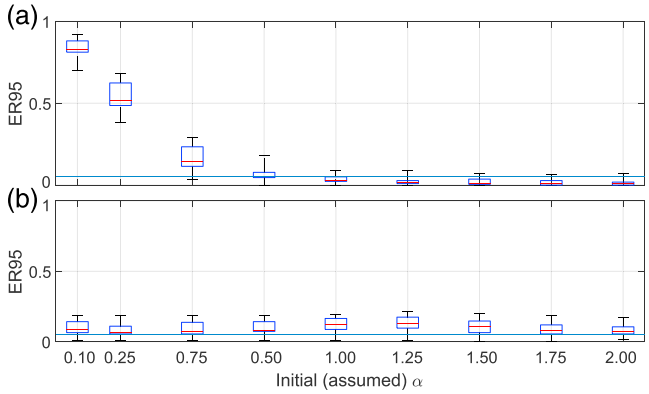

Figure 6. The 95% exceedance ratio for (a) ensemble Kalman filter, and (b) adaptive ensemble Kalman filter with deliberately wrong initial α_Q values. ER95 = 95% exceedance ratio.

To find the optimal β, 19 different values of β for each of the nine different assumed α_Q (in total 171 combinations) were tested. Figure 5b shows the RMSE of these experiments. The results showed that the optimal value of β is between 0.6 and 0.9. However, the relatively constant RMSE values indicate that the estimation efficiency is not sensitive to this value between these limit. Therefore, a value of 0.75 was selected as the optimal β. Reichle et al. (2008) also used the β parameter in the adaptive filtering. Also an optimal β improved the results marginally but the filter was able to improve the majority of estimations with $\beta = 1$. Only the optimal value β was different ($\beta = 1.06$ comparing to 0.75) which can be a result of the difference in the hydrological model structure. Therefore, based on the degree and form of the nonlinearity, the optimal β value is expected to be case specific.

In addition to the deterministic measures used to assess the performance of the EnKF and AEnK, we also used a probabilistic measure, the ER95. Figure 6 shows the ER95 for different degrees of misspecification in error

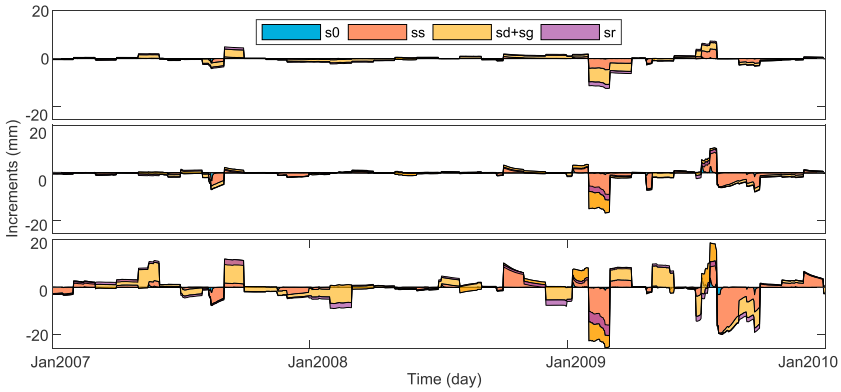

Figure 7. Daily increments obtained by different data assimilation approaches from 2007 to 2009 for a sample location (indicated with a cross in Figure 1). The values of hydrological response units (shallow-rooted and deep-rooted vegetation area inside each pixel) were averaged. s0 = surface soil moisture; ss = shallow soil moisture; sd+sg = deep soil moisture and groundwater; sr = surface storage.

magnitude of the EnKF (Figure 6a) and AEnKF (Figure 6b) and $\beta = 1$. Each boxplot in this figure represents 12 ER95 values from 12 pixels.

The results support the conclusion from the deterministic measures. An ER95 greater than 5% indicates that the ensemble spread is too wide, while a smaller value suggests an ensemble with too narrow spread.

Similar as for RMSE, the EnKF showed sensitivity to error misspecification in terms of ER95. ER95 was very high when α was underestimated, then decreased to approximately 5% (its ideal value) when assumed α was closer to its correct value, and when α was overestimated small values of ER95 resulted. Conversely, the results from AEnKF suggests that the approach can correct the incorrect error misspecification.

Figure 7 shows that in comparison with EnKF with correct error magnitude, the magnitude of the increments was considerably higher when a doubled error magnitude was used. However, applying the adaptive strategy even with this double error magnitude generated increments in a range similar to the EnKF with correct error magnitudes. The same figure was generated for six more points across Australia, supporting the conclusion. The increment sign could change from month to month when a new observation was introduced to the model. The largest impacts were observed for the deep soil moisture and groundwater store, while the surface soil moisture store received the smallest update from the observation.

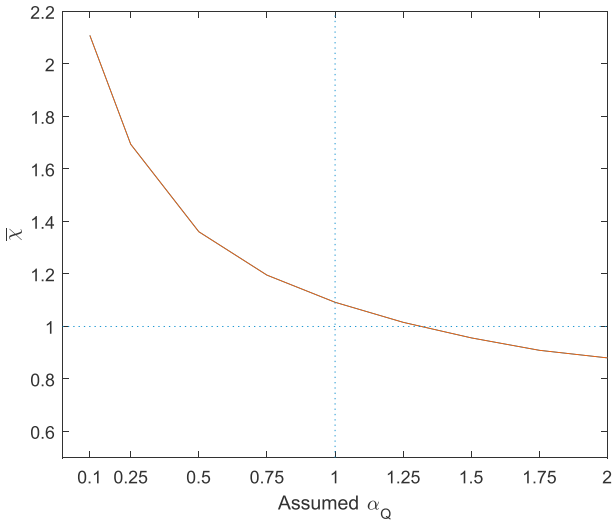

Figure 8. Spatiotemporally averaged $\overline{\chi}$ for test problems with different degrees of error misspecification. The vertical dotted line shows the correct value of error magnitude ($\alpha_Q = 1$).

5.3. Adequacy of $\overline{\chi}$

The expected value of normalized innovation means, $\overline{\chi}$, was used to determine any misspecification in error magnitude. So far it was shown that using this measure and the AEnKF approach, the negative effect of error misspecification is mitigated. However, to gain a better understanding about the adequacy of $\overline{\chi}$, its temporally averaged values for nine different values for α_Q was calculated (Figure 8). This figure shows the $\overline{\chi}$ obtained from the EnKF integrations (averaged across the domain and over the simulation period). The correct magnitude of errors to generate the observations was $\alpha_Q = 1$, but different deliberately erroneous α_Q values were introduced to the filter. In theory, for a linear system, a value for α_Q lower than its optimal value should lead to a value of $\overline{\chi}$ greater than unity. Conversely, when α_Q is too high, $\overline{\chi}$ should be smaller than unity. Figure 8 shows that $\overline{\chi}$ is obtained with an α_Q equal to approximately 1.25. This can be explained by the nonlinearity of the system. This observation is also consistent with the optimal value of 0.75 for β and explains the need for the tuning factor.

5.4. Tuning of the AEnKF Parameters

In the process of AEnKF, in addition to β, two more parameters were included, δ_f and γ (equations (10) and (11)). These parameters control the

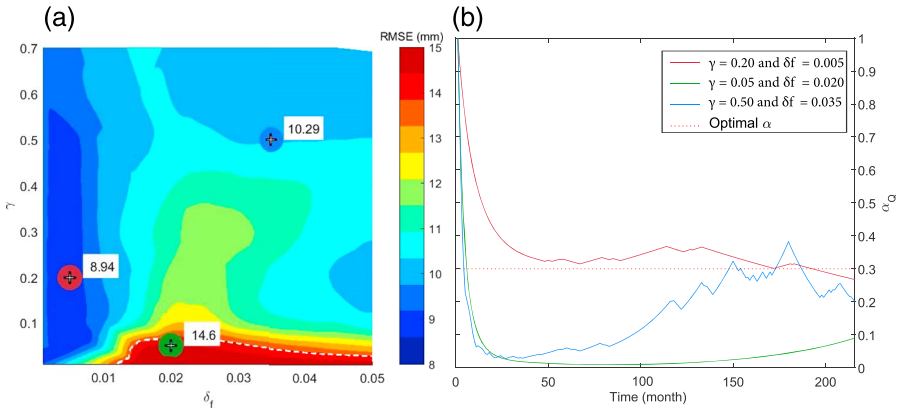

Figure 9. Performance of the Adaptive EnKF with different combinations of γ and δ_f with a wrong initial error magnitude α_Q (twice the correct α_Q). (a) The response surface of the model with the dashed line contour showing the performance of the regular ensemble Kalman filter (with the wrong α_Q). (b) Three different α_Q time series being tuned from using different sets of γ and δ_f (the parameters are color coded as per the points in the left panel).

convergence speed and smoothing degree of the error magnitude tuning. A higher value of γ leads to a longer memory of the moving average operator, removing the noise more efficiently, but with slower convergence speed.

To avoid a negative effect of spikes in the $\overline{\chi}$ values, δf was embedded in the process to filter them out. Larger values of δ_f increased the speed of convergence, but the large spikes negatively affected the tuning process. Therefore, the sensitivity of the estimation accuracy to these parameters was analyzed, and the optimal values found (Figure 9). These parameters do not necessarily have to be constant in time; they can change during the simulation period. The analysis has showed that a higher value of δf and consequently a nonrestricted tuning of the error magnitude at initial steps, can be helpful in experiments with high inaccuracy in the assumed error magnitudes. The reason is that the initial high values of the correction factor (f_τ) are likely to be real and not random spikes. However, after removing the large potential gap between the true and assumed α_Q, a more restricted correction is achieved. Therefore, in the current analysis, a time-varying δf was used. The pattern of this variation was inspired by the simulated annealing optimization algorithm (Kirkpatrick et al., 1983; Černý, 1985). In the current approach, δ_f started from a high value and during a number of predefined steps reaches a final value and retains this for the remaining time steps. This gradual reduction is achieved by

$$\delta f_\tau = max[\delta f_{\tau-1}(\frac{1}{k})^{\frac{1}{n}}, \delta f_T], \tag{14}$$

where τ and T refer to time steps and target values of δf, n is the number of time steps required to reach to the target value, and k is the ratio of the initial to target value of α_Q. After some trial and error, we selected process values $k = 16$ and $n = 40$. To select these parameters, different combinations of k in the range of 2 to 32 and n in a range of 1 to 80 were tested and the best combination in terms of accuracy of the estimates was selected.

Figure 9a shows the response surface for an experiment with correct $\alpha_Q = 1$ and assumed $\alpha_Q = 2$. In this figure, the RMSE of the regular EnKF (without adaptive error tuning) is shown as a dashed line. Figure 9b illustrates the time series of the error magnitude using three samples of AEnKF parameter sets. Most parameter sets led to an AEnKF performance that was better than the EnKF. The parameter sets with poor RMSE all were located in the region with $\gamma < 0.1$. Small γ values indicate a high memory of the moving average, suggesting that the filter needs a very long time to forget the initial (wrong) values and thus correct the α_Q. The green point (in Figure 9a) and line (in Figure 9b), illustrates one sample from this region. The initial simulated annealing process allowed a rapid change from a high to a low value of α_Q, but after reaching to the target value of δf, the restriction caused a slow and smooth convergence Although the α_Q values were moving toward optimality, the correct α_Q could not be reached within the simulation period. The results suggest that even using a nonoptimal parameter set, eventually a more suitable α_Q was found. Therefore, the whole process could be re-applied with the final value of α_Q as a new initial value.

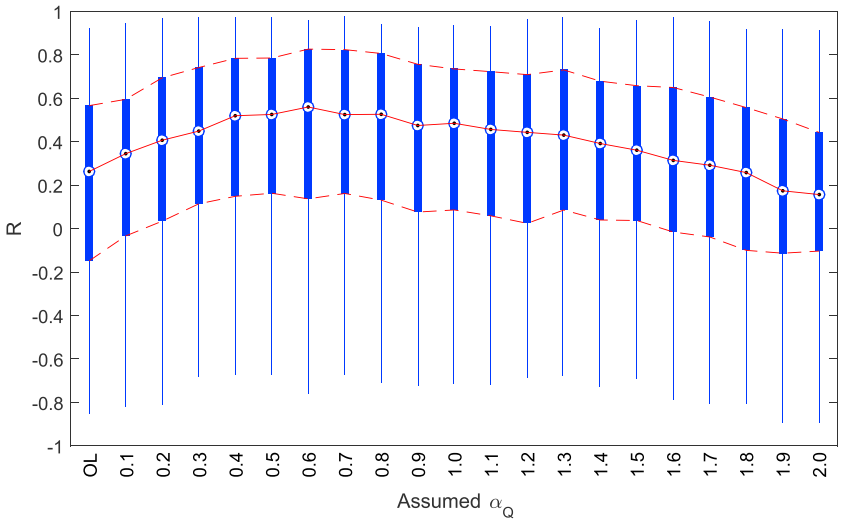

Figure 10. Boxplot of correlation coefficient between the daily groundwater observations (for all selected observation bores) and daily modeled storage from the open loop and ensemble Kalman filter with different assumed error magnitude (α_Q).

A region containing the best parameter sets can also be detected in Figure 9. For the target $\delta_f < 0.01$ good performance was observed for a wide range of γ values. However, for a very large value of γ (more than 0.6), the accuracy reduced marginally. The red point and line show the optimal parameter set. A rapid change and then some limited fluctuation around the optimal α_Q was observed. The restriction from δ_f prevents the fluctuation. It can be compared with the blue point and line which have both high γ and δ_f. Uncontrolled fluctuations and short memory degraded the estimation accuracy, but still produce a better result was achieved than for the EnKF without the adaptation component.

5.5. Application of the Real GRACE Observations

The AEnKF was also evaluated using real GRACE observations. In this case, since the observations were not generated based on perfectly known state variables, it is more difficult to assess the accuracy of all estimated state and flux variables. However, the performance of the filter was evaluated against streamflow records, groundwater level, and soil moisture point measurements. To understand the efficiency of the AEnKF, first, the error magnitude was manually optimized. The correlation with groundwater levels was selected for this analysis because previous studies (e.g., Girotto et al., 2016; Shokri, Walker, van Dijk & Pauwels 2018; Zaitchik et al., 2008) found that this water store was the most strongly influenced by GRACE TWS assimilation.

Figure 10 shows the skill of the OL model and EnKF with different magnitude of errors in terms of the correlation coefficient for groundwater.

Comparing to the OL, the EnKF generally improved the groundwater estimation. In the experiments with high overestimation of error magnitudes a slight degradation in the results was observed. An obvious optimum point can also be detected between $\alpha_Q = 0.6$ and 0.7. This figure shows that an optimization of the error parameters can improve the filter skill. However, given the computationally demanding process of the EnKF, it seems inefficient to do the manual optimization in this way. Therefore, the same problem was also solved with the AEnKF with a wrong error magnitude. As the worst result from the manual error magnitude optimization process, the value of 2 was selected for α_Q. Figure 11 shows the AEnKF skill for different variables. The skill of EnKF with optimal error magnitude ($\alpha_Q = 0.75$), and both EnKF and AEnKF with a suboptimal initial error magnitude ($\alpha_Q = 2.00$) are shown in Figure 11.

The accuracy of surface and root zone soil moisture estimates were evaluated against OzNet soil moisture observations (at 0–5 cm and 0–90 cm), streamflow against gauged observations in unregulated catchments (Zhang et al., 2013), groundwater against bore data, and TWS against the GRACE retrievals. Results indicate that using the EnKF with optimized error magnitude improves the estimation accuracy of groundwater level, surface and root zone layers, and streamflow. Due to the narrow range of storage dynamics in surface soil moisture, the improvement was less significant in this layer with the groundwater level, as the

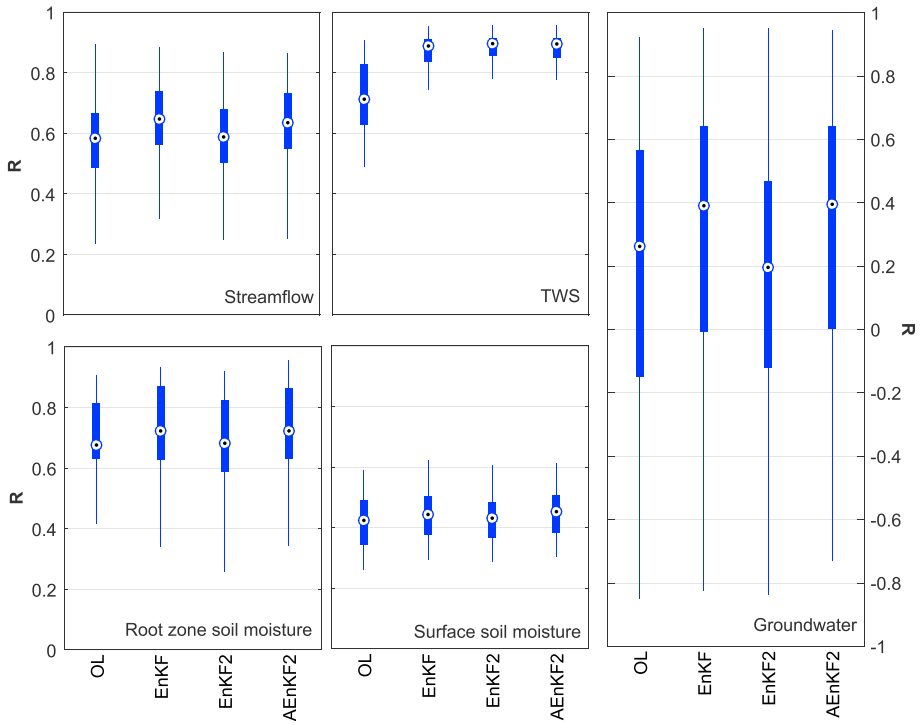

Figure 11. The skill of the open loop, optimal ensemble Kalman filter (EnKF), ensemble Kalman filter with $\alpha_Q = 2$ (EnKF2), and adaptive EnKF with $\alpha_Q = 2$ (AEnKF2) in estimating daily streamflow, terrestrial water storage (TWS), root zone and surface soil moisture, and groundwater.

most significant contributor, receiving the greatest improvement. This corroborates the results from previous studies (e.g., Girotto et al., 2016; Khaki, Hoteit, et al., 2017; Khaki, Ait-El-Fquih, et al., 2017; Schumacher et al., 2018; Shokri, Walker, van Dijk & Pauwels 2018; Tian et al., 2017; Zaitchik et al., 2008). Conversely, the accuracy of the EnKF with initial $\alpha_Q = 2$ (twice the reference error magnitude) slightly degraded in both soil layers and groundwater, comparing to OL. The streamflow estimates were not changed significantly. Finally, the AEnKF with the same wrong assumed error magnitude and without any prior optimization performed similarly to the optimized EnKF. This suggests that deriving the optimal error magnitude does not require several runs of the model; the optimization could be performed using a limited amount of internal information and a limited number of time steps in the initial stages of the AEnKF.

6. Conclusion

The EnKF is widely adapted as a reliable tool for assimilating GRACE low spatiotemporal resolution observations into fine-resolution hydrological models. An EnKF structure to assimilate coarse scale (3°) monthly GRACE observations into a fine scale (0.05°) daily AWRA-L model were applied. Shokri, Walker, van Dijk, and Pauwels (2018) proposed an optimal structure for assimilating the GRACE observations which was applied here. The results confirmed that if the EnKF uses optimal error specification it can improve the consistency between modeled TWS and GRACE observation in both synthetic and in a real case scenario. Other states and flux variables of the model were also improved; groundwater level received the most updates and consequently improvements, and due to the low impact of the surface soil moisture in the TWS, this storage was not affected by the assimilation process considerably.

Our experiments clearly showed that poorly determined error magnitudes reduces the accuracy of the model results. The efficiency reduction in experiments with severely overestimated error magnitudes could lead to a performance worse than the OL simulations. For example, a twofold increase in the error magnitude in the EnKF led to worse estimates than the OL. Results showed that the filter performed optimally with slightly underestimated error magnitudes. This can be explained by the nonlinearity of the model. A tuning parameter β was able to compensate for this effect, leading to a filter performance that is nearly identical to the performance of the optimal EnKF.

If an optimal performance of the EnKF is desired, the error magnitudes needed to be estimated manually, but this would be a very tedious task. The AEnKF can perform this estimation efficiently in one single assimilation run, being almost insensitive to the initial error estimates, which are adjusted by the AEnKF. If the β parameter is adequately chosen, a model performance similar to the optimal EnKF is obtained.

Here, $\overline{\chi}$ was selected as the internal diagnostic for the misspecification of error magnitude. The expected trend of diminishing values of $\overline{\chi}$ with increasing values of assumed α_Q was confirmed. In addition to $\overline{\chi}$ other metrics exist (e.g., three more metrics are introduced by Durand & Margulis, 2008). The investigation of their application for GRACE TWS retrievals was beyond our objective, but could be considered in future studies.

An important consideration before adding any new process to an estimation system is the impact of system complexity. An indicator of the complexity is the number of required parameters. The current adaptive process imposed three more parameters to EnKF, δf, γ, and β. The first two parameters control the speed of convergence and the rate of fluctuation of the tuned error magnitudes, while β compensates for the nonlinearity of the system. Avoiding the use of β marginally affected the filter accuracy, and hence it can be removed in cases with limited benchmarking or computational resources.

Our overall conclusion is that there is a strong potential to improve GRACE data assimilation systems through optimization of the error estimates. An adaptive EnKF can be used for this purpose. Moreover, despite the nonlinearities in hydrologic models, neglecting the nonlinearity (i.e., $\beta = 1$) within the AEnKF led to a filter performance that was almost similar to that if a manually tuned optimal filter performance. These conclusions were confirmed using both a synthetic study and a real-world data assimilation study.

Acknowledgments
Ashkan Shokri acknowledges the support from Monash University in the form of a Monash Graduate Scholarship (MGS) and Monash International Postgraduate Research Scholarship (MIPRS). Valentijn Pauwels was supported through an ARC Future Fellowship Grant (FT130100545). The research is supported through an ARC Discovery Grant (DP140103679). This research is also supported in part by the Monash eResearch Centre and eSolutions-Research Support Services through the use of the Monash Campus HPC Cluster. The meteorological data are obtained from the Australian Water Availability Project (http://www.bom.gov.au/jsp/awap/index.jsp). The GRACE observations are obtained online (https://podaac.jpl.nasa.gov/dataset/TELLUS_GRACE_MASCON_CRI_GRID_RL05_V2; Watkins et al., 2015; Wiese et al., 2016). The groundwater level measurements are obtained online (http://www.bom.gov.au/water/groundwater/explorer/). The OZnet soil moisture observations are obtained online (http://www.oznet.org.au/; A. B. Smith et al., 2012).

References

Aubert, D., Loumagne, C., & Oudin, L. (2003). Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall-runoff model. *Journal of Hydrology*, *280*(1), 145–161. Retrieved from http://www.sciencedirect.com/science/article/pii/S0022169403002294

Barrett, A. P. (2003). National Operational Hydrologic Remote Sensing Center Snow Data Assimilation System (SNODAS) products at NSIDC (*NSIDC Special Report 11*): National Snow and Ice Data Center, Cooperative Institute for Research in Environmental Sciences.

Bateni, S. M., & Entekhabi, D. (2012). Surface heat flux estimation with the ensemble Kalman smoother: Joint estimation of state and parameters. *Water Resources Research*, *48*, W08521. https://doi.org/10.1029/2011WRO11542

Berry, T., & Sauer, T. (2013). Adaptive ensemble Kalman filtering of non-linear systems. *Tellus A: Dynamic Meteorology and Oceanography*, *65*(1), 20331. https://doi.org/10.3402/tellusa.v65i0.20331

Černý, V. (1985). Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *Journal of Optimization Theory and Applications*, *45*(1), 41–51.

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., et al. (2008). Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. *Advances in Water Resources*, *31*(10), 1309–1324. Retrieved from http://www.sciencedirect.com/science/article/pii/S0309170808001012

Crow, W. T., & Ryu, D. (2009). A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals. *Hydrology and Earth System Sciences*, *13*(1), 1–16.

Crow, W. T., & Van Loon, E. (2006). Impact of incorrect model error assumptions on the sequential assimilation of remotely sensed surface soil moisture. *Journal of Hydrometeorology*, *7*(3), 421–432. https://doi.org/10.1175/JHM499.1

Daley, R. (1992). Estimating model-error covariances for application to atmospheric data assimilation. *Monthly Weather Review*, *120*(8), 1735–1746. https://doi.org/10.1175/1520-0493(1992)120<1735:EMECFA>2.0.CO;2

De Lannoy, G. J. M., Houser, P. R., Verhoest, N. E. C., & Pauwels, V. R. N. (2009). Adaptive soil moisture profile filtering for horizontal information propagation in the independent column-based CLM2.0. *Journal of Hydrometeorology*, *10*(3), 766–779. https://doi.org/10.1175/2008JHM1037.1

Dechant, C., & Moradkhani, H. (2011). Radiance data assimilation for operational snow and streamflow forecasting. *Advances in Water Resources*, *34*(3), 351–364.

Dee, D. P. (1995). On-line estimation of error covariance parameters for atmospheric data assimilation. *Monthly Weather Review*, *123*(4), 1128–1145. https://doi.org/10.1175/1520-0493(1995)123<1128:OLEOEC>2.0.CO;2

Durand, M., & Margulis, S. A. (2008). Effects of uncertainty magnitude and accuracy on assimilation of multiscale measurements for snowpack characterization. *Journal of Geophysical Research*, *113*, D02105. https://doi.org/10.1029/2007JD008662

Ellett, K. M., Walker, J. P., Western, A. W., & Rodell, M. (2006). A framework for assessing the potential of remote-sensed gravity to provide new insight on the hydrology of the Murray-Darling basin. *Australasian Journal of Water Resources*, *10*(2), 125–138. https://doi.org/10.1080/13241583.2006.11465286

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, *99*(C5), 10,143–10,162.

Evensen, G. (1997). Advanced data assimilation for strongly nonlinear dynamics. *Monthly Weather Review*, *125*(6), 1342–1354.

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

Forman, B. A., & Reichle, R. (2013). The spatial scale of model errors and assimilated retrievals in a terrestrial water storage assimilation system. *Water Resources Research*, *49*, 7457–7468. https://doi.org/10.1002/2012WR012885

Forman, B. A., Reichle, R. H., & Rodell, M. (2012). Assimilation of terrestrial water storage from GRACE in a snow-dominated basin. *Water Resources Research*, *48*, W0150. https://doi.org/10.1029/2011WR011239

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

Girotto, M., De Lannoy, G. J., Reichle, R. H., Rodell, M., Draper, C., Bhanja, S. N., & Mukherjee, A. (2017). Benefits and pitfalls of GRACE data assimilation: A case study of terrestrial water storage depletion in India. *Geophysical Research Letters*, *44*, 4107–4115. https://doi.org/10.1002/2017GL072994

Hendricks Franssen, H.-J., Zhang, H., Kurtz, W., Kollet, S., & Vereecken, H. (2017). Improved characterization of root zone soil moisture in land surface models by assimilation of groundwater level data. an example with TerrSysMP. In *Egu general assembly conference abstracts* (Vol. 19, p. 15653). Vienna, Austria.

Hong, L., Eugenia, K., & Takemasa, M. (2009). Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, *135*(639), 523–533. https://doi.org/10.1002/qj.371

Houborg, R., Rodell, M., Li, B., Reichle, R., & Zaitchik, B. F. (2012). Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations. *Water Resources Research*, *48*, W07525. https://doi.org/10.1029/2011WR011291

Houser, P. R., Shuttleworth, W. J., Famiglietti, J. S., Gupta, H. V., Syed, K. H., & Goodrich, D. C. (1998). Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resources Research*, *34*(12), 3405–3420.

Khaki, M., Ait-El-Fquih, B., Hoteit, I., Forootan, E., Awange, J., & Kuhn, M. (2017). A two-update ensemble Kalman filter for land hydrological data assimilation with an uncertain constraint. *Journal of Hydrology*, *555*, 447–462. Retrieved from http://www.sciencedirect.com/science/article/pii/S0022169417307072

Khaki, M., Hoteit, I., Kuhn, M., Awange, J., Forootan, E., van Dijk, A. I. J. M., et al. (2017). Assessing sequential data assimilation techniques for integrating GRACE data into a hydrological model. *Advances in Water Resources*, *107*, 301–316. Retrieved from http://www.sciencedirect.com/science/article/pii/S0309170816307564

Khaki, M., Schumacher, M., Forootan, E., Kuhn, M., Awange, J. L., & van Dijk, A. I. J. M. (2017). Accounting for spatial correlation errors in the assimilation of GRACE into hydrological models through localization. *Advances in Water Resources*, *108*, 99–112. Retrieved from http://www.sciencedirect.com/science/article/pii/S0309170816308168

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, *220*(4598), 671–680.

Kumar, S. V., Zaitchik, B. F., Peters-Lidard, C. D., Rodell, M., Reichle, R., Li, B., et al. (2016). Assimilation of gridded GRACE terrestrial water storage estimates in the North American land data assimilation system. *Journal of Hydrometeorology*, *17*(7), 1951–1972.

Lee, H., Seo, D.-J., & Koren, V. (2011). Assimilation of streamflow and in situ soil moisture data into operational distributed hydrologic models: Effects of uncertainties in the data and initial model soil moisture states. *Advances in Water Resources*, *34*(12), 1597–1615. Retrieved from http://www.sciencedirect.com/science/article/pii/S0309170811001680

Li, B., & Rodell, M. (2015). Evaluation of a model-based groundwater drought indicator in the conterminous US. *Journal of Hydrology*, *526*, 78–88.

Li, B., Rodell, M., Zaitchik, B. F., Reichle, R. H., Koster, R. D., & van Dam, T. M. (2012). Assimilation of GRACE terrestrial water storage into a land surface model: Evaluation and potential value for drought monitoring in western and central Europe. *Journal of Hydrology*, *446*, 103–115.

Mehra, R. (1970). On the identification of variances and adaptive Kalman filtering. *IEEE Transactions on Automatic Control*, *15*(2), 175–184.

Moghaddamjoo, A. R., & Kirlin, R. L. (1993). Robust adaptive Kalman filtering. In *Approximate Kalman Filtering* (pp. 65–85). Singapore: World Scientific. https://doi.org/10.1142/9789814317399_0006

Moradkhani, H., Hsu, K., Hong, Y., & Sorooshian, S. (2006). Investigating the impact of remotely sensed precipitation and hydrologic model uncertainties on the ensemble streamflow forecasting. *Geophysical Research Letters*, *33*, L12401. https://doi.org/10.1029/2006GL026855

Pauwels, V. R. N., Hoeben, R., Verhoest, N. E. C., & De Troch, F. P. (2001). The importance of the spatial patterns of remotely sensed soil moisture in the improvement of discharge predictions for small-scale basins through data assimilation. *Journal of Hydrology*, *251*(1), 88–102.

Pigram, J. (2007). *Australia's water resources: From use to management*. Melbourne: Csiro Publishing.

Pipunic, R. C., Walker, J. P., Western, A. W., & Trudinger, C. M. (2013). Assimilation of multiple data types for improved heat flux prediction: A one-dimensional field study. *Remote Sensing of Environment*, *136*, 315–329.

Reichle, R. H., Crow, W. T., & Keppenne, C. L. (2008). An adaptive ensemble Kalman filter for soil moisture data assimilation. *Water Resources Research*, *44*, W03423. https://doi.org/10.1029/2007WR006357

Reichle, R. H., Koster, R. D., Dong, J., & Berg, A. A. (2004). Global soil moisture from satellite observations, land surface models, and ground data: Implications for data assimilation. *Journal of Hydrometeorology*, *5*(3), 430–442.

Reichle, R. H., Walker, J. P., Koster, R. D., & Houser, P. R. (2002a). Extended versus ensemble Kalman filtering for land data assimilation. *Journal of Hydrometeorology*, *3*(6), 728–740. https://doi.org/10.1175/1525-7541(2002)003<0728:EVEKFF>2.0.CO;2

Reichle, R. H., Walker, J. P., Koster, R. D., & Houser, P. R. (2002b). Extended versus ensemble Kalman filtering for land data assimilation. *Journal of Hydrometeorology*, *3*(6), 728–740. https://doi.org/10.1175/1525-7541(2002)003<0728:EVEKFF>2.0.CO;2

Renzullo, L. J., van Dijk, A. I. J. M., Perraud, J. M., Collins, D., Henderson, B., Jin, H., et al. (2014). Continental satellite soil moisture data assimilation improves root-zone moisture analysis for water resources assessment. *Journal of Hydrology*, *519, Part D*, 2747–2762.

Schumacher, M., Forootan, E., van Dijk, A. I. J. M., Müller Schmied, H., Crosbie, R. S., Kusche, J., & Döll, P. (2018). Improving drought simulations within the Murray-Darling basin by combined calibration/assimilation of GRACE data into the WaterGAP Global Hydrology Model. *Remote Sensing of Environment*, *204*, 212–228. Retrieved from http://www.sciencedirect.com/science/article/pii/S0034425717304923

Shokri, A., Walker, J. P., van Dijk, A. I. J. M., & Pauwels, V. R. N. (2018). Performance of different ensemble Kalman filter structures to assimilate GRACE terrestrial water storage estimates into a high-resolution hydrological model: A synthetic study. *Water Resources Research*, *54*, 8931–8951. https://doi.org/10.1029/2018WR022785

Shokri, A., Walker, J. P., van Dijk, A. I. J. M., Wright, A. J., & Pauwels, V. R. N. (2018). Application of the patient rule induction method to detect hydrologic model behavioural parameters and quantify uncertainty. *Hydrological Processes*, *32*(8), 1005–1025. https://doi.org/10.1002/hyp.11464

Slater, A. G., & Clark, M. P. (2006). Snow data assimilation via an ensemble Kalman filter. *Journal of Hydrometeorology*, *7*(3), 478–493.

Smith, A. B. (2013). Soil moisture monitoring with ground-based gravity data (PhD thesis). Monash University.

Smith, A. B., Walker, J. P., Western, A. W., Young, R. I., Ellett, K. M., Pipunic, R. C., et al. (2012). The Murrumbidgee soil moisture monitoring network data set. *Water Resources Research*, *48*, W07701. https://doi.org/10.1029/2012WR011976

Sun, C., Walker, J. P., & Houser, P. R. (2004). A methodology for snow data assimilation in a land surface model. *Journal of Geophysical Research*, *109*, D08108. https://doi.org/10.1029/2003JD003765

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

van Dijk, A. I. J. M. (2010). The Australian Water Resources Assessment system. technical report 3. landscape model (version 0.5) technical description.: CSIRO: Water for a Healthy Country National Research Flagship.

van Dijk, A. I. J. M., Peña-Arancibia, J. L., & Bruijnzeel, L. A. (2012). Land cover and water yield: Inference problems when comparing catchments with mixed land cover. *Hydrology and Earth System Sciences*, *16*(9), 3461–3473.

van Dijk, A. I. J. M., Renzullo, L. J., & 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 Resources Research*, *47*, W11524. Author Affiliation: Black Mountain Laboratories, CSIRO Land and Water, Canberra, ACT, Australia https://doi.org/10.1029/2011WR010714

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

Van Dijk, A. I. J. M., & Warren, G. (2010). The Australian Water Resources Assessment system. Version 0.5 , 3 (5).

Van Geer, F. C., Te Stroet, C. B. M., & Yangxiao, Z. (1991). Using Kalman filtering to improve and quantify the uncertainty of numerical groundwater simulations: 1. The role of system noise and its calibration. *Water Resources Research*, *27*(8), 1987–1994.

Verlaan, M., & Heemink, A. W. (2001). Nonlinearity in data assimilation applications: A practical method for analysis. *Monthly Weather Review*, *129*(6), 1578–1589.

Walker, J. P., & Houser, P. R. (2001). A methodology for initializing soil moisture in a global climate model: Assimilation of near-surface soil moisture observations. *Journal of Geophysical Research*, *106*(D11), 11,761–11,774.

Watkins, M. M., Wiese, D. N., Yuan, D.-N., Boening, C., & Landerer, F. W. (2015). Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. *Journal of Geophysical Research: Solid Earth*, *120*, 2648–2671. https://doi.org/10.1002/2014JB011547

Wiese, D. N., Landerer, F. W., & Watkins, M. M. (2016). Quantifying and reducing leakage errors in the JPL RL05M GRACE mascon solution. *Water Resources Research*, *52*, 7490–7502. https://doi.org/10.1002/2016WR019344

Xu, T., Bateni, S. M., Neale, C. M. U., Auligne, T., & Liu, S. (2018). Estimation of turbulent heat fluxes by assimilation of land surface temperature observations from GOES satellites into an ensemble Kalman smoother framework. *Journal of Geophysical Research: Atmospheres*, *123*, 2409–2423. https://doi.org/10.1002/2017JD027732

Yang, S.-C., Kalnay, E., & Hunt, B. (2012). Handling nonlinearity in an ensemble Kalman filter: Experiments with the three-variable Lorenz model. *Monthly Weather Review*, *140*(8), 2628–2646. https://doi.org/10.1175/MWR-D-11-00313.1

Zaitchik, B. F., Rodell, M., & Reichle, R. H. (2008). Assimilation of GRACE terrestrial water storage data into a land surface model: Results for the Mississippi River basin. *Journal of Hydrometeorology*, *9*(3), 535–548.

Zhang, S.-W., Qiu, C.-J., & Xu, Q. (2004). Estimating soil water contents from soil temperature measurements by using an adaptive Kalman filter. *Journal of Applied Meteorology*, *43*(2), 379–389.

Zhang, Y. Q., Viney, N., Frost, A., Oke, A., Brooks, M., Chen, Y., & Campbell, N. (2013). Collation of Australian modellers streamflow dataset for 780 unregulated Australian catchments. Water for a Healthy Country National Research Flagship, 115pp. Catchment Management.