CrossMark

ORIGINAL ARTICLE

A systematic impact assessment of GRACE error correlation on data assimilation in hydrological models

Maike Schumacher[1] · Jürgen Kusche[1] · Petra Döll[2]

Abstract Recently, ensemble Kalman filters (EnKF) have found increasing application for merging hydrological models with total water storage anomaly (TWSA) fields from the Gravity Recovery And Climate Experiment (GRACE) satellite mission. Previous studies have disregarded the effect of spatially correlated errors of GRACE TWSA products in their investigations. Here, for the first time, we systematically assess the impact of the GRACE error correlation structure on EnKF data assimilation into a hydrological model, i.e. on estimated compartmental and total water storages and model parameter values. Our investigations include (1) assimilating gridded GRACE-derived TWSA into the WaterGAP Global Hydrology Model and, simultaneously, calibrating its parameters; (2) introducing GRACE observations on different spatial scales; (3) modelling observation errors as either spatially white or correlated in the assimilation procedure, and (4) replacing the standard EnKF algorithm by the square root analysis scheme or, alternatively, the singular evolutive interpolated Kalman filter. Results of a synthetic experiment designed for the Mississippi River Basin indicate that the hydrological parameters are sensitive to TWSA assimilation if spatial resolution of the observation data is sufficiently high. We find a significant influence of spatial error correlation on the adjusted water states and model parameters for all implemented filter variants, in particular for subbasins with a large discrepancy between observed and initially simulated TWSA and for north–south elongated sub-basins. Consid-

ering these correlated errors, however, does not generally improve results: while some metrics indicate that it is helpful to consider the full GRACE error covariance matrix, it appears to have an adverse effect on others. We conclude that considering the characteristics of GRACE error correlation is at least as important as the selection of the spatial discretisation of TWSA observations, while the choice of the filter method might rather be based on the computational simplicity and efficiency.

1 Introduction

Since March 2002, the Gravity Recovery And Climate Experiment (GRACE) satellite mission, which consists of two satellites in tandem formation, has been continuously monitoring the Earth's time-variable gravity field. GRACE time-variable level-2 gravity products can be converted into total water storage anomalies (TWSA; Wahr et al. 1998; Tapley et al. 2004) with temporal resolution of 1 month to even 1 day (Kurtenbach et al. 2009) depending on the analysis technique and spatial resolution of down to a few hundred kilometres (Schmidt et al. 2008). GRACE level-2 products have been used in various environmental studies to estimate water storage changes within the Earth system (see Kusche et al. 2012; Famiglietti and Rodell 2013; Wouters et al. 2014, and references therein).

Several recent studies suggested the use of GRACE data products to improve the simulation skills of hydrological models (e.g. Zaitchik et al. 2008; Van Dijk et al. 2014; Eicker et al. 2014). Merging GRACE TWSA and hydrologi-

✉ Maike Schumacher
schumacher@geod.uni-bonn.de

[1] Institute of Geodesy and Geoinformation, University of Bonn, Nussallee 17, 53115 Bonn, Germany

[2] Institute of Physical Geography, University of Frankfurt/Main, Altenhöferallee 1, 60438 Frankfurt am Main, Germany

🖄 Springer

cal models provides a twofold opportunity. From the geodetic point of view, model-derived TWSA simulations that are consistent with time-variable mass estimations derived from GRACE could be very beneficial for applications that require the reduction of short-term gravity change, e.g. dealiasing of GRACE level-2 products (Zenner et al. 2014) and computation of loading effects in geometrical techniques (e.g. Collilieux et al. 2011; Fritsche et al. 2012). From the hydrological point of view, adjusting model-derived water states to GRACE observations helps overcoming limited simulation skills of models caused by uncertainties of input data (in particular climate forcings), model structure and model parameters. Therefore, besides the traditional calibration of hydrological models against discharge measurements (Gupta et al. 1998), multi-criteria calibration against river discharge and GRACE TWSA for large river basins was performed by adjusting sensitive model parameters (Werth and Güntner 2010). Recently, a number of studies have suggested assimilation of GRACE TWSA into hydrological models (Zaitchik et al. 2008; Su et al. 2010; Forman et al. 2012; Houborg et al. 2012; Li et al. 2012; Van Dijk et al. 2014; Eicker et al. 2014; Tangdamrongsub et al. 2015).

Assimilating GRACE TWSA into hydrological and land surface models is challenging because of (i) the temporal and spatial resolution mismatch between model-derived simulated water states and GRACE TWSA, (ii) the difficulty in describing model errors due to forcing, model parameters and model structure (e.g. Reichle and Koster 2003; Crow and Van Loon 2006; Moradkhani et al. 2006; Liu et al. 2012), and finally (iii) the difficulty to appropriately describe errors of GRACE TWSA. In particular, GRACE level-2 products, represented in terms of potential spherical harmonics, contain correlated errors, which result from instrumental noise (K-band ranging system, Pierce et al. 2008), anisotropic spatial sampling of the mission (Schrama et al. 2007), and temporal aliasing caused by incomplete reduction of short-term mass variations by models (Flechtner et al. 2010; Forootan et al. 2014). These errors manifest themselves as "striping" patterns in GRACE-derived TWSA (Kusche 2007). Although striping is reduced after applying de-correlation filters (Swenson and Wahr 2006; Klees et al. 2008; Kusche et al. 2009), correlated errors still exist even after spatial aggregation (Longuevergne et al. 2010; Sakumura et al. 2014).

The assumption of uncorrelated Gaussian distributed errors has been usually made in previous studies for assimilating (sub-)basin-averaged (Zaitchik et al. 2008; Forman et al. 2012; Houborg et al. 2012; Li et al. 2012) or gridded GRACE TWSA (Van Dijk et al. 2014; Tangdamrongsub et al. 2015) into hydrological models. Beyond this point, Forman and Reichle (2013) investigated the effect of spatial aggregation of GRACE TWSA in a data assimilation

framework, assuming white noise for simulated TWSA. They concluded that TWSA observations should be assimilated at the smallest spatial scale for which the observation errors can be considered uncorrelated. For the first time, Eicker et al. (2014) investigated the potential of assimilating gridded GRACE TWSA (5° × 5° grids) with their full error information into the WaterGAP Global Hydrology Model (WGHM, Döll et al. 2003), exemplarily for the Mississippi River Basin. Their study used the full covariance matrix of level-2 products to estimate correlated errors of TWSA. These were then considered in a calibration and data assimilation (C/DA) framework, which was built based on the standard ensemble Kalman filter (EnKF) technique (Evensen 1994).

Assimilation of GRACE TWSA into hydrological models has been usually performed with the ensemble Kalman filter or smoother (EnKF/S, Evensen 1994; Evensen and Van Leeuwen 2000) techniques, since these are easy to implement and well suited for representing model prediction and update errors. Application of EnKF/S avoids the costly computation of gradients of highly nonlinear model equations or the generation of adjoint code as it is required in variational methods (Le Dimet and Talagrand 1986). However, for practical implementation of EnKF/S, the ensemble size is inevitably limited due to computational constraints, causing problems like ensemble inbreeding or artificial model state correlations (see, e.g. Liu et al. 2012, and references therein). Our first motivation for considering variants of the filter algorithm is that the standard EnKF approach uses an observation ensemble that introduces an additional source of sampling errors to the algorithm (Evensen 2004). Whitaker and Hamill (2002) showed that for small ensemble sizes the sampling errors are smaller when using square root analysis (SQRA, Evensen 2004) methods (Tippett et al. 2003 and references therein). The second motivation is to reduce computation time in the update step. When applying the singular evolutive interpolated Kalman (SEIK) filter (Pham et al. 1998), the analysis is performed in the ensemble space instead of the observation space, unlike for the EnKF and SQRA methods. Therefore, especially the assimilation of large numbers of observations (i.e. much larger than the ensemble size) is usually better handled by the SEIK filter. In addition, a range of tuning techniques exist that seek to optimise the generation of ensembles, e.g. applying variance inflation factors (Hamill and Snyder 2002) to avoid filter convergence. It is worth mentioning that so far no single technique has been found that always leads to superior assimilation results for different models and case studies.

Building on the approach presented in Eicker et al. (2014), in this study, the effect of spatially correlated errors in GRACE TWSA products is investigated while assimilating synthetic GRACE TWSA into WGHM. Our investigations

account for a range of design options inherent in the data analysis: (i) diagonal and full error covariance matrices of GRACE level-2 products (as in Eicker et al. 2014) are considered to investigate the effect of spatially correlated errors on the results of the C/DA approach. (ii) Spatial aggregation (as studied in Forman and Reichle 2013) is performed to investigate how correlated GRACE errors affect C/DA results when introducing observations at different spatial scales. (iii) SQRA and SEIK techniques are implemented to understand whether the updated water states and parameters react with a different degree of sensitivity to the assumed observation errors. (iv) Finally, tuning by considering variance inflation is performed for representing errors in model structure and avoiding ensemble convergence.

To design the synthetic experiment, WGHM simulations with two different types of forcing fields and parameter sets were set up, from which one run served as the "truth" and the other as the perturbed model version. Synthetic TWSA observations were generated by adding spatially correlated GRACE-like TWSA errors to the simulated truth.

Within the C/DA analysis steps, either the full observation error covariance matrix or only its diagonal elements, i.e. the assumption of white noise, were introduced in the EnKF variants. The influence of the observation error covariance information on the updated water states and calibration parameters was then assessed by comparing the model outputs with the simulated truth. In the following, we will show that correlated GRACE errors have a significant influence, regardless of the implemented filter approach, on water states in the majority of sub-basins and on sensitive calibration parameters. Those sub-basins that are elongated in north–south direction, and those with high mismatch between modelled and observed TWSA were affected the most.

The remaining part of the paper begins with a description of the hydrological model WGHM and the GRACE TWSA errors in Sect. 2. The mathematical relationship between various EnKF variants, including their similarities and differences are described in Sect. 3. Our experimental set up is introduced in Sect. 4, comprising a description of the study area (Mississippi River Basin) and a summary of the generation of observation errors and model ensembles. Various experiments with varying observation error assumptions (with and without correlations) in the filter variants, including the standard EnKF, SQRA and SEIK, and the effect of spatial discretisation of observations are discussed in Sect. 5. The assessments are performed for the individual water state changes and calibration parameters, as well as of the model-derived total water storage changes after performing calibration/data assimilation. In Sect. 6, we conclude the paper with our main findings.

2 Model and data

2.1 WaterGAP Global Hydrology Model (WGHM)

WGHM simulates daily continental water flows and storages with a spatial resolution of $0.5° \times 0.5°$ for the global land area excluding Antarctica (Döll et al. 2003). Here, we used the model version WaterGAP 2.2, which is calibrated against mean annual river discharge at 1319 gauging stations, of which 84 are located in the Mississippi Basin (Müller Schmied et al. 2014). Water storage in ten individual compartments (canopy, snow, soil, groundwater, local wetlands, global wetlands, local lakes, global lakes, global reservoirs, and rivers) is computed for each grid cell. Local lakes and wetlands receive only local runoff, while global surface water bodies including rivers receive inflow from the upstream grid cells, too. The vertical water balance describes the transport of water through the canopy, snow, and soil compartment, partitioning precipitation into evapotranspiration and runoff. Water transport as runoff from the land area is partitioned into fast surface and subsurface runoff, which flows directly into the surface water bodies and groundwater recharge. The latter first flows into the groundwater and subsequently as groundwater outflow into surface water bodies. In addition, precipitation over surface water is added to the lake, wetland, reservoir, and river compartments, while evaporation reduces the storages. The river compartment is the final storage of the grid cells. The outflow for each cell and, thus, the inflow of the lake and wetland or river compartment of the next cell is directed laterally on the basis of the global Drainage Direction Map DDM30. Furthermore, the impact of human water use as simulated by WaterGAP water use submodels is taken into account in WGHM. Net water use (water abstractions minus return flows) is abstracted from surface water bodies (including river) or groundwater (Döll et al. 2012).

The model can be forced by several climate input data sets. Here, monthly time series of the number of wet days in month, temperature and cloudiness were used from the data set CRU TS 3.2 (Climate Research Unit's Time Series; Harris et al. 2013), whereas monthly precipitation input fields were taken from the GPCC (Global Precipitation Climatology Centre data, Version 6) precipitation monitoring product (Schneider et al. 2014). In WGHM, precipitation values are equally partitioned to the number of wet days in a month, while the wet days were distributed using a first-order Markov chain. Daily short- and long-wave radiation are determined from the cloudiness information. Alternatively, daily time series of precipitation, temperature, short- and long-wave radiation from the WFDEI meteorological data set (WATCH Forcing Data methodology applied to ERA-Interim data; Weedon et al. 2014) were used in this study. The impact of using these two different climate input data sets on water flows and storage as computed by WaterGAP

2.2 was reported in Müller Schmied et al. (2014). A detailed decription of WGHM can be found in Döll et al. (2003) and Müller Schmied et al. (2014).

2.2 GRACE TWSA errors

In this study, we generated synthetic GRACE TWSA values using WGHM simulations (see Sect. 4.3). In order to generate error samples of TWSA in the C/DA procedure, five methods might be used, three of them resulting in white noise and two of them resulting in correlated errors (Fig. 1): the assumption of white noise can be made by either (1) using standard deviations based on literature, e.g. Wahr et al. (2006) (used in Zaitchik et al. 2008; Su et al. 2010; Forman et al. 2012; Forman and Reichle 2013), (2) propagating errors from standard deviations of GRACE level-2 potential coefficients, or (3) propagating errors from the full covariance matrix of GRACE level-2 potential coefficients to standard deviations of TWSA. Alternatively, correlated error samples can be generated from (4) error propagation of standard deviations of potential coefficients or from (5) propagation of the full error covariance matrix of potential coefficients to a full covariance matrix of TWSA (as in Forootan and Kusche 2012; Eicker et al. 2014).

In this study, we simulated "true" TWSA using our hydrological model. GRACE-like TWSA was then generated by adding correlated noise that was derived from the full ITG-GRACE2010 (http://www.igg.uni-bonn.de/apmg/index.php?id=itg-grace2010) error covariance matrix of potential coefficients (August 2003, up to degree/order 60),

which was propagated to the full error covariance matrix of TWSA (option 5 in Fig. 1). In the filter update step, then two assumptions on GRACE TWSA errors were considered: (i) The full ITG-GRACE2010 error covariance matrix of TWSA, which has also been used to simulate GRACE-like TWSA; and (ii) a diagonal error covariance matrix was assumed that considered only the main diagonal elements from the full error covariance matrix of TWSA in (i), which corresponds to Option 3 in Fig. 1. The generated errors according to (ii), therefore, can be considered as white noise.

3 Methodology

In this study, the C/DA framework based on the standard EnKF (Evensen 1994) introduced in Eicker et al. (2014) has been extended by the SQRA (Evensen 2004) and SEIK filters (Pham et al. 1998). In our test case, the number of GRACE observations assimilated into WGHM per epoch is smaller than the ensemble size, which means that SEIK is not necessarily the most efficient choice. Nevertheless, the SEIK is included in our study, since we may as well analyse larger river basins or more observations (e.g. river discharge, lake level, soil moisture or snow water equivalent) in future work.

The two-step procedure of our C/DA includes (i) the ensemble prediction step, i.e. the forward integration of the model for each ensemble member (that is basically independent of the applied filter algorithm), and (ii) the update (or analysis) step that merges model states and observations. To perform a simultaneous calibration of model parameters,

Fig. 1 TWSA errors description: (*1*) using standard deviations based on literature; propagating standard deviations of potential coefficients c_{nm} and s_{nm} to (*2*) standard deviations or (*4*) to correlated errors of TWSA, and propagating correlated errors of potential coefficients to (*3*) standard deviations or (*5*) correlated errors of TWSA
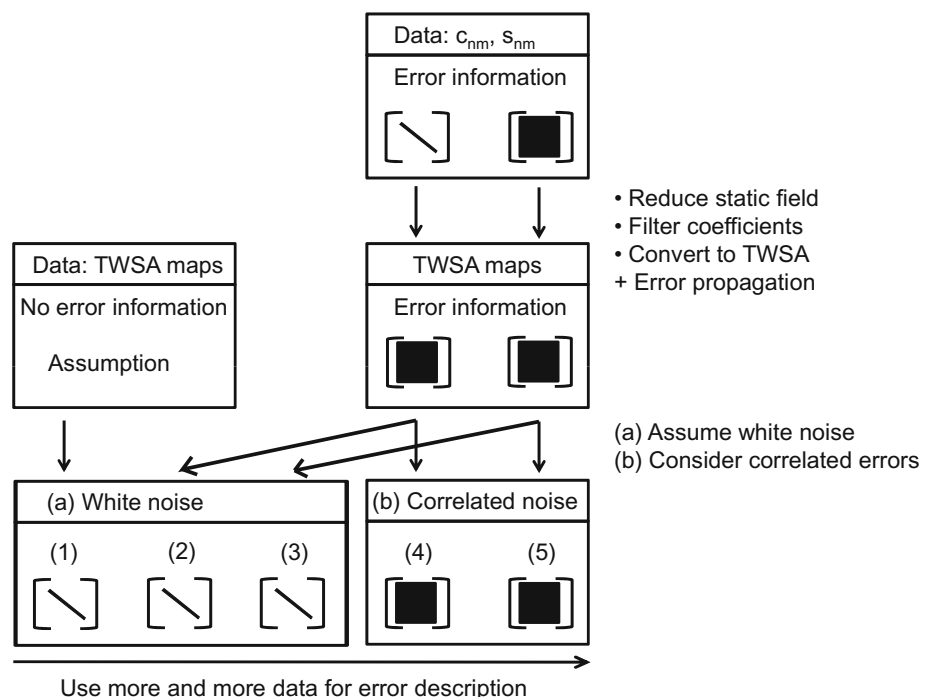
state vector augmentation is introduced (as in Eicker et al. 2014). Additionally, an inflation factor for tuning the model ensemble, and the measurement and mapping operators for merging model states and observations are considered, which will be described in the following.

3.1 Ensemble prediction

The model forward integration is implemented by evaluating the non-linear dynamical model equations, denoted by $f(.)$,

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{u}_k, \mathbf{p}) + \mathbf{q}_{k-1}. \tag{1}$$

The model states $\mathbf{x}_k$ of the current time step k depend nonlinearly on the model states $\mathbf{x}_{k-1}$ of the previous time step $(k-1)$, time-dependent input forcing fields $\mathbf{u}_k$ and constant model parameters $\mathbf{p}$, as well as on unknown model errors $\mathbf{q}_{k-1}$. In linear approaches, the error covariance matrix of the model is obtained from error propagation of the previous model state covariance matrix $\mathbf{C}(\mathbf{x}_{k-1})$ to the current time step, $\mathbf{C}(\mathbf{x}_k) = \mathbf{F}\mathbf{C}(\mathbf{x}_{k-1})\mathbf{F}^T + \mathbf{Q}_{k-1}$. Herein, $\mathbf{F}$ is the transition matrix that relates the model states of time step $(k-1)$ and k. The model error covariance matrix, $\mathbf{Q}_{k-1} = E(\mathbf{q}_{k-1}\mathbf{q}_{k-1}^T)$, in which $E(.)$ denotes the expectation value, should be given.

In ensemble-based data assimilation, the model equations are evaluated for each of the $i = 1, \dots, N_e$ ensemble members (e.g. Evensen 2007):

$$\mathbf{x}_k^{(i)-} = f(\mathbf{x}_{k-1}^{(i)}, \mathbf{u}_k^{(i)}, \mathbf{p}^{(i)}). \tag{2}$$

The model states $\mathbf{x}_k^{(i)-}$ of the current time step k, referred to as model predictions, are denoted with the superscript "−". In this work, $\mathbf{q}_{k-1}$ is neglected, i.e. no realisations of the model errors are generated, due to the difficulty in specifying the matrix $\mathbf{Q}_{k-1}$ (an alternative strategy to consider these errors is introduced in Sect. 3.4).

3.2 Filter update

3.2.1 Ensemble Kalman filter

In the EnKF, the error statistics of the model prediction are represented by the ensemble mean $\bar{\mathbf{x}}_k = \frac{1}{N_e} \sum_{i=1}^{N_e} \mathbf{x}_k^{(i)-}$ and the empirical error covariance matrix (e.g. Ripley 2006)

$$\mathbf{C}^e(\mathbf{x}_k^-) = \frac{1}{N_e - 1} \Delta\mathbf{X}_k^- (\Delta\mathbf{X}_k^-)^T \tag{3}$$

determined from the ensemble spread. Here, the matrix $\Delta\mathbf{X}_k^-$ stores the ensemble perturbations $\Delta\mathbf{x}_k^{(i)-} = \mathbf{x}_k^{(i)-} - \bar{\mathbf{x}}_k$ in its columns. We define $\Delta\mathbf{X}_k^- = \mathbf{X}_k^-\mathbf{W}$ with $\mathbf{X}_k^- = (\mathbf{x}_k^{(1)-}, \dots, \mathbf{x}_k^{(N_e)-})$ and the idempotent ($N_e \times N_e$)-projection

matrix $\mathbf{W}$ with elements equal to $1 - N_e^{-1}$ on its diagonal and $-N_e^{-1}$ as off-diagonal entries. Introducing $\mathbf{W}$ in the mentioned way, with rank ($N_e - 1$), results in the formulation of the model covariance matrix as

$$\mathbf{C}^e(\mathbf{x}_k^-) = \frac{1}{N_e - 1} \mathbf{X}_k^- \mathbf{W}(\mathbf{X}_k^-)^T. \tag{4}$$

In the update (or analysis) step of the standard EnKF (Evensen 1994), each model prediction sample $\mathbf{x}_k^{(i)-}$ is informed by a perturbed version $\mathbf{y}_k + \delta\mathbf{y}_k^{(i)}$ of the observation data. By introducing the perturbations $\delta\mathbf{y}_k^{(i)}$ the observation vector is treated as a random variable in a way to keep the update error covariance matrix within the ensemble unbiased. Burgers et al. (1998) showed that, when neglecting the perturbations, the variance of the updated ensemble is too low. The ensemble of EnKF updated states $\mathbf{X}_k^+ = (\mathbf{x}_k^{(1)+}, \dots, \mathbf{x}_k^{(N_e)+})$ is denoted with superscript "+" and obtained from

$$\mathbf{X}_k^+ = \mathbf{X}_k^- + \mathbf{K}_k((\mathbf{Y}_k + \Delta\mathbf{Y}_k) - \mathbf{A}\mathbf{X}_k^-), \tag{5}$$

with

$$\mathbf{K}_k = \mathbf{C}^e(\mathbf{x}_k^-)\mathbf{A}^T(\mathbf{A}\mathbf{C}^e(\mathbf{x}_k^-)\mathbf{A}^T + \mathbf{\Sigma}_{yy})^{-1}. \tag{6}$$

Herein, $\mathbf{Y}_k$ contains the observation vector $\mathbf{y}_k$ in each of its columns, while $\Delta\mathbf{Y}_k$ stores the realisations of the observation perturbations $\delta\mathbf{y}_k^{(i)}$. The difference between the measured (and perturbed) and the predicted observations $((\mathbf{Y}_k + \Delta\mathbf{Y}_k) - \mathbf{A}\mathbf{X}_k^-)$ is weighted and used to correct the predicted model ensemble $\mathbf{X}_k^-$. In Eq. (6), $\mathbf{A}$ is the design matrix that relates model states to observations. The gain matrix $\mathbf{K}_k$ weights the empirical ensemble covariance matrix of the model prediction $\mathbf{C}^e(\mathbf{x}_k^-)$ and the observation error covariance matrix $\mathbf{\Sigma}_{yy} = E(\delta\mathbf{y}_k\delta\mathbf{y}_k^T)$. From Eq. (6) it becomes obvious that the EnKF uses the same update equation as the Kalman filter (KF; Kalman 1960) but the ensemble representation $\mathbf{C}^e(\mathbf{x}_k^-)$ of the analytical positive definite model prediction covariance matrix $\mathbf{\Sigma}_{x^-x^-}$.

The update error covariance matrix $\mathbf{C}^e(\mathbf{x}_k^+)$ is given by

$$\mathbf{C}^e(\mathbf{x}_k^+) = (\mathbf{I} - \mathbf{K}_k\mathbf{A})\mathbf{C}^e(\mathbf{x}_k^-), \tag{7}$$

in which $\mathbf{I}$ denotes the identity matrix.

3.2.2 Square root analysis scheme for EnKF

The SQRA update (Evensen 2004, 2007) consists of two parts: (1) the update of the ensemble mean, and (2) the update of the ensemble perturbations. In contrast to the EnKF, the SQRA does not perform the update for each sample individually [Eq. (5)] but separately for the ensemble mean of the model predictions (e.g. Tippett et al. 2003)

$$\overline{\mathbf{x}_k^+} = \overline{\mathbf{x}_k^-} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{A}\overline{\mathbf{x}_k^-}) \tag{8}$$

and for the perturbations. Here, only the observation vector $\mathbf{y}_k$ is used for correcting the predicted ensemble mean $\overline{\mathbf{x}_k^-}$.

Yet, since an ensemble of updated model states $\mathbf{X}_k^+$ is needed for the next model forward integration, updating the model ensemble perturbations is required. In this paper, the simple and straightforward version of the SQRA introduced by Evensen (2004) was implemented. As we will show in the following derivation, generating perturbations [the $\Delta\mathbf{Y}_k$ in Eq. (5)] of the observations (as in the standard EnKF) is not required, mitigating another source of sampling errors (see also Whitaker and Hamill 2002).

Similarly as in Eq. (3), we now introduce the ensemble version of the error covariance matrix of the model update as $\mathbf{C}^e(\mathbf{x}_k^+) = \frac{\Delta\mathbf{X}_k^+(\Delta\mathbf{X}_k^+)^T}{N_e-1}$. Then, the ensemble versions of $\mathbf{C}^e(\mathbf{x}_k^-)$ [defined in Eq. (3)] and $\mathbf{C}^e(\mathbf{x}_k^+)$ are inserted in Eq. (7) to compute $\Delta\mathbf{X}_k^+$ depending on the ensemble perturbations of the predictions

$$
\begin{aligned}
&\Delta\mathbf{X}_k^+(\Delta\mathbf{X}_k^+)^T \\
&= \Delta\mathbf{X}_k^-(\mathbf{I} - (\Delta\mathbf{X}_k^-)^T\mathbf{A}^T(\mathbf{A}\Delta\mathbf{X}_k^-(\Delta\mathbf{X}_k^-)^T\mathbf{A}^T \\
&\quad + (N_e-1)\mathbf{\Sigma}_{yy})^{-1}\mathbf{A}\Delta\mathbf{X}_k^-)(\Delta\mathbf{X}_k^-)^T.
\end{aligned}
\tag{9}
$$

Eigenvalue decomposition is applied to $(\mathbf{A}\Delta\mathbf{X}_k^-(\Delta\mathbf{X}_k^-)^T\mathbf{A}^T + \mathbf{\Sigma}_{yy})^{-1} = \mathbf{Z}\mathbf{\Lambda}^{-1}\mathbf{Z}^T$, and Eq. (9) is then reorganised to

$$
\begin{aligned}
&\Delta\mathbf{X}_k^+(\Delta\mathbf{X}_k^+)^T \\
&= \Delta\mathbf{X}_k^-(\mathbf{I} - \underbrace{(\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{Z}^T\mathbf{A}\Delta\mathbf{X}_k^-)^T}_{\mathbf{D}^T}\underbrace{(\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{Z}^T\mathbf{A}\Delta\mathbf{X}_k^-)}_{\mathbf{D}})(\Delta\mathbf{X}_k^-)^T.
\end{aligned}
\tag{10}
$$

The singular value decomposition of $\mathbf{D} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ is inserted into Eq. (10)

$$
\begin{aligned}
\Delta\mathbf{X}_k^+(\Delta\mathbf{X}_k^+)^T &= \Delta\mathbf{X}_k^-(\mathbf{I} - (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T)^T(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T))(\Delta\mathbf{X}_k^-)^T \\
&= \Delta\mathbf{X}_k^-\mathbf{V}(\mathbf{I} - \mathbf{\Sigma}^T\mathbf{\Sigma})\mathbf{V}^T(\Delta\mathbf{X}_k^-)^T
\end{aligned}
\tag{11}
$$

Using the square root of the diagonal matrix $(\mathbf{I} - \mathbf{\Sigma}^T\mathbf{\Sigma})$, Eq. (11) becomes

$$\Delta\mathbf{X}_k^+(\Delta\mathbf{X}_k^+)^T = (\Delta\mathbf{X}_k^-\mathbf{V}\sqrt{\mathbf{I}-\mathbf{\Sigma}^T\mathbf{\Sigma}})(\Delta\mathbf{X}_k^-\mathbf{V}\sqrt{\mathbf{I}-\mathbf{\Sigma}^T\mathbf{\Sigma}})^T. \tag{12}$$

Equation (12) represents a symmetric expression that can be used to generate normally distributed perturbation vectors with zero mean and covariance matrix $\mathbf{C}^e(\mathbf{x}_k^+)$. Finally, the updated ensemble perturbations are added to the updated ensemble mean

$$\mathbf{X}_k^+ = \overline{\mathbf{X}_k^+} + \underbrace{\Delta\mathbf{X}_k^-\mathbf{V}\sqrt{\mathbf{I}-\mathbf{\Sigma}^T\mathbf{\Sigma}}}_{\Delta\mathbf{X}_k^+}\mathbf{\Theta}^T \tag{13}$$

In Eq. (13), $\mathbf{\Theta}^T$ represents a random orthonormal matrix, which contains the right-hand side eigenvectors of a matrix that holds uniformly distributed random numbers. By multiplying $\Delta\mathbf{X}_k^+$ with $\mathbf{\Theta}^T$, realisations of ensemble perturbations are generated from the update error covariance matrix $\mathbf{C}^e(\mathbf{x}_k^+)$ by Monte Carlo sampling (e.g. Kusche 2003). A detailed derivation of the algorithm and a comparison to the standard EnKF can be found in Evensen (2004, 2007).

3.2.3 Singular evolutive interpolated Kalman filter

In the SEIK filter (Pham et al. 1998), the ensemble representation of the model prediction error covariance matrix is given in form of

$$\mathbf{C}_{\text{SEIK}}^e(\mathbf{x}_k^-) = \mathbf{L}_k^e\mathbf{G}^e\mathbf{L}_k^{e^T}, \tag{14}$$

where the matrix $\mathbf{L}_k^e = \mathbf{X}_k^-\mathbf{T}$ is of dimension $m \times (N_e - 1)$, m is the number of entries in the model prediction vectors $\mathbf{x}_k^{(i)-}$, and N_e is the ensemble size. Here, $\mathbf{T}$ is a full rank matrix with zero column sums, which consists of the first $(N_e - 1)$ columns of the matrix $\mathbf{W}$ in Eq. (4): $\mathbf{W} = [\mathbf{T}|\mathbf{t}]$ with $\mathbf{t}$ representing the last column of $\mathbf{W}$. $\mathbf{G}^e = \frac{1}{N_e}(\mathbf{T}^T\mathbf{T})^{-1}$ is normalised by the ensemble size N_e. Using Eq. (14), the model prediction errors are represented in the space that is spanned by the columns of $\mathbf{L}_k^e$.

As for the EnKF, the formulation of the SEIK filter update can be derived from the KF equations. Here, however, we replace the model prediction error covariance matrix in Eq. (6) by the ensemble representation defined in Eq. (14)

$$\mathbf{K} = \mathbf{L}_k^e\mathbf{G}^e\mathbf{L}_k^{e^T}\mathbf{A}^T(\mathbf{A}\mathbf{L}_k^e\mathbf{G}^e\mathbf{L}_k^{e^T}\mathbf{A}^T + \mathbf{\Sigma}_{yy})^{-1}. \tag{15}$$

By applying the matrix identity $\mathbf{QW}(\mathbf{Z} + \mathbf{VQW})^{-1} = (\mathbf{Q}^{-1} + \mathbf{WZ}^{-1}\mathbf{V})^{-1}\mathbf{WZ}^{-1}$ (Koch 1997, p. 37, Eq. (134.7)) for invertible matrices $\mathbf{Q}$ and $\mathbf{Z}$ and arbitrary matrices $\mathbf{V}$ and $\mathbf{W}$ to Eq. (15), the formulation of the gain matrix becomes

$$\mathbf{K}_k = \mathbf{L}_k^e\underbrace{[(\mathbf{G}^e)^{-1} + \mathbf{L}_k^{e^T}\mathbf{A}^T\mathbf{\Sigma}_{yy}^{-1}\mathbf{A}\mathbf{L}_k^e]^{-1}}_{N_e \times N_e}\mathbf{L}_k^{e^T}\mathbf{A}^T\mathbf{\Sigma}_{yy}^{-1}. \tag{16}$$

This is the SEIK ensemble formulation implemented in our study. Here, the observation error covariance matrix $\mathbf{\Sigma}_{yy}$ is transformed to the ensemble space by applying $\mathbf{A}\mathbf{L}_k^e$ to $\mathbf{\Sigma}_{yy}^{-1}$. It becomes obvious that the size of the matrix to be inverted depends on the model ensemble size N_e. The update is performed in the ensemble space, and if the number of observations is much larger than the ensemble size, the application of SEIK is efficient. We would like to stress that the

formulation of the Kalman gain matrix based on the EnKF ensemble representation $\mathbf{C}^e(\mathbf{x}_k^-)$ in Eq. (3) and on the SEIK ensemble representation $\mathbf{C}_{\text{SEIK}}^e(\mathbf{x}_k^-)$ in Eq. (14) of the model prediction error covariance matrix is only identical during the first update (identical model configuration and initial state estimate and covariance matrix implied). However, the EnKF and SEIK updated model state vectors differ from each other, since the EnKF relies on an observation ensemble but the SEIK considers an update of the ensemble mean of the model prediction vector similar to the SQRA method. Therefore, the sequence of updates will numerically differ in both approaches. However, in the limit $N_e \rightarrow \infty$, assuming ergodicity, the two ensemble representations fall back to the conventional Kalman filter and thus would lead to identical data assimilation results. By defining

$$\mathbf{U}_k = ((\mathbf{G}^e)^{-1} + (\mathbf{AL}_k^e)^T \mathbf{\Sigma}_{yy}^{-1} \mathbf{AL}_k^e)^{-1} \tag{17}$$

and $\mathbf{a}_k = \mathbf{U}_k (\mathbf{AL}_k^e)^T \mathbf{\Sigma}_{yy}^{-1} (\mathbf{y}_k - \mathbf{A}\overline{\mathbf{x}_k^-})$, and inserting these and Eq. (16) into Eq. (8), the formulation of the model update is finally converted to the common notation of the SEIK filter

$$\overline{\mathbf{x}_k^+} = \overline{\mathbf{x}_k^-} + \mathbf{L}_k^e \mathbf{a}_k. \tag{18}$$

Basically, one projects the errors of the updated states onto the space spanned by the columns of $\mathbf{L}_k^e$, which results in the formulation of the model update covariance matrix $\mathbf{C}^e(\mathbf{x}_k^+)$ as

$$\mathbf{C}^e(\mathbf{x}_k^+) = \mathbf{L}_k^e \mathbf{U}_k \mathbf{L}_k^{e^T}. \tag{19}$$

A detailed derivation of Eq. (19) can be found in Pham et al. (1998).

Finally, the update of the ensemble perturbations is performed. To this end, the minimum second-order exact sampling is used (Pham et al. 1998, Appendix, pp. 17–21). Ensemble perturbations are generated from the eigenvalue-decomposed error covariance matrix of the filter update. The ensemble mean and the ensemble covariance matrix need to match exactly the updated ensemble mean $\overline{\mathbf{x}_k^+}$ and the updated error covariance matrix $\mathbf{C}(\mathbf{x}_k^+)$

$$\frac{1}{N_e} \sum_{i=1}^{N_e} \mathbf{x}_k^{(i)} = \overline{\mathbf{x}_k} \equiv \overline{\mathbf{x}_k^+}, \tag{20}$$

$$\mathbf{L}_0 \mathbf{C}_0^T \mathbf{\Omega}_0^T \mathbf{\Omega}_0 \mathbf{C}_0 \mathbf{L}_0^T = \mathbf{S}_0 \equiv \mathbf{C}(\mathbf{x}_k^+). \tag{21}$$

This is realised by determining a low $(N_e - 1)$-rank approximation of the covariance matrix, using the leading eigenvalues and eigenvectors (or dominant orthogonal modes) of the ensemble update error covariance matrix $\mathbf{C}^e(\mathbf{x}_k^+)$, whose eigenvectors and eigenvalues are stored in $\mathbf{L}_0$

and $\mathbf{U}_0 = \mathbf{C}_0^T \mathbf{C}_0$, respectively. In Eq. (21), $\mathbf{\Omega}_0$ is an orthonormal matrix. Its columns are orthogonal to a vector that contains only ones. This matrix can, for example, be determined by Householder transformation (Hoteit et al. 2002, Appendix, pp. 125–126). The update ensemble $\mathbf{X}_k^+$ is determined by adding the generated perturbations to the updated ensemble mean, which is stored in each column of $\overline{\mathbf{X}_k^+}$:

$$\mathbf{X}_k^+ = \overline{\mathbf{X}_k^+} + \sqrt{N_e} \mathbf{L}_0 \mathbf{C}_0^T \mathbf{\Omega}_0^T. \tag{22}$$

A comparison of the standard EnKF and SEIK filter can also be found e.g. in Nerger (2003).

3.3 Parameter estimation

In hydrological modeling it is common to calibrate basin-wide empirical model parameters that are usually assumed to be temporally constant. Some of these parameters describe physio-geographic characteristics, e.g. average lake depth, while other parameters appear as conceptual such as the groundwater outflow coefficient in WGHM (Döll et al. 2003). In data assimilation, the model ensemble prediction vector is augmented by model parameters for a simultaneous calibration in the EnKF analysis step. Therefore, in our approach, the prediction vector $\mathbf{x}_k^-$ is composed of two parts

$$\mathbf{x}_k^- = \begin{pmatrix} \mathbf{v}_k^- \\ \mathbf{w}_k^- \end{pmatrix}, \tag{23}$$

in which $\mathbf{v}_k^-$ contains the model state values and $\mathbf{w}_k^-$ comprises the model calibration parameters. The latter cannot be observed, and they are, therefore, updated via the cross-correlations of model states and parameters. In contrast to model calibration as common in hydrology, the parameters are updated as soon as observations become available and, therefore, their values change over time. Schumacher et al. (2015), for instance, showed how GRACE observations contribute in calibrating WGHM parameters. This is effective whenever large correlations exist between model states and parameters.

3.4 Tuning techniques: inflation

Estimation of the emprical model covariance matrix $\mathbf{C}^e(\mathbf{x}_k^-)$ might be too optimistic when neglecting errors in the model structure [$\mathbf{q}_{k-1}$ in Eq. (1)]. In the absence of reliable information about these errors, alternative strategies to enlarge the ensemble spread have been developed: Hamill and Snyder (2002) introduced the so-called inflation factor. Here, the ensemble perturbations are multiplied by a constant inflation factor m_c

$$\mathbf{X}_k^{'-} = m_c(\mathbf{X}_k^- - \overline{\mathbf{X}_k^-}) + \overline{\mathbf{X}_k^-}, \tag{24}$$

prior to the introduction of the predicted model states into the standard EnKF or SQRA. As a result, $\mathbf{X}_k^{'-}$ appears as the predicted ensemble with increased perturbations. The factor helps avoiding fast ensemble convergence due to the reduction of the variances with each filter update, i.e. it preserves the ensemble spread. In the SEIK filter, the inverse matrix $\mathbf{G}^{-1}$ in Eq. (17) is replaced by $\frac{1}{m_c}\mathbf{G}^{-1}$, where $\frac{1}{m_c}$ is denoted as forgetting factor in Pham et al. (1998).

3.5 Measurement and mapping operator

To merge WGHM outputs with GRACE TWSA, the design matrix is split according to $\mathbf{A} = \mathbf{BH}$, which includes a vertical aggregation operator $\mathbf{H}$ and a horizontal mapping operator $\mathbf{B}$ (Fig. 2). The vertical sum of all modelled storage compartments is determined for each grid cell by incorporating $\mathbf{H}$. Due to the coarser spatial resolution of GRACE data, TWSA are spatially averaged through $\mathbf{B}$. Thus, the design matrix $\mathbf{A}$ in Eqs. (5–7) (EnKF), Eq. (8) (SQRA), and Eqs. (16–18)

(SEIK) is replaced by the product of the measurement $\mathbf{H}$ and the mapping operator $\mathbf{B}$ (see also Eicker et al. 2014).

4 Twin experiment set-up

A synthetic experiment was designed to study the impact of GRACE error correlations on the C/DA results when merging water state outputs and parameters of WGHM with GRACE TWSA. Our twin experiment started with the definition of "true" hydrological water states. These serve as the basis to assess the C/DA results. In addition, GRACE-like errors, to be added to TWSA observations, were generated as described in Sect. 4.3. An imperfect representation of the truth was realised by replacing the forcing, parameters and initial water states in the model simulation. Errors of the model simulation were represented by an ensemble of N_e randomly perturbed precipitation and temperature input fields, calibration parameters and initial water states. Open loop (OL) simulations were performed without integrating GRACE TWSA observations and compared to model simulation after the C/DA process. An overview of the twin experiment set-up is given in Fig. 3. The details of the procedure are described in this section.

Fig. 2 Operators that allow for combination of the model-derived storage compartments and GRACE TWSA. **a** Observations of a vertical sum: the measurement operator $\mathbf{H}$ adds the vertical layers, i.e. the compartmental water storage values, together to compute TWSA. **b** Horizontally aggregated measurements: mapping operator $\mathbf{B}$ determines spatial averages (e.g. TWSA)
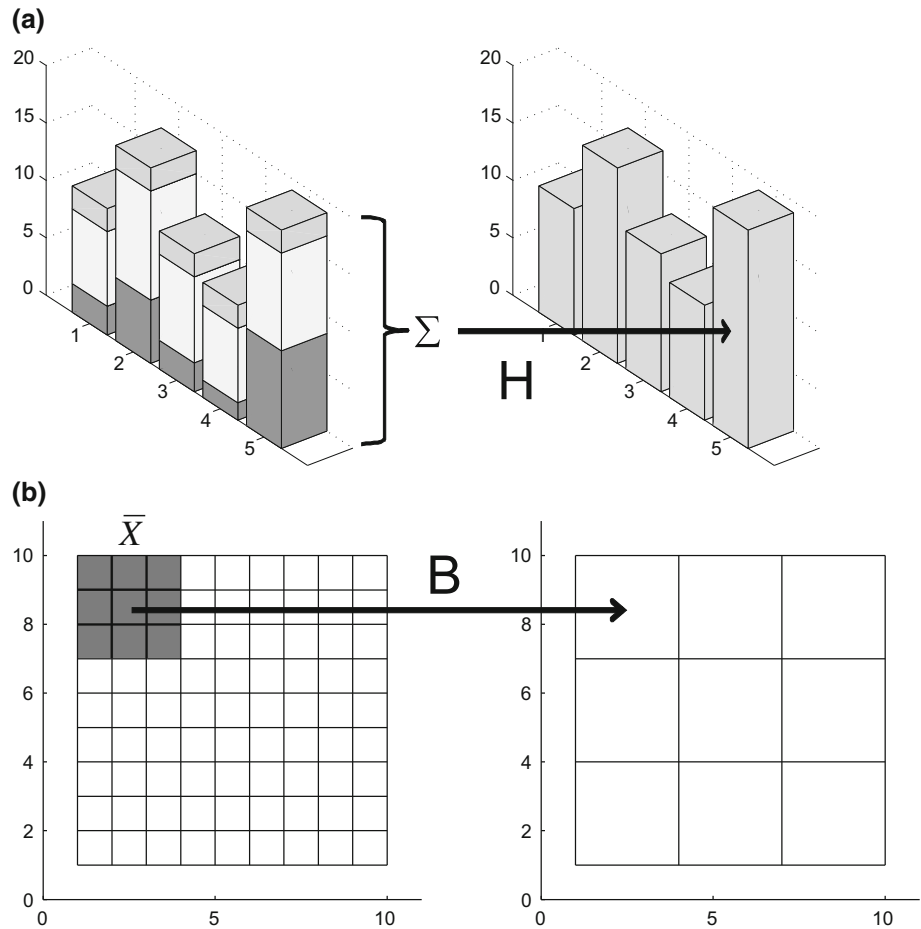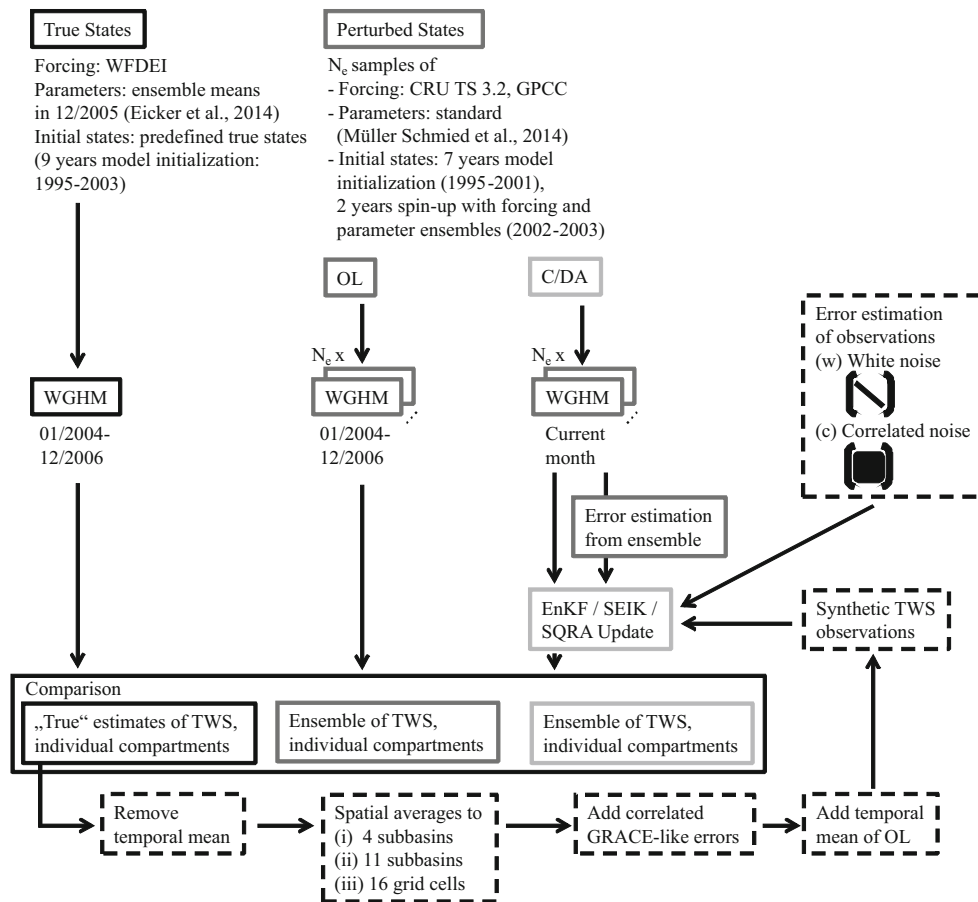
Fig. 3 Twin experiment set-up: definition of true and perturbed model states (*first row*). Model prediction in open loop (OL) mode (*second column*), i.e. without integrating GRACE data, and in calibration and data assimilation (C/DA) mode (*third column*). Generation of synthetic GRACE-like observations (*last row*). OL and all C/DA variants are compared to the true states. The performance of C/DA variants is analysed compared to the OL performance, and compared to each other

4.1 Study area

The Mississippi River Basin is located in the eastern part of the United States of America. It covers large parts of the High Plains aquifer (HPA), where groundwater is abstracted for irrigation purposes resulting in groundwater depletion (e.g. Rodell et al. 2007; Strassberg et al. 2009; Döll et al. 2012, 2014). In order to study the impact of different spatial discretisation of TWSA observations from GRACE on the C/DA results, the entire basin of the size of 2.9×10^6 km^2 was divided into (i) four sub-basins (similar to Zaitchik et al. 2008), (ii) 11 sub-basins and (iii) sixteen $5° \times 5°$ grid cells (similar to Eicker et al. 2014), with areas varying between 50,000 and 1.17×10^6 km^2 (for details see Fig. 4; Table 1).

4.2 Synthetic true and perturbed model states

For defining "true" hydrological states, WGHM was driven by daily time series from the WFDEI meteorological data

set (Fig. 3). The applied model parameters were calibrated values derived from the first C/DA of the Mississippi Basin by Eicker et al. (2014), i.e. the ensemble means in December 2005. Since model parameters and climate input data are the major sources of uncertainties in hydrological modelling, the perturbed model, into which we will assimilate GRACE data, used the monthly time series from CRU TS 3.2 and GPCC as climate forcing fields and the model parameters reported in Döll et al. (2003), Kaspar (2004) and Hunger and Döll (2008). Both model versions were initialised over a period of nine years (1995–2003). The annual amplitudes of the perturbed model water storage in snow, soil, river and groundwater were larger than the true water storages as can be seen in Fig. 5 for our three-year study period (2004–2006).

4.3 Synthetic TWSA observations

The generation of synthetic GRACE-like TWSA observations involved three steps: (1) $0.5° \times 0.5°$ gridded monthly

Fig. 4 Sub-basins within the Mississippi River Basin. The four sub-basin definition is chosen similar to Zaitchik et al. (2008) and is shown with *different colors*. Eleven sub-basins are shown with the *thick grey polygons* and numbered for identification in the results section (Sect. 5). The grid definition is chosen similar to Eicker et al. (2014) and is shown using the *thin black lines*. Names and areas of the basins can be found in Table 1. The *orange dots* indicate the extent of the High Plains aquifer (HPA)
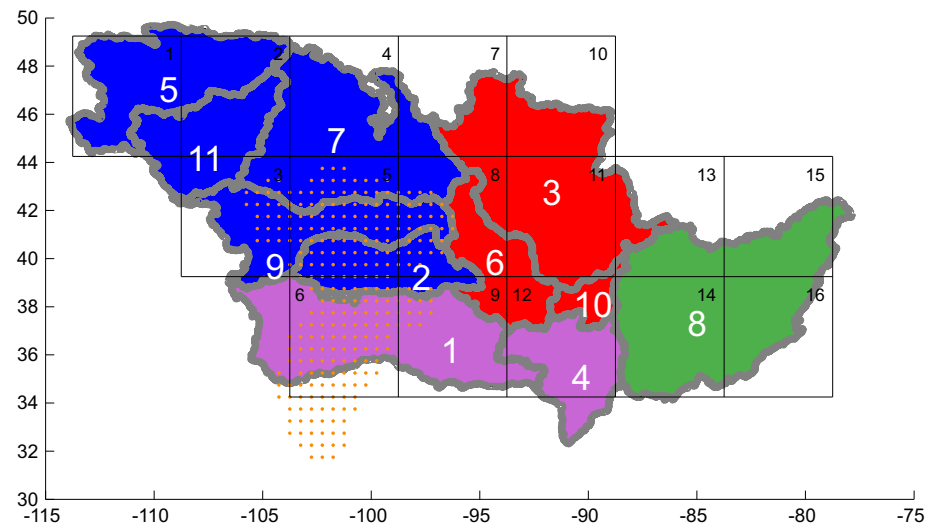
Table 1 Subbasins defined within the Mississippi River Basin. Area, standard deviation of TWSA observations (Std), and signal-to-noise-ratio (SNR, i.e. ratio of annual amplitude and error standard deviation) are reported for each subbasin of Fig. 4. Standard deviations are estimated after full error propagation of GRACE level-2 products while considering the square root of the main diagonal elements of the full error covariance matrix of August 2003. Numbers and colours for identification are given according to Fig. 4

Basin	Area (10^6 km^2)	Std (mm)	SNR
4 subbasins			
1—Missouri (blue)	1.17	9	2.4
2—Ohio Tennesse/Metropolis (green)	0.53	14	2.0
3—Red Arkansas/Lower Mississippi (purple)	0.57	9	3.6
4—Upper Mississippi (red)	0.66	17	1.2
11 subbasins			
1—Arkansas Little Rock	0.41	12	1.2
2—Kansas River Desoto	0.16	18	0.8
3—Mississippi Grafton	0.45	18	1.2
4—Mississippi Vicksburg	0.16	24	3.1
5—Missouri Culbertson	0.24	16	1.4
6—Missouri Hermann	0.16	25	0.6
7—Missouri Sioux City	0.37	17	1.3
8—Ohio Metropolis	0.53	14	2.0
9—Platte River Louisville	0.22	14	1.8
10—Thebel	0.05	27	1.3
11—Missouri Yellowstone	0.18	22	1.1
5 °× 5° grid (16 observations)	0.11–0.23	22–25	0.5–2.2

means of TWS outputs of the true model were reduced by their temporal mean over the C/DA period from 2004 to 2006. These values were then spatially averaged to 4 and 11 sub-basin means, and sixteen 5 °× 5° grid cells, where the boundaries were taken from Fig. 4. (2) Spatially correlated errors of TWSA were generated by error propagation of the full ITG-GRACE2010 error covariance matrix (see Sect. 2.2) in August 2003. In this study, we assumed a time-constant observation error covariance matrix. The generated correlated errors were added to the TWSA time series derived in step 1 (Fig. 6). In the EnKF update, either the analytical TWSA error covariance matrix was used or a diagonal error covariance matrix considering the main diagonal ele-

ments from the analytical TWSA error covariance matrix. (3) For merging TWSA from the perturbed model states (from Sect. 4.2) and the synthetic observations (derived in step 1 and 2), they need to have the same temporal mean. Therefore, the temporal means of the OL simulations (described in Sect. 4.4.1) were added to the synthetic TWSA. As a result, corresponding to the number of sub-basins, the observation vector $\mathbf{y}_k$ in Eqs. (5), (8) and (18) included four, 11 or 16 sub-basin/grid cell averaged TWSA values. Standard deviations of the generated observations (Fig. 7 shows 11 sub-basin means, black dots) and the signal-to-noise ratios (SNR) are reported in Table 1. In Fig. 6 the correlations ρ between the GRACE TWSA errors are shown for (a) four, (b) 11 or (c) 16
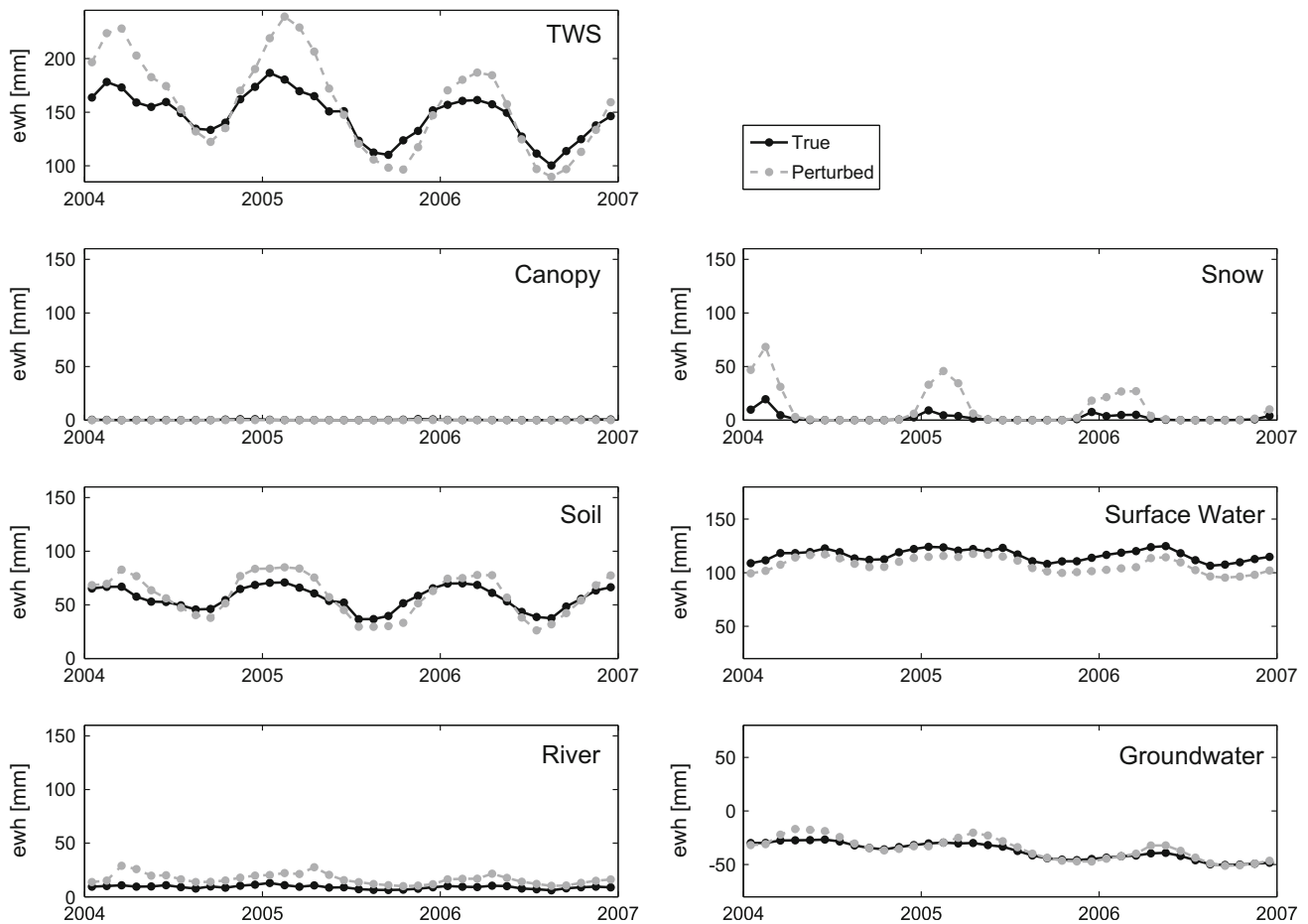
Fig. 5 Monthly time series of simulated true and perturbed total water storage (TWS) and individual water storages, averaged over the whole Mississippi Basin, in millimeter of equivalent water heights (ewh)
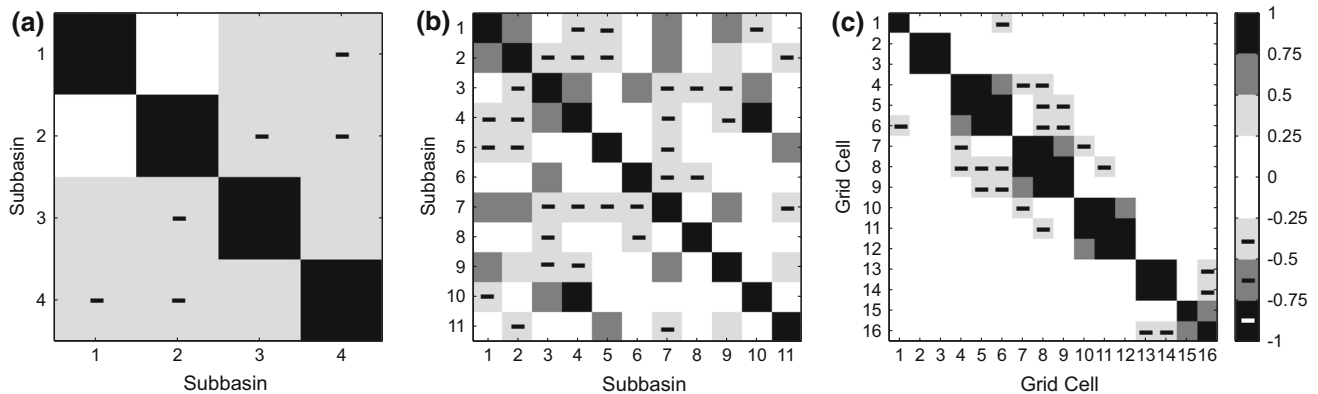

Fig. 6 Correlations between the GRACE TWSA errors after aggregating to **a** four and **b** 11 sub-basins, as well as **c** 16 grid cells. Numbers for identification are given according to Table 1 and Fig. 4. Here, the full error covariance matrix of the potential coefficients from the ITG-GRACE2010 solution of Bonn University in August 2003 was used for error propagation and generating correlated TWSA errors

observations. In case (a) modest correlations between TWSA errors in almost all sub-basins exist, reaching −0.5 between errors in sub-basin 1 and 4. When using 11 observations $|\rho| > 0.25$ in half of the cases. The highest correlation of almost 0.9 appears between the errors in sub-basin 4 and 10. In case (c) positive correlations >0.5 exist between errors of TWSA in sub-basins that are located in north–south direction to each other, i.e. in grid cells located in one column of
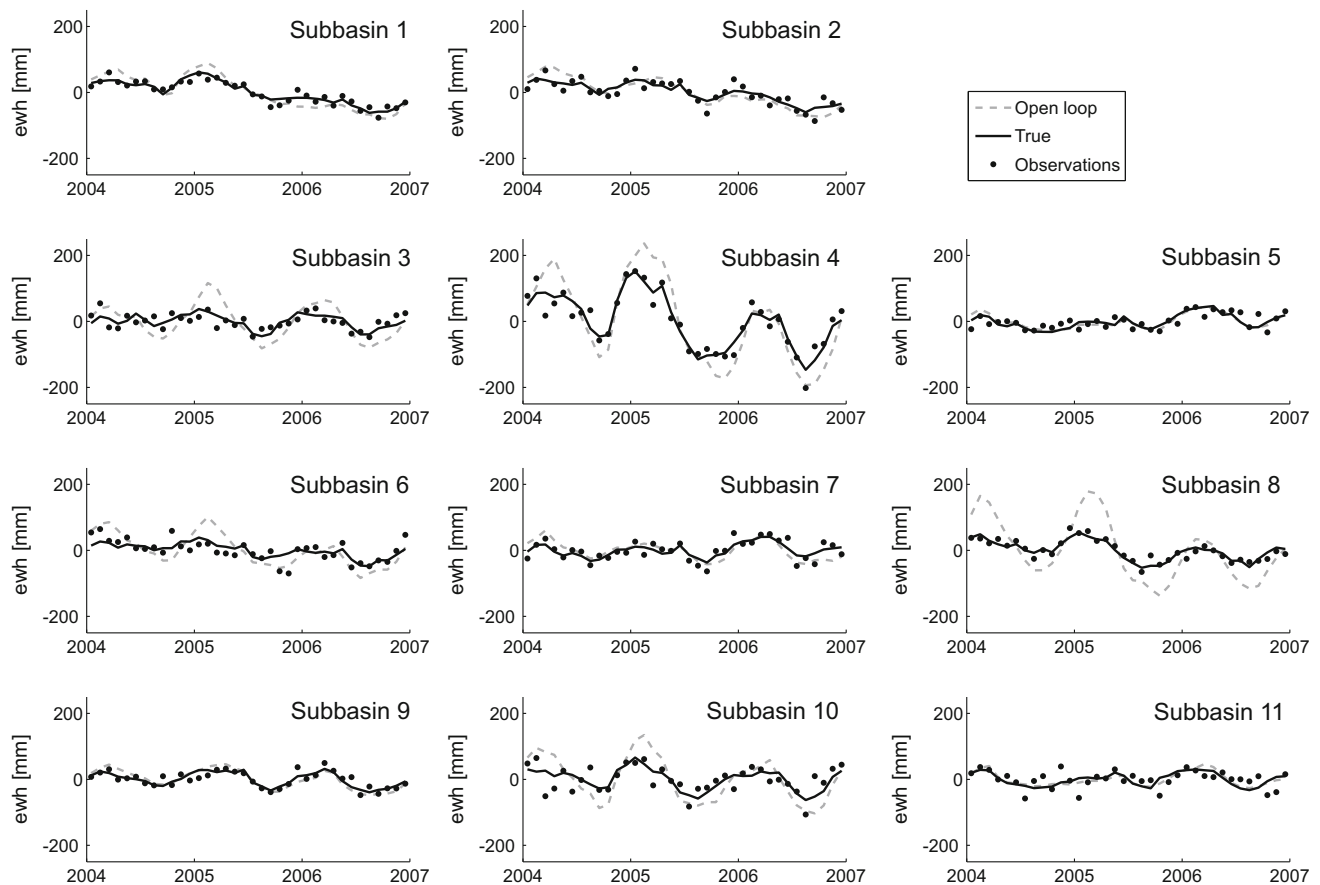
Fig. 7 Monthly TWSA time series of the ensemble mean of open loop (OL) simulations, true model and GRACE-like observations, spatially averaged over the 11 sub-basins of the Mississippi Basin

the grid in Fig. 4 (e.g. grid cells 10, 11 and 12). Errors of TWSA between grid cells located in neighbouring columns of the grid in Fig. 4 are mostly negatively correlated (up to −0.4) or have small positive correlations (<0.25). The subbasin/grid cell size influences the number of grid cells with error correlations, as well as the magnitude of correlations, which increases with increasing spatial resolution.

4.4 EnKF design

4.4.1 Ensemble of model states

An ensemble size of 30 samples was defined as a trade-off between computational costs, storage capacity and representative error statistics, and in accordance with previous GRACE data assimilation studies in hydrology [from five ensemble members in Van Dijk et al. (2014) to 25 in Su et al. (2010) and 30 in Eicker et al. (2014)]. To generate the initial model ensemble, 20 calibration parameters were sampled using the Latin-Hypercube method (Iman 2008), with a priori probability density functions as listed in Table 2. To account for uncertainties in climate forcing, precipita-

tion and temperature fields were perturbed using random Monte Carlo sampling from triangular probability density functions. An additive error model was assumed for temperature, centered at 0 °C with the maximum limits of ±2 °C, and a multiplicative error model was introduced for precipitation, centred at 1.0 with the maximum limits of 0.7 and 1.3. In fact we found that using an ensemble of perturbed precipitation grids did not result in a multiplicative (area-average) bias in monthly fields. This justifies that this spatial precipitation error model may be considered as independent of the error model implicitly realised through perturbing the area-average WGHM precipitation multiplier defined in Table 2 (otherwise, our ensemble-based representation of the area-average precipitation uncertainty would be misspecified too low). For generating an ensemble of initial water states, the model initialisation phase was shortened to seven years and a spin-up phase of two years (2002–2003) was performed with the parameter and climate input ensembles. The water storage outputs for canopy, snow, soil, local and global wetland, local and global lake, reservoir, river and groundwater were introduced as inital values at the beginning of the C/DA phase. It is worth mentioning that for implementing the SEIK filter

Table 2 WGHM parameters that are calibrated within the ensemble filter variants with identification number (IN), true value according to Eicker et al. (2014), as well as value that is used in WaterGAP version 2.2 (mode) and limits (Döll et al. 2003; Kaspar 2004; Hunger and Döll 2008) used for ensemble generation. To generate ensembles of the parameters, either triangular or uniform distributions were assumed, indicated by $\triangle$ and $\circ$ in the first column, respectively. Units of parameters are given in the second column

IN	Calibration parameter	True	Mode	Limits
1$\triangle$	Root depth multiplier	0.68	1	[0.5 2.0]
2$\triangle$	River roughness coefficient multiplier	0.49	1	[0.5 2.0]
3$\triangle$	Surface water outflow coefficient (day^{-1})	0.09	0.01	[0.001 0.1]
4$\triangle$	Net radiation multiplier	0.52	1	[0.5 2.0]
5$\triangle$	Priestley–Taylor coefficient (humid)	1.03	1.26	[0.885 1.65]
6$\triangle$	Priestley–Taylor coefficient (arid)	2.06	1.74	[1.365 2.115]
7$\triangle$	Max. daily potential evapotranspiration (mm/day)	7.31	15	[7.25 22.5]
8$\triangle$	Max. canopy water height per leaf area (mm)	0.81	0.3	[0.1 1.4]
9$\triangle$	Specific leaf area multiplier	0.69	1	[0.5 2.0]
10$\triangle$	Snow freeze temperature (°C)	−0.51	0	[−1.0 3.0]
11$\triangle$	Snow melt temperature (°C)	−0.30	0	[−3.75 3.75]
12$\triangle$	Degree day factor multiplier	1.88	1	[0.5 2.0]
13$\triangle$	Temperature gradient (°C/m)	0.01	0.006	[0.004 0.01]
14$\triangle$	Groundwater factor multiplier	0.47	1	[0.5 2.0]
15$\triangle$	Max. groundwater recharge multiplier	1.04	1	[0.5 2.0]
16$\circ$	Critical precipitation for groundwater Recharge (mm/day)	20.05	10	[2.5 20.0]
17$\triangle$	Groundwater outflow coefficient (day^{-1})	0.007	0.006	[0.006 0.018]
18$\triangle$	Net abstraction surfacewater multiplier	0.49	1	[0.5 2.0]
19$\triangle$	Net abstraction groundwater multiplier	0.99	1	[0.5 2.0]
20$\triangle$	Precipitation multiplier	0.80	1	[0.8 1.2]

the minimum second-order exact sampling is widely used to generate intial water states. However, to focus on the effect of spatially correlated observation error information on the C/DA results, here the initial states were kept identical for all implemented filter variants.

OL simulations, i.e. model runs without introducing TWSA observations, were performed for 2004 to 2006 for each of the initial model ensemble members. The ensemble mean of the OL is shown in Fig. 7 (grey curves), and this was used for comparison with the C/DA simulations, where synthetic GRACE-like TWSA observations were assimilated (black dots). The OL simulations resulted in large annual amplitudes of TWS in sub-basin 3, 4, 8 and 10, which especially in sub-basin 8 overestimated the "observed" annual amplitude. Sub-basins located in the HPA (1, 2, and 9) exhibited negative trends in TWS, caused by the negative trend in groundwater storage. The amplitude of annual TWS changes was found similar to the observations for these sub-basins, as well as for the sub-basins 5, 6, 7 and 11. However, in sub-basin 6, the OL TWS changes overestimate the true TWS changes.

The model prediction vector [see Eq. (2)] in this study is composed of the model outputs of monthly means of water states in the ten individual water compartments for each of the

1262 grid cells in the Mississippi Basin and the 20 WGHM calibration parameters

$$\mathbf{x}_k^{(i)-} = \begin{pmatrix} \text{storage compartments in cell } 1^{(i)} \\ \vdots \\ \text{storage compartments in cell } 1262^{(i)} \\ \text{WGHM calibration parameters}^{(i)} \end{pmatrix}. \quad (25)$$

This resulted in $1262 \times 10 + 20$ entries of $\mathbf{x}_k^{(i)-}$, with 10 being the number of the storage compartments, for each of the $i = 1, \ldots, 30 = N_e$ model ensemble members that were merged with the synthetic TWSA observations.

4.4.2 EnKF variants

For our investigations, a range of design options were defined: (i) diagonal or full GRACE observation error covariance matrices, (ii) spatial aggregation of the observations to four, 11 or 16 sub-basin/grid cell averages and (iii) EnKF, SQRA or SEIK as filter algorithm. Additionally, an inflation factor of 10 % was used for representing errors in model structure to mitigate ensemble convergence. This factor was chosen as small as possible as to avoid a strong influence on

Table 3 Calibration and data assimilation (C/DA) variants used in this study. For each case, 30 samples and an inflation factor of 10 % were used

Filter method	Discretisation	Variant name	
		White noise	Correlated noise
EnKF	4 subbasins	4 w	4 c
EnKF	11 subbasins	11 w	11 c
EnKF	16 grid cells	16 w	16 c
SQRA	11 subbasins	Sq w	Sq c
SEIK	11 subbasins	Se w	Se c

the model ensemble, and large enough to ensure that a contribution of the GRACE observations to the model update is guaranteed over the entire study period. For each of the EnKF variants the full error covariance matrix of the model was considered. An overview of the variants used in this study is given in Table 3.

4.5 Validation of results

To validate our results, we determined the ensemble mean estimates of monthly water storage values for each $0.5° \times 0.5°$ grid cell and aggregated them to 11 sub-basin means (see Figs. 4, 7). Water storage changes in local and global lake, local and global wetland, as well as global reservoir were accumulated and defined as surface water storage changes. River storage was evaluated separately. Several metrics were determined for assessing TWSA and anomalies of water storage in snow, soil, surface water, river and groundwater of the OL model run, and the C/DA variants for each of the sub-basins in comparison to the simulated truth (Fig. 5): (1) root mean square error (RMSE); (2) correlation between residual curves after subtracting a linear trend, as well as annual and semi-annual cycles; (3) ratio of the annual amplitudes reduced by 1 (i.e. zero represents equal amplitudes); (4) introduced or removed water mass (sum of filter update increments over the C/DA period); and (5) absolute value of water mass change in the model (sum of absolute values of filter update increments over the C/DA period). The metrics (1)–(3) show the agreement of the C/DA results with the truth, while metrics (4) and (5) describe the degree of violation of mass conservation due to assimilated TWSA. The first three months were defined as run-in period of the filter and, therefore, the metrics were determined with respect to the period from April 2004 to December 2006.

5 Results and discussion

This section starts with quantifying the impact of implementing only the diagonal (white noise) or the full observation

error covariance matrix (correlated errors) in the filter update step on the C/DA results using the standard EnKF approach; in other words, we investigate whether the GRACE spatial error correlations may be neglected. This is then compared with the results after application of the SQRA and the SEIK algorithms. The section is concluded with a discussion of the calibrated parameters.

5.1 Does the observation error model influence the C/DA results?

First, the results for sub-basin 8 (the largest of the 11 sub-basins, see Fig. 4) are presented, for which the modelled (OL) annual amplitude of TWSA overestimates the true one. Correlations between GRACE TWSA errors of up to -0.5 exist when assimilating four sub-basin-averaged observations, almost 0.9 in case of 11 sub-basin averages, and exceeds 0.9 in case of gridded observations (Fig. 6). The five metrics (RMSE, correlation between residual curves, ratio of amplitudes, mass change and absolute mass change) are shown in the columns in Fig. 8 with respect to the synthetic truth. Metrics associated with TWSA are shown along the top row, while the following rows correspond to the individual water compartment changes (snow, soil, surface water, river and groundwater). Each individual subplot contains the results from OL (shown in grey) and C/DA indicating the discretisation level of assimilated TWSA observations. White bars correspond to white observation noise introduced in the EnKF update step (additionally indicated by "w"), while black bars indicate results from considering correlated observation errors (indicated by "c"). For clarity, we repeat here that the synthetic GRACE observations have been simulated by adding correlated noise in all cases. All assimilated variants outperform OL regarding the ratio of amplitude for all compartments. Regarding RMSE and correlation, this is not the case for the surface water and groundwater compartment. In addition, correlation in soil is not generally higher than OL. While integrating GRACE data into the model guarantees an improved simulation of TWSA, this is not true for individual compartments. Insufficiently resolved or numerically introduced correlations between the individual storages, as reflected in the error covariance matrix of the model (that is rank deficient and shows large condition numbers), might result in a deterioration of individual water compartment estimates.

We focus on the first three columns on the top row in Fig. 8 and on just the assimilation of TWSA observations aggregated to 16 grid cells, while considering correlated errors (the right-most bars labelled with 16 c). The introduction of TWSA into WGHM considerably reduced the RMSE (from about 62 to 20 mm) and the ratio of amplitudes (from 3.5 to 1.5). It also improved the correlation of the residuals (from 0.6 to 0.9). These improvements were also achieved for the
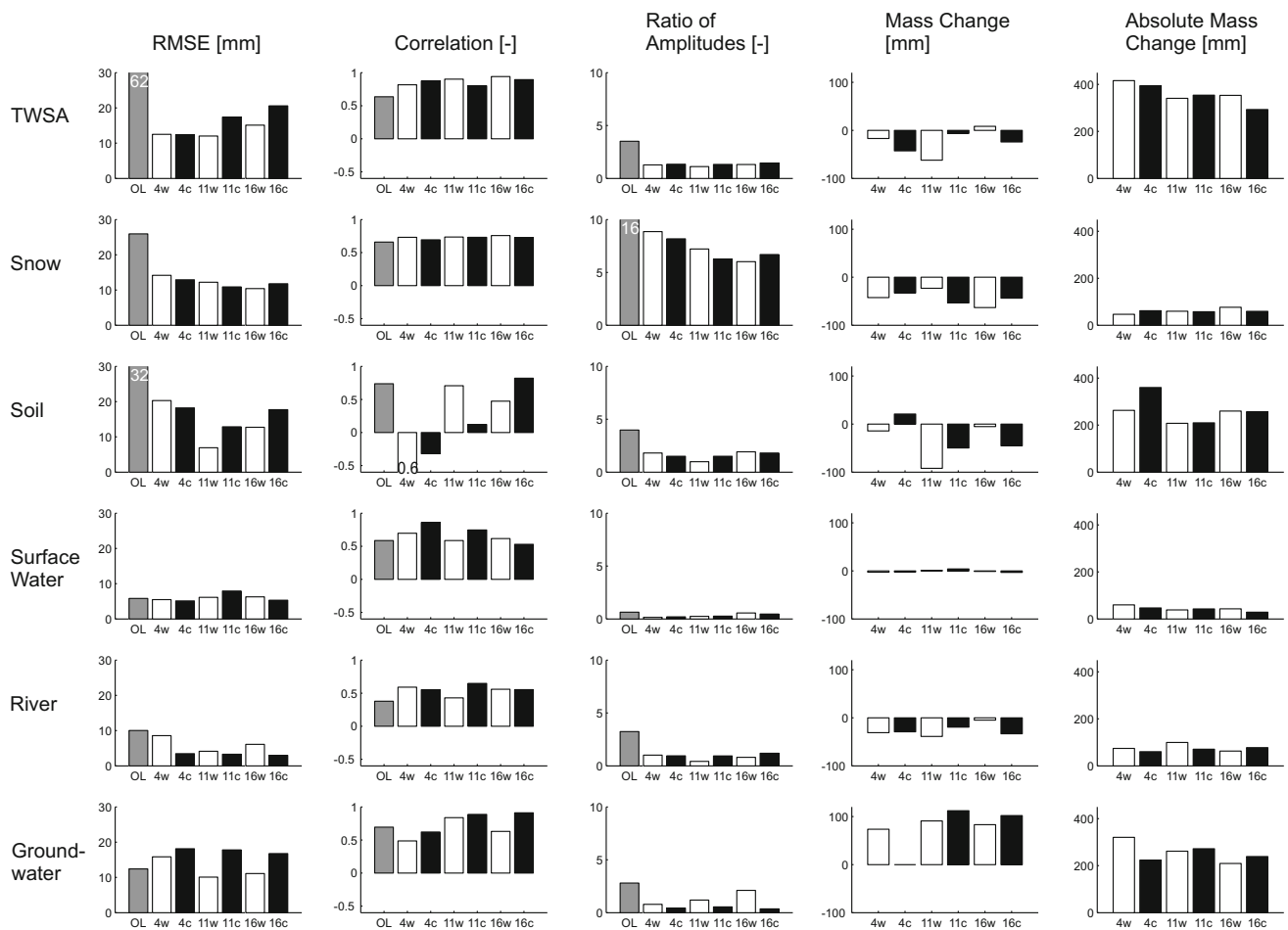
Fig. 8 Metrics for area mean of sub-basin 8 (see Fig. 4) for open loop run (OL) and calibration and data assimilation variants (names can be found in Table 3). Please note that the ratio of amplitudes is reduced by one, so that zero represents equal amplitudes. Some *bars* are truncated to fit the shown range. For these, the metric value is displayed at the *top* (or *bottom*) of the *bar*

individual water compartments snow, soil, surface water and river. For groundwater storage, only correlation and the ratio of amplitudes were improved. The biggest part of the added water mass affected the storage of soil and groundwater, as well as the snow storage during winter, which resulted in higher values for the mass changes (right columns in Fig. 8). Altogether, TWSA water mass was reduced resulting in a smaller annual amplitude that fitted considerably better to the annual amplitude of the synthetic TWSA observations (see Fig. 7).

When considering the same TWSA observations but introducing a diagonal observation error covariance matrix to the EnKF (case 16 w in Fig. 8), the RMSE of TWSA was even improved to 15 mm, mostly due to the smaller RMSE in soil and groundwater changes (13 and 10 mm, respectively). However, the correlation of soil and groundwater changes decreased compared to the OL and was found 0.3 lower compared to case 16 c. Note that in contrast to RMSE, the computation of correlations was based on the residual curves

after subtracting the linear trend, annual and semi-annual cycles. Water mass was added to the model over the complete C/DA phase (mass change of TWSA on top row in Fig. 8).

These results indicate that the chosen observation error model had a considerable impact on the C/DA results for TWSA and several individual water storages. Some metrics indicate that it is helpful to consider the full GRACE error covariance matrix (e.g. RMSE of surface water and river and correlation of soil and groundwater), while it has an adverse impact on others (e.g. RMSE of TWSA, soil and groundwater and correlation of TWSA). In summary, this experiment does not allow to unambiguously decide whether considering observation error correlations improves the C/DA results or not. We note that, in case of the white noise assumption, the GRACE data have a higher weight and, therefore, the model update should be pulled closer towards GRACE TWSA than with the correlated noise model; yet this does not always mean that our metrics improve.

5.2 Do the correlated GRACE errors affect C/DA when assimilating observations of different spatial scales?

To be consistent with the previous section, again, we performed the analyses for sub-basin 8. When introducing synthetic TWSA that were aggregated to 11 sub-basin means, case 11 w and 16 w yielded similar values for RMSE, correlation of residual curves and the ratio of annual amplitude for TWSA and the individual water compartments (first to third column in Fig. 8). The same holds for case 11 c and 16 c. Only the correlation of soil changes was considerably reduced to 0.1 in case 11 c. These results indicate that the change of the spatial discretisation from 16 grid cells to 11 sub-basins has a smaller impact on C/DA results compared to the switch from a diagonal (white noise) to a full observation error covariance matrix (correlated errors) in the filter update step (compare e.g. 11 w and c in Fig. 8).

When assimilating synthetic TWSA aggregated to four sub-basins (see Fig. 4) the effect of changing the diagonal to a full observation error covariance matrix in the EnKF on RMSE, correlation and the ratio of amplitudes is less than the effect of changing the spatial discretisation of the introduced TWSA (case 4 w and c in first to third column in Fig. 8). For both cases 4 w and c the RMSE is reduced for TWSA and all individual compartments (except groundwater) compared to the open loop simulation. However, the residual correlation for soil is negative, while the correlation for TWSA and the individual compartments (again except groundwater) increases. It seems that interannual changes of the soil storage are rather harmed for the EnKF variants 4w and 4c by introducing monthly means of GRACE TWSA, while the annual cycle is captured quite well (reflected in the RMSE and ratio of amplitudes). The amount of water that is introduced to the model in case 4 w and c depends clearly on the choice of the observation error model (fifth column in Fig. 8): the amount of absolute mass change in case 4 c is about 100 mm higher for the soil storage but about 100 mm smaller for the groundwater compartment.

These comparisons indicated that the observation error model affected C/DA on the three selected spatial scales. The effect of changing the observation error model was found to be large, when assimilating TWSA with a fine spatial discretisation, for which the correlations at least for several observation errors appeared high. In this case the impact was seen at least as big as the impact of the chosen spatial discretisation of observations on the C/DA results (compare e.g. RMSE for soil in case 16 w and c, where the error model changed, and in case 11 c and 16 c, where the discretisation changed, in Fig. 8). One might conclude that in cases of high observation error correlations, the choice of the observation error model has at least the same importance as the choice of the spatial discretisation of observations. In summary, we cannot provide a final answer whether, and under what circumstances, implementing observation error correlation in data assimilation—i.e. applying a model of spatial error correlation in the analysis step—will lead to improved results in a general sense. For GRACE assimilation, the problem is further intricate since the spatial scales of error correlation (several 100 km along-track) are similar to the scales of physical correlation of land surface and groundwater variables. From an estimation-theoretical point of view, accounting for correlated errors is considered helpful since it aims at decreasing the variance of the estimator. This is, on repeating the same assimilation experiment with many realisations of data errors, the estimate will be closer to reality in the mean. On the other hand, it is easy to show that disregarding observation correlations does not cause the estimate to be biased. Moreover, disregarding correlations in data assimilation means that the data get a higher weight compared to model forecasts. As a result, any evaluation metrics that (implicitly) assumes the data as true will appear favourable in this case. It is thus difficult to directly compare experiments with and without (or with partly) implementing error correlations. Moreover, for the original GRACE data, unlike for many remote sensing observations, it is not possible to define a "natural" grid resolution. It thus is tempting to simply work with the grid resolution applied in hydrological modelling and rely to error correlations, but this may easily lead to numerical stability problems in the gain matrix. In fact, an ensemble of limited size results in a model error covariance matrix that is rank-defect. Therefore, a non-singular error covariance matrix of the observations is required to enable a numerically stable solution of the ensemble Kalman filter update equation. As a result, not (or only partly) implementing error correlations may lead to a stabilising effect. In summary, we believe that assessing the effect of error correlations must be studied on a case-base, through simulations as realistic as possible. We are aware, of course, that this may limit the general applicability of our results somewhat.

5.3 Are the findings transferable to other regions?

We analysed the results of case 11 w and c of Sect. 5.2 for the different regions within the Mississippi Basin. Here, three representative sub-basins were chosen based on their location, shape and area, as well as observation error correlation, annual amplitude and signal-to-noise ratio (SNR) of observations: (i) the smallest of the 11 sub-basins (sub-basin 10) with large annual amplitude, (ii) one sub-basin located in the HPA (sub-basin 9) with east–west expansion and an overall good agreement between modelled and observed TWSA and (iii) the sub-basin with the lowest SNR (sub-basin 6) and north–south spatial expansion. These sub-basins also represent fairly good, average and poor performances of the C/DA results. High correlations to sub-basins in the north and south,

i.e. located in one column of the grid in Fig. 4, for each of the presented sub-basins were found (Fig. 6). In addition, we present the metrics averaged for the entire Mississippi Basin. The results are shown in Fig. 9. Here, each individual subplot contains the results for the Mississippi Basin as a whole, as well as sub-basin 8, 9, 6 and 10, ordered by decreasing areas. The OL results are shown by grey horizontal lines, while the white and black bars refer to the assumed white noise and correlated observation errors in the EnKF, respectively.

Regarding TWSA (top row in Fig. 9), sub-basin 6 and 8 showed noticeable differences in RMSE when considering white noise or correlated errors (9 and 12 mm, respectively) and sub-basin 6 in the ratio of amplitudes (0.6 and 1, respectively). However, less differences of metrics for TWSA were visible for sub-basin 9 and 10, as well as for the average over the entire Mississippi Basin. In case of the assumption of white observation noise in the EnKF water was subtracted from the model (up to −90 mm in case 10 w), while water

was introduced to the model (up to 30 mm in case 6 c) when assessing correlated errors.

Only a small volume of water was introduced into the sub-basin 9 (fifth column in Fig. 9: absolute water mass change less than 100 mm), which was less than 50 % of the absolute water mass change in sub-basin 6, and only about 25 % of sub-basin 8. Therefore, the effect of C/DA itself appeared smaller in sub-basin 9 compared to the other sub-basins and the sensitivity to the observation error model in the EnKF was rather small.

Sub-basins 6 and 10 appeared quite sensitive to the chosen observation error model in the EnKF for the soil compartment, which was found in all metrics and for which the white noise showed better agreements with the simulated truth (first and second column in Fig. 9: 6 mm RMSE instead of 11 mm, correlation of 0.7–0.9 instead of 0.2–0.5 in case of correlated errors). However, the amplitude of snow and groundwater was clearly improved when considering correlated errors in the EnKF update (third column in Fig. 9: ratio of amplitudes
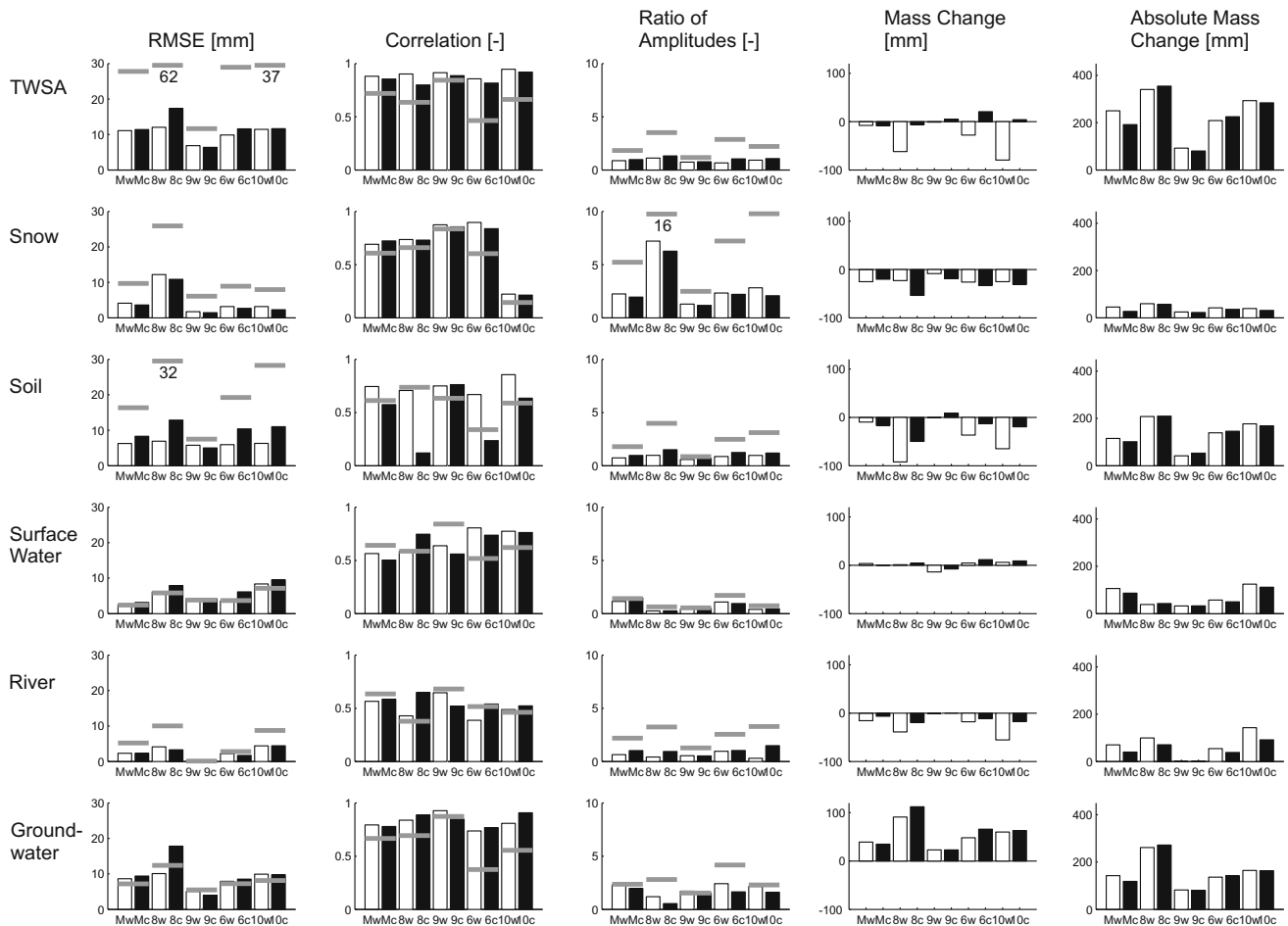

Fig. 9 Metrics for area mean of the entire Mississippi Basin (Mw and Mc), as well as sub-basin 8, 9, 6 and 10 (see Fig. 4), sorted by decreasing area, for open loop run (*grey horizontal lines*) and calibration and data assimilation variants. Please note that the ratio of amplitudes is reduced by one, so that zero represents equal amplitudes. Some of the RMSE values of the open loop run exceed the shown range. These values are displayed at the *grey horizontal lines*

of snow for case 10 c is 2 instead of 3, and ratio of amplitudes of groundwater is 1 instead of 2). Also, the average for the entire Mississippi Basin showed differences in the metrics for the soil compartment (for which the white noise showed again better agreements with the simulated truth). The metrics of the other individual water compartments appeared less sensitive.

In summary, sub-basins for which the EnKF update increments were high (due to high discrepancy between modelled and observed TWSA and small standard deviation of the observation), and sub-basins that are elongated in north–south direction were predominantly affected by the chosen observation error model in the EnKF.

5.4 Do the filter algorithms show a different sensitivity with respect to correlated GRACE errors?

The experiment of Sect. 5.1 was repeated here, this time for 11 sub-basin observations and considering the SQRA and SEIK methods. The results for TWSA for the sub-basin 8 are shown in Fig. 10, where the plots for SQRA (labelled by Sq) and SEIK (labelled by Se) are compared with those of the standard EnKF. Grey bars show the results of OL, while the others are assigned to the specified observation error model in the filter variant, i.e. assumption of white noise (white bars) or consideration of correlated errors (black bars).

Results of C/DA were found to be significantly improved after application of both SQRA and SEIK when compared to the OL simulation. The RMSE was reduced up to 11 mm, the ratio of amplitude up to 1.0 and the correlations of the residual curves increased up to 0.9 in case of the SEIK filter, when considering correlated errors (case Se c). The water mass that was introduced into the model was similar for all cases (about 350 mm in absolute terms, except Sq w), while the net introduced water mass differed more strongly depending on which observation error covariance matrix was applied in the update, compared to the effect of the filter variants (see Fig. 10, fourth and fifth columns). The application of the SQRA and SEIK algorithms had only a small influence

on the RMSE with respect to the standard EnKF when considering white noise in the update step (less than 2 mm). In case of SQRA the correlation was even degraded by 0.1, while the consideration of correlated errors in the SEIK filter update improved RMSE by 6 mm and the correlation by 0.1.

The EnKF showed the biggest differences between the assumption of white noise or the consideration of correlated errors in the filter, especially in terms of RMSE (5 mm less in case of white noise) and correlation (0.1 larger in case of white noise). This might be due to the fact that the EnKF relies on an ensemble of observation perturbations. The results for TWSA for both cases (w and c) were quite similar when applying SQRA and SEIK, whereas the individual water compartments were affected by the correlated errors, especially that of the soil compartment (not shown here).

The investigations indicate that correlated GRACE errors affected the results of all filter variants. In our test case the SEIK filter, which provides the best numerical efficiency among the analysed algorithms, was found to perform slightly better than the standard EnKF and SQRA methods, especially in terms of RMSE.

5.5 Does the choice of the filter variant affect linear trend estimation?

We examined linear trend estimations from the EnKF variants and compared them to the linear trend of OL simulation, the synthetic GRACE observations and the synthetic truth. We analysed the trends in TWSA averaged over the 11 sub-basins (Table 4), as well as trends in total and individual storages averaged over the entire Mississippi Basin (Table 5). Clearly, a linear trend estimated over 3 years has to be considered with caution, especially in real data analysis, since it cannot be considered as a long-term trend. However, in our synthetic experiment, linear trend estimation addresses the question as to how far data assimilation may alter the trends that are present in either the open loop simulation or in the GRACE data. When comparing the trend estimations from
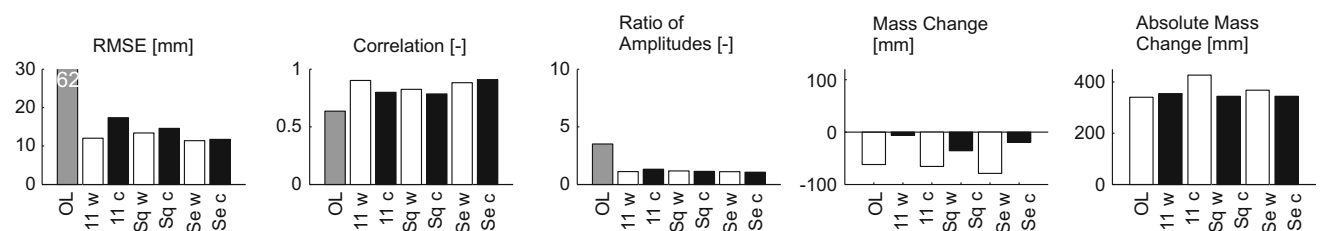

Fig. 10 Metrics of TWSA for area mean of sub-basins 8 (see Fig. 4) for open loop run (OL, *grey bars*) and calibration and data assimilation variants when applying the standard EnKF (11 w and c), the SQRA (Sq w and c) or the SEIK approach (Se w and c). Designations of the variants can also be found in Table 3. Please note that the ratio of amplitudes is reduced by one so that zero represents equal amplitudes. The RMSE value of the OL run exceeds the shown range. The value is displayed at the *top* of the *bar*

Table 4 Linear trend estimation in mm/year for TWSA in the 11 sub-basins **S** of the Mississippi Basin for open loop (**OL**) model simulation, the synthetic truth (**T**), synthetic GRACE observations (**y**) and the ensemble filter variants. Names of sub-basins can be found in Table 1 and names of the ensemble filter variants in Table 3

S	OL	T	y	4 w	4 c	11 w	11 c	16 w	16 c	Sq w	Sq c	Se w	Se c
1	−49	−34	−33	−28	−35	−38	−34	−33	−34	−38	−34	−36	−35
2	−46	−29	−30	−28	−28	−31	−32	−23	−32	−26	−31	−29	−33
3	−17	−7	−5	−2	−7	−8	−3	−9	−11	−9	−8	−9	−5
4	−85	−54	−51	−70	−55	−54	−50	−66	−46	−58	−51	−56	−54
5	5	9	11	11	11	13	12	13	15	16	15	14	15
6	−36	−17	−21	−9	−7	−16	−10	−14	−22	−18	−16	−17	−16
7	−13	6	5	−10	−3	1	−1	3	0	0	2	2	0
8	−55	−18	−21	−19	−17	−20	−25	−20	−26	−18	−21	−21	−24
9	−21	−10	−9	−10	−9	−9	−11	−5	−9	−9	−9	−10	−11
10	−31	−13	−6	−10	−5	−7	−9	−10	−24	−10	−10	−9	−10
11	−1	0	−3	8	0	3	0	0	−1	4	4	2	2

Table 5 Linear trend estimation in mm/year for total and individual water storage changes averaged over the entire Mississippi Basin for open loop (**OL**) model simulation, the synthetic truth (**T**), synthetic GRACE observations (**y**) and the ensemble filter variants. Names of the ensemble filter variants can be found in Table 3

Storage	OL	T	y	4 w	4 c	11 w	11 c	16 w	16 c	Sq w	Sq c	Se w	Se c
TWSA	−32	−11	−14	−14	−14	−15	−15	−14	−16	−15	−14	−15	−15
Snow	−7	−3	–	−5	−5	−6	−5	−5	−6	−6	−5	−6	−6
Soil	−7	−3	–	7	8	−2	1	−1	4	−9	−10	−6	−8
Surface water	0	0	–	0	−1	0	0	0	0	0	0	0	0
River	−3	0	–	−6	−1	−2	−2	−2	−2	−5	−2	−2	−3
Groundwater	−12	−4	–	−10	−14	−5	−10	−6	−10	3	1	−3	−2

the EnKF variants with the OL simulation, differences of 15 mm/year on average up to 40 mm/year exist (in sub-basins 4 and 8). A comparison of the estimated trends from EnKF variants with GRACE observations showed differences of 5 mm/year on average up to 20 mm/year (in sub-basins 4 and 10), while a good agreement was achieved in sub-basins 8 and 9. Hence, the linear trends of the EnKF variants are mostly closer to the trend estimated from GRACE compared to the linear trends of the OL simulation. Furthermore, a comparison with the synthetic truth shows that in nine of the 11 sub-basins the estimated trend from all ensemble filter variants are closer to the truth than the trend of the OL. Only in sub-basins 5 and 11 the OL simulation represents the true trend better than most of the ensemble filter variants. Both sub-basins are located in the north-west of the Mississippi Basin and show rather small trends compared to the other sub-basins. We averaged the TWSA from the EnKF variants over the entire Mississippi Basin and estimated the linear trend. Differences of about 20 mm/year were found in comparison to the trend from OL. In contrast, the trends agreed quite well with the trend from the synthetic observations and the synthetic truth, i.e. the differences were smaller than 5 mm/year. Therefore, we conclude that GRACE C/DA affects the estimation of linear trends positively in our par-

ticular experiments. Additionally, we determined the linear trends for compartmental water storages averaged over the entire Mississippi Basin. The individual compartments show differences of 5 mm/year on average up to 20 mm/year in the soil and groundwater storages compared to the OL simulation. A comparison to the synthetic truth shows that surface water is not affected by GRACE data assimilation, which results from the fact that OL and synthetic truth do not show any trend. Also, only a small influence on the linear trend in snow and river is visible for all filter variants, which seems to be justified, since both storages experience only small negative trends (or no trend in case of the synthetic truth of the river storage). In contrast, linear trends in soil water and groundwater are clearly affected by GRACE assimilation. In case of the filter variants 11 w and c, as well as Se w and c, introduction of GRACE TWSA pulls the trends (mostly) closer to the true trend. For the other variants, GRACE assimilation might also have the effect that the sign of trend changes, e.g. in case 4 w and c for soil, and in case Sq w and c for groundwater. The trends in soil and groundwater seem to compensate each other. Therefore, we assume that the vertical disaggregation between soil water and groundwater might be more difficult compared to the other individual compartments.

Table 6 Metric A: the percentage of parameters that are sensitive to the assimilation of TWSA. Metric B: the percentage of sensitive parameters that were also found in Schumacher et al. (2015). Names of the EnKF variants can be found in Table 3

Metric	4 w	4 c	11 w	11 c	16 w	16 c	Sq w	Sq c	Se w	Se c
A	0	0	15	35	30	55	35	40	20	40
B	–	–	0	60	60	80	40	60	60	100

5.6 Does the choice of the observation error model affect parameter calibration?

First, we identified those parameters that were sensitive to TWSA assimilation. Parameters whose standard deviation (i.e. ensemble spread) σ was reduced to less than 25 % of their initial value after 18 months (50 % of update steps) were defined as sensitive. Results are reported in Table 6 (Metric A). First, we analysed the results when applying the standard EnKF (cases 4, 11 and 16). When using a coarse observation discretisation (case 4 w and c), TWSA assimilation did not affect the parameter estimation. With increasingly finer discretisation of TWSA observations, the influence of assimilation was increased, i.e. the number of sensitive parameters increased from 15 % (in case 11 w) to 55 % (in case 16 c). We believe this is likely due to the fact that water states were constrained more when using more detailed observation information in space. Therefore, parameters were constrained more via their cross-correlations to the water states. The number of sensitive parameters was found to be higher in the cases with correlated TWSA errors (cases indicated by c) compared to the cases when assuming white noise for TWSA (cases indicated by w).

The application of the SQRA and SEIK filter increased the number of sensitive parameters up to 40 % (case Sq c and Se c). Here as well, the number of sensitive parameters was found to be larger in case of assuming correlated observation errors (see Metric A in last four columns in Table 6).

Additionally, those parameters that were found as sensitive to TWSA assimilation in this study were compared to the five sensitive parameters that were found in Schumacher et al. (2015), in which Spearman's rank correlation coefficient was used (Table 6, Metric B, and Table 7). Our results indicated that 40–100 % of the sensitive parameters in Schumacher et al. (2015) were also found as sensitive in the simulations performed here. The root depth multiplier (parameter 1) was found to be sensitive in all filter variants (except 16 w, see Table 7), but was not identified as sensitive in Schumacher et al. (2015).

We cannot claim that parameter values are individually improved (closer to "true" values) after C/DA since different parameter combinations may result in a similar optimal simulation of water storages. In summary, our results indicated that with increasingly finer discretisation of observations,

Table 7 Parameters that are sensitive to TWSA assimilation, and sensitive parameters found in Schumacher et al. (2015) for comparison. Names of ensemble filter variants can be found in Table 3. Parameter names according to identification numbers (IN) are given in Table 2

Filter variant	Sensitive parameters (IN)
4(w)	–
4(c)	–
11(w)	1, 3, 8
11(c)	1, 2, 3, 4, 6, 17, 20
16(w)	3, 6, 11, 17, 19, 20
16(c)	1, 2, 3, 4, 5, 6, 10, 15, 19, 20
SQ(w)	1, 4, 8, 16, 17, 18, 20
SQ(c)	1, 2, 3, 4, 13, 15, 17, 20
SE(w)	1, 4, 11, 20
SE(c)	1, 2, 4, 11, 12, 17, 19, 20
Schumacher et al. (2015)	2, 4, 11, 19, 20

or when implementing error correlations in the filter, the number of parameters that can be calibrated by GRACE increases.

6 Conclusions

We discuss a flexible calibration and data assimilation (C/DA) framework that allows for the integration of gridded and basin averaged GRACE TWSA observations into WGHM while simultaneously estimating calibration parameters. We extended the framework based on the standard EnKF while considering computationally efficient variants such as the SQRA and SEIK algorithms. In addition, an inflation factor was introduced to account for model errors. After implementing the modifications, a synthetic twin experiment was conducted to investigate the effect of GRACE TWSA error correlations on the C/DA results. In addition to the true and open loop (OL) simulations, a total of ten C/DA variants were implemented including the options of (i) diagonal or full GRACE observation error covariance matrices in the filter update step, (ii) spatial aggregation of the observations to four, 11 or 16 sub-basin/grid cell averages and (iii) EnKF, SQRA or SEIK as filter algorithm. We summarise our main findings as follows:

1. Consideration of GRACE error correlation affects anomalies of total and compartmental water storages determined by C/DA that is based on TWSA observations. The impact increases with increasing error correlations and thus higher spatial resolution of TWSA observations. It is particularly high in basins that are elongated in north–south direction and in basins in which TWSA simulated without C/DA is very different from the observed TWSA.

2. Considering these correlated observation errors does not generally improve the results. Some metrics indicate that it is helpful to consider the full GRACE error covariance matrix, while it appears to have an adverse influence on others.

3. The C/DA results of the EnKF algorithm are more sensitive to the chosen observation error model than the results of the SQRA and SEIK algorithms.

4. C/DA leads to adjustment of the model parameters only in case of sufficient spatial resolution of the TWSA observations. The number of sensitive parameters increases with increasing spatial resolution of the TWSA observations and if GRACE error correlation is taken into account.

Based on these findings, we conclude that the observation error model is at least as important as the choice of discretisation of observations. We recommend to consider GRACE error correlations, since they characterise the error structure of GRACE products; even so there appears no general rule as to whether applying spatial error correlations in the data assimilation update step will lead to improved results. We found also promising results when applying alternative methods. We could show that by considering, e.g. the SEIK filter and correlated GRACE errors in the update step, the RMSE and correlation coefficients of TWSA were improved by 6 mm and 0.1, respectively, with respect to the EnKF (see case 11 c and Se c in Fig. 10). This is likely caused by avoiding sampling errors, since no observation ensemble has to be generated, and applying the minimum second-order exact sampling for generating updated ensemble perturbations in the filter update. Therefore, we will investigate the effect of alternative methods on C/DA results in more detail in our future work.

This study was built on a synthetic experiment that enabled us to validate the OL and C/DA results with predefined true hydrological states. In parallel activities, our framework was transferred to real GRACE data application (Eicker et al. 2014). In the future, an extensive validation with various independent data sets (e.g. river discharge, groundwater, lake level, soil water equivalent) will be carried out. In addition, extending the application of the proposed C/DA framework to other river basins with other climatic and anthropogenic characteristics will be considered in future studies.

References

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

Collilieux X, Van Dam T, Ray J, Coulot D, Metivier L, Altamimi Z (2011) Strategies to mitigate aliasing of loading signals while estimating GPS frame parameters. J Geod 86:1–14. doi:10.1007/s00190-011-0487-6

Crow WT, Van Loon E (2006) Impact of incorrect model error assumptions on the sequential assimilation of remotely sensed surface soil moisture. J Hydrometeor 7:421–432. doi:10.1175/JHM499.1

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, Hoffmann-Dobrev H, Portmann FT, Siebert S, Eicker A, Rodell M, Strassberg G, Scanlon B (2012) Impact of water withdrawals from groundwater and surface water on continental water storage variations. J Geodyn 59–60:143–156. doi:10.1016/j.jog.2011.05.001

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(7):5698–5720. doi:10.1002/2014WR015595

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. Surv 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

Evensen G, Van Leeuwen PJ (2000) An ensemble Kalman smoother for nonlinear dynamics. Mon Wea Rev 128:1852–1867. doi:10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2

Evensen G (2004) Sampling strategies and square root analysis schemes for the EnKF. Ocean Dynam 54:539–560. doi:10.1007/s10236-004-0099-2

Evensen G (2007) Data assimilation. The Ensemble Kalman Filter. Springer, Berlin

Famiglietti JS, Rodell M (2013) Water in the balance. Science 340:1300–1301. doi:10.1126/science.1236460

Flechtner F, Thomas M, Dobslaw H (2010) Improved non-tidal atmospheric and oceanic de-aliasing for GRACE and SLR satellites. Adv Technol Earth Sci 2:131–142. doi:10.1007/978-3-642-10228-8_11

Forman BA, Reichle RH, Rodell M (2012) Assimilation of terrestrial water storage from GRACE in a snow-dominated basin. Water Resour Res 48:W01507. doi:10.1029/2011WR011239

Forman BA, Reichle RH (2013) The spatial scale of model errors and assimilated retrievals in a terrestrial water storage assimilation system. Water Resour Res 49:7457–7468. doi:10.1002/2012WR012885

Forootan E, Kusche J (2012) Separation of global time-variable gravity signals into maximally independent components. J Geod 86(7):477–497. doi:10.1007/s00190-011-0532-5

Forootan E, Didova O, Schumacher M, Kusche J, Elsaka B (2014) Comparisons of atmospheric mass variations derived from ECMWF reanalysis and operational fields, over 2003 to 2011. J Geod 88(5):503–514. doi:10.1007/s00190-014-0696-x

Fritsche M, Döll P, Dietrich R (2012) Global-scale validation of model-based load deformations from water mass and atmospheric pressure variations using GPS. J Geodyn 59–60:133–142. doi:10.1016/j.jog.2011.04.001

Gupta HV, Sorooshian S, Yapo PO (1998) Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. Water Resour Res 34(4):751–763. doi:10.1029/97WR03495

Hamill TM, Snyder C (2002) Using improved background-error covariances from an ensemble Kalman filter for adaptive observations. Mon Wea Rev 130:1552–1572. doi:10.1175/1520-0493(2002)130<1552:UIBECF>2.0.CO;2

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

Hoteit I, Pham DT, Blum J (2002) A simplified reduced order Kalman filtering and application to altimetric data assimilation in Tropical Pacific. J Marine Syst 36:101–127. doi:10.1016/S0924-7963(02)00129-X

Houborg R, Rodell M, Li B, Reichle RH, Zaitchik BF (2012) Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations. Water Resour Res 48:W07525. doi:10.1029/2011WR011291

Hunger M, Döll P (2008) Value of river discharge data for global-scale hydrological modeling. Hydrol Earth Syst Sci 12:841–861. doi:10.5194/hess-12-841-2008

Iman RL (2008) Latin Hypercube sampling. III, encyclopedia of quantitative risk analysis and assessment. doi:10.1002/9780470061596.risk0299

Kalman RE (1960) A new approach to linear filtering and prediction problems. Trans ASME J Basic Eng 82(D):35–45

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

Klees R, Revtova EA, Gunter BC, Ditmar P, Oudman E, Winsemius HC, Savenije HHG (2008) The design of an optimal filter for monthly GRACE gravity models. Geophys J Int 175(2):417–432. doi:10.1111/j.1365-246X.2008.03922.x

Koch KR (1997) Parameterschätzung und Hypothesentests (in German). Dümmler, Bonn

Kurtenbach E, Mayer-Gürr T, Eicker A (2009) Deriving daily snapshots of the Earth's gravity field from GRACE L1B data using Kalman filtering. Geophys Res Lett 36:L17102. doi:10.1029/2009GL039564

Kusche J (2003) A Monte-Carlo technique for weight estimation in satellite geodesy. J Geod 76(11–12):641–652. doi:10.1007/s00190-002-0302-5

Kusche J (2007) Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models. J Geod 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

Kusche J, Klemann V, Bosch W (2012) Mass distribution and mass transport in the Earth system. J Geodynam 59–60:1–8. doi:10.1016/j.jog.2012.03.003

Le Dimet FX, Talagrand O (1986) Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. Tellus 38A:97–110. doi:10.1111/j.1600-0870.1986.tb00459.x

Li B, Rodell M, Zaitchik BF, Reichle RH, Koster RD, Van Dam TM (2012) Assimilation of GRACE terrestrial water storage into a land surface model: evaluation and potential value for drought monitoring in western and central Europe. J Hydrol 446–447:103–115. doi:10.1016/j.jhydrol.2012.04.035

Liu Y, Weerts AH, Clark M, Hendricks Franssen HJ, Kumar S, Moradkhani H, Seo DJ, Schwanenberg D, Smith P, Van Dijk AIJM, Van Velzen N, He M, Lee H, Noh SJ, Rakovec O, Restrepo P (2012) Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. Hydrol Earth Syst Sci 16:3863–3887. doi:10.5194/hess-16-3863-2012

Longuevergne L, Scanlon BR, Wilson CR (2010) GRACE hydrological estimates for small basins: evaluating processing approaches on the high plains aquifer. USA. Water Resour Res 46:W11517. doi:10.1029/2009WR008564

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. Geophys Res Lett 33:L12401. doi:10.1029/2006GL026855

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

Nerger L (2003) Parallel filter algorithms for data assimilation in oceanography. PhD thesis, University of Bremen, Germany

Pham DT, Verron J, Roubaud MC (1998) A singular evolutive extended Kalman filter for data assimilation in oceanography. J Marine Syst 16(3–4):323–340. doi:10.1016/S0924-7963(97)00109-7

Pierce R, Leitch J, Stephens M, Bender P, Nerem R (2008) Intersatellite range monitoring using optical interferometry. Appl Optics 47(27):5007–5019. doi:10.1364/AO.47.005007

Reichle RH, Koster RD (2003) Assessing the impact of horizontal error correlations in background fields on soil moisture estimation. J Hydrometeorol 4(6):1229–1242. doi:10.1175/1525-7541(2003)004<1229:ATIOHE>2.0.CO;2

Ripley BD (1987) Stochastic simulation. Wiley, New York

Rodell M, Chen J, Kato H, Famiglietti JS, Nigro J, Wilson CR (2007) Estimating groundwater storage changes in the Mississippi River basin (USA) using GRACE. Hydrogeol J 15(1):159–166. doi:10.1007/s10040-006-0103-7

Sakumura C, Bettadpur S, Bruinsma S (2014) Ensemble prediction and intercomparison analysis of GRACE time-variable gravity field models. Geophys Res Lett 41:1389–1397. doi:10.1002/2013GL058632

Schmidt R, Flechtner F, Meyer U, Neumayer KH, Dahle C, Koenig R, Kusche J (2008) Hydrological signals observed by the GRACE satellites. Surv Geophys 29:319–334. doi:10.1007/s10712-008-9033-3

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

Schrama EJO, Wouters B, Lavallee DD (2007) Signal and noise in Gravity Recovery and Climate Experiment (GRACE) observed surface mass observations. J Geophys Res 112:B08407. doi:10.1029/2006JB004882

Schumacher M, Eicker A, Kusche J, Müller Schmied H, Döll P (2015) Covariance analysis and sensitivity studies for GRACE assimilation into WGHM. IAG Symp 143. doi:10.1007/1345_2015_119

Strassberg G, Scanlon BR, Chambers D (2009) Evaluation of groundwater storage monitoring with the GRACE satellite: case study of

the High Plains aquifer, central United States. Water Resour Res 45:W05410. doi:10.1029/2008WR006892

Su H, Yang ZL, Dickinson RE, Wilson CR, Niu GY (2010) Multisensor snow data assimilation at the continental scale: the value of gravity recovery and climate experiment terrestrial water storage information. J Geophys Res 115:D10104. doi:10.1029/2009JD013035

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

Tangdamrongsub N, Steele-Dunne SC, Gunter BC, Ditmar PG, Weerts AH (2015) Data assimilation of GRACE terrestrial water storage estimates into a regional hydrological model of the Rhine River basin. Hydrol Earth Syst Sci 19:2079–2100. doi:10.5194/hess-19-2079-2015

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

Tippett MK, Anderson JL, Bishop CH, Hamill TM, Whitaker JS (2003) Ensemble Square Root Filters. Mon Wea Rev 131:1485–1490. doi:10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2

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

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

Wahr J, Swenson S, Velicogna I (2006) Accuracy of GRACE mass estimates. Geophys Res Lett 33:L06401. doi:10.1029/2005GL025305

Weedon GP, Balsamo G, Bellouin N, Gomes S, Best MJ, Viterbo P (2014) The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. Water Resour Res 50(9):7505–7514. doi:10.1002/2014WR015638

Werth S, Güntner A (2010) Calibration analysis for water storage variability of the global hydrological model WGHM. Hydrol Earth Syst Sci 14:59–78. doi:10.5194/hess-14-59-2010

Whitaker JS, Hamill TM (2002) Ensemble data assimilation without perturbed observations. Mon Weather Rev 130:1913–1924. doi:10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2

Wouters B, Bonin JA, Chambers DP, Riva REM, Sasgen I, Wahr J (2014) GRACE, time-varying gravity, Earth system dynamics and climate change. Rep Prog Phys 77:116801. doi:10.1088/0034-4885/77/11/116801

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

Zenner L, Bergmann-Wolf I, Dobslaw H, Gruber T, Güntner A, Wattenbach M, Esselborn S, Dill R (2014) Comparison of daily GRACE gravity field and numerical water storage models for de-aliasing of satellite gravimetry observations. Surv Geophys 35(6):1251–1266. doi:10.1007/s10712-014-9295-x