Data assimilation of GRACE terrestrial water storage estimates into a regional hydrological model of the Rhine River basin

N. Tangdamrongsub[1], S. C. Steele-Dunne[2], B. C. Gunter[1,3], P. G. Ditmar[1], and A. H. Weerts[4,5]

[1]Department of Geoscience and Remote Sensing, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands

[2]Department of Water Resources, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands

[3]School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, The United States of America

[4]Operational Water Management, Deltares, Delft, The Netherlands

[5]Hydrology and Quantitative Water Management Group, Department of Environmental Sciences, Wageningen University, The Netherlands

Correspondence to: N. Tangdamrongsub (n.tangdamrongsub@tudelft.nl)

Abstract

The ability to estimate Terrestrial Water Storage (TWS) realistically is essential for understanding past hydrological events and predicting future changes in the hydrological cycle. Inadequacies in model physics, uncertainty in model land parameters, and uncertainties in meteorological data commonly limit the accuracy of hydrological models in simulating TWS. In an effort to improve model performance, this study investigated the benefits of assimilating TWS estimates derived from the Gravity Recovery And Climate Experiment (GRACE) data into the OpenStreams wflow_hbv model using an Ensemble Kalman Filter (EnKF) approach. The study area chosen was the Rhine River basin, which has both well-calibrated model parameters and high-quality forcing data that were used for experimentation and comparison. Four different case studies were examined which were designed to evaluate different levels of forcing data quality and resolution including those typical of other less well-monitored river basins. The results were validated using in situ groundwater and stream gauge data. The analysis showed a noticeable improvement in groundwater estimates when GRACE data were assimilated, with a best-case improvement of 71% in correlation coefficient (from 0.31 to 0.53) and 35% in RMS error (from 8.4 to 5.4 cm) compared to the reference (ensemble open-loop) case. The correlation and RMSE improvements in groundwater estimates for the data-sparse case were up to 33% and 35%, respectively, while the average improvements for all four cases evaluated were 13 % and 14%, respectively. Only a slight overall improvement was observed in streamflow estimates when GRACE data were assimilated. Further analysis suggested that this is likely due to sporadic short-term, but sizeable, errors in the forcing data and the lack of sufficient constraints on the soil moisture component. Overall, the results highlight the benefit of assimilating GRACE data into hydrological models, particularly in data-sparse regions, while also providing insight on future refinements of the methodology.

1 Introduction

Terrestrial Water Storage (TWS) is the integrated sum of all surface water, soil moisture, snow water, and groundwater availability, and is a metric critical for monitoring the water supply for domestic, industrial, and agricultural sectors. The ability to estimate TWS is useful for understanding past events and predicting future changes in the hydrological cycle, streamflow and water availability, as well as their impact on the occurrence of droughts, heat waves, and floods (Hirschi, et al., 2007). The individual components of TWS influence the climate system in different ways. Soil moisture is a major source of water for the atmosphere in the terrestrial water cycle (Jung et al, 2010) and plays a particularly important role in the climate system (Seneviratne et al., 2010). Soil moisture estimates are also useful for seasonal predictions, and have been shown to improve predictions of air temperature in North America (Koster et al., 2010) and Europe (van den Hurk et al., 2012). Similarly, realistic estimation of the snowpack can improve the prediction of near surface temperature at high latitude regions at 15-30 day scales (Orsolini et al., 2013). Finally, groundwater variability influences soil moisture and evapotranspiration, and is related to long-term water availability and climate changes (Bierkens and van den Hurk, 2007; Green et al., 2011).

Despite the importance of having reliable estimates of TWS, knowledge about the spatial and temporal variations of TWS and its components is generally lacking. This is particularly true at large scales, due to the absence of global monitoring systems. Ground-based measurements, while very accurate, only provide point-wise estimates (Dorigo et al., 2011; Lettenmaier and Famiglietti, 2006). Large spatial coverage can be achieved using satellite remote sensing observations, but these often measure only one component of the total storage and suffer from additional limitations. For example, in the case of soil moisture, satellite observations are limited to the top few centimetres of the soil column and to areas free from dense vegetation cover (e.g., de Jeu et al., 2008; Entekhabi et al., 2010; Kerr et al., 2012). Variations in surface water can be observed with satellite altimetry but this technique is currently limited to large target areas (>10 km) (Phan et al., 2012; Schwatke et al., 2013, Kleinherenbrink et al., 2014).

Since measurements alone are not sufficient to comprehensively monitor all components of TWS, hydrological models are often employed. A strong point of hydrological models is their ability to obtain spatially distributed estimates, differentiate TWS components, and simulate changing boundary conditions. Many hydrological models are available, which vary in terms of process description, temporal resolution, spatial resolution, and the detail in process representation (Koster et al., 2000; Rodell et al., 2004). Models vary in terms of which TWS components are included in the model, and how they are represented. The performance of hydrological models is also influenced by the accuracy of the input forcing data and the quality of the model calibration. The existence of model uncertainties motivates the need to combine the model with independent observations to obtain a better representation of the system's behaviour.

Changes in TWS can also be estimated by observing variations of the regional gravity field over time. The idea is that changes in water storage, including those deep underground, induce a gravitational signature proportional to the amount of (water) mass redistribution. Since 2002, these variations have been measured by the Gravity Recovery and Climate Experiment (GRACE) satellite mission (Tapley et al., 2004). GRACE allows temporal variations of Earth's gravity field to be observed at spatial scales ranging in the hundreds of kilometres, and at time scales as short as one month. As part of the GRACE data processing, atmospheric and ocean related time-variable gravity effects are removed from the data, leaving the remaining gravity signal over the continents mostly representing changes in TWS (in some areas, additional removal of other nuisance signals is needed, such as those due to glacier melting, glacial isostatic adjustment, and megathrust earthquakes). The GRACE mission has enabled the first direct observations of large-scale TWS, and studies to date have shown high correlation with modelled TWS in terms of seasonal dynamics and regional spatial patterns (Syed et al., 2008; Becker et al., 2011; Longuevergne et al., 2013). A unique feature of satellite gravimetry is that it observes the total column of mass variations (including groundwater) while other remote sensing techniques can only penetrate to a very limited depth, often just a few centimetres. In contrast to hydrological modelling, it is not possible to identify which layer the inferred mass variations can be attributed (Rodell et al., 2009).

Several earlier studies have employed data assimilation to combine the strengths of hydrological modelling and GRACE observations and to mitigate their respective weaknesses (Zaitchik et al., 2008; Su et al., 2010; Houborg et al., 2012; Li et al., 2012; Forman et al., 2012). In data assimilation, the model states are constrained by observations, taking into account the estimated uncertainties for both the model states and the observations (Evensen, 2003; Reichle, 2008). Employing data assimilation provides a mechanism to downscale the coarse GRACE TWS variations to the temporal and spatial resolution of the model as well as providing insight from the hydrological model into the distribution of TWS between the individual storage terms. Zaitchik et al. (2008) assimilated GRACE into the Catchment Land Surface Model to estimate the TWS over the Mississippi River Basin. Houborg et al. (2012) and Li et al. (2012) applied a similar strategy to improve the drought indicator over North America and Europe, respectively. Su et al. (2010) and Forman et al. (2012), extended the work of Zaitchik et al. (2008) to improve the estimated snow water equivalent over North America and northwestern Canada, respectively. All results from earlier studies reported that assimilating GRACE improved, or at least did not degrade, the hydrology model's performance. In particular, good agreements between estimated state variables, e.g., groundwater and streamflow, and the in situ measurements were observed. This study adds to these prior works by examining how GRACE assimilation performs when the hydrological model is not well calibrated or when unreliable meteorological data are used to force the model. This focus of the study is on the Rhine River basin (Fig. 1), which is significantly smaller than the large basin or continent scale studies of these prior works, so the analysis presented here provides new insight into the performance of GRACE assimilation over smaller basins. And, while previous data assimilation studies have been performed in the Rhine and neighbouring basins (e.g. Weerts and Serafy, 2006; Rakovec et al., 2012), this study is the first to incorporate GRACE observations in the assimilation scheme for this region.

The primary goal of this study was to understand the impact of GRACE assimilation on the estimated TWS, groundwater (GW) variations and streamflow in the Rhine basin. The second goal was to investigate the potential value of assimilating GRACE observations in data-sparse regions. Four scenarios were considered in which the model parameters used were either calibrated (high quality) or basin-averaged (poor quality) values, and the forcing data were obtained from either local (high quality) or global (poorer quality) datasets. In this context, comparison of the four scenarios provides insight into how GRACE can be used to constrain hydrological models when limited data are available.

2 Hydrological modelling

The hydrological model employed in this study is the OpenStreams wflow_hbv model (Schellekens, 2014). This is a distributed version of the HBV-96 model, named after the Hydrologiska Byråns Vattenbalansavdelning (Hydrological Bureau Waterbalance-section). The HBV model was originally developed at this former section of the Swedish Meteorological and Hydrological Institute (SMHI) in the early 1970's. Since then, the HBV model has been used in over 40 countries. In 1996, a comprehensive re-evaluation of the HBV model routines was carried out (Lindström et al., 1997), which resulted in the HBV-96 version. The OpenStreams wflow_hbv model is a variant of this model, programmed in the PCRaster-Python environment (Karssenberg et al., 2009), but using a kinematic wave for hydrological routing. It is publicly available through the OpenStreams project (https://code.google.com/p/wflow/, last access 18 January 2015). The defined grid resolution used in this study was 1 km. A schematic representation of OpenStreams wflow_hbv is given in Fig. 2 (a) and the key parameters of the soil moisture and runoff response routines are listed and described in Table 1.

OpenStreams wflow_hbv consists of three main routines: (*i*) precipitation and snow, (*ii*) soil moisture, and (*iii*) runoff. The water from either precipitation or snow first enters the interception storage and snow routine. The remaining liquid water (from rainfall and snow melt) after the snow routine infiltrates into the soil. The soil moisture storage term (SM in [mm]), which includes both surface and root zone soil moisture is controlled by three main parameters fc, lp, and β (see also Table 1). When the amount of water exceeds the maximum capacity (fc), the excess water becomes available for direct runoff. Within the soil layer, seepage is generated and controlled by an empirical parameter β. The volume of water available for runoff (direct runoff and seepage) is transferred to the runoff response routine. Additionally, some percentage of the soil moisture evaporates, which is controlled by a defined threshold (fc×lp).

Two linear reservoirs are defined in the runoff routine, namely the upper and lower zones (UZ and LZ). The excess water from SM recharges the upper zone, and some of the water in UZ percolates to LZ, as determined by the perc parameter. At the same time, capillary flow from UZ to SM also occurs, controlled by cflux. The runoff generation in UZ is controlled mainly by two main parameters, the recession constant (khq) and the non-linearity parameter (α). LZ contributes the water to the base flow through the recession constant (k4). The amount of base flow is simply the multiplication between k4 and the amount of LZ. Runoff from UZ and LZ then enters the routing model to determine the streamflow.

For reference, TWS is defined here as the sum of SM, UZ and LZ. Groundwater storage (GW) is defined as the sum of UZ and LZ. These storage terms are calculated in the soil moisture and runoff response routines. Fig. 2 (b) shows the simulated SM, UZ and LZ from a nominal model run (i.e. using the calibrated parameters and local forcing data). The main source of TWS variation in this model is SM, with the variations in LZ and UZ an order of magnitude smaller. Extraction of groundwater for irrigation is considered to be small over our study region. It accounts for less than 1 km^3/year. Industry is the largest user (Wada et al. (2014). However, The net removal is small as only 10% of the total water withdrawal over the Rhine is from groundwater and the water is re-introduced to the system after being used for industry. This is markedly different to the extraction of groundwater for irrigated agriculture observed in India (Ferrant et al., 2014). Therefore, this impact on TWS is not considered in this study.

The OpenStreams wflow_hbv model was calibrated for the Rhine river basin using observations from in situ streamflow gauges (Mülders et al., 1999; Eberle et al., 2002; 2005; Photiadou et al., 2011). The spatial distribution of the calibrated model parameters is shown in Fig. 3.

In data-sparse regions, a lack of in situ (meteorological and streamflow) data makes it difficult to calibrate hydrological models (Sivapalan et al., 2003; Hrachowitz et al., 2013). Therefore, we decided to add “non-calibrated” cases to our simulations. In those cases, we defined the non-calibrated parameters as the areally-averaged values of the calibrated parameters in the entire basin, and used these for every grid cell in the basin.

3 Datasets

3.1 GRACE observation

The most recent release (RL05) of the GRACE gravity model product, generated by the University of Texas at Austin's Center of Space Research (CSR: Bettadpur, 2012), was used in the analysis. The CSR RL05 models represent a time-series of Stokes coefficients up to a maximum spherical harmonic degree and order of 60, and are provided monthly. Following the GRACE conventional processing steps, degree-1 coefficients provided by Swenson et al. (2008) were added, and the degree-2 coefficients were replaced by the values estimated from satellite laser ranging (Cheng and Tapley, 2004). Variations in the gravity field were computed by removing the long-term mean (computed over the entire study period, see Sect. 5) from each monthly solution. The TWS variations over the Rhine basin were then produced using the approach described by Wahr et al. (1998). Because of strong noise artefacts present in the high degree coefficients, a de-striping filter similar to that described in Swenson and Wahr (2006) was applied to each monthly solution. The filter used a 5th degree polynomial (Savitsky-Golay) over a 5-point window to remove the correlations, and orders below 8 remained unchanged. Further, an additional 250-km radius Gaussian smoothing (Jekeli, 1981) was applied. While this process helps to mitigate noise in the solution, it also attenuates genuine signal, so a scale factor is often applied in an effort to restore some of the signal that gets "leaked" out of the basin due to the spatial filtering. To that end, scale factors using the Global Land Data Assimilation System (GLDAS) hydrological model (Rodell et al., 2004) were computed following the method described by Landerer and Swenson (2012). The sum of four soil moisture layers (0 to 2 m) and a snow water equivalent layer from a monthly GLDAS NOAH Version 1 model was defined as the TWS. We (least squares) fitted the time series between the original and filter GLDAS at every grid node over the Rhine using only one scale factor. The estimated filtering scale factors varied between 0.98 and 1.02 over the Rhine River basin. The correction for glacial isotactic adjustment, which has been shown in other regions to affect the interpretation of long-term trends (Peltier, 2004), was determined to be small in our study, so the corresponding correction was not applied.

3.2 Forcing data

The forcing data required to drive the OpenStreams wflow_hbv model are precipitation, temperature and potential evapotranspiration (PET). Two types of forcing data were used in this study. “Local” forcing data indicates the best available data, and “global” forcing data indicates a lower quality dataset but one which is available globally or nearly globally.

In this study of the Rhine basin, local forcing data refer to meteorological data from the network of local weather stations, providing higher spatial and temporal resolution. Local precipitation and temperature data were retrieved from the European Climate Assessment & Data set (ECA&D) and ENSEMBLE project, known as E-OBS data (Haylock et al., 2008). Data collected from several hundred ground stations were combined to produce a daily grid of precipitation and mean surface temperature at a 0.25-degree spatial resolution. Local PET data were derived from climatological data obtained from the Commission for the Hydrology of the Rhine basin (CHR) and the German Meteorological Service (DWD) (Weerts et al., 2008). The daily local PET was interpolated from a monthly mean value with a fixed annual cycle and was available at a 1-km spatial resolution (Weerts et al., 2008; Photiadou et al., 2011).

Global precipitation and temperature data were obtained from Sheffield et al. (2005). These data are constructed based on the long-term near-surface meteorological variables from the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP/NCAR) reanalysis product. The daily global precipitation and temperature data were provided at a spatial resolution of 0.5-degree. For global PET, the 1-degree daily product generated by Senay et al. (2008) was used.

Fig. 4 shows a comparison between mean daily precipitation, temperature and PET in 2006 from the local and global forcing datasets. For the mean temperature, aside from the resolution difference, the spatial distribution and magnitude is very similar between the two datasets. On the other hand, significant differences can be seen between the local and global precipitation data, especially over the High Rhine. Differences are also observed in the PET products, with the global dataset having generally higher values than the local one, in addition to the much coarser spatial resolution of the global product.

3.3 Validation data

Groundwater and streamflow measurements from various networks are used to validate our estimated results.

3.3.1 Groundwater data

In situ groundwater measurements were obtained from 3 different networks:

1) Ministerium für Klimaschutz, Umwelt, Landwirtschaft, Natur- und Verbraucherschutz des Landes Nordrhein-Westfalen (http://www.elwasweb.nrw.de, last access: 5 March 2014)
2) Bayerisches Landesamt für Umwelt (http://www.gkd.bayern.de, last access: 5 March 2014)
3) Portail national d'Accès aux Donnéessur les Eaux Souterraines (ADES, http://www.ades.eaufrance.fr, last access: 17 March 2014)

Measurements that did not exhibit seasonal variations were flagged as belonging to confined aquifers, and were excluded. Data from stations with weekly measurements (e.g., ADES) were interpolated to daily intervals. A total of eighteen wells were used for validation. Their locations are shown in Fig. 1, and their names are provided in Table A1.

The in situ groundwater measurements were provided in the form of piezometric head. The variations in piezometric head can be related to variations in groundwater storage if the specific yield is known (Rodell et al., 2007). As the latter data were unavailable, the piezometric head was scaled to the units of GW storage based on other GW data. Previous studies have demonstrated that subtracting SM derived from GLDAS from GRACE was able to extract the groundwater component from GRACE in several regions e.g., North America (Rodell et al., 2006; 2007), Australia (Tregoning et al., 2012), the Middle East (Longuevergne et al., 2013), etc. We adopt a similar idea by using the relationship between ΔTWS-ΔSM (TWS variation from GRACE minus SM variation) and the observed head to scale the observed head. Ideally, we would prefer to use in-situ soil moisture data to represent the SM term, but they are not available at the well locations, and the nearest station from the International Soil Moisture Network (ISMN: Dorigo et al., 2011) does not have data covering the GRACE observation period. The soil moisture estimated from remote sensing was also not appropriate because the penetration depth depends on frequency and would not be the same as that in OpenStreams wflow_hbv. Therefore, we decided to use GLDAS-derived SM in this study. The SM variation from GLDAS (ΔSM_{GLDAS}) was computed by removing its long-term mean value. The long-term mean value was produced from all GLDAS SM data over the same period as the GRACE observations (see Sect. 5). The groundwater variations from GRACE ($\Delta \mathrm{GW}_{\mathrm{GRACE}}$) were obtained by removing ΔSM_{GLDAS} from the GRACE observations every month. $\Delta \mathrm{GW}_{\mathrm{GRACE}}$ was interpolated to daily values in order to compare it to the daily head variations Δh. The comparison was done using the following relationship:

$$\Delta \mathrm{GW}_{\mathrm{GRACE}} + e = a + b \cdot \Delta h \tag{1}$$

where e indicates the observation error. The two parameters a and b were estimated by least-squares regression. The scaled in situ GW variation ($\Delta GW_{\text{in situ}}$) were then obtained from the observed variations in piezometric head using:

$$\Delta \mathrm{GW}_{\mathrm{in-situ}} = \hat{a} + \hat{b} \cdot \Delta h \tag{2}$$

where $\hat{a}, \hat{b}$ are the parameters estimated from Eq. (1).

3.3.2 Streamflow data

Streamflow was validated using observations from the thirteen in situ gauges indicated in Fig. 1. Time-series were provided by the Hydrological Modelling Basis in the Rhine Basin (HYMOG; Bader et al., 2013). The hourly data were aggregated to daily data for this study.

4 Data assimilation

4.1 Ensemble Kalman Filter

The Ensemble Kalman Filter (EnKF) is used here to assimilate GRACE TWS into the OpenStreams wflow_hbv model. The EnKF uses a Monte Carlo approach: an ensemble of model states is integrated forward in time using the forward model. The update equation from the classical Kalman filter is used to update the model estimate, where the Kalman gain is determined using the error covariances calculated from the ensemble (Evensen, 1994). The EnKF and its variants are widely used because they are efficient, easy to implement and allow great flexibility in terms of model uncertainty (Evensen, 2003). In this study, we implement a so-called 1D-EnKF (De Lannoy et al., 2009) in which each grid cell is updated individually. The state equation in discrete form is given as:

$$\psi(t+1) = f\big(\psi(t), u(t+1), \alpha, w(t)\big) \tag{3}$$

where f is the model operator, ψ is the state variables, u is the forcings, α is the model parameters, and w is the model error. In this paper, the state variables (ψ) are an n×1 vector of TWS from OpenStreams wflow_hbv. The observations available at a measurement time t are gathered in a vector of observations d (TWS from GRACE):

$$d(t) = H\psi(t) + \epsilon; \epsilon \sim \mathcal{N}(0, R) \tag{4}$$

where d is an m×1 vector containing the observations, H is measurement operator which relates the state $\psi(t)$ to the measured variables $d(t)$. In this study, the observation and the state vector are TWS, so n=m=1 and H is the unit matrix. The uncertainties in the observations are given in the random error ϵ, which is assumed to have zero mean and covariance matrix R. In the initialization phase, the EnKF is initialized by generating an ensemble (i) of N realizations of the state vector $\psi_i(t)$, i=1,…,N around a nominal $\psi(t)$. This reflects the prior knowledge of the state at the initial time. The EnKF moves sequentially from one observation time to the next and works in two steps, a forecast step and an update step. At the updated time t (when the observation is available), an ensemble of perturbed observations, $d_i(t)$ is generated as:

$$d_i(t) = d(t) + \epsilon_i(t), \tag{5}$$

where ϵ_i denotes the perturbation of the error of each ensemble member i. If the ensembles of the variables are stored in a matrix $A = (\psi_1, \psi_2, \psi_3, \dots, \psi_N)$, the ensemble perturbation matrix can be defined as $A' = A - \bar{A}$ where $\bar{A}$ is the mean computed from all ensemble members. Similarly, the ensemble members of the observation and perturbations are gathered into the matrices $D = (d_1, d_2, d_3, \dots, d_N)$ and $\gamma = (\epsilon_1, \epsilon_2, \epsilon_3, \dots, \epsilon_N)$. The analysis equation can be expressed as (Evensen, 2003):

$$A^a(t) = A(t) + A'(t){A'}^T(t)H^T\left(HA'(t){A'}^T(t)H^T + \gamma\gamma^T\right)^{-1}\left(D(t) - HA(t)\right) \tag{6}$$

where A^a is the analyzed model state

4.2 Assimilating GRACE observations

Several steps must be taken before GRACE TWS can be assimilated into OpenStreams wflow_hbv. GRACE observations represent average TWS variations over one month, while the OpenStreams wflow_hbv model has a daily time step. In this study, it is assumed that the average TWS corresponds to the middle of the month. Then, spline interpolation between consecutive months is used to generate a time series of GRACE TWS variations at five-day intervals. The five-day interval was chosen through trial-and-error to be a good compromise between allowing the ensemble to grow between updates and avoiding implausible jumps. As in any land surface assimilation application, the update results in discontinuities as mass is added or removed from the state but these are not large enough to be obvious when a five-day interval is used (see Sect. 5.1). If the update took place at larger time interval (e.g., once a month) and the entire increment was applied on one day, more significant artefacts or temporal discontinuities would occur (Widiastuti, 2009). In order to convert GRACE variations to absolute values the mean TWS in the study period was calculated from the nominal OpenStreams wflow_hbv run and added to the GRACE time series.

GRACE observes total TWS, some components of which can be neglected (e.g., nominal OpenStreams wflow_hbv simulations indicate that surface water and interception storage contributed by less than 1 % to the estimated TWS). Snow is also small averaged over the study area (approximately 2% to the estimated TWS in winter). Only over the Alps (see Fig. 1) is the snow contribution greater (approximately 7%). Therefore, we decided to exclude the snow from the state vector. To reconcile GRACE to OpenStreams wflow_hbv TWS, we then removed the snow component estimated from the nominal run from the GRACE prior to assimilation. Note that in catchments where the snow component is more significant, it should not be excluded from the state vector.

In the EnKF, the GRACE TWS are calculated and assimilated at each 1-km model grid cell every five days. Because the analyzed model state $A^{a}(t)$ was an integrated value of TWS, the increment ($\Delta A(t) = A^{a}(t) - A(t)$) for every ensemble member needed to be disseminated among the three stores, SM, UZ, and LZ. The information about the distribution of the increment among the different model compartments could be obtained directly from the Kalman filter. However, we chose to carry out the vertical distribution in the way consistent with the OpenStreams wflow_hbv model (Fig. 2). While the SM and LZ stores have upper bounds determined by model parameters, UZ does not. As a result, allowing it to update freely in the EnKF runs the risk that it becomes excessively large, which would also have a detrimental effect on runoff. Therefore, the increment is used to adjust the SM first, subject to the upper and lower limits of zero and fc. Any remaining increment is applied in turn to LZ, up to its upper limit, and only then to UZ.

The GRACE observation error is assumed to be 20 mm and horizontal observation error correlations are not considered. The 20 mm value is considered realistic as it was suggested by several independent assessments e.g., Klees et al. (2008), Wahr et al. (2006), Schmidt et al.

(2008) and it also had been applied in previous GRACE assimilation studies (Zaitchik et al., 2008; Houborg et al., 2012). Our philosophy was to set the GRACE errors to realistic values determined from independent studies, so that the solutions were not guided towards any particular outcome.

4.3 Uncertainty in model forcing data and parameters

In the EnKF, stochastic noise can be included in model forcing data and parameters to account for model uncertainty. An earlier sensitivity study (Widiastuti, 2009) was conducted to identify the parameters of the OpenStreams wflow_hbv model that had a significant impact on TWS. Six such parameters, which include *fc*, *lp*, *β*, *cflux*, *khq*, and *perc* were found. Therefore, the soil moisture routine parameters, *fc*, *lp* and *β*, as well as the runoff routine parameters, *cflux*, *khq* and *perc*, were perturbed. For the “calibrated” case, the calibrated model parameters in each grid cell were perturbed using additive Gaussian noise, with a mean of zero and a standard deviation equal to 10% of the range of values that occurred over the whole Rhine basin. In the “non-calibrated” case, the mean parameter value in each grid cell was set to the average calibrated value across the whole basin, and the standard deviation was set to that of the calibrated parameter across the whole basin. This was considered as a proxy for assigning approximate values based on the land cover type, topography, and climatology from the globally available databases. Averaging each parameter across the entire Rhine basin is intended merely to reflect this kind of first-order assumption. Though not all OpenStreams wflow_hbv parameters can be gleaned from such global databases, we the averaged values could be compared to those in the Food and Agriculture Organization of the United Nations (FAO) database (http://www.fao.org/geonetwork/srv/en/main.home, last access: 5 December 2014). The areally averaged parameter values over the Rhine were found to be within the range the provided by FAO. For example, the areally averaged soil moisture field capacity over the Rhine FAO provided is mostly between 150 and 200 mm, while the areally averaged value of approximately 180 mm is used as a mean in this study with a standard deviation of 33 cm. The meteorological forcing data were also varied, with the temperature data being perturbed with additive Gaussian noise, and the precipitation and PET being perturbed with additive lognormal noise. In the “local forcing data” case, noise with standard deviation based on 10 % of the nominal value was added to precipitation while 15 % noise was added to temperature and PET. For the “global forcing data” case, we assumed that the local forcing data were accurate and reliable, and the differences between the local and global forcing data represent the errors of global forcing data. The errors were assumed to be spatially correlated, so an exponential correlation function was applied to the covariance matrix for each variable. The correlation lengths for precipitation, temperature and PET were determined using variogram analysis (Widiastuti, 2009) and found to be 21 km, 21 km, and 59 km, respectively.

Recall from Sect. 1 and Sect. 3.2, that four cases are considered in this study: 1) calibrated parameters with local forcing data (CL), 2) calibrated parameters with global forcing data (CG), 3) non-calibrated parameters with local forcing data (NCL), and 4) non-calibrated parameters with global forcing data (NCG). Comparison of the four scenarios provides insight into the benefit of GRACE assimilation under different degrees of uncertainty. The lowest and highest levels of uncertainty are associated with the CL and the NCG cases.

5 Results and discussion

Using the EnKF approach described above, GRACE observations were assimilated into the OpenStreams wflow_hbv model. An ensemble of 100 model states was propagated forward from 1 Jan 2001 to 30 Nov 2003 to spin up the model. The ensemble state at the end of the spin-up period provided the initial state for the assimilation. The study period is from 1 Dec 2003 to 31 Oct 2007 because the observed streamflow was only available until Autumn 2007.

5.1 Impact of GRACE assimilation on TWS estimates

First, the impact of assimilating GRACE on the temporal and spatial patterns of the estimated TWS is considered. For the temporal pattern, the areal mean of the estimated TWS over the entire Rhine River basin was computed. The time series of TWS variations from the ensemble open loop (EnOL, ensemble run without GRACE assimilation), EnKF, and GRACE observations are shown in Fig. 5.

As expected, there is a seasonal cycle in the TWS estimates, which varies between ±75 mm. The high frequency variations in TWS in the CL and NCL that are not apparent in CG and NCG are due to the coarser spatial resolution of the global precipitation product. Lower spatial variability of the global data causes smoother averaged TWS presented in the CG and NCG time series. During the summer of 2006 (June, July, August: JJA), the areal mean global and local precipitation and temperature products agree. However, the global PET product estimates an areal mean PET of 4.10 mm/day while the local PET data suggest it was 2.89 mm/day. As the result, the minimum TWS in the CL and NCL cases in the EnOL is -69 mm while CG and NCG are close to -90 mm. In this period, GRACE assimilation has little impact on CL and NCL, but results in a significant (25 mm) update in TWS in the CG and NCG cases. The largest difference between the EnOL and EnKF occurs when TWS is increasing (for example, October 2005). This is apparent in all cases, but is greatest in the two non-calibrated cases. In all cases, Fig. 5 shows that assimilation draws the TWS estimate toward the GRACE observation.

The impact of GRACE assimilation also varies within the basin. Fig. 6 shows the spatial distribution of the average increment (posterior minus prior) in TWS during winter (December, January, February: DJF, 2005-2006) and summer (JJA) of 2006. During the winter (left), the EnKF estimated wetter conditions over entire Rhine River basin when the local forcing data were used. In the Alps, the global precipitation product is approximately 35% higher than the local precipitation product. Therefore, GRACE assimilation reduced the TWS estimate over the Alps in the CG and NCG cases. During the summer (right), GRACE assimilation reduced the TWS estimate over the Alps and Neckar basin when local forcing data were applied, but adds moisture in the global data case. In this period, the local PET product is 66% lower than the global product over the Alps and 44% lower over the Lahn basin. This is consistent with the increase in areal averaged TWS observed in the CG and NCG cases in Fig. 5. Since the local precipitation data are generally considered to be more accurate, the adjustment of the TWS estimates towards those produced by the local product is an excellent example of the benefit of GRACE assimilation, particularly in data sparse areas.

In the Regnitz basin (east of domain), GRACE assimilation leads to a significant increase in TWS in both calibrated cases during the winter months. In this basin, the upper zone recession coefficient (*khq*) is 0.52 in the calibrated case, compared to 0.3 in the non-calibrated case. This results in almost twice as much fast runoff in the calibrated case, which depletes the terrestrial water storage in the winter months. GRACE assimilation adds moisture to the UZ and LZ stores, drawing the TWS closer to the GRACE observations.

In the summer, an average of 0.7 and 1.07 mm was removed in each update from the southern part of Moselle basin in the CL and CG cases, respectively (Fig. 6(b) and (d)), compared to 0.74 mm and 1.25 mm added per update in the NCL and NCG cases. In the two calibrated cases, the evaporation threshold value (the product of *fc* and *lp*) is approximately 11 % less than that in the non-calibrated cases. This leads to less soil evaporation and higher soil moisture in the calibrated cases. GRACE assimilation reduces the SM in the calibrated cases, and increases it in the non-calibrated cases to draw the TWS closer to the GRACE observations in all cases.

5.2 Impact of GRACE assimilation on GW estimates

The TWS and GW variations from OpenStreams wflow_hbv were computed at every grid cell. The estimates at the Sundern and A319C wells are shown in Fig. 7 and 8. The two stations represent the behaviour of the other 16 stations (detailed below). For example, stations 2, 3, 4, 6, 9, 10, 11, 13, and 18 have similar behaviour to Sundern, while the rest have similar behaviour to A319C station. Recall that GW is defined as the sum of UZ and LZ, so the difference between the left and right columns is the SM term. GRACE measures monthly variations, so the monthly mean of TWS, GW estimates and the in situ data are shown. Similar to the areal mean values, the TWS from the EnKF in the individual grid cells (left column) is generally between the values from the EnOL and those observed by GRACE.

At Sundern (Fig. 7) in the CG and NCG cases, the impact of the forcing data was seen in the summer of every year. Table 2 shows that the precipitation, temperature and PET at Sundern were higher in the global forcing data than in the local data. Fig. 7(c) and 7(g) suggest that this leads to a more negative estimate of TWS in the EnOL for the CG and NCG cases. In the EnKF results, these TWS estimates are drawn towards the GRACE observations. The corresponding updates in terms of GW are larger in the global forcing data case than in the local forcing data cases - assimilation added approximately 5-10 mm of water to GW in the global data cases. Similar behaviour was also seen in CL and NCL cases in summer 2005.

At Sundern, the estimated GW in the CL case agrees quite well with the in situ values, suggesting that the distribution between the SM and GW components is reasonable in the calibrated cases. The fact that a good estimate of TWS does not result in an improved GW estimate indicates that the non-calibrated parameters are leading to an incorrect distribution of the TWS between the different stores. In the NCL and NCG cases, *fc* is just 179 mm compared to the calibrated value of 239 mm. So, for the same TWS value, the non-calibrated cases have more water in GW than the calibrated cases. As a result, despite the agreement in TWS in the winter months, the GW variation is considerably overestimated.

In every case at the A319C well location (Fig. 8), the EnOL estimated lower TWS in the first half of 2004 and 2006, and higher in the second half of the same years. Assimilation updated the TWS toward GRACE observation in these periods and resulted in better agreement between the assimilated and observed GW. In late-2005, the estimated TWS from the EnOL and EnKF are very close to the GRACE observations. However, the estimated GW in both cases is a lot lower than that observed in situ. As discussed, the difference between the two is soil moisture. The model is predicting a significant increase in soil moisture in all four cases. However, given there is little to improve in terms of TWS, the GW estimate from the EnKF is as bad as that from the EnOL.

The impact of the forcing data used is also presented. In CG and NCG cases, on 3 Oct and 23 Oct 2006, underestimated global precipitation caused the underestimated GW. GRACE could not correct such a high frequency event due to the limitation of its temporal resolution.

The choice of the parameters plays a role in the estimated GW magnitude (as seen in Fig. 7), but now the non-calibrated parameters (compared to the calibrated ones) provided closer values to the in situ data (Fig. 8(f) and (h)). Higher non-calibrated fc parameter (see Table 3 for the values) was responsible for smaller GW estimates.

Tables 4 and 5 show the correlation coefficient and RMS error (RMSE) between the estimated and in situ GW for all eighteen well locations indicated on Fig. 1. These were calculated based on the monthly mean, but similar results were obtained using the daily values. In most cases, assimilation leads to an increase in correlation coefficient and a reduction in RMSE.

The results varied across the wells. The highest correlation coefficients in the EnOL simulations were typically found in the CL case, followed by the NCL. Clearly, using the local forcing data has a significant impact in resolving features at a single grid cell. An exception is the Main basin (wells 5, 7-10) where the global forcing data produce TWS more consistently with the GRACE observations and hence result in a better agreement with the GW. The highest correlation coefficients in the EnKF cases are also found in the two local data cases. The improvements in correlation coefficient are seen in all four cases. The CL and NCL cases also yield the lowest RMSE values in the EnOL case, and the results with the EnKF are very mixed.

It is important to note that at many wells, the NCL and NCG cases yield higher correlation coefficients than the CL and CG cases, respectively. Recall that the model is calibrated using streamflow, not groundwater data. So, while assimilation draws the modelled TWS towards the GRACE observations, the model parameters have a significant impact on whether or not this translates to an improvement in GW estimate.

One of the objectives was to examine the potential value of GRACE assimilation in data-sparse regions. In the NCG case, it is encouraging that GRACE assimilation consistently leads to an increase in correlation coefficient (up to 33 %) and reduction in RMSE (up to 35 %). In other scenarios, assimilation of GRACE observations also leads into an increase in correlation coefficient (up to 71%, at station 11 in the CG case) and a decrease in RMSE (up to 35 %, at station 1 in the NCG case). In average, correlation and RMSE improvements in groundwater estimates for all cases evaluated are 13 % and 14 %, respectively.

5.3 Impact of GRACE assimilation on streamflow estimates

The estimated and observed streamflows at Maxau (upstream) and Wessel (downstream) gauge stations are shown in Fig. 9 and 10. Accurate forcing data, particularly precipitation, are essential for reproducing the observed streamflow. The high frequency variations in streamflow associated with fast response to local precipitation are often reproduced reasonably well in the CL case, but not in the CG case (compare Fig. 9(a) to 9(b) and 10(a) to 10(b)).

Use of the global data frequently underestimates the streamflow. This is clear on 5 Jun 2004, 24 Aug 2005, 6 Oct 2006, and 10 Aug 2007 in Fig. 9(b) and 9(d). Comparing Fig. 9(a) to 9(b), it is clear that the larger peaks in streamflow are poorly estimated when the global data are used. Because GRACE observations describe monthly variations over a larger area, they can do little to capture these individual streamflow events. By correcting TWS, GRACE assimilation mainly influences the longer term variations. The difference between EnOL and EnKF is very small in the CL case. The largest differences are observed in the CG and NCG cases, where TWS is updated to correct for errors in forcing data (e.g., summer 2004 and 2006 in Fig. 9).

Fig. 11 shows the impact of GRACE assimilation on the correlation coefficient, Nash-Sutcliffe coefficient (NS) (Nash and Sutcliffe, 1970), and RMSE in streamflow. Results are shown for four gauge stations along the main channel, as well as the average value across all thirteen stations. These results underscore the importance of forcing data and calibration for estimating streamflow. By far, the highest correlation coefficients and Nash-Sutcliffe coefficients and lowest RMSEs are obtained when local forcing data are used. Use of global forcing data leads to a significant loss in performance. For example, using global rather than local forcing data with the calibrated model results in a decrease in correlation coefficient from 0.89 to 0.65, a decrease in Nash-Sutcliffe coefficient from 0.76 to 0.35 and an increase in RMSE of 71 % in the EnKF results. Using the non-calibrated model rather than the calibrated model also leads to poorer performance, though to a lesser degree. For example, using the non-calibrated rather than calibrated model with the local forcing data results in a decrease in correlation coefficient from 0.89 to 0.88, a decrease in Nash-Sutcliffe coefficient from 0.76 to 0.65 and an increase in RMSE of 23 % in the EnKF results.

Compared to the differences due to forcing data and calibration, GRACE assimilation leads to a relatively modest improvement in streamflow estimates. In terms of correlation coefficient, the largest improvements on average (Avg column) are found when the global forcing data are used. The correlation coefficient increased from 0.64 to 0.65 in the CG case, and 0.65 to 0.66 in the NCG case. The largest improvement at an individual station was found at Maxau where assimilation resulted in an increase in correlation coefficient from 0.54 to 0.59 in the NCG case.

Similarly, GRACE assimilation leads to a modest improvement in terms of NS coefficient. The largest average improvement was from 0.62 to 0.65 in the NCL case. GRACE assimilation slightly reduced the RMSE in all 4 cases. The greatest reduction is 4 % in the NCL case.

Though it is encouraging that GRACE assimilation improved the estimated streamflow, these results demonstrate that it clearly cannot replace high quality forcing data or good model calibration.

6 Conclusions

The first goal of this study was to investigate the impact of assimilating GRACE into the OpenStreams wflow_hbv model on the estimated terrestrial water storage, groundwater storage and streamflow in the Rhine river basin. GRACE observations were assimilated into each grid cell of the model with an EnKF to update the soil moisture and upper and lower zone storage terms of the model. In general, assimilation drew the EnOL estimated TWS closer to the GRACE observations. In the absence of independent TWS observations, a qualitative analysis of the increments in TWS indicated that GRACE assimilation could partially correct the TWS estimate for the influence of errors in the meteorological forcing data and model parameters. As result, an improvement in groundwater estimate after assimilating GRACE data was noticeable, with an overall improvement up to 71% (correlation coefficient) and 35% (RMS error) over the EnOL case. However, it is found that the improvement in TWS estimates did not always translate to an improved agreement between the estimated and observed groundwater storage variation at certain well locations. The differences may be due to the OpenStreams wflow_hbv parameters: if the upper limit on soil moisture storage is too high (low), then the groundwater variations could be under (over)-estimated. This is particularly relevant in the type of model where the calibration is per sub-basin. This does not allow for local differences on the order of single or a few grid cells. The issue of scale is also significant because GRACE observes monthly variations on the order of hundreds of kilometres. Groundwater variations can be influenced by local features at finer scales. When the basin average is considered, validation against a denser network of well data or an independent groundwater model could be used to determine if an improvement occurs at the scale of the entire basin.

Furthermore, the considered model was used to simulate runoff. The groundwater terms, UZ and LZ, primarily serve as reservoirs for quick and base runoff generation. Due to the coarse resolution of the observations, GRACE assimilation resulted in only a modest improvement in streamflow estimates. Correlation coefficients increased by up to 2 %, Nash Sutcliffe coefficients increased by up to 4 % and RMSE was reduced by up to 4 %.

The second goal of this study was to investigate the potential value of assimilating GRACE observations in data-sparse regions. Results from four scenarios were compared in which the ensemble mean model parameters were either calibrated values, or basin average values and the meteorological forcing data were either local (high quality) data or global (poorer quality) data. By comparing the four cases, it was shown that GRACE assimilation could correct for errors in model forcing data and parameter calibration by drawing the estimated TWS toward that observed by GRACE. This also resulted in drawing the estimated groundwater storage closer to the in situ measurement. Given that the most significant improvements were observed in the NCG case, this suggests that GRACE observations are most valuable in data sparse regions. In these regions any additional observations, even those at coarse spatial and/or temporal resolution, are welcome. GRACE can provide essential independent observations for validation, and serves as a constraint for TWS within the assimilation process. In terms of streamflow, a comparison of the four scenarios demonstrates that the ability to capture high flow events is determined largely by the quality of the forcing data and the model parameters. The improvements in streamflow estimates after assimilation are modest. Nevertheless, we consider the obtained results as promising, particularly in data-sparse scenarios, e.g., the NCG case. They indicate that GRACE contains information that can be useful for streamflow estimation. Whether updating TWS is the best way to use this information is an open question. An alternative strategy could be to use GRACE assimilation for parameter estimation at a sub-basin or basin scale and constrain the rainfall-runoff model through assimilation of soil moisture observations.

In conclusion, GRACE assimilation is beneficial, and the largest improvements are generally observed in the NCG (i.e. “data-sparse”) cases. In addition to providing a modest improvement to the estimated streamflow, it may result in a noticeable improvement in TWS estimates, yielding an extra insight into the behaviour of the hydrological model, its forcing data and parameters. Further research will combine assimilation of GRACE and a soil moisture remote sensing product to constrain the SM estimate storage term, and ensure that improved TWS would lead to more consistently improved estimates of groundwater storage variations. Further research will also explore the value of assimilating GRACE into a groundwater model in which the primary processes of interest vary on temporal and spatial scales similar to those of GRACE. In addition, recent studies have explored the effect of spatial aggregation of GRACE TWS prior to assimilation (Forman and Reichle, 2013) as well as inclusion of the full GRACE error structure (Eicker et al., 2014). Combining the advances made in those studies with the assimilation framework presented here is expected to yield even more realistic estimates. As shown by De Lannoy et al. (2009), working with a spatially distributed state vector (3D-EnKF) can lead to an improved estimate. Given the coarse resolution of GRACE, we expect that implementing a 3D-EnKF within the assimilation framework would lead to an improved performance. This could be particularly important in small basins like the Rhine, and can be used to account for the fact that the GRACE overpasses are infrequent and may not sensitive to TWS variations in response to specific events.

Appendix A: Names of well locations

Table A1: Names of the well locations used in this paper.

Location number	Name	Source
1	Sundern	Ministerium für Klimaschutz, Umwelt, Landwirtschaft, Natur- und Verbraucherschutz des LandesNordrhein-Westfalenb(http://www.elwasweb.nrw.de)
2	GEW KOELN 557	
3	SHELL GODORF GW I	
4	LGD BN-BEUEL	
5	Stetten S1	Bayerisches Landesamt für Umwelt(http://www.gkd.bayern.de)
6	Dietersdorf	
7	Haβfurt Q2	
8	Limbach Q1	
9	Rattelsdorf 136	
10	Faulbach	
11	01373X0130/A25	Portail national d’Accès aux Donnéessur les Eaux Souterraines(http://www.ades.eaufrance.fr)
12	02303X0065/P	
13	02307X0281/S	
14	01995X0030/563	
15	02344X0082/326E	
16	02344X0055/319	
17	02348X0009/319C (called A319C in this paper)	
18	03426X0197/136	

Acknowledgements

This research was funded by The Netherlands Organisation for Scientific Research, NWO (project number 842.00.006). The work of S. C. Steele-Dunne was supported by the NWO Veni Grant Program (ALW 863.09.015). We would like to thank Eric Sprokkereef Secretary of the International Commission for the Hydrology of the Rhine basin for granting permission for the use of hydrologic data. Jaap Schellekens (Deltares) is acknowledged for help with the OpenStreams model of the Rhine. The authors also want to thank Edwin Sutanudjaja for productive discussions about the groundwater data.

References

Bader, S., Belz, J. U., Jakob, A., Mathis, C., Morf-Graf, S., Mürlebach, M., and Schürch, M.: International commission for the hydrology of the Rhine basin (CHR): Annual CHR report 2012, Sprokkereef, E, Annual CHR report 2012, Lelystad, the Netherlands, 2013.

Becker, M., Meyssignac, B., Xavier, L., Cazenave, A., Alkama, R., and Decharm, B.: Past terrestrial water storage (1980–2008) in the Amazon Basin reconstructed from GRACE and in situ river gauging data, Hydrol. Earth Syst. Sci., 15, 533–546, doi:10.5194/hess-15-533-2011, 2011.

Bettadpur, S.: Insights into the Earth System mass variability from CSR-RL05 GRACE gravity fields, EGU Meeting, abstract #EGU2012-6409, Vienna, Austria, 2012.

Bierkens, M. F. P, and van den Hurk, B. J. J. M.: Groundwater convergence as a possible mechanism for multi-year persistence in rainfall, Geophys. Res. Lett., 34, L02402, doi:10.1029/2006GL028396, 2007.

Cheng, M., and Tapley, B.: Variations in the Earth's oblateness during the past 28 years, J. Geophys. Res., 109(B09402), doi:10.1029/2004JB003028, 2004.

de Jeu, R. A. M., Wagner, W., Holmes, T. R. H., Dolman, A. J., van de Giesen, N. C., Friesen, J.: Global Soil Moisture Patterns Observed by Space Borne Microwave Radiometers and Scatterometers, Surv. Geophys., 28, 399-420 doi 10.1007/s10712-008-9044-0, 2008.

De Lannoy, G. J. M., Reichle, R. H., Houser, P. R., Arsenault, K. R., Verhoest, N. E. C., and Pauwels, R. N.: Satellite-Scale Snow Water Equivalent Assimilation into a High-Resolution Land Surface Model, J. Hydrometeor, 11, 352–369, doi:http://dx.doi.org/10.1175/2009JHM1192.1, 2009.

Dorigo, W. A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., Drusch, M., Mecklenburg, S., van Oevelen, P., Robock, A., and Jackson, T.: The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements, Hydrol. Earth Syst. Sci., 15, 1675-1698, 2011.

Eberle, M., Buiteveld, H., Beersma, J., Krahe, P., and Wilke, K.: Estimation of extreme floods in the river Rhine basin by combining precipitation-runoff modelling and a rainfall generator, in: Proceedings International Conference on Flood Estimation, Berne, CHR report II-17, International Commission for the Hydrologyof the Rhine basin (CHR), lelystad, The Netherlands, 2002.

Eberle, M., Buiteveld, H., Krahe, P., and Wilke, K.: Hydrological Modelling in the River Rhine Basin, Part III: Daily HBV model for the Rhine basin, Report 1451, Bundesanstalt für Gewässerkunde (BFG), Koblenz, Germany, 2005.

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., Entin, J. K., Goodman, S. D., Jackson, T. J., Johnson, J., Kimball, J., Piepmeier, J. R., Koster, R. D., Martin, N., McDonald, K. C., Moghaddam, M., Moran, S., Reichle, R., Shi, J. C., Spencer, M. W., Thurman, S. W., Leung T., van Zyl, J.: The Soil Moisture Active Passive (SMAP) Mission, Proceedings of the IEEE, 98, 704–716, doi:10.1109/JPROC.2010.2043918, 2010.

Eicker, A., Schumacher, M., Kusche, J., Döll, P., Müller Schmied H.: 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, 2014.

Evensen, G.: Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., 99(C5), 10143–10162, 1994.

Evensen, G.: The ensemble Kalman filter: Theoretical formulation and practical implementation, Ocean Dyn., 53(4), 343-367, doi:10.1007/S10236-003-0036-9, 2003.

Ferrant, S., Caballero, Y., Perrin, J., Gascoin, S., Dewandel, B., Aulong, S., Dazin, F., Ahmed, S., and Maréchal J.-P.: Projected impacts of climate change on farmers' extraction of groundwater from crystalline aquifers in South India, Nature Sci. Rep., 4, 3697, doi:10.1038/srep03697, 2014.

Forman, B. A., Reichle, R. H., and Rodell, M.: Assimilation of terrestrial water storage from GRACE in snow-dominated basin, Water Resour. Res., 48, W01507, doi:10.1029/2011WR011239, 2012.

Forman, B. A. and Reichle, R.: 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, 2013.

Green, T. R., Taniguchi, M., Kooi, H., Gurdak, J. J., Allen, D. M., Hiscock, K. M., Treidel, H., and Aureli, A.: Beneath the surface of global change: Impacts of climate change on groundwater, J. Hydrol., 405, 532-560, doi:10.1016/j.jhydrol.2011.05.002, 2011.

Güntner, A.: Improvement of global hydrological models using GRACE data, Surv. Geophys., 29, 375-397, doi:10.1007/s10712-008-9038-y, 2008.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M.: A European daily high-resolution gridded dataset of surface temperature and precipitation, J. Geophys. Res., 113, D20119, doi:10.1029/2008JD010201, 2008.

Hirschi, M., Seneviratne, S. I., Hagemann, S., and Schär, C.: Analysis of seasonal terrestrial water storage variations in regional climate simulations over Europe, J. Geophys. Res., 112, D22109, doi: 10.1029/2006JD008338, 2007.

Houborg, R., Rodell, M., Li, B., Reichle, R., and Zaitchik, B. F.: 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, 2012.

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E., and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB) - A review. Hydrolog. Sci. J., 58(6), 1198-1255, doi: 10.1080/02626667.2013.803183, 2013.

Jekeli, C.: Alternative methods to smooth the Earth’s gravity field, Rep., 327, Dept. of Geod. Sci. and Surv., Ohio State Univ., Columbus, 1981.

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., Bonan, G., Cescatti, A., Chen, J., de Jeu, R., Dolman, A. J., Eugster, W., Gerten, D., Gianelle, D., Gobron, N., Heinke, J., Kimball, J., Law, B. E., Montagnani, L., Mu, Q., Mueller, B., Oleson, K., Papale, D., Richardson, A. D., Roupsard, O., Running, S., Tomelleri, E., Viovy, N., Weber, U., Williams, C., Wood, E., Zaehle, S., and Zhang, K.: Recent decline in the global land evapotranspiration trend due to limited moisture supply, Nature, 467, 951–954, doi:10.1038/nature09396, 2010.

Karssenberg, D., Schmitz, O., Salamon, P., De Jong, K., and Bierkens, M. F. P.: A software framework for construction of process-based stochastic spatio-temporal models and data assimilation, Environ. Modell. Softw., 25, 489-502, 2009.

Kerr, Y. H., Waldteufel, P., Richaume, P., Wigneron, J. P., Ferrazzoli, P., Mahmoodi, A., Bitar, A. A., Cabot, F., Gruhier, C., Juglea, S. E., Leroux, D., Mialon, A., and Delwart, S.: The SMOS Soil Moisture Retrieval Algorithm, IEEE Trans. Geosci. Remote Sens., 50, 1384-1403, doi:10.1109/TGRS.2012.2184548, 2012.

Klees, R., Liu, X., Wittwe, T., Gunter, B. C., Revtova, E. A., Tenzer, R., Ditmar, P., Winsemius, H. C., and Savenije, H. H. G.: A Comparison of Global and Regional GRACE Models for Land Hydrology, Surv. Geophys., 29, 335-359, doi:10.1007/s10712-008-9049-8, 2008.

Kleinherenbrink, M., Ditmar, P. G., and Lindenbergh, R. C.: Retracking Cryosat data in the SARIn mode and robust lake level extraction, Remote Sens. Environ., 152, 38-50, doi:10.1016/j.rse.2014.05.014, 2014.

Koster, R. D., Suarez, M. J., Ducharne, A., Stieglitz, M., and Kumar, P.: A catchment-based approach to modeling land surface processes in a general circulation model. 1. Model structure, J. Geophys. Res., 105, 24809-24822, 2000.

Koster, R. D., Manhanama, S. P. P., Yamada, T. J., Balsamo, G., Berg, A. A., Boisserie, M., Dirmeyer, P. A., Doblas-Reyes, F. J., Drewitt, G., Gordon, C. T., Guo, Z., Jeong, J.-H., Lawrence, D. M., Lee, W.-S., Li, Z., Luo, L., Malyshev, S., Merryfield, W. J., Seneviratne, S. I., Stanelle, T., van den Hurk, B. J. J. M., Vitart, F., and Wood, E. F.: Contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment, Geophys. Res. Lett., 37, L02402, doi:10.1029/2009GL041677, 2010.

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

Lettenmaier, D. P., and Famiglietti, J. S.: Water from on high, Nature, 444, 562–563, 2006.

Li, B., Rodell, M., Zaitchik, B. F., Reichle, R. H., Koster, R. D., and van Dam, T. M.: 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, 2012.

Lindström, G., Johansson, B., Persson, M., Gardelin, M., and Bergström, S.: Development and test of the distributed HBV-96 hydrological model, J. Hydrol., 201(1-4), 272-288, 1997.

Longuevergne, L., Wilson, C. R., Scanlon, B. R., and Crétaux, J. F.: GRACE water storage estimates for the Middle East and other regions with significant reservoir and lake storage, Hydrol. Earth Syst. Sci., 17, 4817-4830, doi:10.5194/hess-17-4817-2013, 2013.

Mülders, R., Parmet, B., and Wilke, K.: Hydrological Modelling in the river Rhine basin, final report, Report 1215, Bundesanstalt für Gewässerkunde (BFG), Koblenz, Germany, 1999.

Nash, J. E., and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, J. Hydrol., 10(3), 282–290, 1970.

Orsolini, Y. J., Senan, R., Balsamo, G., Doblas-Reyes, F. J., Vitart, F., Weisheimer, A., Carrasco, A., and Benestad, R. E.: Impact of snow initialization on sub-seasonal forecasts, Clim. Dyn., 41, 1969-1982, doi:10.1007/s00382-013-1782-0, 2013.

Peltier, W. R.: Global glacial isostasy and the surface of the ice-age Earth: The ICE-5G (VM2) model and GRACE, Annu. Rev. Earth Planet. Sci., 32, 111-149, 2004.

Phan, V. H., Lindenbergh, R. C., and Menenti, M.: ICESat derived elevation changes of Tibetan lakes between 2003 and 2009, Int. J. Appl. Earth Obs. Geoinf., 17, 12-22, doi:10.1016/j.jag.2011.09.015, 2012.

Photiadou, C. S., Weerts, A. H., van den Hurk, B. J. J. M.: Evaluation of two precipitation data sets for the Rhine River using streamflow simulations, Hydrol. Earth Syst. Sci., 15, 3355–3366, doi:10.5194/hess-15-3355-2011, 2011.

Rakovec, O., Weerts, A. H., Hazenberg, P., Torfs, P. J. J. F., and Uijlenhoet, R.: State updating of a distributed hydrological model with Ensemble Kalman Filtering: effects of updating frequency and observation network density on forecast accuracy, Hydrol. Earth Syst. Sci., 16, 3435-3449, doi:10.5194/hess-16-3435-2012 , 2012.

Reichle, R. H.: Data assimilation methods in the Earth sciences, Adv. Water Resour., 31, 1411-1418, doi:10.1016/j.advwatres.2008.01.001, 2008.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The global land data assimilation system, Bull. Amer. Meteor. Soc., 85(3), 381–394, 2004.

Rodell, M., Chen, J., Kato, H., Famiglietti, J. S., Nigro, J., and Wilson, C. R.: Estimating groundwater storage changes in the Mississippi River basin (USA) using GRACE, Hydrogeol. J., 15, 159-166, 2007.

Rodell, M., Velicogna, I., and Famiglietti, J. S.: Satellite-based estimates of groundwater depletion in India, Nature, 460, 999-1002, 2009.

Schellekens, J.: OpenStream wflow documentation release 1.0RC1, Deltares, http://schj.home.xs4all.nl/html (last access: 5 January 2015), 2014.

Senay, G. B., Verdin, J. P., Lietzow, R., and Melesse, A. M.: Global daily reference evapotranspiration modeling and evaluation, J. Am. Water Resour. Assoc., 44(4), 969-979, doi:10.1111/j.1752-1688.2008.00195.x, 2008.

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., and Teuling, A.: Investigating soil moisture-climate interactions in a changing climate: A review, Earth-Sci. Rev., 99, 125-161, doi:10.1016/j.earscirev.2010.02.004, 2010.

Schwatke, C., Bosch, W., and Dettmering D.: A new Database of Water Level Time Series for Lakes, Rivers, and Wetlands from Multi-Mission Satellite Altimetry, IAG Scientific Assembly 2013, Potsdam, Germany, 2013-09-01/06, 2013.

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-yr high-resolution global dataset of meteorological forcings for land surface modeling, J. Climate, 19 (13), 3088-3111, 2005.

Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J. J., Mendiondo, E. M., O'Connell, P. E., Oki, T., Pomeroy, J. W., Schertzer, D., Uhlenbrook, S., and Zehe, E.: IAHS decade on predictions in ungauged basins (PUB) 2003–2012: shaping an exciting future for the hydrological sciences, Hydrolog. Sci. J., 48,857–880, 2003.

Su, H., Yang, Z.-L., Dickinson, R. E., Wilson, C. R., and Niu, G.-Y.: Multisensor snow data assimilation at the continental scale: The value of Gravity Recovery and Climate Experiment terrestrial water storage information, J. Geophys. Res., 15, D10104, doi:10.1029/2009JD013035, 2010.

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

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

Syed, T. H., Famiglietti, J. S., and Chambers, D.: GRACE-based estimates of terrestrial freshwater discharge from basin to continental scales, J. Micrometeorol., 10, 22-40, doi:10.1175/2008JHM993.1, 2008.

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., and Watkins, M.: GRACE Measurements of Mass Variability in the Earth System, Science, 305 (5683), 503-505, 2004.

Tregoning, P., McClusky, S., van Dijk, A. I. J. M., Crosbie, R. S.,and Peña-Arancibia, J. L.: Assessment of GRACE Satellites for Groundwater Estimation in Australia, National Water Commission, Canberra, 82, 2012.

van den Hurk, B., Doblas-Reyes, F., Balsamo, G., Koster, R. D., Seneviratne, S. I., and Camargo Jr, H.: Soil moisture effects on seasonal temperature and precipitation forecast scores in Europe, Clim. Dyn., 28, 349-362, 2012.

Wada, Y., Wisser, D., and Bierkens, M. F. P.: Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources, Earth Syst. Dynam., 5, 15-40, doi:10.5194/esd-5-15-2014, 2014.

Wahr, J., Molenaar, M., and Bryan, F.: Time variability of the Earth’s gravity field: Hydrological and oceanic effects and their possible detection using GRACE, J. Geophys. Res., 103(B12), 30205–30229, 1998.

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

Weerts, A. H., and Serafy G. Y. H.: Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, Water Resour. Res., 42, W09403, doi:10.1029/2005WR004093, 2006.

Weerts, A. H., Meißner, D., and Rademacher, S.: Input data rainfall-runoff model operational systems FEWS-NL & FEWS-DE, Tech. rep., Deltares, 2008.

Widiastuti, E.: Data assimilation of GRACE terrestrial water storage data into a hydrological model using the ensemble Kalman smoother: A case study of the Rhine river basin, MSc Thesis, TU Delft, 2009.

Zaitchik, B. F., Rodell, M., and Reichle, E. H.: Assimilation of GRACE terrestrial water storage data into a land surface model: Results for the Mississippi basin, Amer. Meteor. Soc., J. Hydrometeor, 9, 535–548, 2008.

Table 1. Parameters of the soil moisture and runoff routines in the OpenStreams wflow_hbv model.

Parameter	Description	Unit
fc	Maximum soil moisture storage	mm
β	Empirical based parameter determines the relative contribution to runoff from soil moisture storage	-
cflux	Maximum value of capillary rise from upper zone storage to soil moisture storage	mm/day
khq	Recession constant of upper zone storage, determines the amount of quick runoff from upper zone storage	1/day
perc	Maximum percolation value (from upper to lower zone storage)	mm/day
lp	Soil moisture fraction above which actual evapotranspiration (ET) equals potential ET	-
k4	Recession constant of lower zone storage, determines the amount of baseflow from lower zone storage	1/day
α	Non-linearity parameter of upper zone storage	-

Table 2. Daily mean values of the forcing data at Sundern during summer (JJA) months.

	Precipitation (mm)		Temperature (°C)		PET (mm)	
	Local	Global	Local	Global	Local	Global
2004	5.48	5.21	16.10	17.65	2.47	3.30
2005	5.25	4.68	16.12	17.46	2.47	3.55
2006	4.96	4.39	17.64	19.25	2.47	3.90
2007	9.97	7.08	16.09	17.52	2.47	3.25

Table 3. Ensemble mean parameter values at the Sundern, and A319C well locations for the calibrated and non-calibrated simulations.

	Sundern		A319C	
Parameter	Calibrated	Non-calibrated	Calibrated	Non-calibrated
fc	239.03	179.12	130.95	181.98
β	2.06	1.65	1.95	1.68
cflux	0.06	0.27	0.41	0.30
khq	0.10	0.12	0.06	0.09
perc	0.67	1.15	0.43	1.09
lp	0.88	0.75	0.67	0.72
*k*4	0.63	0.03	0.01	0.03

Table 4. Correlation coefficient computed between monthly mean estimated GW variation and monthly mean in situ variation. Names of the stations (first column) are provided in Appendix A.

	CL		CG		NCL		NCG	
	EnOL	EnKF	EnOL	EnKF	EnOL	EnKF	EnOL	EnKF
1	0.85	0.83	0.71	0.74	0.79	0.85	0.70	0.78
2	0.57	0.68	0.32	0.52	0.43	0.65	0.38	0.45
3	0.69	0.81	0.46	0.68	0.51	0.73	0.46	0.61
4	0.60	0.67	0.42	0.58	0.50	0.75	0.60	0.67
5	0.71	0.68	0.67	0.76	0.71	0.72	0.72	0.78
6	0.57	0.64	0.58	0.66	0.74	0.78	0.66	0.72
7	0.77	0.80	0.67	0.71	0.80	0.83	0.64	0.72
8	0.75	0.80	0.65	0.78	0.81	0.83	0.62	0.74
9	0.50	0.64	0.54	0.70	0.72	0.78	0.65	0.80
10	0.56	0.58	0.50	0.55	0.66	0.70	0.42	0.46
11	0.41	0.55	0.31	0.53	0.71	0.73	0.72	0.74
12	0.71	0.80	0.64	0.71	0.76	0.85	0.74	0.78
13	0.77	0.80	0.50	0.56	0.72	0.84	0.59	0.66
14	0.71	0.70	0.73	0.74	0.33	0.47	0.51	0.56
15	0.82	0.85	0.67	0.69	0.72	0.83	0.65	0.71
16	0.68	0.80	0.55	0.64	0.77	0.88	0.63	0.71
17	0.67	0.79	0.55	0.60	0.70	0.82	0.57	0.66
18	0.65	0.66	0.64	0.65	0.45	0.54	0.59	0.63
Mean	0.67	0.73	0.56	0.66	0.66	0.75	0.60	0.68

Table 5. RMSE [mm] computed between monthly mean estimated GW variation and monthly mean in situ variation. Names of the stations (first column) are provided in Appendix A.

	CL		CG		NCL		NCG	
	EnOL	EnKF	EnOL	EnKF	EnOL	EnKF	EnOL	EnKF
1	4.16	3.84	5.63	4.02	7.00	6.99	8.37	5.40
2	5.34	4.91	6.66	5.96	6.36	5.73	10.78	8.14
3	3.62	3.06	5.04	4.35	5.96	4.85	10.64	8.13
4	3.79	3.65	4.41	3.55	5.83	5.03	9.00	7.56
5	9.72	8.30	6.89	5.49	8.43	7.83	6.06	5.03
6	6.19	5.19	5.47	5.26	7.56	6.31	5.25	4.29
7	8.30	6.75	7.48	6.99	6.45	5.88	7.36	6.82
8	8.76	6.59	6.63	4.96	5.21	5.20	5.71	4.67
9	5.95	5.38	5.16	5.09	7.33	6.43	5.43	3.91
10	8.95	7.64	6.44	5.66	8.92	8.54	7.62	6.21
11	6.03	5.10	6.21	4.89	9.88	8.32	11.43	8.30
12	7.17	6.37	7.33	7.42	6.24	5.01	6.58	5.95
13	6.25	5.34	6.80	5.91	7.90	7.55	8.84	8.53
14	12.67	10.16	11.43	9.01	9.34	7.97	9.29	7.21
15	8.83	8.28	10.28	10.08	9.78	8.31	10.08	9.89
16	12.74	9.58	13.20	10.60	9.76	8.32	10.44	9.78
17	12.01	9.17	11.10	9.54	7.59	6.14	7.90	6.38
18	7.23	7.50	8.62	7.94	9.59	8.42	9.51	7.88
Mean	7.65	6.49	7.49	6.48	7.73	6.82	8.35	6.89

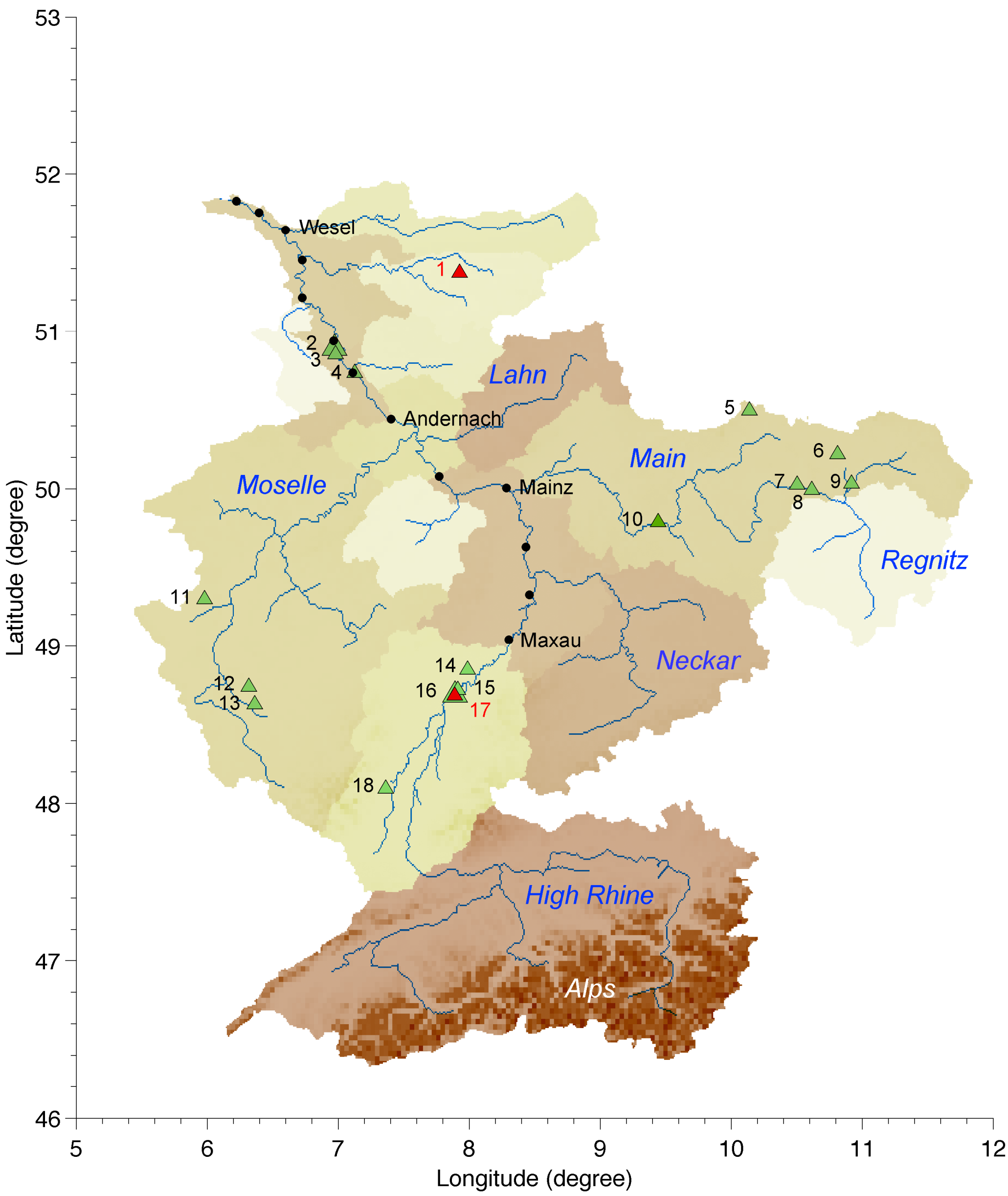
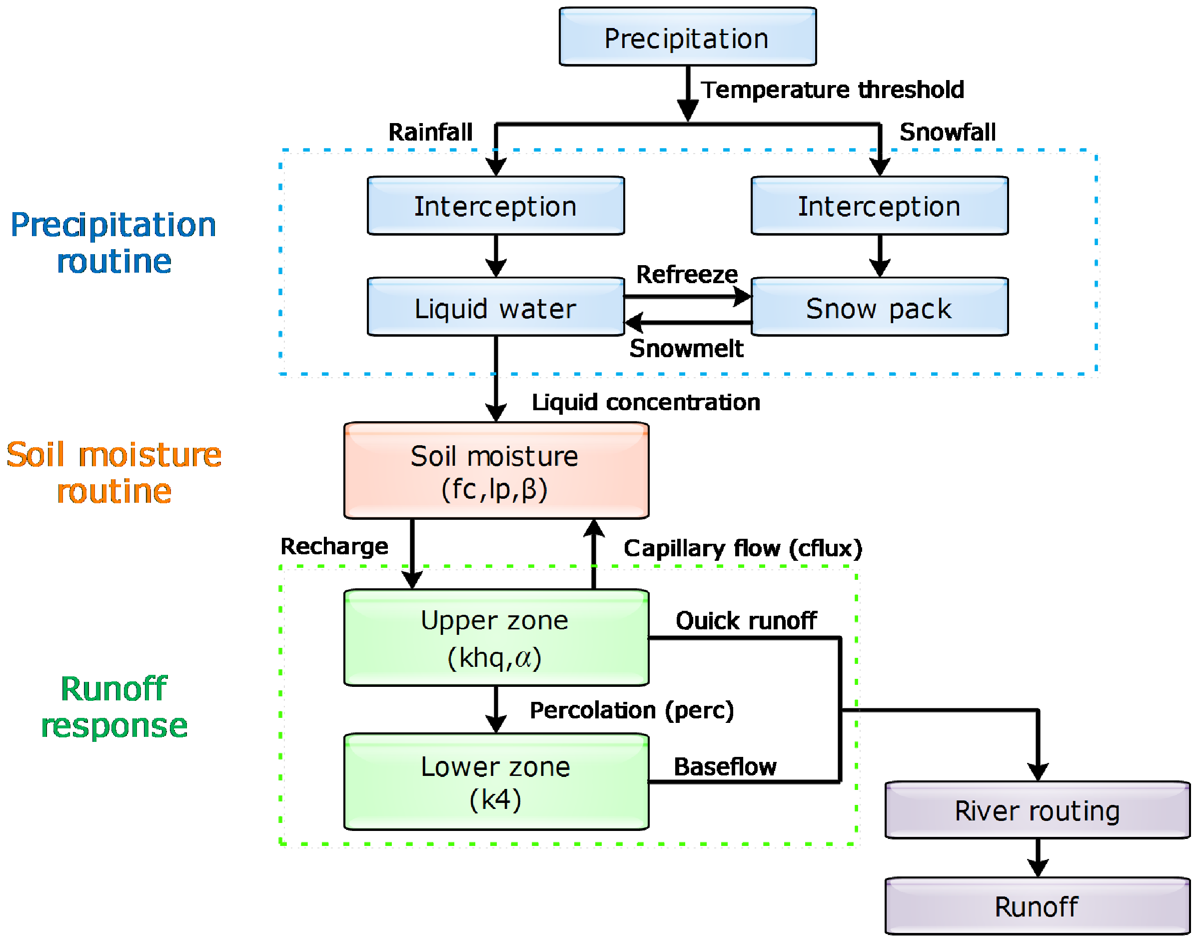


Figure 1. River gauge (circle) and well (triangle) locations over the Rhine River basin used in this paper. Red triangles indicate Sundern (1) and A319C locations (17). Names of all well locations are given in Table A1.

(a)



(b)

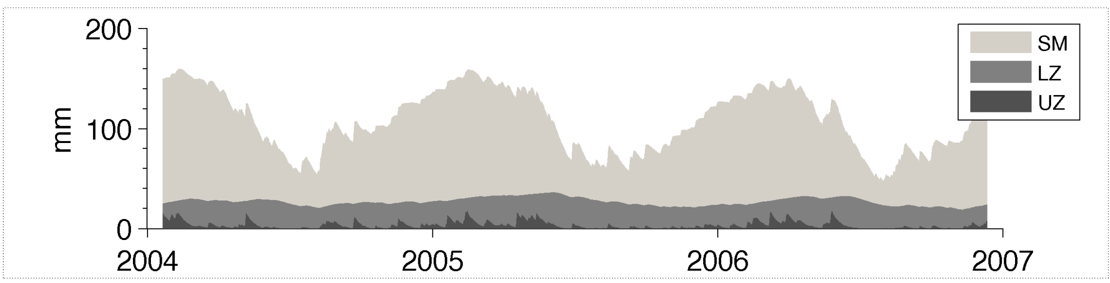


Figure 2. (a) OpenStreams wflow_hbv model structure, adapted from Schellekens (2014). (b) Sample results of the nominal run related to soil moisture (SM), upper grounwater zone (UZ), and lower groundwater zone (LZ) storages averaged over Rhine River basin.

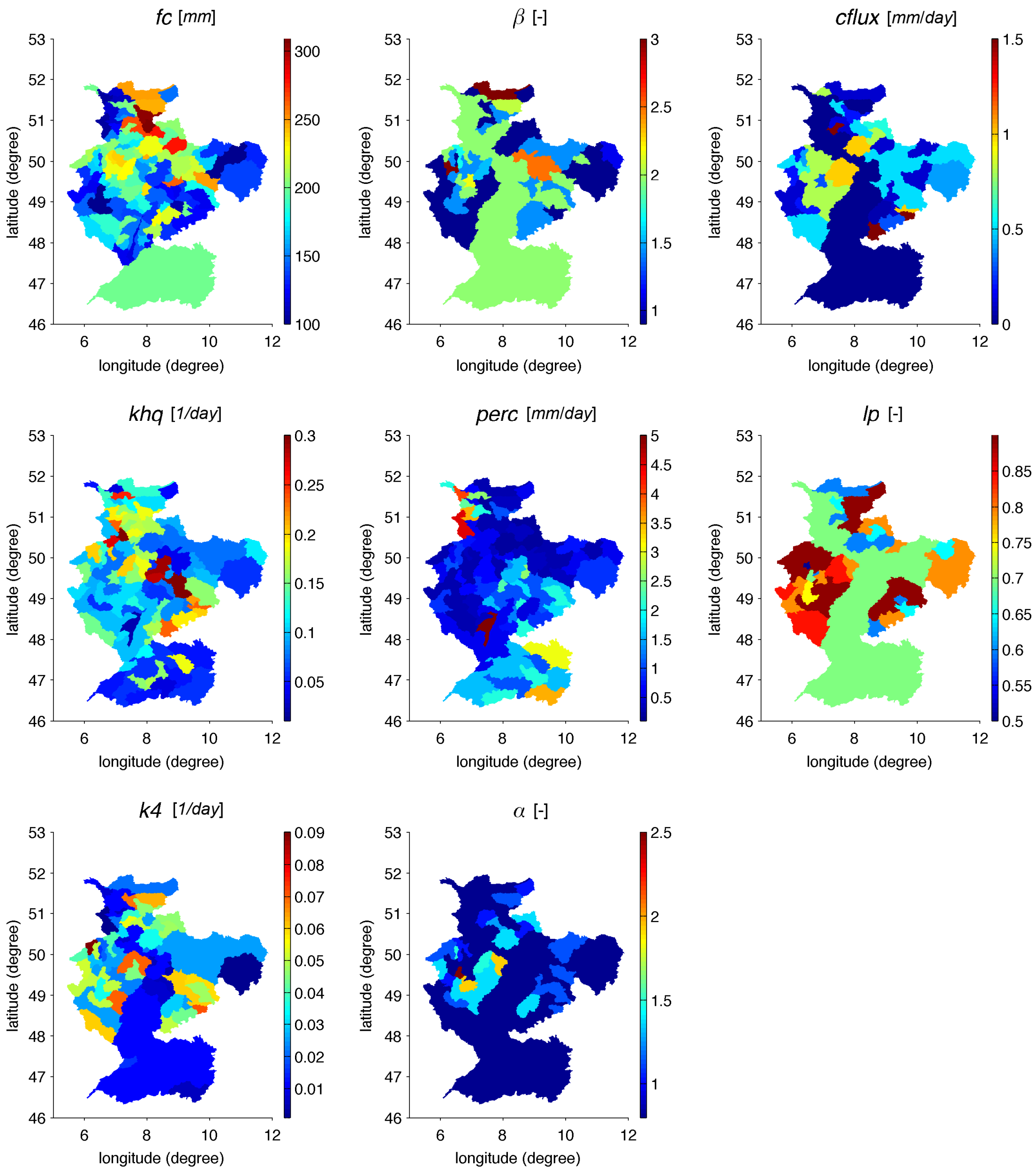


Figure 3. Calibrated parameters of the soil moisture and runoff response routines of the OpenStreams wflow_hbv model.

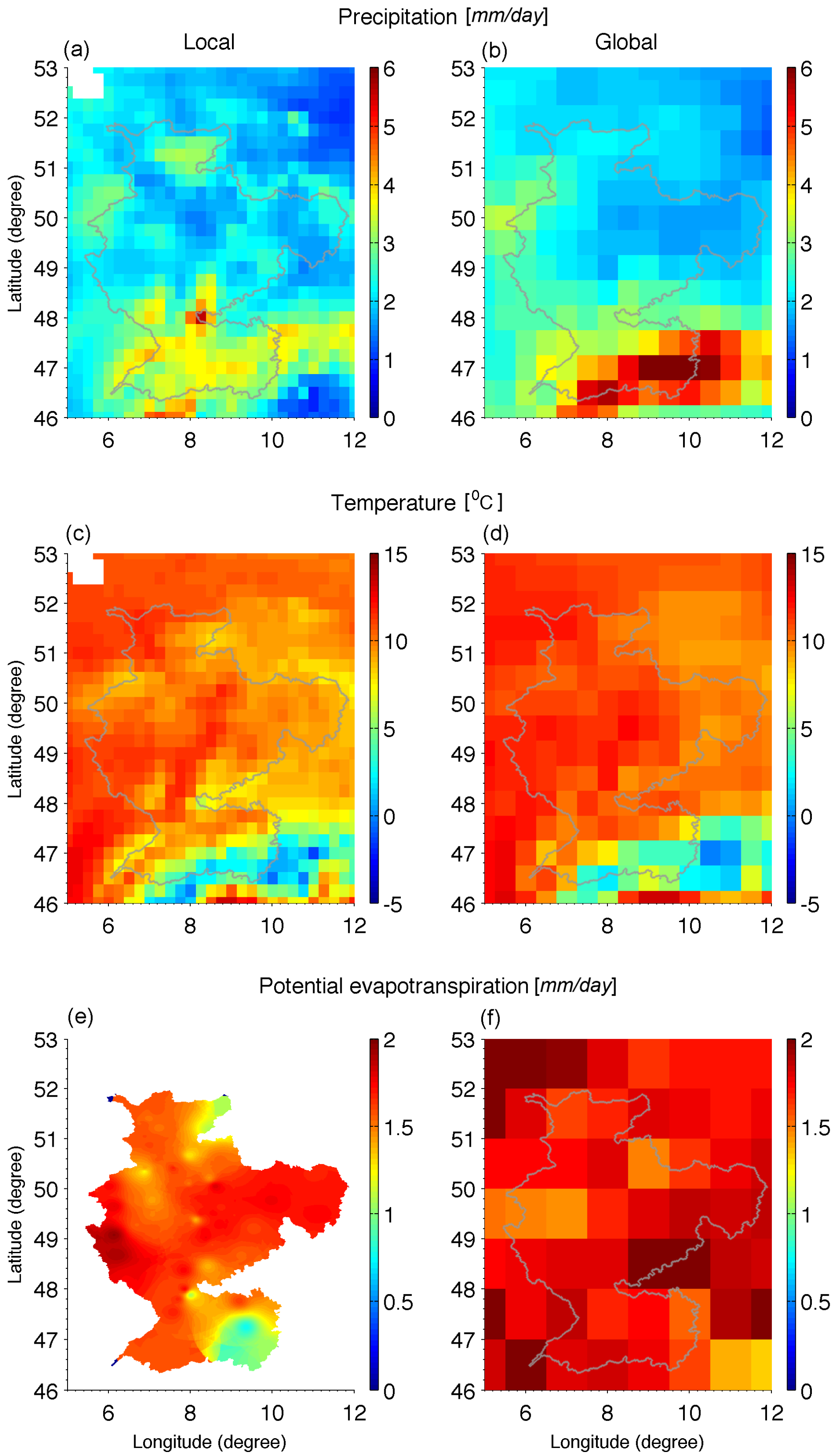


Figure 4. Mean daily precipitation, temperature, and potential evapotranspiration in 2006 from the local (left) and global (right) forcing datasets.

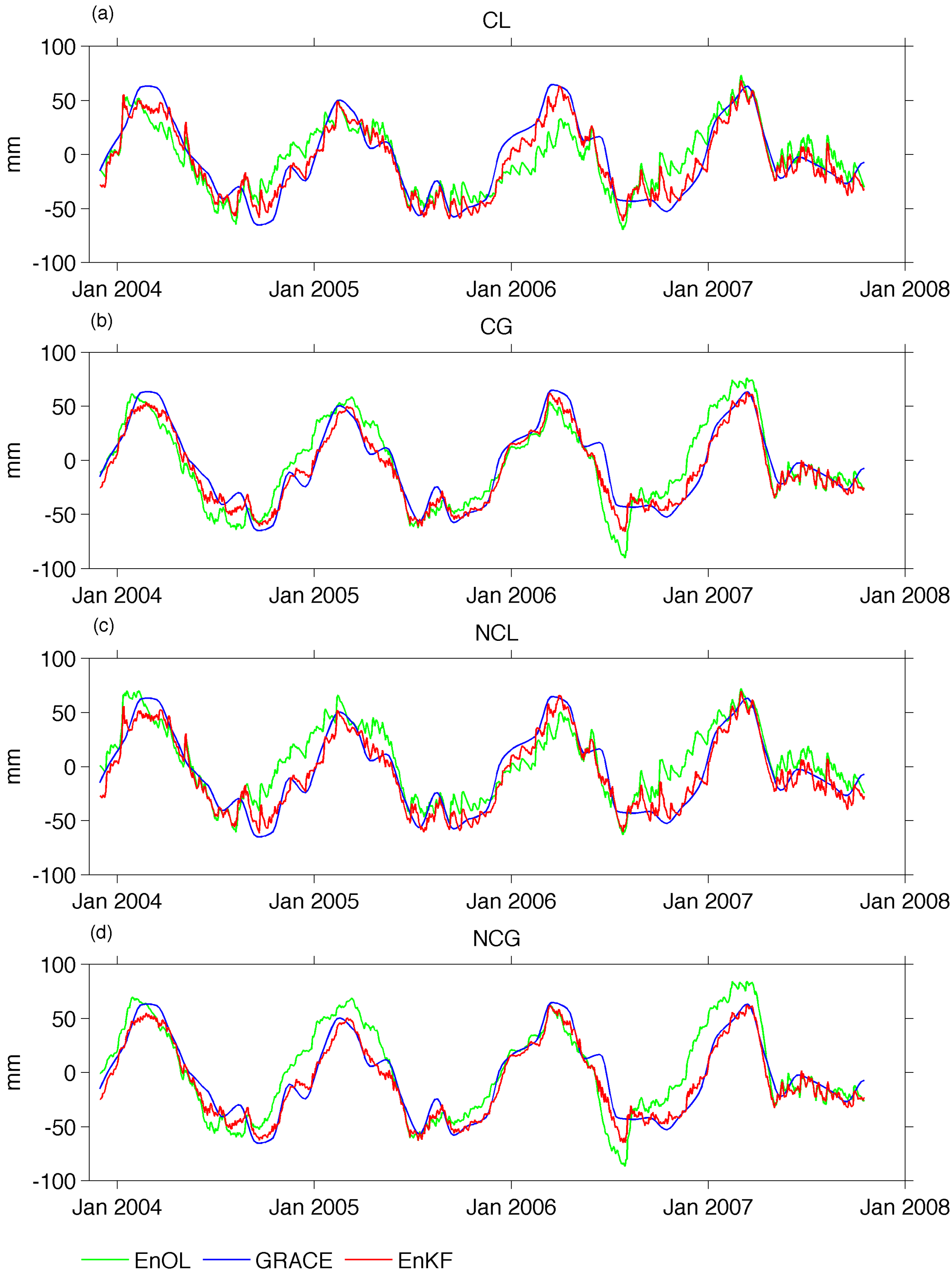


Figure 5. Area-averaged mean terrestrial water storage (TWS) over the Rhine River basin from the EnOL, EnKF and GRACE observations in 4 different scenarios (CL: calibrated parameters with local forcing data, CG: calibrated parameters with global forcing data, NCL: non-calibrated parameters with local forcing data, NCG: non-calibrated parameters with global forcing data).

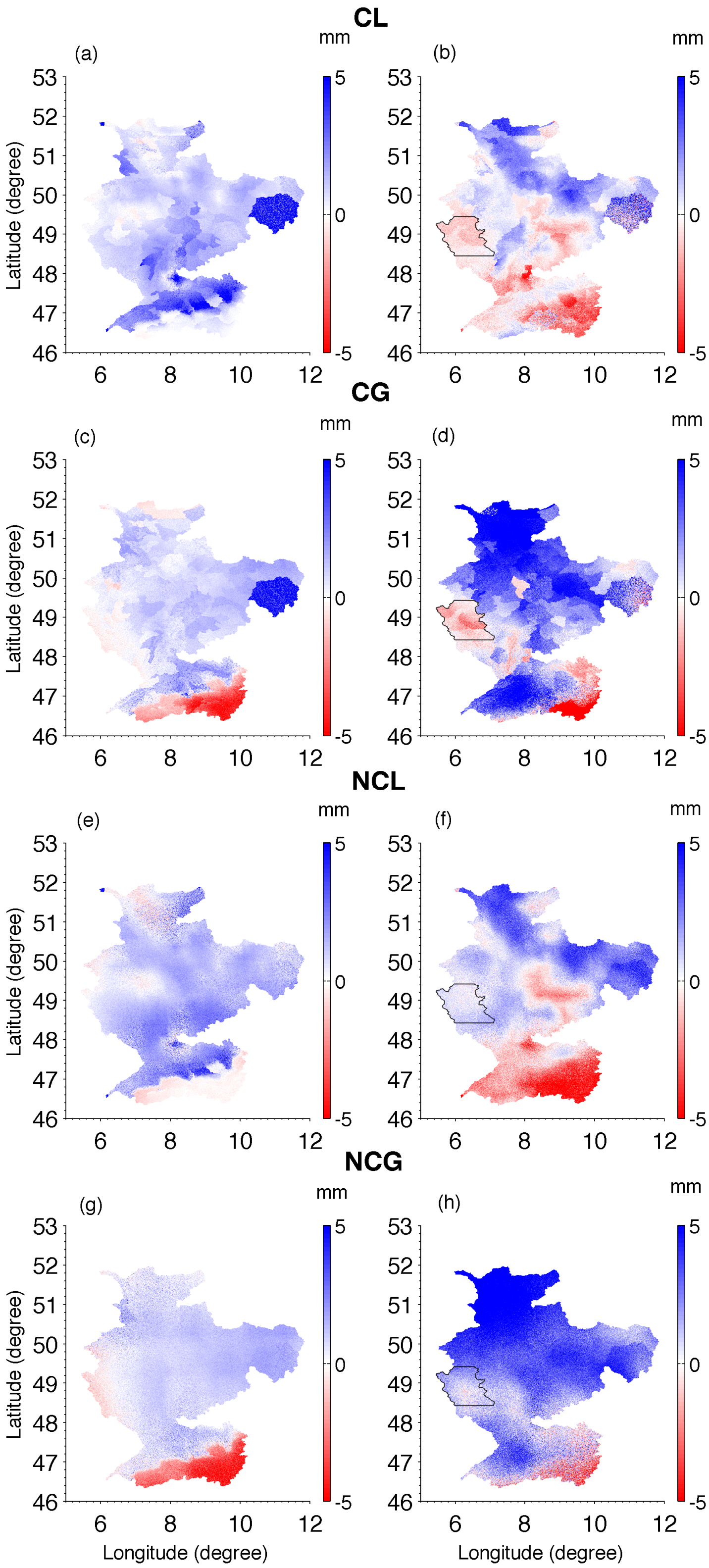


Figure 6. Averaged increment (posterior minus prior) of TWS in mm during the winter 2005-2006 (left) and summer of 2006 (right) in 4 different scenarios (CL: calibrated parameters with local forcing data, CG: calibrated parameters with global forcing data, NCL: non-

calibrated parameters with local forcing data, NCG: non-calibrated parameters with global forcing data). The polygons in the right column define the southern part of Moselle basin.

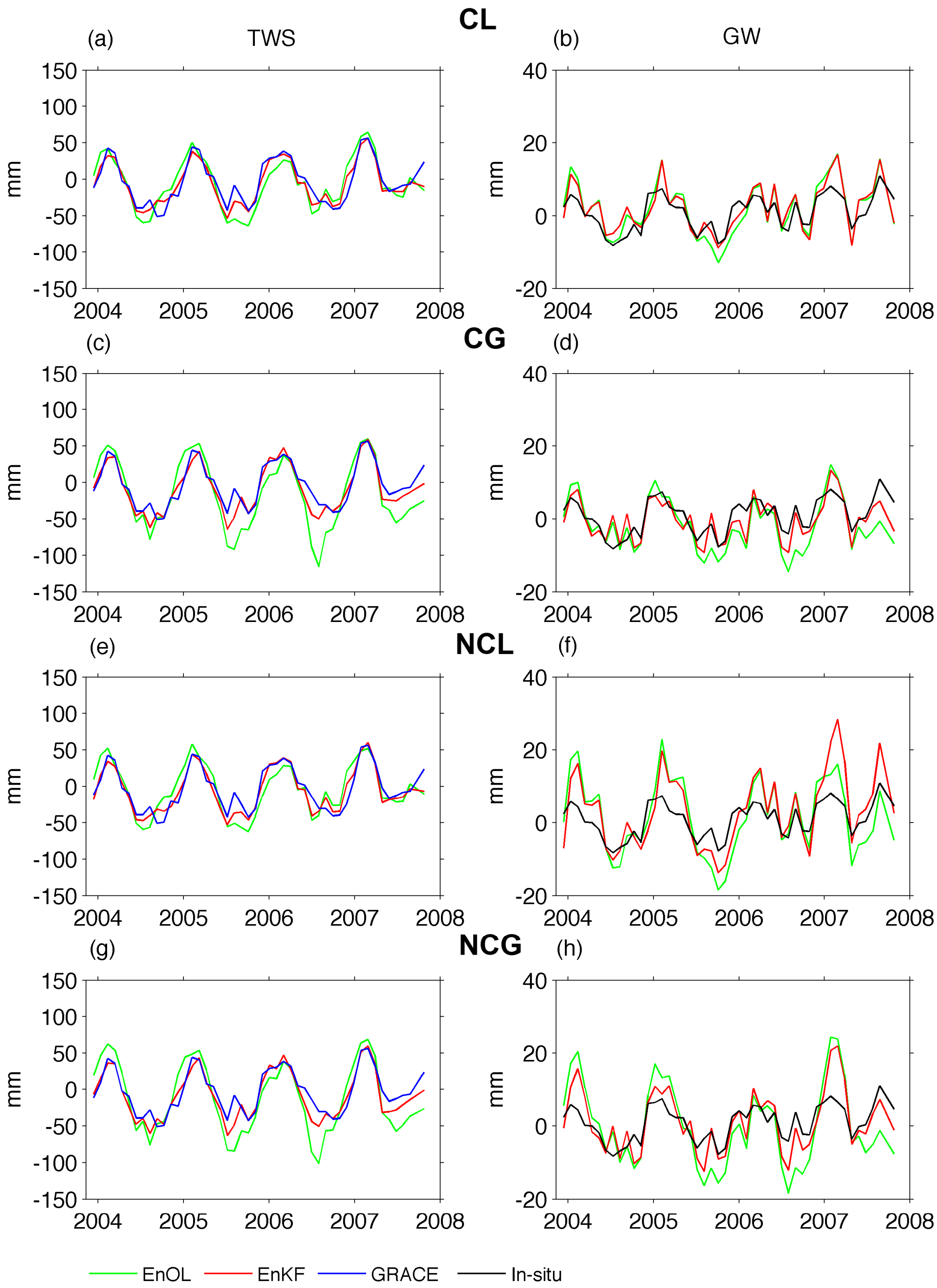
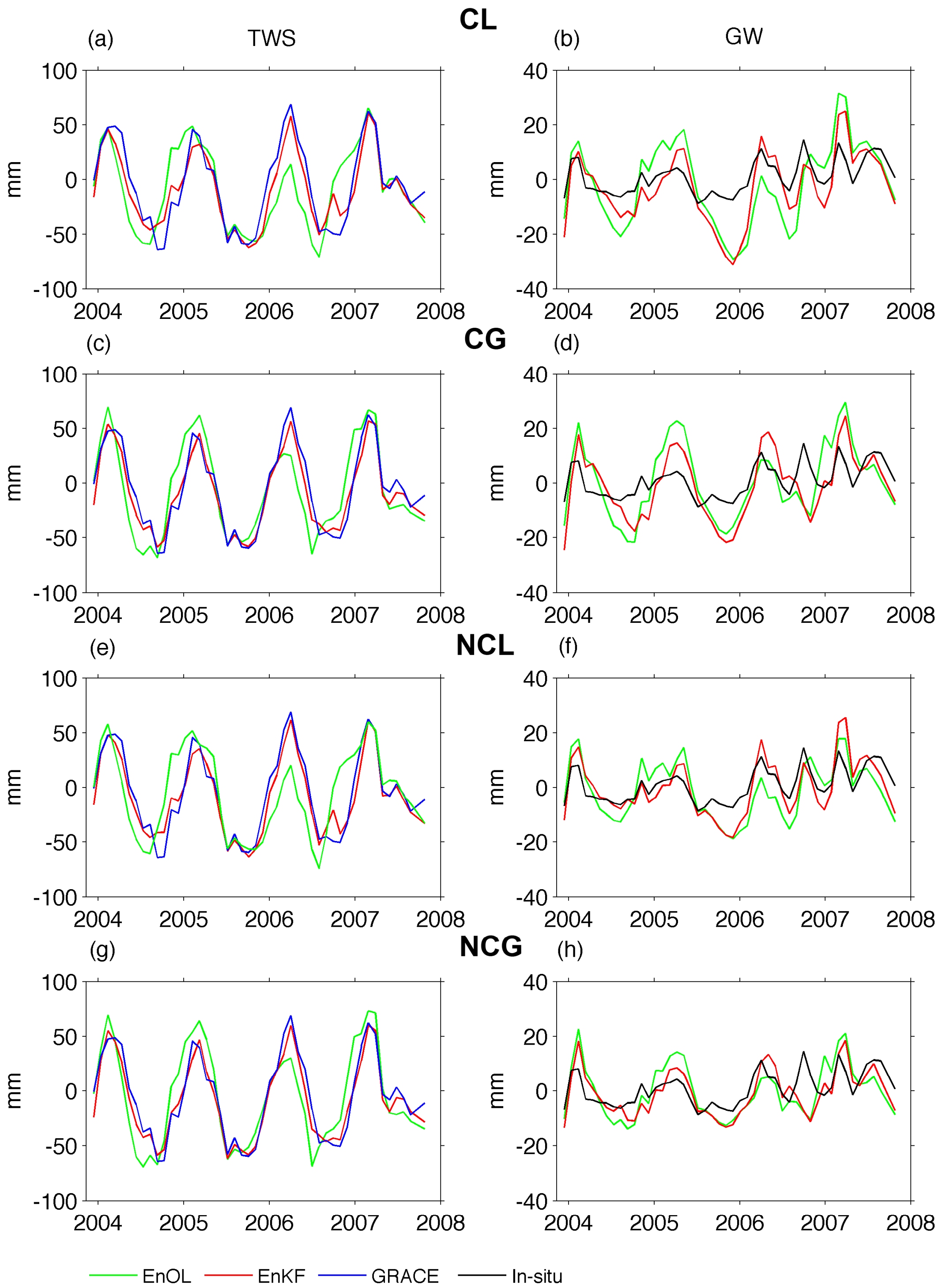
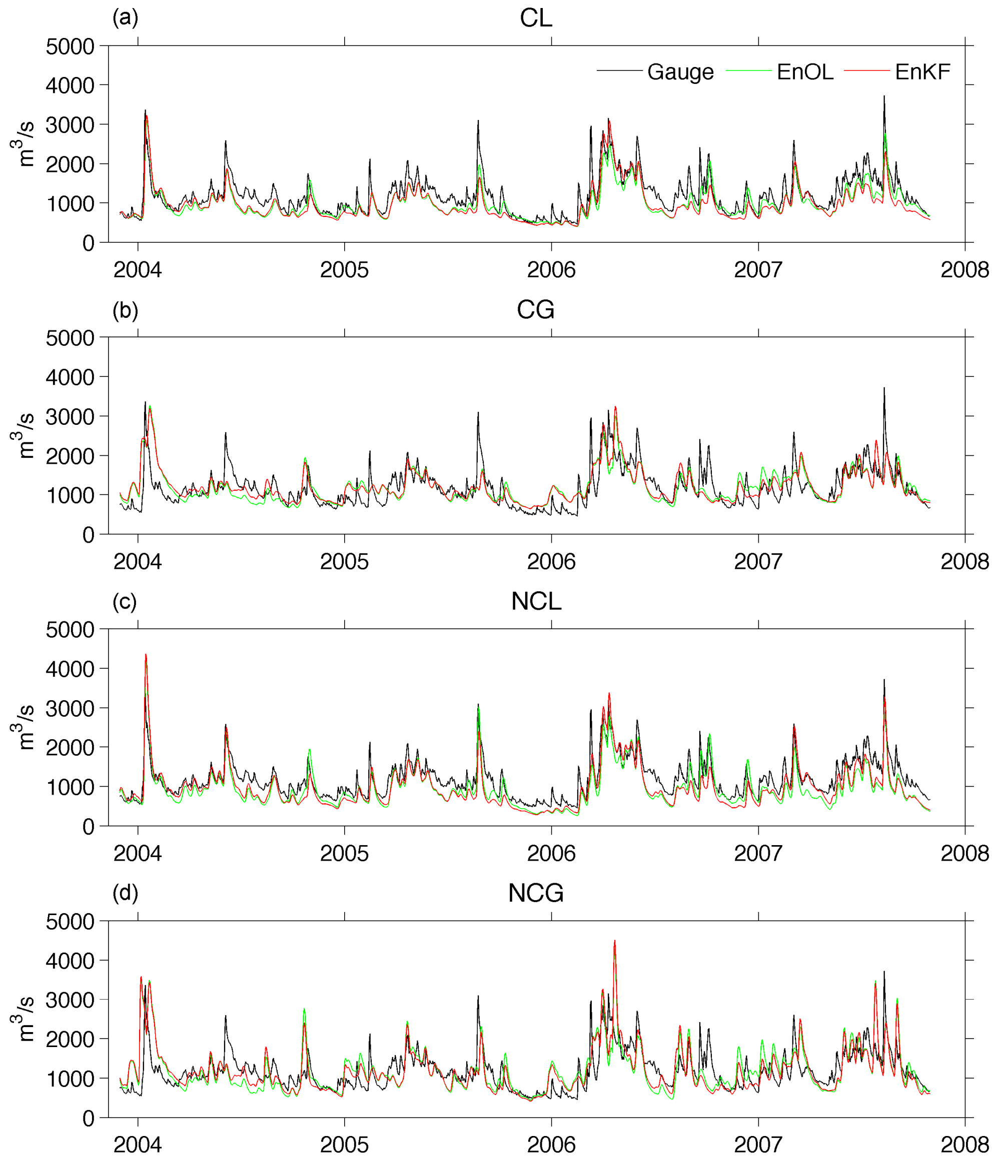


Figure 7. TWS variation (left) and GW variation (right) at the Sundern well location in 4 different scenarios (CL: calibrated parameters with local forcing data, CG: calibrated parameters with global forcing data, NCL: non-calibrated parameters with local forcing data, NCG: non-calibrated parameters with global forcing data).



936

937 Figure 8. TWS variation (left) and GW variation (right) at the A319C well location in 4
938 different scenarios (CL: calibrated parameters with local forcing data, CG: calibrated
939 parameters with global forcing data, NCL: non-calibrated parameters with local forcing data,
940 NCG: non-calibrated parameters with global forcing data).



941

942 Figure 9. Estimated and observed streamflow at the Maxau gauge station in 4 different
943 scenarios (CL: calibrated parameters with local forcing data, CG: calibrated parameters with
944 global forcing data, NCL: non-calibrated parameters with local forcing data, NCG: non-
945 calibrated parameters with global forcing data).

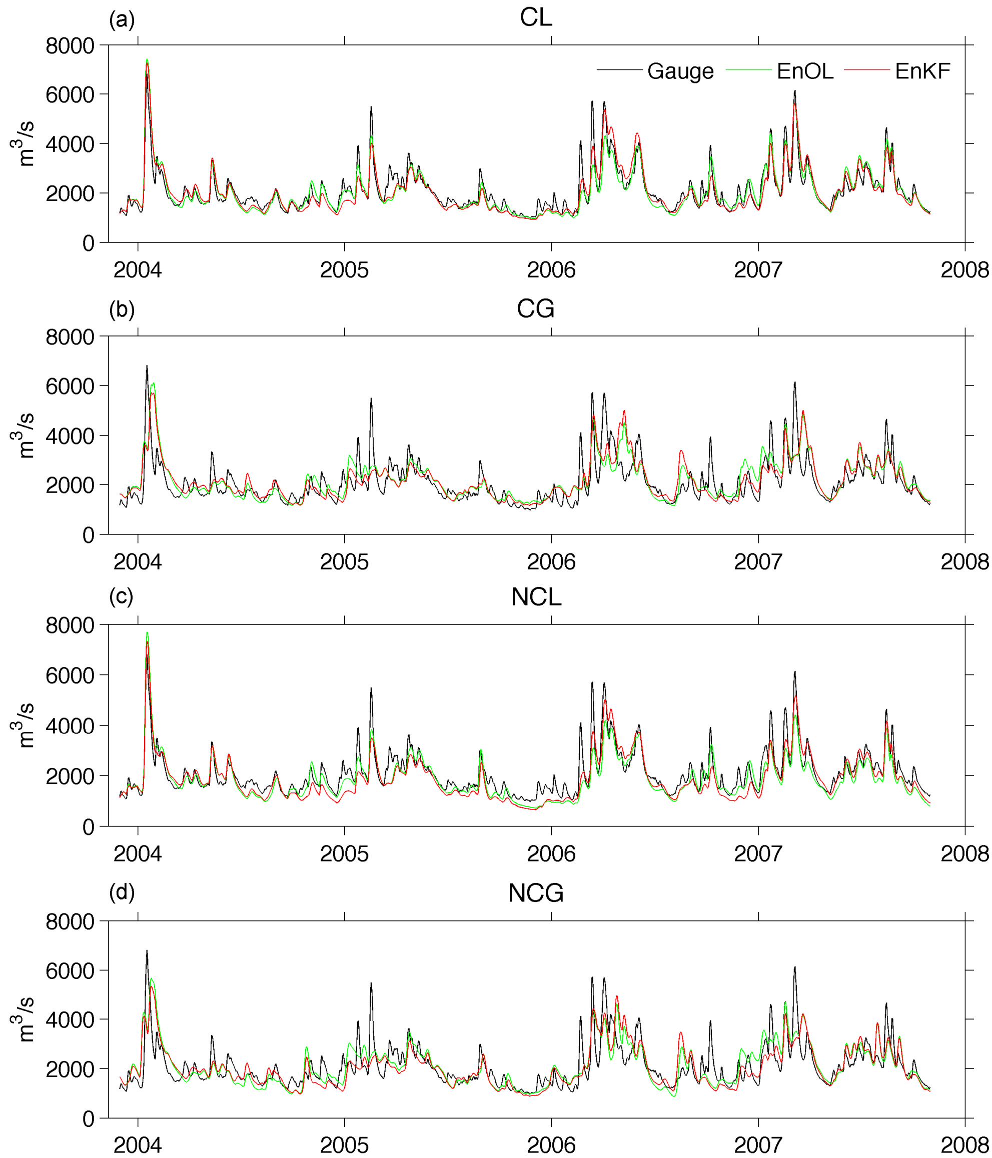


Figure 10. Estimated and observed streamflow at the Wesel gauge station in 4 different scenarios (CL: calibrated parameters with local forcing data, CG: calibrated parameters with global forcing data, NCL: non-calibrated parameters with local forcing data, NCG: non-calibrated parameters with global forcing data).

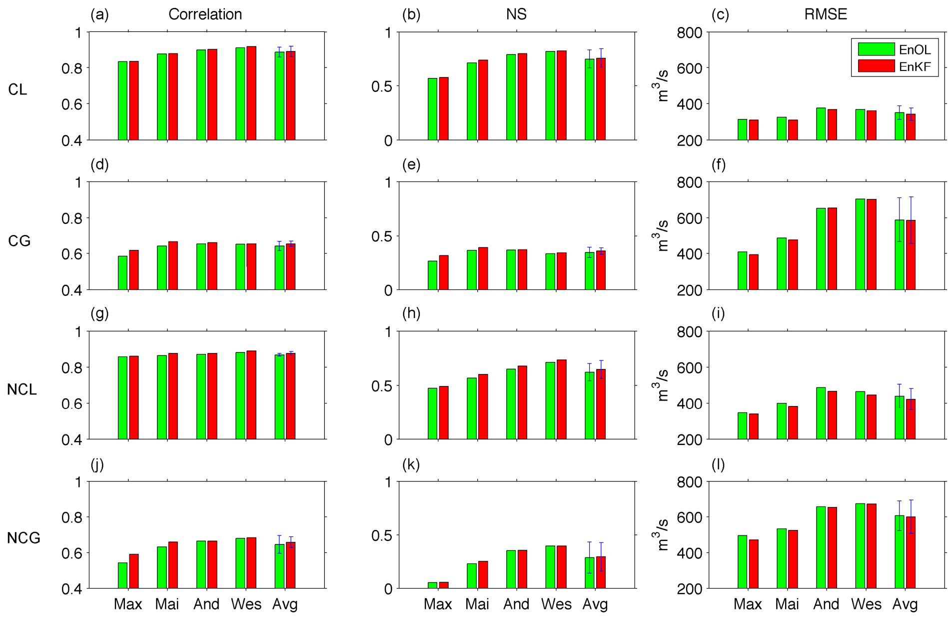


Figure 11. The correlation coefficient (left), Nash-Sutcliffe coefficient (middle) and RMS error (right) computed between estimated streamflows and gauge measurements in 4 different scenarios (CL: calibrated parameters with local forcing data, CG: calibrated parameters with global forcing data, NCL: non-calibrated parameters with local forcing data, NCG: non-calibrated parameters with global forcing data). Results are shown for the Maxau (Max), Mainz (Mai), Andernach (And), and Wesel (Wes) gauge stations. Average values (Avg) calculated across all 13 gauge locations are shown in the rightmost bar of each histogram, with the standard deviations indicated by error bars.