AGU ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE
10.1029/2021WR031549

Assimilation of Groundwater Level and Soil Moisture Data in an Integrated Land Surface-Subsurface Model for Southwestern Germany

Ching Pui Hung[1,2], Bernd Schalge[2,3], Gabriele Baroni[4], Harry Vereecken[1,2], and Harrie-Jan Hendricks Franssen[1,2]

[1]Forschungszentrum Jülich GmbH, Agrosphere (IBG-3), Jülich, Germany, [2]Centre for High-Performance Scientific Computing in Terrestrial Systems, Jülich, Germany, [3]Institute for Geosciences, University of Bonn, Bonn, Germany, [4]Department of Agricultural and Food Sciences, University of Bologna, Bologna, Italy

Key Points:

- Groundwater level (GWL) and soil moisture data assimilation (DA) improve characterization of subsurface states
- GWL DA improves GWL characterization up to 8 km from observation locations
- Evapotranspiration and discharge hardly better estimated, despite better estimates of soil moisture and GWL

Correspondence to:

C. P. Hung,
c.hung@fz-juelich.de

Citation:

Hung, C. P., Schalge, B., Baroni, G., Vereecken, H., & Hendricks Franssen, H.-J. (2022). Assimilation of groundwater level and soil moisture data in an integrated land surface-subsurface model for southwestern Germany. *Water Resources Research*, *58*, e2021WR031549. https://doi.org/10.1029/2021WR031549

Received 4 NOV 2021
Accepted 28 MAY 2022

Abstract Integrated terrestrial system models predict the coupled water, energy and biogeochemical cycles. Simulations with these models are affected by uncertainties of model parameters, initial and boundary conditions, atmospheric forcings and the biophysical processes. Data assimilation (DA) can quantify and reduce the uncertainty. This has been tested intensively for single compartment models, but far less for integrated models with multiple compartments. We constructed a virtual reality (VR) with a coupled land surface-subsurface model under the Terrestrial Systems Modeling Platform, which mimics the Neckar catchment in southern Germany. Soil moisture and groundwater level (GWL) data extracted from the simulated VR are used as measurements to be assimilated with state-only/state-hydraulic parameter estimation. Soil moisture DA improves soil moisture characterization in the vertical profile and the neighboring grid cells, with a 40 ~ 60% reduction of root mean square error (RMSE) over the observation points. In spite of a small ensemble size of 64 members, assimilating soil moisture data improved saturated hydraulic conductivity estimation around the measurement locations. The characterization of evapotranspiration and river discharge only show limited improvements (1% at observation points and less than 0.1% in RMSE at 3 selected gauge locations respectively). GWL DA not only improves the GWL characterization (76 ~ 88% RMSE reduction at observation locations) but also soil moisture for some cases. In addition, a clear improvement in GWL characterization is observed up to 8 km from the observations, and updating the model states of the saturated zone only instead of the complete domain gives better performance.

1. Introduction

Soil moisture plays an important role in the earth system through the coupled water and energy balances (Seneviratne et al., 2010). It is a source of water for the atmosphere through evapotranspiration (ET) from land, which includes plant transpiration and soil evaporation. In wet and dry climate regimes, atmospheric conditions show less dependence on soil moisture, but in a transitional climate regime, soil moisture provides a first-order constraint on ET variability (Koster et al., 2004; Seneviratne et al., 2006). On the other hand, groundwater availability imposes crucial effects on the soil moisture by water exchange between the aquifer and unsaturated zone. Shallow groundwater enhances root zone soil moisture content and therefore also ET (X. Chen & Hu, 2004), and the surface energy budget can be very sensitive to the groundwater level (GWL) (Kollet & Maxwell, 2008). Soil moisture and GWL characterization are therefore important for a better estimate of the water and energy fluxes from the land to the atmosphere.

Various terrestrial system models are developed to simulate water and energy cycles, hydrological processes and characterize soil moisture and GWL. Land surface models were first used in the late 1960s (Manabe, 1969) and then developed into multi-layered descriptions of vegetation and root zone biogeophysical processes (E. Dickinson et al., 1993; R. E. Dickinson, 1986; Sellers et al., 1986). Later also 3D-integrated hydrological models capturing horizontal water redistribution with lateral groundwater flow and surface runoff were developed (Kollet & Maxwell, 2006, 2008; Morita & Yen, 2002; Panday & Huyakorn, 2004; Qu & Duffy, 2007; VanderKwaak & Loague, 2001). Afterward, land surface and integrated hydrological models were combined in integrated land surface-hydrological models, driven by atmospheric forcing data. To further understand the feedbacks between the land and the atmosphere, simulation platforms have been developed (Anyah et al., 2008; Jiang et al., 2009; Mölders & Rühaak, 2002; Seuffert et al., 2002; VanderKwaak & Loague, 2001; York et al., 2002) which led to

fully coupled physically-based models of the terrestrial system from the subsurface across the land surface into the atmosphere (Larsen et al., 2014; Maxwell et al., 2007, 2011; Shrestha et al., 2014). Examples of these modeling platforms include Parflow-WRF (Maxwell et al., 2011), AquiferFlow-SiB2 (Tian et al., 2012), HIRHAM-MIKE-SHE (Butts et al., 2014), WRF-Hydro (McCreight et al., 2021) and Terrestrial Systems Modeling Platform (TSMP) (Shrestha et al., 2014). Terrestrial system modeling with a coupled atmosphere-land surface-subsurface model is now used for a better description and explanation of (bio)physical terrestrial system processes (Gerten et al., 2004), including the exchange of water, energy, carbon and nitrogen between the land and the atmosphere. However, such coupled models with high complexity involve different uncertain parameters, forcing terms as well as boundary conditions which are difficult to estimate. The resulting simulations are therefore associated with high uncertainties. To reduce the uncertainties of the estimated model states and parameters, data assimilation (DA) can be helpful. DA allows correcting the model simulations with help of measurements, by optimally combining model predictions and measurements in a formal modern statistical framework (Moradkhani et al., 2019).

The Ensemble Kalman filter (EnKF) is one of the most common DA algorithms. It belongs to the class of sequential DA techniques, which involve (a) a model propagation step for an ensemble of model simulations to compute changes in state estimates between measurement times, and (b) an update step which modifies state and parameters estimates to account for information contained in new measurements (McLaughlin, 2002). It is based on the classical Kalman filter and was found to be suitable for nonlinear systems and high-dimensional problems (Burgers et al., 1998; Evensen, 1992). Moreover, localized variants of Ensemble Kalman filters have been proposed (Houtekamer & Mitchell, 2001; Hunt et al., 2007; Nerger & Gregg, 2007). Localization is considered as it suppresses spurious long-range correlations between states produced by a limited ensemble size (Anderson, 2001; Hamill et al., 2001; Houtekamer & Mitchell, 2001) and increases the computational efficiency. A local analysis involves much less data than a global analysis, and the local analyzes can be computed independently in parallel (Harlim & Hunt, 2007).

Most of the DA studies are for a single compartment, like for the land surface only or groundwater only, and often for simplified, artificial domains. For example, in Y. Chen and Zhang (2006) and Hendricks-Franssen and Kinzelbach (2008) hydraulic head data are assimilated into a groundwater flow model, to estimate hydraulic head and hydraulic conductivity. Data assimilation studies which comprised multiple compartments of the terrestrial system include for example, Shi et al. (2014), who presented a synthetic multivariate DA study, with assimilation of GWL, soil moisture, discharge, land surface temperature and ET data in the physically based land surface hydrological model Flux-PIHM. They found that model parameters can be strongly constrained, especially by soil moisture, land surface temperature and discharge measurements, and confirmed those findings in a real world case study for the Shale Hills watershed (0.08 km^2) (Shi et al., 2015). D. Zhang et al. (2016) jointly assimilated soil moisture and GWL data in the integrated hydrological model MIKE-SHE, in which they found that localization is important. In their study, assimilation of these data improved the simulation of ET and discharge. Rasmussen et al. (2015); Rasmussen et al. (2016) assimilated with MIKE-SHE GWL and discharge data and analyzed factors which influenced the quality of DA. They found that when few GWL observations (0–35 observations from Karup catchment in Denmark were tested) are assimilated or when the parameters are estimated, more ensemble members are needed, and the ensemble size can be reduced when adaptive localization is used. They were also able to account for the observation bias by using bias-aware Kalman filters in assimilation of GWL data. Hydraulic head assimilation with the integrated model MIKE-SHE was also described in Ridler et al. (2017) and D. Zhang et al. (2015), and the same model was also applied to assimilate surface water and groundwater data (He et al., 2019). It should be noticed that in Shi et al. (2014) and Rasmussen et al. (2015); Rasmussen et al. (2016) uniform hydraulic parameter fields were adopted in the subsurface models. Also, in Botto et al. (2018) soil moisture and pressure head were assimilated, and they showed that EnKF can correct states and parameters in a real application with strong nonlinearities, but multi-source DA degraded model predictions in some cases. Gebler et al. (2019) assimilated soil moisture data in a high resolution CLM-Parflow land surface-subsurface model for a small subcatchment (1.28×1.12 km^2). They concluded that joint updating of soil moisture and hydraulic conductivity with EnKF gives significantly better results than soil moisture updating alone, and also the estimation of discharge could be improved. However, they found a much better performance for synthetic simulation scenarios that mimicked the real-world case, than for the real-world case simulations. It could only be understood partly why this was the case. Furthermore, H. Zhang et al. (2018) studied different assimilation strategies of groundwater levels with CLM-Parflow, over a small artificial domain (4 soil columns). They concluded that

jointly assimilating and updating soil moisture and groundwater data can give the best results in characterizing root zone area soil moisture.

Most of the mentioned established work on soil moisture and GWL DA focused on small domains (~1–~100 km^2), or simplified and artificial domains (e.g., hillslope in Botto et al. (2018), 4 soil columns in (H. Zhang et al., 2018)), and also assumed homogeneity in soil parameters (e.g., Shi et al. (2014); Shi et al. (2015), in H. Zhang et al. (2018) hydraulic conductivity only varies in depth). In this work, the domain of simulation is 57,850 km^2 in southwestern Germany, covering a variety of landscapes and land uses. The land surface and subsurface variables in the reference (integrated) models and the DA ensemble simulation exhibit complex 3-dimensional heterogeneous patterns based on high resolution land use, soil and geological information, which are assumed to be much closer to reality than most used in DA studies. As a result, the more realistic and subtle settings of this work allow us to investigate soil moisture and pressure head DA in integrated land surface-subsurface models. In addition, this is the first time GWL assimilation in a coupled land surface-subsurface model is evaluated for a fully heterogeneous 3D subsurface. This is an important step towards the assimilation of groundwater levels in integrated land surface-subsurface models for large scale real-world cases. The DA experiments in this work are carried out with TSMP-PDAF, which includes a coupled land surface-subsurface model and a DA framework (Kurtz et al., 2016; Nerger & Hiller, 2013; Shrestha et al., 2014). Measurements are extracted from a virtual reality (VR) simulation (hereafter we name it as VR) which mimics southwestern Germany, including the Neckar River catchment (Schalge et al., 2021). The simulation with the coupled model is based on realistic atmospheric forcings, land use including leaf area indices, soil maps and geological maps from this region in southwestern Germany. Given these inputs, for a long time period simulations with the coupled model are made, and the outcome of these simulations is the VR. Virtual observations which are used in DA experiments are extracted from this VR, by sampling soil moisture and GWL values at randomly selected locations, and adding noise corresponding to a measurement error. The VR approach is used as subsurface measurements are scarce, while such a VR reference model allows the extraction of many data. Also, with the VR approach DA can be assessed based on variables not assimilated. This is useful to verify the robustness of the DA method.

The ensemble size has important affects on the accuracy of DA. Related studies were conducted for example, by Y. Chen and Zhang (2006), who performed DA experiments with 2D and 3D transient groundwater flow models and EnKF and found that 200 ensemble members are needed for adequate estimation of states and parameters, and larger ensemble sizes are needed for realistic estimation of the uncertainties of the ensemble means. Keller et al. (2018) found from various experiments with a 2D groundwater flow-tracer transport model and a 2D groundwater flow model that the characterization of unknown hydraulic conductivity fields improved for increasing ensemble sizes, especially until an ensemble size of 250 is reached. Hendricks-Franssen and Kinzelbach (2008) reported an ensemble size of 200–500 is needed for state and parameter estimation with EnKF in a groundwater hydrological model to avoid filter inbreeding. In this work, we will investigate the ensemble spread and examine for the presence of underdispersion and filter inbreeding for the ensemble size we use.

We present our work as follows. In Section 2 we describe the methodology of our work, including the terrestrial system model used (Section 2.1), DA algorithm and framework (Section 2.2) and the VR used as reference (Section 2.3). Then we describe the setup for numerical experiments in Section 3, and the results are presented in Section 4, followed by discussion in Section 5. We conclude in Section 6.

2. Methodology

2.1. The Terrestrial System Modeling Platform (TSMP)

TSMP (Shrestha et al., 2014) is a modular scale-consistent terrestrial systems model consisting of well-established models for the atmosphere, land surface and subsurface. These models are two-way coupled with a coupler, which is included in the framework for communication of variables like exchange of fluxes at the shared boundaries of the different component models. In this work, only the land surface and subsurface components of TSMP were used.

The land surface model used in TSMP is the Community Land Model, version 3.5 (CLM3.5), developed by the National Center for Atmospheric Research (Oleson et al., 2004, 2008). The model simulates the radiative fluxes and transfer through the canopy, water and energy exchange between the land and the atmosphere, snow accumulation and melt, water and energy transport in the soil, further hydrological processes like surface runoff and

interception, photosynthesis and vegetation growth. (The hydrological processes will be handled by the subsurface model Parflow in our simulation with the CLM-Parflow coupled model.) CLM has 10 soil layers extending 3 m below the surface. Water and energy transport in CLM are in the vertical direction only. CLM has up to 16 plant functional types (PFTs) which can grow on soil columns, and has PFT-specific vegetation parameters.

For subsurface modeling, the finite-difference subsurface hydrological model Parflow (version 3.4.0) is used (Ashby & Falgout, 1996; Jones & Woodward, 2001; Kollet & Maxwell, 2006, 2013). It solves the 3D Richards equation for flow in the unsaturated and saturated zone and the kinematic wave equation for overland flow. A central finite difference discretization scheme is used, with Eulerian scheme in space and time, and the linear system of equations is solved by a Newton-Krylov nonlinear solver (Jones & Woodward, 2001). A terrain-following grid transformation is used so that the number of vertical layers is independent of the maximum topographic height of the simulation domain (Maxwell, 2013). This reduces the computation time and allows a finer resolution near the surface of order 10^{-2} m. Parflow can simulate lateral transport of water and replaces the hydrology simulation in CLM.

The coupler OASIS-MCT (version 3) (Ocean-Atmosphere-Sea-Ice-Soil coupler - Model Coupling Toolkit) (Craig et al., 2017; Gasper et al., 2014; Valcke, 2013) controls the exchange of the fluxes and state variables between CLM and Parflow. The variables communicate during the runtime by calling OASIS-MCT routines at explicitly defined time intervals. The implementation of the coupler in the parallel computing architecture of the TSMP framework and the variables exchanged between the model components are described in Kurtz et al. (2016).

2.2. Data Assimilation

2.2.1. Ensemble Kalman Filter With Localization (LEnKF)

In this work, we apply the localized ensemble Kalman filter (LEnKF) for DA. LEnKF is a variant of the sequential DA algorithm EnKF. Starting from some specified initial conditions, the model solution is sequentially updated at every time step where measurements are available. It propagates the previous information which has been assimilated forward in time to increase the knowledge of the system (Evensen, 1994). For LEnKF, updates from the filter are limited to the vicinity of the observation locations. Here we give a brief description of the EnKF algorithm and then the localization procedure imposed.

We start with some physical states describing a system, and a dynamical model $\mathcal{M}$ simulating the evolution of these states. The states and the model parameters to be updated by DA are collected in the state vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ with size n (superscript T denotes the transpose as $\mathbf{x}$ is a column vector), and $\mathbf{P}$ is the corresponding covariance matrix. We also define the observation vector $\mathbf{y}$ and observation operator $\mathcal{H}$ which relates the observations and state vector by

$$\mathbf{y} = \mathcal{H}(\mathbf{x}). \tag{1}$$

For the ensemble Kalman filter (EnKF) (Evensen, 1994), the model is described statistically with an ensemble of states so that the uncertainty is given by the ensemble spread and the covariance matrix can be approximated by the covariance of the ensemble of the state vectors. We denote the ensemble as $\mathbf{x}_k$, $1 \le k \le p$, and p is the ensemble size. Moreover, the observable is treated as random variables with and ensemble (of size p) of observation vectors. This is done by adding a random error ϵ to $\mathbf{y}$, where the variance of ϵ is equal to the error variance of the observations, that is, for each ensemble member k,

$$\mathbf{y}'_k = \mathbf{y} + \epsilon_k, \tag{2}$$

where $\mathbf{y}'_k$ is a vector containing the perturbed observations. For each ensemble member, the forecast is (for clarity we omit the subscript k):

$$\mathbf{x}^f = \mathcal{M}(\mathbf{x}^a), \tag{3}$$

$$\mathbf{P}^f \approx \left\langle \left(\mathbf{x}^f - < \mathbf{x}^f >\right) \left(\mathbf{x}^f - < \mathbf{x}^f >\right)^T \right\rangle, \tag{4}$$

where the superscripts a and f denote analysis and forecast respectively. And the analysis is given by:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}\left[\mathbf{y}' - \mathcal{H}\left(\mathbf{x}^f\right)\right], \tag{5}$$

$$\mathbf{P}^a \approx \left\langle (\mathbf{x}^a - < \mathbf{x}^a >)(\mathbf{x}^a - < \mathbf{x}^a >)^T \right\rangle, \tag{6}$$

where the angle brackets denote averaging over the ensemble members (Burgers et al., 1998). The covariance matrices are approximated with covariance matrices of the ensemble instead of propagating with the model, which is the key to reduce computational cost and improve robustness under non-linear model dynamics. The Kalman gain (at step $i+1$) is given by

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^T \left(\mathbf{H}\mathbf{P}^f \mathbf{H}^T + \mathbf{R}\right)^{-1}. \tag{7}$$

here $\mathbf{H}$ is the tangent linear form of $\mathcal{H}$ and $\mathbf{R}$ is the covariance matrix of the observation errors. If model parameters are included in the state vector, the model $\mathcal{M}$ in Equation 3 will not modify the parameters; the parameters are only modified in the analysis step.

In some simulation scenarios a damping factor (α) was used. Dampening can reduce filter inbreeding, which happens when the covariance matrix is estimated based on a too small ensemble size. The damping factor reduces the modification of the forecast with the Kalman gain and limits the intensity of the perturbation of the log-hydraulic conductivity field (Gebler et al., 2019; Hendricks-Franssen & Kinzelbach, 2008). To achieve this, the analysis of the state vector in Equation 5 is modified as:

$$\mathbf{x}^a = \mathbf{x}^f + \boldsymbol{\alpha}\mathbf{K}\left[\mathbf{y}' - \mathcal{H}\left(\mathbf{x}^f\right)\right], \tag{8}$$

where $\boldsymbol{\alpha}$ is a diagonal matrix with values between 0 and 1 for the diagonal entries.

The formulation of localization follows the approach described in Houtekamer and Mitchell (2001), where the covariance matrix $\mathbf{P}$ is modified with a correlation matrix $\mathbf{A}$. In the evaluation of the Kalman gain in Equation 7, $\mathbf{P}^f$ is replaced by

$$\mathbf{P}'^f = \rho_{\mathbf{A}} \circ \mathbf{P}^f \tag{9}$$

where the operation $\rho_{\mathbf{A}} \circ \mathbf{P}$ denotes the Schur product between $\mathbf{A}$ and the covariance matrix $\mathbf{P}$, defined by $P'^f_{ij} = A_{ij} P^f_{ij}$. Therefore, it requires the dimensions of $\mathbf{A}$ and $\mathbf{P}$ to be the same. The localization effect is imposed on the covariance matrix by specifying a spatial correlation through $\mathbf{A}$, which results in:

$$\mathbf{K} = \rho_{\mathbf{A}_1} \circ \left(\mathbf{P}^f \mathbf{H}^T\right)\left[\rho_{\mathbf{A}_2} \circ \left(\mathbf{H}\mathbf{P}^f \mathbf{H}^T\right) + \mathbf{R}\right]^{-1}. \tag{10}$$

here $\mathbf{A}_1 \in M_{n\times m}(\mathbb{R})$ and $\mathbf{A}_2 \in M_{m\times m}(\mathbb{R})$, and n, m are the sizes of the state and observation vector respectively. For both $\mathbf{A}_1$ and $\mathbf{A}_2$ we use the correlation matrix $A_{i,j} = f(r_{ij})$, where $f(r_{ij})$ is a function with local support, and r_{ij} is the physical distance between the locations $\mathbf{r}_i$ and $\mathbf{r}_j$, corresponding to row i and column j of $\mathbf{P}^f\mathbf{H}^T$ (or $\mathbf{H}\mathbf{P}^f\mathbf{H}^T$) respectively. Therefore, in Equation 10, in the first Schur product r_{ij} is the distance between the locations corresponding to x_i and y_j, while for the second Schur product it is the distance between the locations corresponding to y_i and y_j. The Schur product (Equation 9) between the covariance matrix of the background error and a correlation function $f(r_{ij})$ can filter the small and noisy background-error covariances associated with remote observations (Houtekamer & Mitchell, 2001). Figure 1 shows schematically the update of the ensemble of state vectors and the covariance matrix in an iteration of the DA cycle using Equations 3–10.

In our work, an exponential function is adopted with cut-off distance for f, that is, $f(r_{ij}) = \exp(-r_{ij}/r_s)$ for $r_{ij} \le r_c$ and zero otherwise, where r_s is a length scale where spatial correlation of soil moisture is significant, which determines the extent of localization, r_c is a cut-off radius, and $r_c = 3r_s$ in our experiments. Furthermore, only localization in the horizontal plane is considered, that is, $r_{ij} = \sqrt{(r_{i,x} - r_{j,x})^2 + (r_{i,y} - r_{j,y})^2}$, where $r_{i,x}(r_{j,x})$ and $r_{i,y}(r_{j,y})$ are the physical coordinates on the horizontal plane.

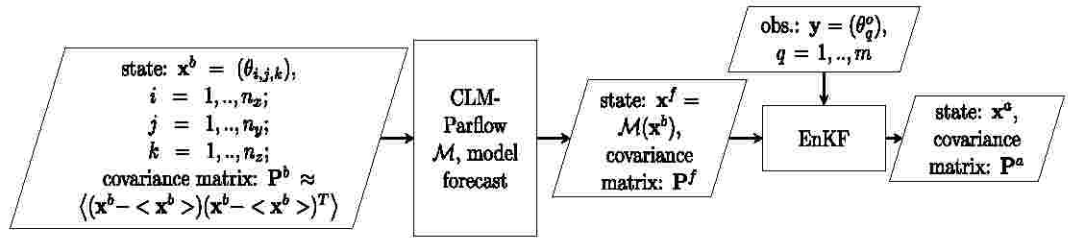

Figure 1. Schematic view of a cycle of soil moisture data assimilation (DA). In the diagram, $\mathbf{x}^b$ is the initial estimate of the state vector, which is also the analysis at the end of the previous DA cycle.

2.2.2. Defining State Vector, Observation Vector and Observation Operator for the Data Assimilation Experiments

In the following experiments, the model $\mathcal{M}$ is the land surface-subsurface model CLM-Parflow, the state and observation vectors are as follows:

1. If soil moisture data are assimilated and only soil moisture states are updated, the state vector contains all soil moisture values for the subsurface, that is, $\mathbf{x} = (\theta_{i,j,k})^T$, $1 \le i \le n_x$, $1 \le j \le n_y$ and $1 \le k \le n_z$, where θ is soil moisture, and n_x, n_y and n_z are the number of grid cells in the x, y and z directions, respectively, the size of the state vector $\mathbf{x}$ is $n = n_x n_y n_z$. If in addition hydraulic conductivity K_s is also updated, also hydraulic conductivity values for those grid cells are included in the state vector, that is, $\mathbf{x} = (\theta_{i,j,k}, \log K_{s\ i,j,k})^T$, $n = 2n_x n_y n_z$. The observation vector contains soil moisture values at the selected measurement locations.
2. If GWL data are assimilated in a weakly coupled DA approach, only saturated subsurface states are updated. In this case, the state vector contains the pressure values for the saturated grid cells. Therefore, $\mathbf{x} = (\psi_{i,j,k})^T$ for all (i, j, k) such that water saturation $S_w = 1$ at (i, j, k), and ψ is the pressure head. H. Zhang et al. (2018) showed, for a small artificial domain, that this can give more stable results. The saturated grid cells are selected based on the model forecast. If in addition hydraulic conductivity is also updated, also hydraulic conductivity values for those grid cells are included in the state vector. The observation vector contains the pressure head data not only at the GWL, but also below the GWL assuming hydrostatic conditions. The domain is of finite depth of 100 m, and only groundwater from the upper unconfined aquifer is considered so the assumption is reasonable. This condition is imposed to enhance the stability of the DA updates (H. Zhang et al., 2018).
3. If GWL data are assimilated in a fully coupled DA approach, all subsurface states are updated. The state vector contains now pressure values for the saturated grid cells and soil moisture values for the unsaturated grid cells. that is, $\mathbf{x} = \left(\psi_{i,j,k}, \theta_{i',j',k'}\right)^T$, for all (i, j, k) such that $S_w = 1$ at (i, j, k), and (i', j', k') are the coordinates of the remaining grid cells, that is, unsaturated ones. H. Zhang et al. (2018) showed that this approach is more stable than updating pressure values in the unsaturated zone, given strongly non-Gausssian distributions of pressure under dry conditions. Saturated grid cells are selected in the same way as for the weakly coupled approach and hydraulic conductivity can also be updated and included in the state vector. The observation vector contains again the pressure head data and possibly also soil moisture data.

The observation operator is linear and extracts the values of soil moisture θ (and/or pressure head ψ) at the selected grid cells. In matrix representation, $[\mathbf{H}]_{n \times N_{obs}}$, $H_{ij} = 1$ when the physical coordinates for state vector component x_i are the same as the coordinates of the j-th measurement, and $H_{ij} = 0$ otherwise. The operator will simply truncate components for the model parameter if the latter is included in the state vector.

The Kalman gain matrix $\mathbf{K}$ is evaluated for the whole domain with localization (Equation 10) and operated on $\mathbf{y}' - \mathcal{H}\left(\mathbf{x}^f\right)$ (which is a column vector with the entries representing each measurement location) to update the estimate of the states or both states and parameters at every grid cell and for all soil layers. In updating the soil moisture by Equation 5 we included further constraints to avoid non-physical solutions. If the analyzed θ is close to the residual soil moisture content, the pressure head evaluated by the van Genuchten model (van Genuchten, 1980) might be extremely negative and the DA updates might become numerically unstable. We imposed a lower bound to the water saturation of the soil to be 0.3 vol. % above the residual value S_{w0}, that is, $S_w \ge S_{w0} + 0.003$. The lower bound of soil moisture $\theta_{i,j,k}$ at any position is:

$$\theta_{i,j,k} \ge (S_{w0}|_{i,j,k} + 0.003)\, \phi_{i,j,k}, \tag{11}$$

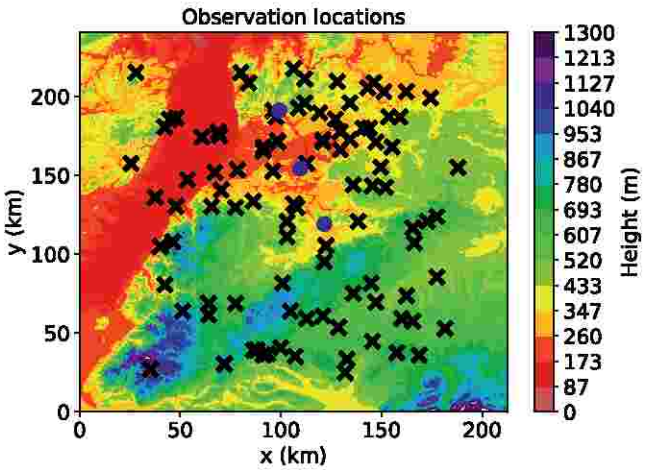

Figure 2. Soil moisture and groundwater level observation locations (100 in total) used in the data assimilation experiments, indicated by black crosses. The river discharge at 3 locations will be evaluated in the experiments, indicated by the blue dots, namely (from north to south), Rockenau, Lauffen, Plochingen. The topography of the simulation domain is shown in the background.

Any value of soil moisture below the lower bound after evaluation will be truncated to this lower bound according Equation 11.

2.2.3. Data Assimilation Framework

The DA framework consists of the coupling of TSMP and the parallel data assimilation framework (PDAF), under TSMP-PDAF. PDAF is a library providing modules of various sequential DA filters with parallel algorithms, and is applicable to any geoscientific model (Nerger & Hiller, 2013). The library also includes the EnKF and the corresponding localized version which are coded as described in Section 2.2.1. The coupling between the model and PDAF is in the so-called online mode so that the data exchange between the model and PDAF is stored in the main memory, contrary to the offline mode which writes the input and output files on disk space which is less efficient. The interface for online coupling between the PDAF library and TSMP was implemented and documented by Kurtz et al. (2016). This involves defining the state vector $\mathbf{x}$ describing the model as the input for PDAF, and the observation operator $\mathbf{y} = \mathbf{H}(\mathbf{x})$ relating the observations $\mathbf{y}$ with the model states.

For the structure of the TSMP-PDAF framework, the land surface and subsurface models (with coupler OASIS-MCT) are treated as libraries, wrapped in the TSMP framework. On the other hand, the PDAF library is connected with a user defined interface. Both the TSMP framework and the PDAF interface are called in the main program (Kurtz et al., 2016).

2.3. The Virtual Reality

2.3.1. The Neckar River Catchment

The whole rectangular computational domain is 57,850 km^2 and covers southwestern Germany and adjacent areas in France and Switzerland. The Neckar catchment is covered by the domain and has a size of 13,928 km^2, and simulations are conducted for the year 2015. The topography of the simulation domain is shown in Figure 2. The horizontal spatial resolution is 800 m for both of the land surface (CLM) and subsurface (Parflow). For the vertical profile, there are 10 soil layers for CLM up to a depth of 3 m, and 50 for Parflow up to a depth of 100 m, while the upper 10 layers coincide with the 10 CLM soil layers. The Neckar catchment possesses various features of typical central European catchments so that it is of interest to construct a VR model based on this river catchment.

The catchment is located at the east of the Black Forest and north of the Swabian Jura, and the VR also includes surrounding regions like the upper Rhine valley and the foothills of the Alps. The catchment features distinct variability in terms of hydrology, climate, soil properties and land uses including grassland, cropland, broad leaf forest and needle leaf forest. Mountainous regions like the Black Forest and the Swabian Jura are mostly forested, while grasslands are rare and the majority of the land is used for agriculture. In most regions the groundwater table is only at a depth of a few meters or less except in the mountains, and therefore coupling between the ET and groundwater table can be expected. Annual rainfall is 1,000 mm on average with minimum values of 600 mm in the Rhine valley and peaks of up to 2,000 mm in the Black Forest. The annual temperature in the center of the domain in Stuttgart is 8.6°C and monthly average temperatures vary between −0.4°C in January and 17.7°C in July. Further information on the Neckar catchment can be found in Schalge et al. (2021).

2.3.2. The VR as a Reference Model

The reference model is a VR produced by a simulation with the CLM-Parflow components of TSMP, driven by atmospheric forcing data, which are obtained from another fully coupled model including also an atmospheric component. Spin-up simulations with the land surface-subsurface model are conducted in order to obtain equilibrated initial conditions. This is done by recycling the atmospheric forcings for the year 2015 for 10 consecutive years. The end of the spin-up is used as initial condition for the VR simulation. Soil moisture and GWL measurements are taken from this VR and used in the DA experiments.

The atmospheric forcing data, used to drive the VR and also the model ensemble, are obtained from an ensemble simulation with the fully coupled model with the atmospheric model compartment. The atmospheric model used is Consortium for Small-scale Modeling (COSMO), and a brief description of the model and equations of COSMO are given by Shrestha et al. (2014). The fully coupled model COSMO-CLM-Parflow is at a resolution of 2.8 km for the atmosphere and 800 m for the land and subsurface. The ensemble consists of only 16 members, there are 4 members with lateral forcing originating from different global atmospheric circulation models (Integrated Forecasting System, (Gregory et al., 2000), Global Forecast System (White et al., 2018), Global Spectral Model (Mizuta et al., 2006) and Globales Modell (GME Majewski et al. (2002))) while 4 additional members use the same lateral forcing (GME) but have different parametrizations of shallow convection, cloud microphysics, turbulence and boundary layer processes. The remaining 8 members use the atmospheric setup of one of the first 8 but are combined with additional perturbations of the Leaf Area Index (LAI) and soil texture and associated soil hydraulic parameters. The time step is 15 min for the land surface and subsurface components, and only 5 s for the atmosphere. The coupling frequency of atmosphere-land surface and land surface-subsurface is 15 min. The operational COSMO setup of the German national weather service (Deutscher Wetterdienst, DWD) is used. Lateral boundary forcing and constant fields include topology and landmask, and the lateral relaxation zone is 12 km.

Land use in land surface modeling was taken from the 2,006 Corine Land Cover Data Set of EEA (European Environment Agency). The LAI was computed from MODIS (Myneni et al., 2002) as monthly average, with four land uses: broad leaf forest, needle leaf forest, cropland and grassland.

The 1:1,000,000 soil map (BUEK1000) from the Federal Institute for Geosciences and Natural Resources (BGR) was used. The saturated hydraulic conductivity, Mualem-van Genuchten parameters, porosity and residual water content, were estimated with pedotransfer functions based on soil texture data derived from the soil map (Cosby et al., 1984; Tóth et al., 2015) down to 3 m depth. For grid cells at or next to river cells gravel layers were introduced with high hydraulic conductivity of 2.78×10^{-4} m s^{-1}, down to 8 m depth, and van-Genuchten parameters $n = 2$ and $\alpha = 4$ m^{-1}. The porosity and residual saturation were respectively 0.6 and 0.06. The Manning's coefficient was 1.99 s/m$^{1/3}$ and the specific storage was 10^{-3} m^{-1}. This chosen Manning's coefficient gives a realistic base flow for the local rivers and accounts for scale mismatches between model and reality. Moreover, a layer of weathered bedrock starts below the soil, defined by a higher porosity (0.4) and hydraulic conductivity (2.78×10^{-5} m s^{-1}). The layer increased subsurface flow and countered the too shallow simulated groundwater levels. Furthermore, a 3D geologic model of the geological survey of Baden-Württemberg defined by 11 rock types was used, which should provide a realistic representation of the deeper subsurface of the domain. Using this setting, Schalge et al. (2021) demonstrated that the simulated GWL and river discharge were in good correspondence with the observed values.

3. Experimental Setup

3.1. Ensemble Simulations

The localized EnKF is used in the DA experiments with a total of 64 ensemble members. For each member random perturbations are introduced for (a) atmospheric forcing terms, (b) LAI and (c) soil properties. Correlations among these parameters and forcing terms are prescribed when imposing these perturbations:

1. Atmospheric forcings: In order to impose meaningful correlations we analyzed the 16 model runs with the fully coupled atmosphere-land surface-subsurface model (see Section 2.3.2). We evaluated the auto-variograms and cross-variograms, in space and time of air temperature, incoming shortwave radiation, incoming longwave radiation and precipitation with the library Gslib (Deutsch & Journel, 1992). The fitted semivariograms and cross-variograms are detailed in Appendix A. In total 64 spatially distributed fields of the four atmospheric variables were generated with help of these semivariograms and cross-variograms, using the GCOSIM3D code for sequential Gaussian co-simulation (Gómez-Hernández & Journel, 1993). Each variogram is composed of two spherical models having different ranges (see Isaaks and Srivastava (1994) for technical details). Let f_i be unperturbed forcing, and f_i' be an ensemble of perturbed forcings, $i \in \{1, 2, 3, 4\}$ represents one of the four atmospheric variables. The ensemble averages for the different atmospheric forcings $< f_i' >$ are consistent with those of the 16 fully coupled model runs. The standard deviations of the

perturbations are also rescaled to those of the 16 fully coupled model runs. We perturb the forcing either additively or multiplicatively. Let ϵ_i be the perturbation field. An ensemble member is additively perturbed according:

$$f_i' = f_i + \epsilon_i, \tag{12}$$

where ϵ_i is a normal distribution with $< \epsilon_i > = 0$ and the same standard deviation $\sigma(\epsilon_i)$ as for the specific variable over the 16 model runs. An ensemble member is multiplicatively perturbed according:

$$f_i' = f_i \epsilon_i, \tag{13}$$

where ϵ_i is a log-normal distribution with $< \epsilon_i > = 1$ and the same standard deviation as for the specific variable over 16 model runs. The precipitation and shortwave radiation are perturbed multiplicatively, while the longwave radiation and temperature are perturbed additively (Reichle et al., 2007).

2. LAI: The LAI in the domain is computed from MODIS for the four vegetated land use classes. The MODIS LAI values are perturbed to generate 64 different monthly spatial LAI distributions. The perturbations for the four different land use classes are perfectly correlated with $r = 1.0$. The LAI peak and the length of the growing season are perturbed and the perturbations are assumed to be directly proportional to each other so that a long growing season is associated with a high LAI peak. The standard deviations for the LAI peak are 1 for broad leaf forest, and 0.75 for the other PFT's. The standard deviations for the start and the end of the growing season are 0.5 months. The LAI ensemble average and standard deviation for the four PFTs are given in Figure 3.
3. Soil properties: The soil properties are based on the soil map also used for the VR simulation (Section 2.3). The ensemble of soil properties is generated according to the methodology described in Baroni et al. (2017). In summary, the map is sampled approximately at 1 point per 50 km^2 based on a stratified approach, and the soil properties, including sand and clay content, organic carbon content, and dry soil bulk density for each point are extracted. The variograms and covariograms of these soil properties are calculated and fitted with exponential models. These models are used for kriging the depth and carbon content of each soil layer, while geostatistical co-simulation is used to generate an ensemble of spatial fields for the other soil properties. The soil parameters such as saturated hydraulic conductivity, Mualem-van Genuchten parameters and porosity are estimated by applying pedotransfer functions on the generated spatial distributions of soil texture. Eventually, we have generated 64 ensemble members of spatially distributed soil properties and parameters. The ensemble average and standard deviation for four soil parameters at 10 cm depth are given in Figure 4, and the spatially averaged ensemble mean and standard deviation of the subsurface hydraulic parameters at different depths are shown in Figure 5.

With this ensemble of parameters and forcing terms, spin-up simulations are carried out for the ensemble of model runs for a period of 10 years, using the same atmospheric forcings for the year 2015. This results in an ensemble of equilibrated initial conditions for the DA experiments. The total calculations cost 3.7M core-hours and 6TB of disk space, on supercomputer JUWELS at Jülich Supercomputing Center (JSC).

3.2. Settings for Data Assimilation

We conducted open loop runs (i.e., forward model runs without DA) with CLM-Parflow for the year 2015 and compared the results with the VR simulation and more importantly, with the results of the DA experiments. This allows us to evaluate to which degree certain DA configurations can improve the characterization of soil moisture, groundwater levels, river discharge and ET.

For the DA experiments, time series of soil moisture and/or GWL data were assimilated. For soil moisture, observations at 100 randomly selected locations at 5 cm and/or 50 cm depth which cover the simulation domain are extracted from the VR for soil moisture DA. Similarly, the groundwater levels at the same 100 locations are extracted from the VR for assimilation. These soil moisture (and GWL) measurements are assimilated every 3 days, mimicking Cosmic Ray Neutron Sensor data (similar footprint size) (Baatz et al., 2014; Bogena et al., 2013; Franz et al., 2015; Schrön et al., 2017; Zreda et al., 2008), remotely sensed soil moisture data (similar temporal resolution) (Entekhabi et al., 2010) or networks of soil moisture sensors which cover one grid cell (Bogena et al., 2010). The locations never include river grid cells, and are displayed in Figure 2. For soil moisture

Table 1
List of Conducted Data Assimilation Experiments

Experiment (abbrev.)	Observations (at 100 locations)	State vector	Dampening (α)
0 (OL)	–	–	–
1 (SM_5)	θ at 5 cm	θ	–
2 (SM_{50})	θ at 50 cm	θ	–
3 ($SM_{5,50}$)	θ at 5 cm, 50 cm	θ	–
4 (SM_5_PAR)	θ at 5 cm	θ, log K_s	0.1
5 (SM_{50}_PAR)	θ at 50 cm	θ, log K_s	0.1
6 ($SM_{5,50}$_PAR)	θ at 5 cm, 50 cm	θ, log K_s	0.1
7 (WC_GWL)	GWL	ψ_{sat}	–
8 (WC_GWL_PAR)	GWL	ψ_{sat} and log $K_{s,sat}$	0.1
9 (FC_GWL)	GWL	ψ_{sat}, θ_{unsat}	–
10 (FC_GWL_SM)	GWL and θ at 5 cm	ψ_{sat}, θ_{unsat}	–
11 (FC_GWL_SM_PAR)	GWL and θ at 5 cm	ψ_{sat}, θ_{unsat}, log K_s	0.1

Note. The subscripts sat and unsat denote the saturated and unsaturated zones respectively, for soil moisture θ, pressure head ψ and saturated hydraulic conductivity K_s.

observations, the measurement error is 0.02 cm^3/cm^3 (Rosenbaum et al. (2010) give some examples of soil moisture sensors with this measurement error), and for GWL the error is 1 cm, corresponding to instrumentation error (e.g., British Geological Survey, https://www2.bgs.ac.uk/groundwater/datainfo/levels/measuring_errors.html). There are other sources of error which affect the GWL measurement uncertainty such as the uncertainty in terrain height, which are not considered here.

Soil moisture and GWL DA are studied with a series of experiments according to Table 1. EnKF with a horizontal localization radius of 12 km is used. In experiments 1–6 only soil moisture data are assimilated (measurements at 5 cm and/or 50 cm depth, with and without parameter estimation), to update soil moisture. In experiments 7–9 GWL data are assimilated (weakly coupled, fully coupled and with and without parameter estimation) and in experiments 10–11 both soil moisture and GWL measurements are jointly assimilated (again with and without parameter estimation). For the abbreviations of the different scenarios see also Table 1.

The root mean square error (RMSE) is introduced as a measure to evaluate the impact of DA on the characterization of soil moisture, ET, river discharge and GWL:

$$\mathrm{RMSE}(t) = \sqrt{\frac{\sum_{i=1}^{N_{obs}} \left[y_i^f(t) - y_i^o(t)\right]^2}{N_{obs}}}, \tag{14}$$

where $y_i^f(t)$ is the ensemble average of the forecasted values of the output variable under investigation (soil moisture content, ET, river discharge or GWL), at time t, $y_i^o(t)$ is the corresponding value taken from the VR, and N_{obs} is the total number of observations or verification points in the study. Notice that the verification points were not included as observation points in the DA. This is an advantage of the VR approach, which allows an extensive verification of the simulation results, not being limited to the observation locations only. We evaluated the DA performance not only for vertical profiles, but also in the horizontal plane. This was done by selecting four grid cells North, South, West and East of the observation locations (see Figure 6b). This results in $100 \times 4 = 400$ verification points in a sample when RMSE is evaluated. The comparison was made for grid cells located 0.8 km, 4 and 8 km away from the observed grid cell.

An additional evaluation metric was:

$$f(t) = \frac{\mathrm{RMSE}_{OL}(t) - \mathrm{RMSE}_{DA}(t)}{\mathrm{RMSE}_{OL}(t)}, \tag{15}$$

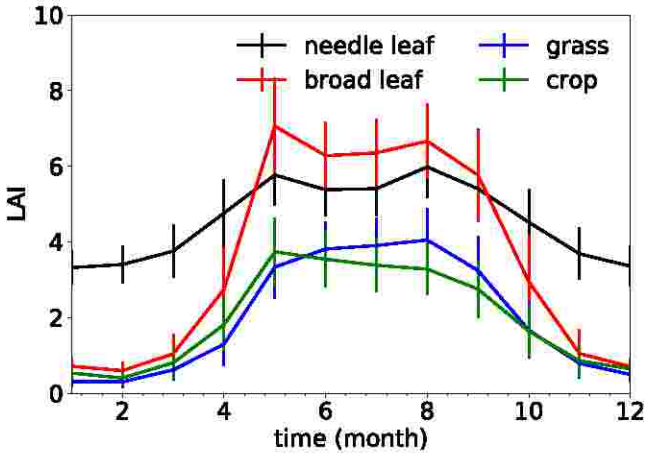

Figure 3. Monthly ensemble average of the Leaf Area Index's for the 4 plant functional types, namely, needle leaf forest (black), broad leaf forest (red), grassland (blue) and cropland (green). The ensemble standard deviations are shown as vertical bars.

where $\mathrm{RMSE}_{OL}(t)$ and $\mathrm{RMSE}_{DA}(t)$ are the RMSE of the open loop and DA experiment. This is the fractional reduction of the estimation error (in RMSE). The time average of this improvement over the year 2015 was then evaluated, that is,:

$$F = < f(t) >_{T_{2015}} . \tag{16}$$

In order to study the ensemble statistics, the estimation error of soil moisture θ at location i, $e_{\theta,i}$ is defined as the difference between the ensemble average and the value from the VR reference model such that $e_{\theta,i} = < \theta_i >_{\mathrm{ens}} - \theta_{i,o}$. The ensemble spread $\sigma_{\theta,i}$ is defined as the ensemble standard deviation of θ_i. For GWL, similar qualities $e_{d,i}$ and $\sigma_{d,i}$ are defined and will be referred to in later sections.

4. Results

4.1. Assimilation of Soil Moisture Data

We focus first on the DA experiments where soil moisture data were assimilated, measured at different depths. The mean and standard deviation of the estimation errors in soil moisture at the measurement locations are displayed in Table 2. It indicates that the average error approaches zero with reduced standard deviation. The statistics show that the initial guess is unbiased or slightly biased for these experiments, and after DA the widths of the distributions decrease.

On the other hand, when considering individual locations, there is substantial bias for soil moisture. The root mean squares of the time averaged bias over samples of 100 observation points and 400 verification points before and after DA for experiment $\mathrm{SM}_{5,50}$ are shown in Table 3. The table shows that the time series of soil moisture at most individual locations are biased, and DA reduces the bias.

Next, the ensemble spread before and after DA is examined for underestimation of the variance. To do this, the ratio of the yearly averaged square of estimate error to the ensemble spread is evaluated pointwise:

$$c_{\theta,i} = \frac{< e^2_{\theta,i} >_t}{< \sigma^2_{\theta,i} >_t}, \tag{17}$$

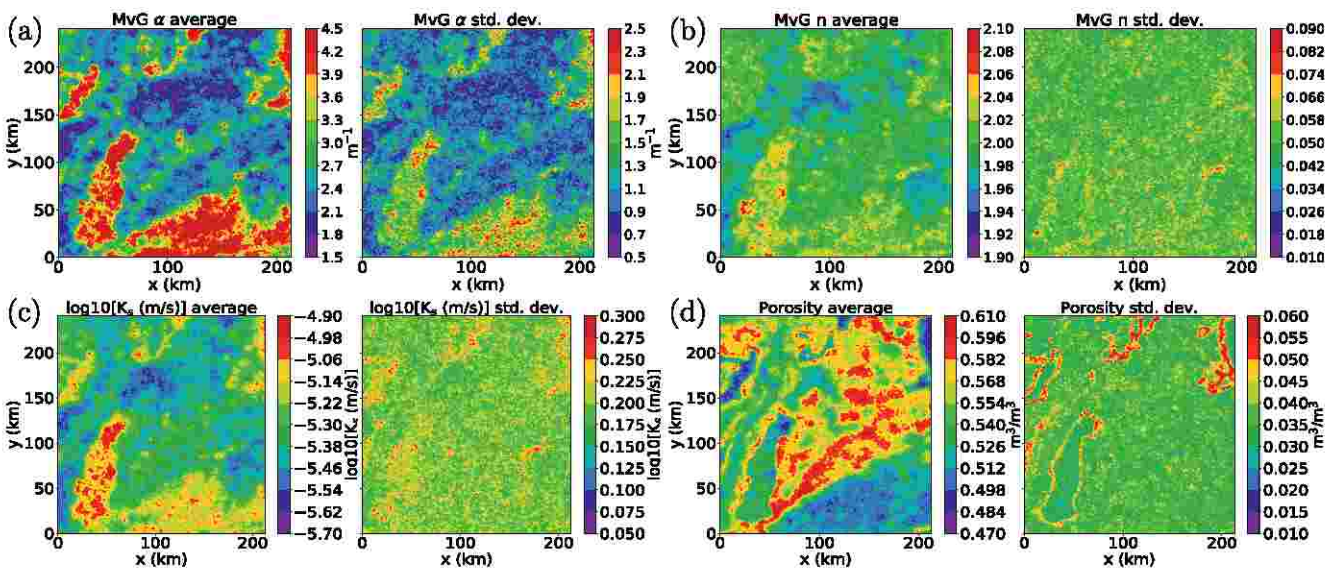

Figure 4. Ensemble mean (left) and standard deviation (right) of soil hydraulic parameters, at 10 cm depth in the model: (a) van Genuchten parameter α, (b) van Genuchten parameter n, (c) logarithmic saturated hydraulic conductivity, and (d) porosity.

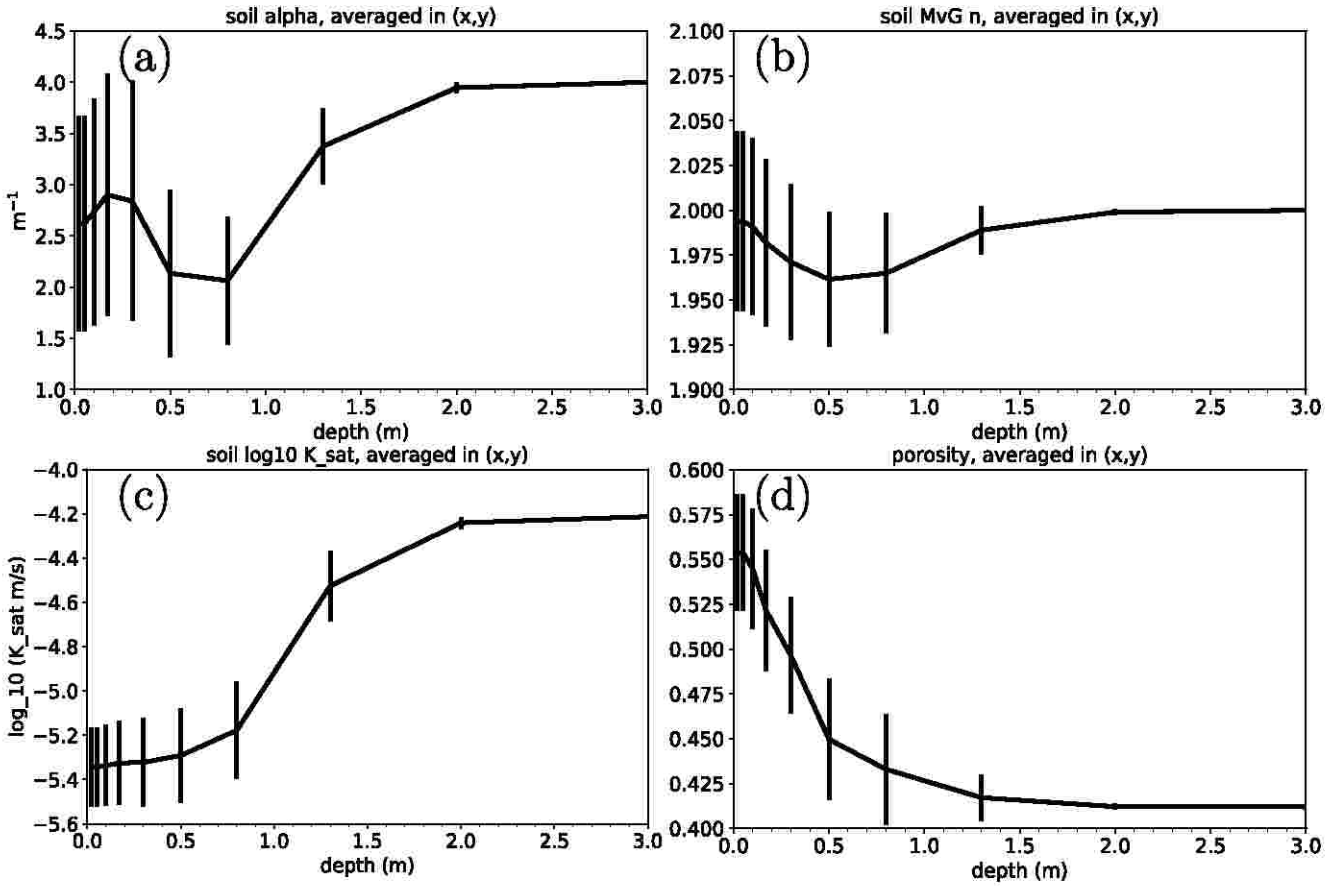

Figure 5. Ensemble mean (left) and standard deviation (right) of soil hydraulic parameters, averaged over the x-y plane, as a function of depth up to 3 m: (a) van Genuchten parameter α, (b) van Genuchten parameter n, (c) logarithmic saturated hydraulic conductivity, and (d) porosity.

where i is the location being evaluated. The spread is considered to be appropriate if the ratio is of order unity or smaller; otherwise the spread is too small (De Lannoy et al., 2006). In Figure 7 the distributions of the ratio $c_{\theta,i}$ for the open loop and experiment $SM_{5,50}$ are shown, for the 100 observation locations, at 5 and 50 cm depth, and also for 400 verification points at 8 km distance from the observation locations. In each sample, $c_{\theta,i}$ is less than unity at more than half of the locations, which shows that the ensemble spread is appropriate and without underdispersion.

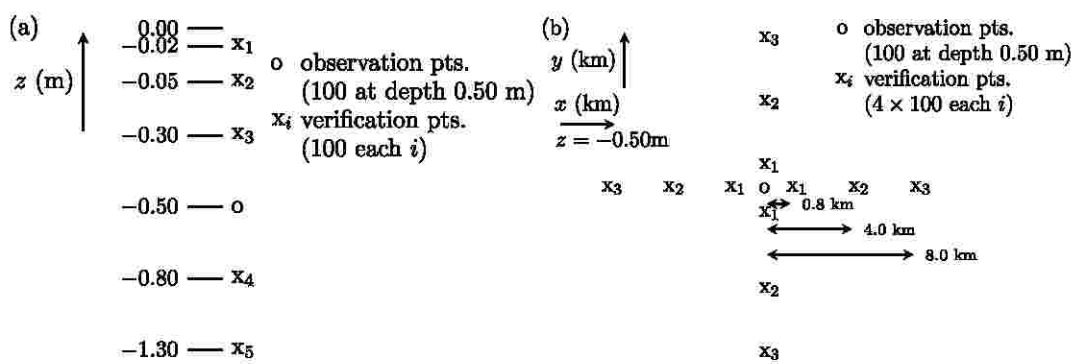

Figure 6. Locations sampled to calculate performance statistics like root mean square error with and without data assimilation, in panel (a) for the vertical profile at measurement locations and in panel (b) for the horizontal plane around measurement locations. Here we show the case where observations at depth 50 cm are assimilated.

Table 2
Average and Standard Deviation of the Soil Moisture Estimation Errors Calculated Over the Measurements and the Year 2015, Both for Open Loop Simulations and Data Assimilation Experiments

Experiment	Depth of observations (cm)	Average $\theta^f - \theta^o$ (m^3/m^3)	Standard deviation $\theta^f - \theta^o$ (m^3/m^3)
0. OL	5	0.0522	0.200
0. OL	50	0.110	0.326
0. OL	5 + 50	0.0812	0.272
1. SM_5	5	0.0108	0.0996
2. SM_{50}	50	0.0107	0.0936
3. $SM_{5,50}$	5 + 50	0.0112	0.0978
4. SM_5_PAR	5	0.0157	0.0960
5. SM_{50}_PAR	50	0.0188	0.0948
6. $SM_{5,50}$_PAR	5 + 50	0.0169	0.101
10. FC_GWL_SM	5	0.00324	0.0880
11. FC_GWL_SM_PAR	5	0.0107	0.0864

Table 3
Root Mean Square of the Time Averaged Bias of Soil Moisture Over the Measurement Locations and Verification Locations, for OL and $SM_{5,50}$

Experiment	Bias at depth (m) (obs. pts.)					
	0.05			0.5		
OL	7.14			14.5		
$SM_{5,50}$	2.75			2.92		
	Bias at depth (m) (verif. pts.)					
	0.05			0.05		
Horizontal distance (km)	0.8	4	8	0.8	4	8
OL	8.33	8.01	8.70	13.4	13.9	13.0
$SM_{5,50}$	6.60	7.06	8.12	10.3	11.8	12.0

Note. The values are normalized with the observation errors.

Then, the reduction in estimation error in soil moisture estimate is evaluated in detail. In Figure 8 the improvements in soil moisture characterization for experiment $SM_{5,50}$ on day 365 are shown. Small improvements are observed in most grid cells. The middle and right panels show the ensemble standard deviation with and without DA, and the uncertainties decrease in most of the grid cells when soil moisture data are assimilated.

Figure 9 illustrates the characterization of the soil moisture content. It shows the temporal evolution of the RMSE for soil moisture (at the locations defined in Figure 6). For all cases, the RMSE's at the observation points and most of the verification points are lower than the corresponding values in the open loop simulation. Note that for upper soil layers the RMSE shows larger temporal fluctuations, as the soil moisture at the upper layer is affected more by precipitation and drying events than the deeper layers. The DA reduces the forecast error for verification points in the soil profile above and below the observation points, as well as the verification points at some horizontal distance from the observations. For the deeper soil profile, the joint assimilation of soil moisture measurements at 5 and 50 cm depth gives the best results.

We evaluate the reduction in RMSE over the verification points (defined in Figure 6) relative to the open loop simulation and summarized the results as the yearly averages in Table 4, that is, the measure F defined in Equation 16 for θ. Table 4 clearly shows that when moving away horizontally from the observation locations, the improvement in forecast drops, which is also shown in Figure 9b. For SM_5, the measure F for improvement at 5 cm depth drops from 13% at 0.8 km horizontal distance to 3% at 8 km, and a similar trend is observed at 50 cm depth and for other soil moisture DA experiments (2–6). Results show that characterization of soil moisture for the upper 5 cm of the soil profile is improved more by the assimilation of soil moisture measurements taken at 5 cm depth. On the other hand, soil moisture characterization for deeper layers improves more with soil moisture measurements taken at 50 cm depth. The best overall results are obtained if soil moisture observations from both 5 and 50 cm depth are assimilated; this is especially the case for the deeper soil layers.

Then, we analyze the impact of the soil moisture DA on the ET modeling. The RMSE for each month is shown in Figure 10. The relative improvement in the characterization of ET is far smaller than for the soil moisture, regardless of the observations used, and from Table 4 the improvement is greater for experiments assimilating soil moisture taken at 5 cm depth, and in these cases the estimation errors are still only 1% lower than the case without DA. The improvement is more obvious in July, but the RMSE-reduction by DA is still smaller than 0.02 mm/month. The ET characterization in summer is a bit more improved by DA because the soil moisture content in most of the simulations is so low that ET is reduced by soil water availability, as a result the sensitivity of ET with respect to soil moisture content is larger (e.g., Seneviratne et al. (2006)). Table 4 summarizes the improvement in the ET characterization over the year for all the experiments.

Next, we evaluate the river discharge at the river gauge stations at Rockenau, Lauffen and Plochingen (indicated in Figure 2), simulated for the VR, open loop and DA experiments. These locations have been selected because they define catchment of increasing size within the overall domain. Baroni et al. (2017) found that river discharge is only sensitive to large scale perturbation of subsurface conditions. Therefore, assimilating river discharge data may not improve the characterization of small scale states and parameters. In contrast, we assimilated small scale observations with the hypothesis that river discharge estimate could have been affected and improved. This is done by evaluating the discharge of individual grid cells across the Neckar River at these three locations. Figure 11 shows the discharge time series (Only the results of $SM_{5,50}$ are shown for clarity, as the results for other experiments are very similar). The discharge in the VR is very close to the ensemble

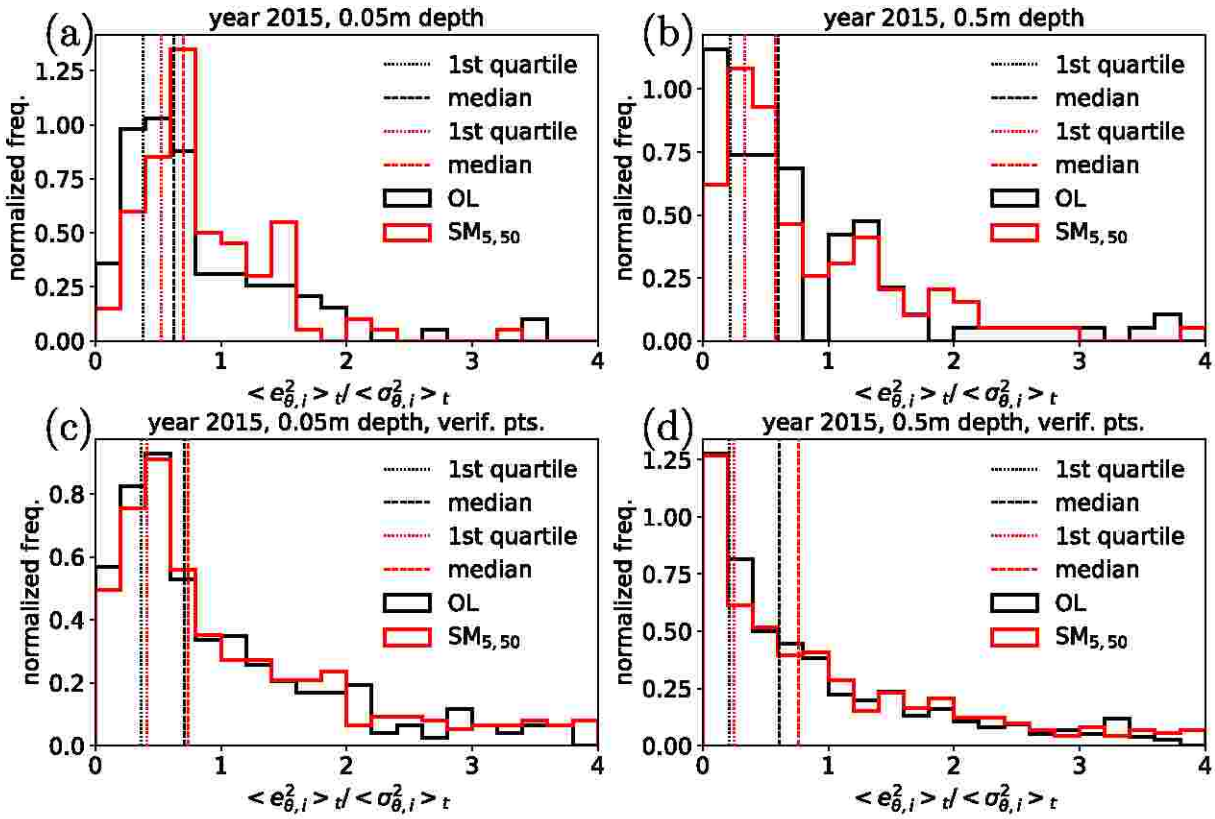

Figure 7. Distribution of $c_{\theta,i}$ for ensemble spread evaluation of soil moisture, for open loop and experiment $SM_{5,50}$. The locations are (a) observation locations at 5 cm depth, (b) observation locations at 50 cm depth, (c) verification locations at 5 cm depth, 8 km from the observation locations, (d) verification locations at 50 cm depth, 8 km from the observation locations.

average of the open loop simulation except for some discharge peaks, for example, around day 90 and day 230 in Plochingen, when the value for the open loop is 80 m^3/s higher (on day 90) (or 40 m^3/s lower on day 230) than the reference value. Careful evaluation shows that there are still tiny improvements in the estimates compared with the open loop for all tests (a fraction of $\sim 10^{-4}$ reduction in estimation error), summarized in Table 4. The perturbation of precipitation and soil parameters resulted only in a very limited spread of discharge, as the local effects averaged out over the larger catchment. The limited improvement in discharge characterization is therefore also related to the too limited ensemble spread.

We also analyze whether soil moisture DA could improve the characterization of the groundwater table depth. During a filter update, a new soil moisture analysis of the whole subsurface domain is obtained, and the pressure head is modified according to the Mualem-van Genuchten model (van Genuchten, 1980). As a result, the GWL will also be modified. The GWL of different simulation settings is compared with that of the VR. The RMSE at the observation locations are shown in Figure 12. The estimation errors are not reduced so assimilating soil moisture data does not improve the characterization of the groundwater table depth. One possible explanation is that the soil parameters α and n used in the ensemble simulation are also unknown and initialized with values different from the VR, and are not part of the state vector. As the pressure head is estimated from the forecasted soil moisture in the model, this could also have limited the improvement in the analysis. Another important factor will be that 100 soil moisture measurements are not informative enough to improve large scale GWL characterization.

4.2. Assimilation of Groundwater Level Data

The means and standard deviations of differences between forecasted and measured groundwater levels at measurement locations for the year 2015 are tabulated in Table 6. The ensemble water table depth is slightly shallower than the VR (see column one in Table 6). The small bias is reduced by DA, resulting in a narrower distribution.

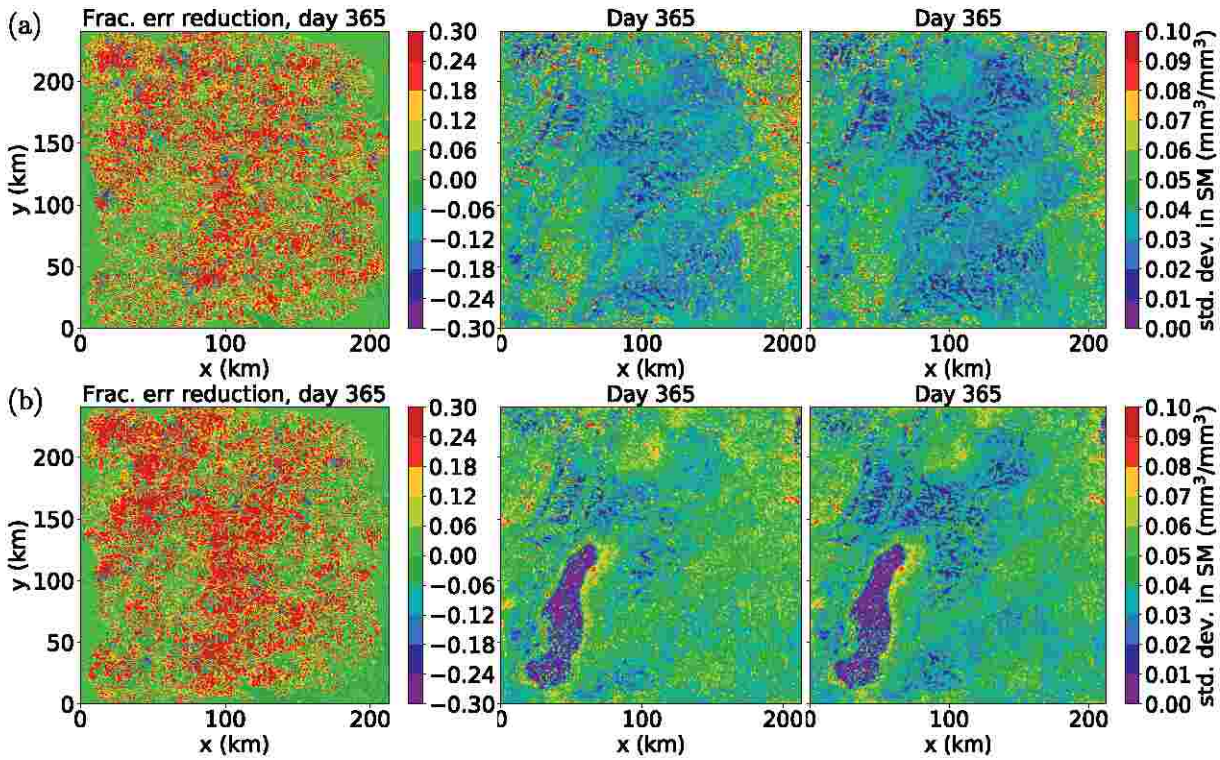

Figure 8. (a) Left: relative reduction of error after data assimilation according to Equation 15 on day 365 for $SM_{5,50}$ at 5 cm depth. Middle: the standard deviation of soil moisture of open loop simulation on day 365 at 5 cm depth. Right: the standard deviation of soil moisture of $SM_{5,50}$ on day 365 at 5 cm depth. (b) Same plots as (a) at 50 cm depth.

In all cases the GWL estimates are improved compared with the open loop simulation. Moreover, weakly coupled DA results in a better characterization of GWL compared with fully coupled DA, with smaller bias and standard deviations.

Again, similar to soil moisture DA, the root mean squares of time averaged bias of GWL over observation and verification points are evaluated before and after DA for experiment WC_GWL. The results are shown in Table 5. There are biases in the pointwise time series of GWL and the biases are reduced by DA for the observation points and most of the verification points.

We also evaluated underdispersion of GWL in the ensemble before and after DA. Similar to Equation 17, the ratio $c_{d,i} = < e_{d,i}^2 >_t / \sigma_{d,i}^2$ can be defined for GWL. The distributions of this ratio at the observation locations and verification locations (8 km from the observation locations) are shown in Figure 13, for the open loop and experiment WC_GWL. Compared with Figure 7, the spread is smaller and the medians are greater than unity but still less than 2. The median increases after DA, showing the ensemble spread decreases in general after GWL DA.

The absolute error of the GWL estimates over the domain for the open loop simulation and the weakly coupled case WC_GWL are compared. The fractional reduction of the estimation error after DA at the end of the assimilation period, is shown in Figure 14a. For 40% of the domain the characterization of the GWL improves significantly with DA, with a reduction of the estimation error of more than 1 cm (which is the observation error assumed in virtual GWL measurements). For 23% of the domain, on the other hand, there are more than 1 cm increase in the error, and the changes in the estimate are small for the rest of the domain (37%). The ensemble standard deviations before and after DA at the end of the year are also shown in Figures 14b and 14c respectively, which clearly indicates a reduction in uncertainties in the majority of the simulation domain.

The RMSE in GWL estimates at the observation locations and the neighboring grid cells are shown in Figure 15. In all cases, there is improvement in the forecast of the GWL at the measurement locations as well as in the surrounding of these locations, and the benefits decrease further away from the measurement locations. The weakly coupled cases show that GWL estimation is still improved 8 km away from the measurement locations

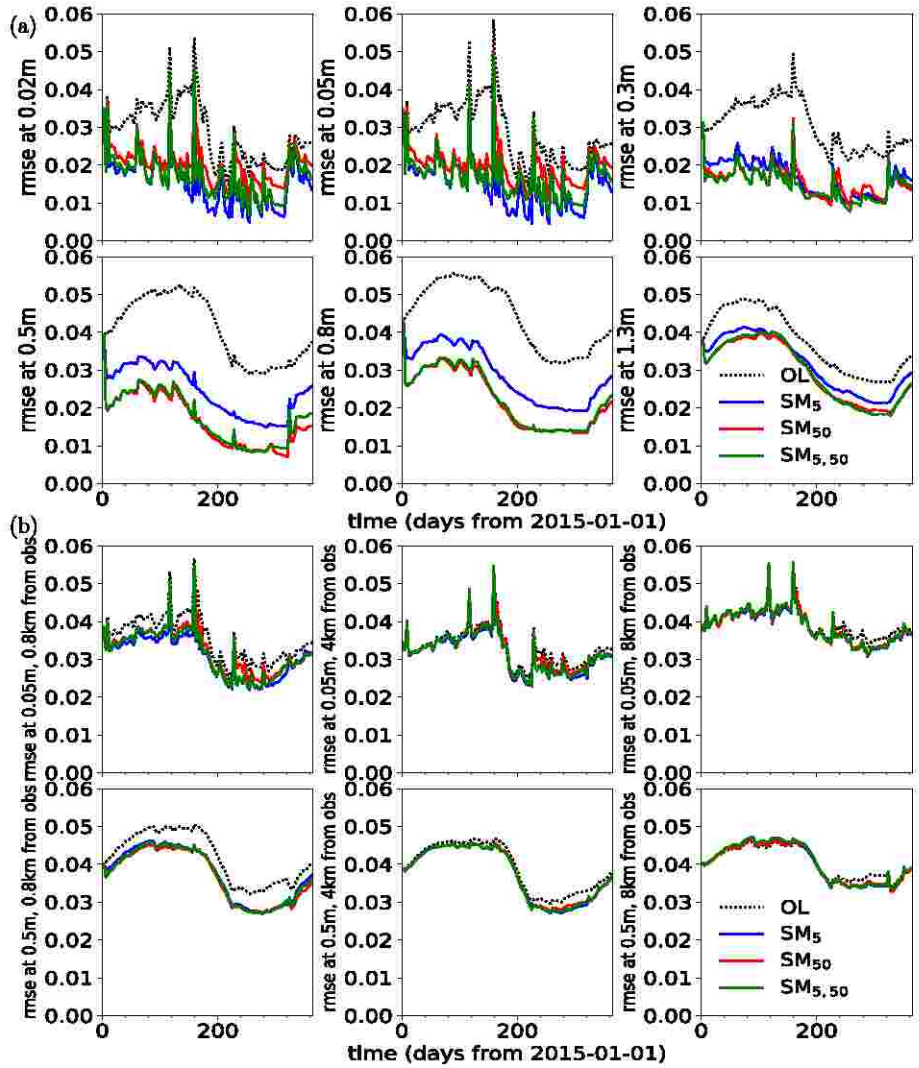

Figure 9. (a) Temporal evolution of root mean square error (RMSE) of soil moisture at six different depths (ranging between 0.02 and 1.30 m), at 100 observation locations, for the year 2015, for the open loop run and different data assimilation runs with assimilation of soil moisture data measured at 5 cm and/or 50 cm depth. (b) Temporal evolution of RMSE of soil moisture at 5 cm depth (upper row) or 50 cm depth (lower row), at verification points which are 800 m, 4 km or 8 km separated from observations.

(around 10% RMSE reduction) whereas for fully coupled simulations the GWL estimation is still improved 4 km away from measurement locations, but not anymore for 8 km separation. For a separation distance 800 m from the measurement locations, RMSE reductions are around 50% at the end of the simulation period for the different scenarios.

We analyze now whether GWL assimilation also improved soil moisture characterization. The RMSE along the vertical profile of the observation locations and in the neighborhood of the observation locations are shown in Figure 16 (The results in terms of RMSE for GWL DA with parameter update are very close to the results for only updating groundwater levels, so these are not shown in Figure 16 for clarity). For weakly coupled cases, soil moisture RMSE reduces by 2 $\sim$ 6%. For fully coupled DA where the soil moisture in the unsaturated zone is also directly updated, there is even not any improvement compared with the open loop simulation.

Table 4 also shows results for ET characterization. Similar to soil moisture DA, we get improved estimates compared with the open loop simulation in some experiments, but the improvements are again small. The relative improvements in river discharge estimation compared with open loop simulation are even smaller.

Table 4

Yearly Average Error Reduction F (Equation 16) for the Estimated Soil Moisture, Groundwater Level and Evapotranspiration for the Year 2015, for All the Experiments

Experiment	1 SM_5	2 SM_{50}	3 $SM_{5,50}$	4 SM_{5}_PAR	5 SM_{50}_PAR	6 $SM_{5,50}$_PAR	7 WC_GWL	8 WC_GWL_PAR	9 FC_GWL	10 FC_GWL_SM	11 FC_GWL_SM_PAR
Soil moisture/θ											
Depth (m)(obs loc)											
0.02	0.456	0.273	0.394	0.514	0.298	0.434	0.0247	0.0240	−0.222	0.516	0.561
0.05	0.479	0.286	0.411	0.537	0.308	0.453	0.0259	0.0251	−0.241	0.540	0.587
0.30	0.449	0.479	0.512	0.453	0.511	0.541	0.0109	0.0109	−0.296	0.433	0.443
0.50	0.406	0.605	0.585	0.385	0.625	0.592	0.0120	0.0121	−0.315	0.354	0.364
0.80	0.357	0.511	0.501	0.341	0.520	0.500	−0.00497	−0.00504	−0.340	0.280	0.294
1.30	0.158	0.220	0.222	0.159	0.227	0.226	0.0125	0.0134	−0.0575	0.128	0.102
Horizontal distance (km)											
at depth 5 cm											
0.8	0.132	0.0865	0.113	0.139	0.0912	0.119	0.0587	0.0587	−0.102	0.0155	−0.0375
4.0	0.0566	0.0300	0.0457	0.0580	0.0258	0.0459	0.0398	0.0397	−0.170	−0.158	−0.2065
8.0	0.0274	0.0188	0.0203	0.0258	0.0197	0.0185	−0.0278	−0.0294	−0.0824	−0.0559	−0.0718
Horizontal distance (km)											
at depth 50 cm											
0.8	0.109	0.126	0.119	0.104	0.128	0.121	0.0266	0.0270	−0.255	−0.108	−0.161
4.0	0.0469	0.0404	0.0464	0.0480	0.0376	0.0456	0.0178	0.0169	−0.272	−0.288	−0.317
8.0	0.0156	0.0106	0.00448	0.0137	0.0102	0.00311	−0.0391	−0.0404	−0.219	−0.201	−0.192
Groundwater level (m)											
Horizontal distance (km)											
0 (obs loc)	−0.126	−0.164	−0.152	−0.125	−0.164	−0.152	0.761	0.761	0.863	0.875	0.872
0.8	−0.0809	−0.0824	−0.0840	−0.0830	−0.0793	−0.0821	0.345	0.345	0.441	0.535	0.541
4.0	−0.121	−0.133	−0.131	−0.121	−0.134	−0.127	0.165	0.172	0.127	0.123	0.126
8.0	0.0124	0.0187	0.0135	0.0138	0.0196	0.0285	0.0930	0.101	−0.175	−0.101	−0.100
River discharge											
Rockenau (10^{-4})	4.54	4.18	4.05	5.03	4.49	4.49	5.12	5.10	1.21	3.57	3.69
Lauffen (10^{-4})	3.85	3.66	3.60	4.08	3.71	3.70	3.14	3.02	0.486	2.21	2.10
Plochingen (0^{-4})	1.51	1.34	1.35	1.63	1.30	1.41	1.37	1.36	0.791	1.47	1.41
Evapotranspiration (obs loc)	0.0129	−0.00351	0.00319	0.0126	−0.00433	0.00314	0.00485	0.00587	−0.0228	0.0131	0.0121

Note. The metric *F* is evaluated for soil moisture along the vertical profile (rows 1–6) (labeled as depth) and at a horizontal distance from the observation at 5 cm depth (rows 7–9) and 50 cm depth (rows 10–12). The metric *F* is also evaluated for groundwater level and are shown at the measurement locations and as a horizontal distance from the location as well (rows 13–16). Similarly the values of *F* for discharge at the 3 locations (rows 17–19) and for the ET (the last row) are shown.

4.3. Joint Assimilation of Groundwater Level and Surface Soil Moisture Data

For the fully coupled cases FC_GWL_SM and FC_GWL_SM_PAR, GWL and surface soil moisture data are jointly assimilated, resulting in improvement in the soil moisture forecast for observation locations along the vertical profile, but for other verification points the soil moisture forecast is worse than for the open loop run. Also, similar to cases with univariate soil moisture or GWL DA, there are only very small improvements in the estimates of ET. This implies that the additional assimilation of GWL data, besides soil moisture data, does not improve soil moisture characterization further, at least not for this relatively small ensemble size.

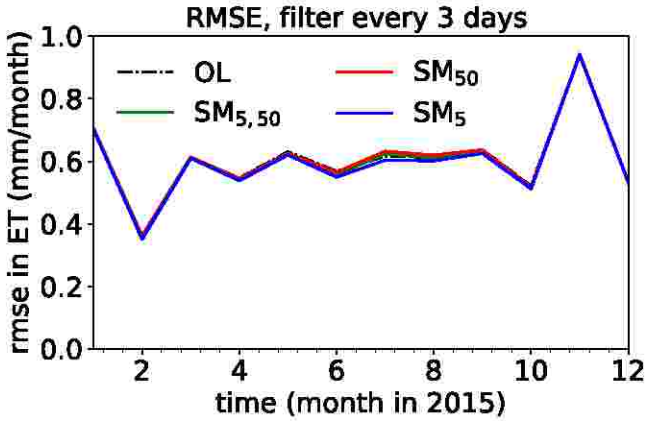

Figure 10. Monthly root mean square error of evapotranspiration at the 100 observation locations, for data assimilation-experiments with soil moisture observations at 5 cm (blue), 50 cm (red), and both (green).

4.4. Joint State-Parameter Estimation

If parameters are estimated with help of the assimilation of soil moisture data, soil moisture characterization is more improved than for the cases with state updating alone, without parameter estimation. This is shown in the summary of results on the basis of yearly statistics of the error measure F in Equation 16 presented in Table 4. Such a result is expected as K_s is an important parameter to characterize θ in the subsurface model. Figure 17 shows the color contour plot of log K_s of the VR, open loop run and DA runs, over the Neckar river catchment. It shows that DA starts with a higher K_s average than in the open loop. Eventually, at the end of year 2015, the mean spatial average of K_s is reduced but still systematically higher than for the VR. The time evolution of the estimation error in log K_s over the observation locations and other neighboring locations is shown in Figure 18, and it shows that the estimation errors are decreasing asymptotically as time evolves. Among the cases examined, assimilation soil moisture data measured at 50 cm depth gives the best estimate of K_s. Figure 18b indicates the estimation error of log K_s at the end of the assimilation period, subtracted by the error before DA. The RMSE is evaluated over samples of verification points with increasing distance from the observations (Figure 6). There is improvement in the estimate of the hydraulic conductivity locally around the observation locations, at a vicinity of 4 km from the observations, further than 4 km the estimation error increases to the same levels as the initial estimated K_s and there is no benefit from DA. This range is shorter than the filter localization radius (12 km), and also shorter than the range for which improved soil moisture estimation is achieved (at least 8 km). Overall, at the end of the DA period, the RMSE of log K_s at the observation locations is reduced by 11%, for both 5 and 50 cm depths.

On the other hand, updating log K_s with GWL data does not improve GWL estimation, when compared to the corresponding experiments where log K_s is not updated (Figure 15). This is the case for the weakly coupled and fully coupled DA scenarios. Figure 19 also illustrates that there is no improvement in the estimate of log K_s at 5 cm and 2.0 m depth at the observation locations, after DA. For the weakly coupled case, the components of the state vector for the unsaturated zone will not be modified during filtering, and the changes in log K_s near the surface are limited. For the fully coupled case, the estimate does not approach the values in the VR, compared with the case where only soil moisture data at 5 cm depth are assimilated. Therefore, unlike soil moisture, GWL DA is ineffective to estimate hydraulic conductivity, possibly related to the small ensemble size.

In all experiments, additional estimation of hydraulic conductivity does not improve the ET-modeling. The effectiveness of estimating K_s is hindered by (a) the small ensemble size used and (b) the uncertainties of soil parameters other than K_s which are not estimated.

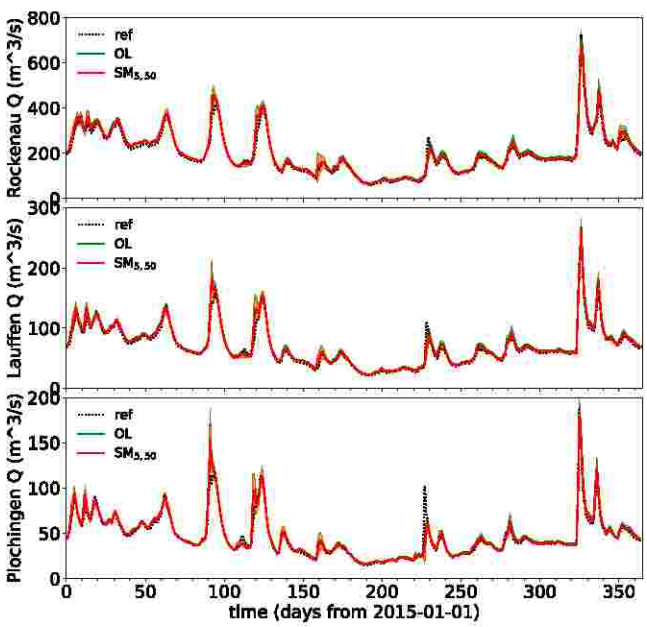

Figure 11. Temporal evolution of the river discharge at Rockenau, Lauffen and Plochingen in the year 2015, for the virtual reality (black dotted), open loop simulation (green solid), $SM_{5,50}$ (red solid) respectively, and ensemble standard deviations are shown in light green and light red shaded area respectively (barely noticeable at some peaks).

5. Discussion

Without DA, model states and fluxes differ between the VR and the ensemble members as the atmospheric forcing terms, vegetation parameters and the subsurface hydraulic parameters are uncertain. We showed that these variabilities in forcings and parameters contribute to the differences between the reference and the ensemble model runs, in soil moisture, GWL, river discharge and ET. Data assimilation is capable of reducing these differences in soil moisture and GWL between the reference and ensemble model runs significantly, but benefits are less obvious for river discharge and ET.

Assimilation of soil moisture data over the network of 100 locations can improve the characterization of soil moisture for the whole domain. The

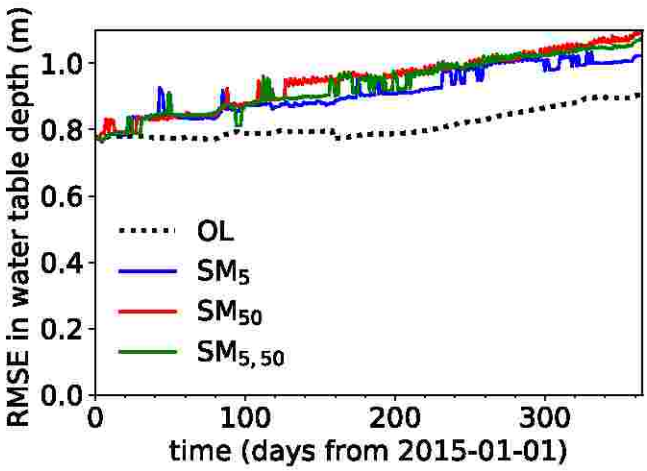

Figure 12. Root mean square error of water table depth for the 100 observation locations over the year 2015, for the open loop and different data assimilation experiments where soil moisture measurements were assimilated.

estimation error is higher for locations horizontally further away from the observations, but still better than the estimate from the open loop simulation. Data assimilation also improves the ET characterization, compared to the open loop simulation. However, the average RMSE-reduction is limited to 1.2% over the year 2015, and here soil moisture observations near the surface are more effective in reducing RMSE of ET than deeper soil moisture observations. Larger improvements are found when conditions are drier (summer and beginning autumn). When soil moisture content is high, the ET is close to the potential ET and insensitive to variations in soil moisture content. In addition, the latent heat flux is not the observation used in the experiments, and only soil moisture content at one location in the soil profile is assimilated. Therefore, possibilities to improve ET-characterization would be to assimilate further data types (latent heat flux, but also LAI). In this work, the land surface model CLM is driven by atmospheric forcings, instead of being coupled to an atmospheric model. Data assimilation will not alter the forcing terms such as incident radiation and the latter also affects the ET. This also limits the performance regarding the ET estimation. Therefore, a possible improvement is to extend the model to a fully coupled system with an atmospheric model integrated in the DA framework. In general, the experiments show that assimilation of soil moisture alone will often not be enough to substantially improve the ET-characterization.

On the other hand, soil moisture DA does not improve the GWL characterization. The new estimate of GWL from soil moisture DA gives a slightly higher RMSE compared with the open loop simulation, though the differences are small compared with the absolute value of the error. The values of the river discharge in the VR at the 3 locations investigated are close to the values of the ensemble average, subject to small uncertainties, and the effect of DA is also limited. These results point to the difficulties to characterize the states of different compartments of an integrated land surface-subsurface model with just one specific measurement type. It is also possible that the relatively small ensemble size compromises the effectiveness of the method. However, DA studies with integrated models will in general have to face that the ensemble size will be smaller than desired.

If saturated hydraulic conductivity is updated together with the states, there is slightly further improvement in soil moisture estimation compared with the cases soil moisture is updated alone, but the effect is relatively local, along the vertical profile, and the further benefit is less than 6% along the vertical profile. The new estimate of the K_s is also approaching the reference K_s, but the effect is again locally around the soil moisture observation locations. The ensemble size (64) is not large enough to improve further the simulations, and is associated with (too) large sampling errors for the covariances. Studies found that for successful joint state-parameter estimation with groundwater hydrological models ~200–~1,000 realizations are needed (Hendricks-Franssen & Kinzelbach, 2008; Zhou et al., 2011). Moreover, other soil parameters used in the ensemble simulations are different from the VR setup (as described in Section 3) and these parameters are not included in the state vector. Reasonably estimating these parameters as well by DA also requires enlargement of the ensemble size. However, in this study a larger ensemble size was not feasible given the needed computation times with the coupled high-resolution land surface-subsurface model.

Table 5
Root Mean Square of the Time Averaged Bias of Groundwater Level Over the Measurement Locations and Verification Locations, for OL and Experiment WC_GWL

		Verif. pts. Horizontal distance from obs. pts. (km)		
Experiment	Obs. pts	0.8	4	8
OL	79.1	84.5	77.2	60.0
WC_GWL	15.8	93.2	61.8	52.7

Note. The values are normalized with the observation errors.

Assimilation of GWL data helped to improve soil moisture estimation in the weakly coupled scenarios, to a lesser extent than assimilating directly soil moisture data, as could be expected. In these scenarios, the soil moisture is not directly modified by the EnKF, instead, the effect of updating the pressure head under the watertable propagates upwards by the model Parflow and alters the pressure head and therefore also soil moisture content above the groundwater table. In the fully coupled scenarios, the soil moisture above the watertable is directly updated by the EnKF. However, there is no improvement in the estimate of soil moisture for the case only GWL data are assimilated (FC_GWL). If besides GWL data also surface soil moisture data are assimilated, the soil moisture estimate is only improved along the

Table 6
Average and Standard Deviation of the Estimation Errors of Groundwater Level at the Measurement Locations for the Year 2015, and for the Experiments Where Groundwater Level Data Are Assimilated

Experiment	Average GWL^l-GWL^o (m)	Standard deviation GWL^l-GWL^o (m)
0. OL	−0.395	0.568
7. WC_GWL	0.0464	0.291
8. WC_GWL_PAR	−0.0465	0.388
9. FC_GWL	−0.102	0.525
10. FC_GWL_SM	−0.0981	0.480
11.FC_GWL_SM_PAR	−0.0940	0.475

Note. The corresponding values for open loop simulation are also shown for comparison.

vertical profile at the observation location. At other verification points, the RMSE is higher than the estimate from the open loop simulation. The joint assimilation of soil moisture and GWL data results in a worse soil moisture estimation than the univariate assimilation of soil moisture data. This conclusion differs from the work by H. Zhang et al. (2018), who used the same coupled model CLM-Parflow and stated that jointly assimilating surface soil moisture and GWL data with update of all subsurface states resulted in a better soil moisture characterization than other scenarios. This is because in their work the simulation domain consisted of only 2×2 soil columns, and a larger ensemble could be used. In this work, groundwater levels are also strongly controlled by topography variations and lateral groundwater flow, so that soil moisture data are probably less informative about groundwater levels and vice versa. In addition, as indicated before, the small ensemble size may have prevented better results, but no significant filter inbreeding was observed, suggesting that the small ensemble size has not been a major factor preventing better results. Currently, a single DA experiment with 64 ensemble members takes 0.55M core-hours (not including the spin up) and 6TB of disk space (the disk usage is the same for the open loop simulation). The cost is at least directly proportional to the ensemble size.

Both the weakly and fully coupled assimilation of groundwater levels result in an improved characterization of groundwater levels. The weakly coupled cases also perform better than the fully coupled cases, with a narrower distribution of estimation error over the yearly measurements. The weakly coupled cases bring a longer range of improvement around the observation locations compared with the fully coupled cases. It is for the first time that for a larger scale integrated land surface-subsurface model, and a fully 3D heterogeneous subsurface, it could be shown that GWL assimilation is able to improve GWL characterization also substantially between measurement locations. Assimilating GWL data can hardly bring any better estimate of ET in all scenarios, which is related to the very limited improvement in soil moisture characterization, already discussed earlier. This is partly related to the fact that in a substantial part of the domain the watertable is quite deep, so that soil moisture content near the surface and ET are not better characterized there.

We assumed hydrostatic conditions for assimilating groundwater levels. In a real-world case multiple aquifers separated by aquitards might be present, so that the hydrostatic assumption would not be valid. In that case, for the updating step in DA a hydrostatic assumption for each aquifer layer could be made, and measurements could be used to just update pressures in a single aquifer layer.

The coarse model discretization is a limitation in this work, for example, for the representation of aquifers in narrow river valleys. We are in current research addressing the issue of a scale mismatch between reality and the simulation model. We agree that in case of a coarse discretization also the comparison of a GWL measurement and simulated groundwater levels is more complicated, because the simulated groundwater levels need to be inter-

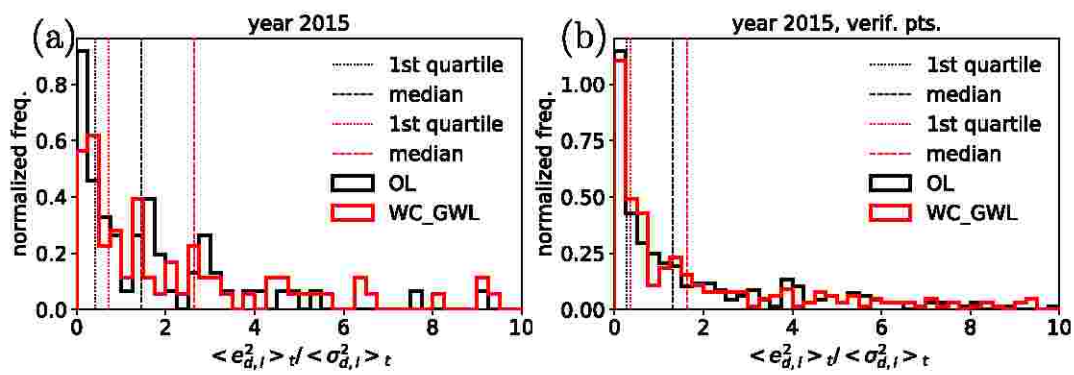

Figure 13. Distribution of $c_{d,l}$ for ensemble spread evaluation of groundwater level, for open loop and experiment WC_GWL. The plotted locations are (a) observation locations, (b) verification locations, 8 km from the observation locations.

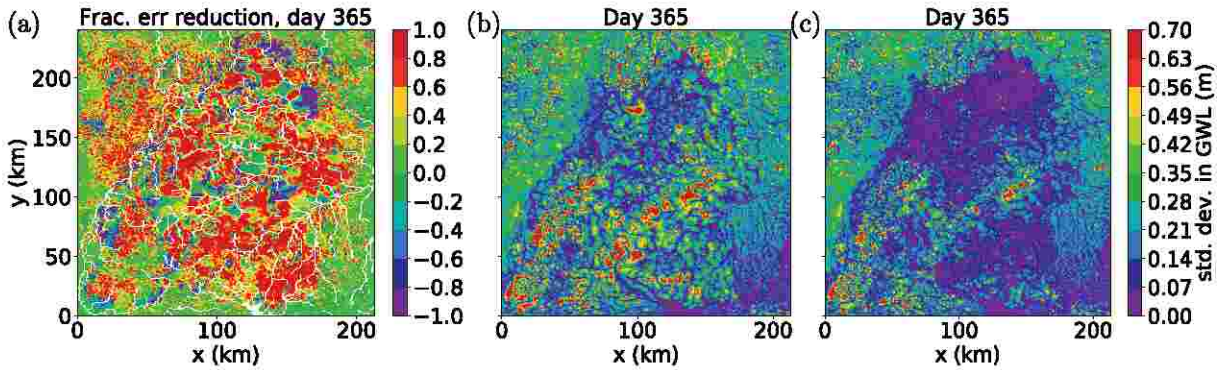

Figure 14. (a) Fractional reduction of estimation error in groundwater level (GWL) after data assimilation at the end of the simulation period (scenario WC_GWL). (b) Ensemble standard deviation of the GWL for the open loop run. (c) Ensemble standard deviation of the GWL for WC_GWL scenario.

polated to make a direct comparison with the measured value, exactly at the measurement locations. This interpolation will also be associated with uncertainty. In this work this issue was avoided by assigning the measurement locations to the grid cell centers so that measurement locations and simulated values can be directly compared. In real-world cases, more complications can be expected. In addition, the effect of the prior on the results is not studied in this work, which can be important. Examples of related studies include Crow and Loon (2006), in which they conducted a series of twin experiments of soil moisture DA with EnKF, and they found that when the model error is intentionally under-estimated, or when incorrect model error assumptions are imposed, the accuracy of the results would be reduced or the same as that of the open loop simulations. Therefore, the sensitivity with respect to the prior will also be another follow-up study.

Although the Mualem-van Genuchten parameters α and n can also be considered for estimation (e.g., Chaudhuri et al. (2018)), this was not done in this work. The problem size did not allow us to use a large ensemble, and the estimation of α and n is prone to (numerical) instabilities under those conditions. On the other hand, Gebler et al. (2019) illustrated that uncertainty of α and n could have little impact on the performance of the DA, though this finding might have been case specific.

Another important future follow-up of this work could be the assimilation for a real-world case, including remotely sensed soil moisture data and GWL measurements. In addition, the effect of the biased ensemble can be investigated in detail, and using a bias-aware filter is also a way to reduce such a bias.

Overall, it can be observed that the benefits of DA in the above experiments are relatively local in the vicinity of the measurement locations. When compared with the open loop simulation, the improvements in soil moisture or GWL characterization are limited to a distance shorter than the localization radius of the filter, and the effect is more obvious for parameter estimation. Nevertheless, given the relatively dense network of groundwater wells domain wide improvements in the GWL characterization could be achieved, which was shown in this work for the first time for a coupled land surface-subsurface model and a complex topography with many small valleys.

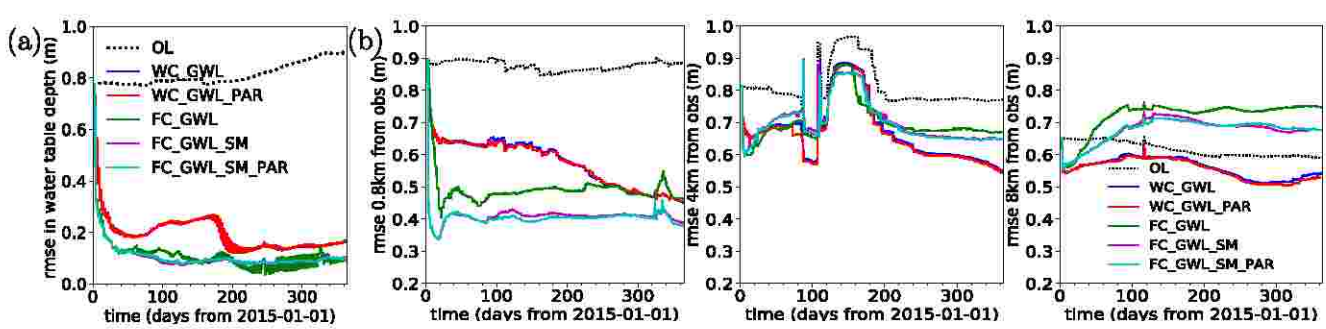

Figure 15. (a) Temporal evolution of root mean square error (RMSE) of groundwater level (GWL) at 100 observation locations, for the year 2015, for the open loop run (dotted line) and different data assimilation runs for Experiments WC_GWL (blue), FC_GWL (green), and FC_GWL_SM (magenta). (b) Temporal evolution of RMSE of GWL at verification points which are 800 m, 4 km or 8 km separated from observations.

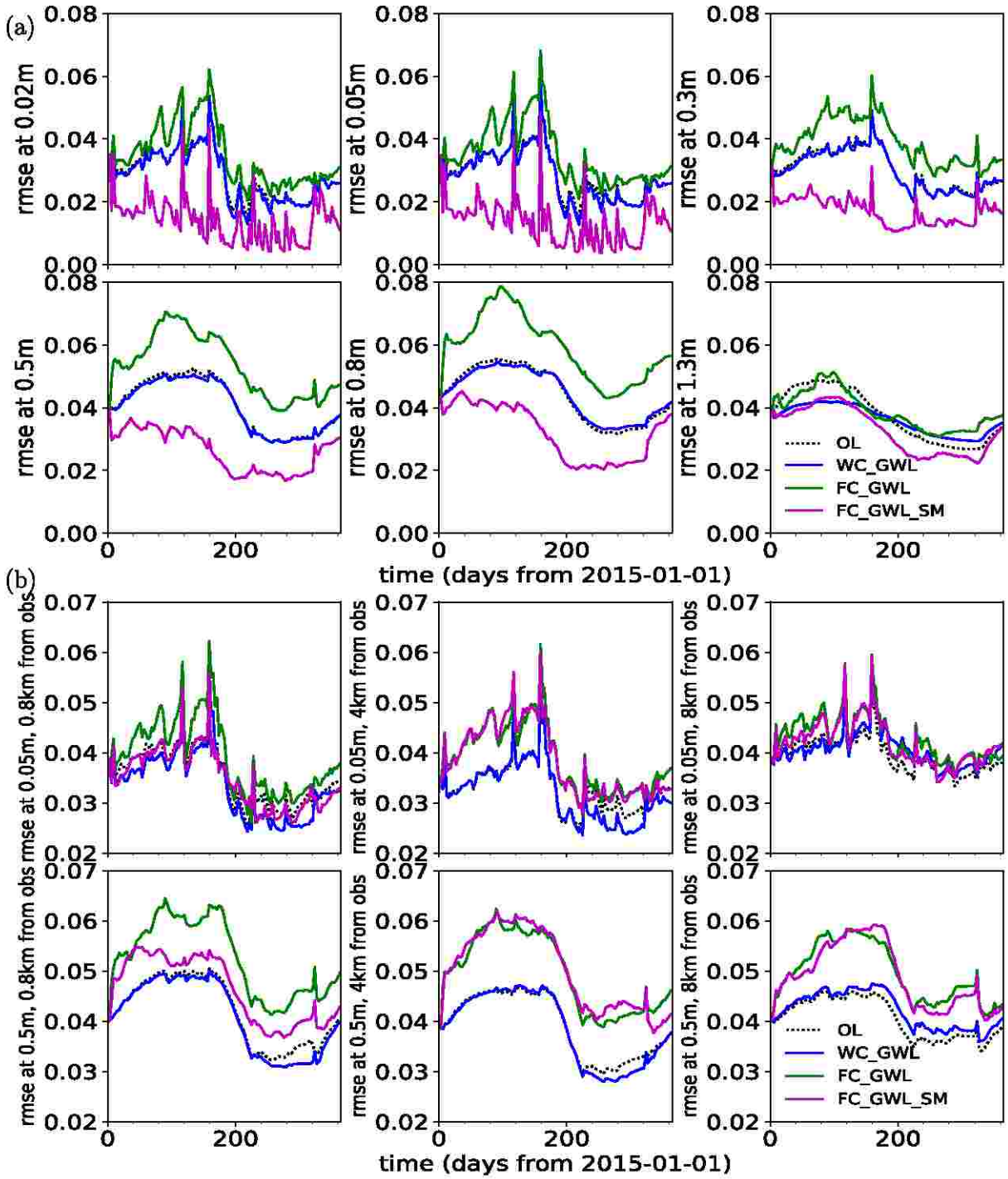

Figure 16. (a) Temporal evolution of root mean square error (RMSE) of soil moisture at six different depths (ranging between 0.02 and 1.30 m), at 100 observation locations, for the year 2015, for the open loop run (dotted black) and for the different scenarios, WC_GWL (blue), FC_GWL (green) and FC_GWL_SM(magenta). (b) Temporal evolution of RMSE of soil moisture at 5 cm depth (upper row) or 50 cm depth (lower row), at verification points which are 800 m, 4 km or 8 km separated from observations.

Therefore, a denser measurement network can definitely improve the characterization of the subsurface states and parameters of large land surface-subsurface simulation domains. More computation power to increase the ensemble size is also important for achieving better results.

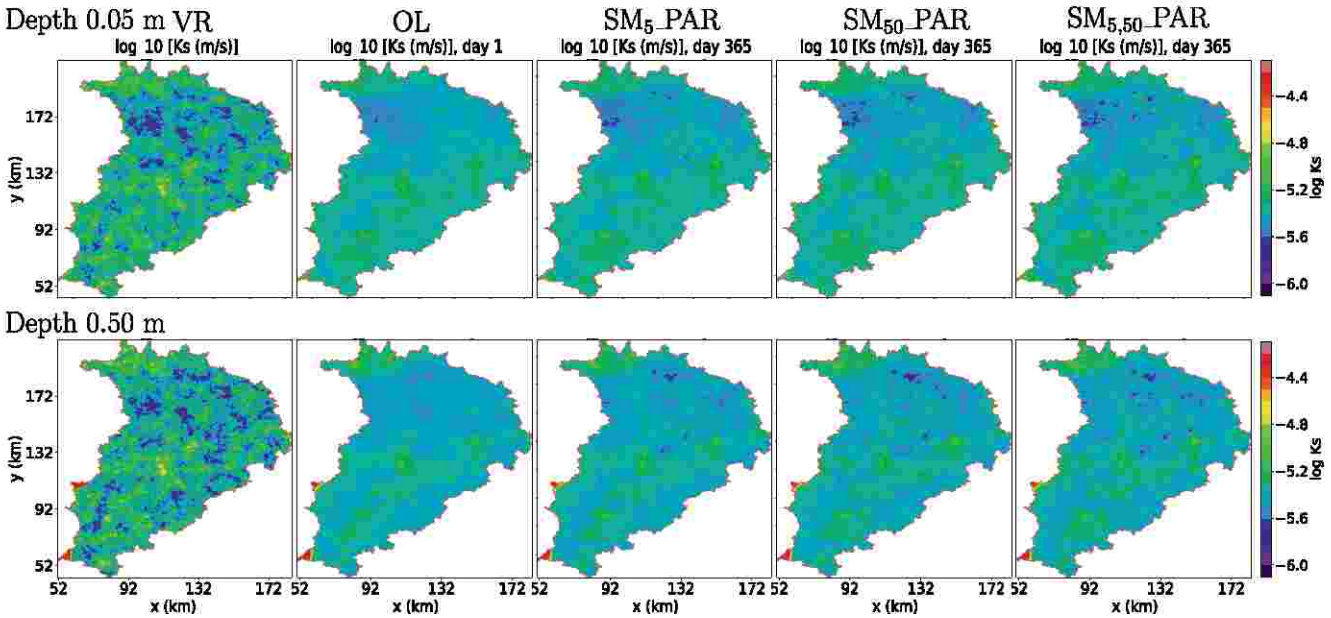

Figure 17. Color contour plots of the saturated hydraulic conductivity K_s. The panels at the top show K_s at depth 5 cm, and the panels at the bottom at depth 50 cm. From left to right: (a) virtual reality, (b) open loop simulation, (c) data assimilation using soil moisture observations (a) SM_5_PAR, (b) SM_{50}_PAR, and (c) $SM_{5,50}$_PAR.

6. Conclusion

This study deals with soil moisture DA and GWL DA over the Neckar catchment (57,850 km^2) in Germany using a VR approach. A VR is constructed for the area with the coupled land surface-subsurface model TSMP, and at a high horizontal spatial resolution of 800 m for the land surface and subsurface components. Soil moisture and GWL observations are extracted from the VR and used in DA experiments. Data assimilation experiments are done with the coupled land surface and subsurface components of TSMP, and the EnKF including localization. The ensemble simulation are based on perturbed atmospheric forcings, LAI and soil hydraulic parameters. The DA experiments are carried out with 64 ensemble members and for the year 2015. Soil moisture characterization

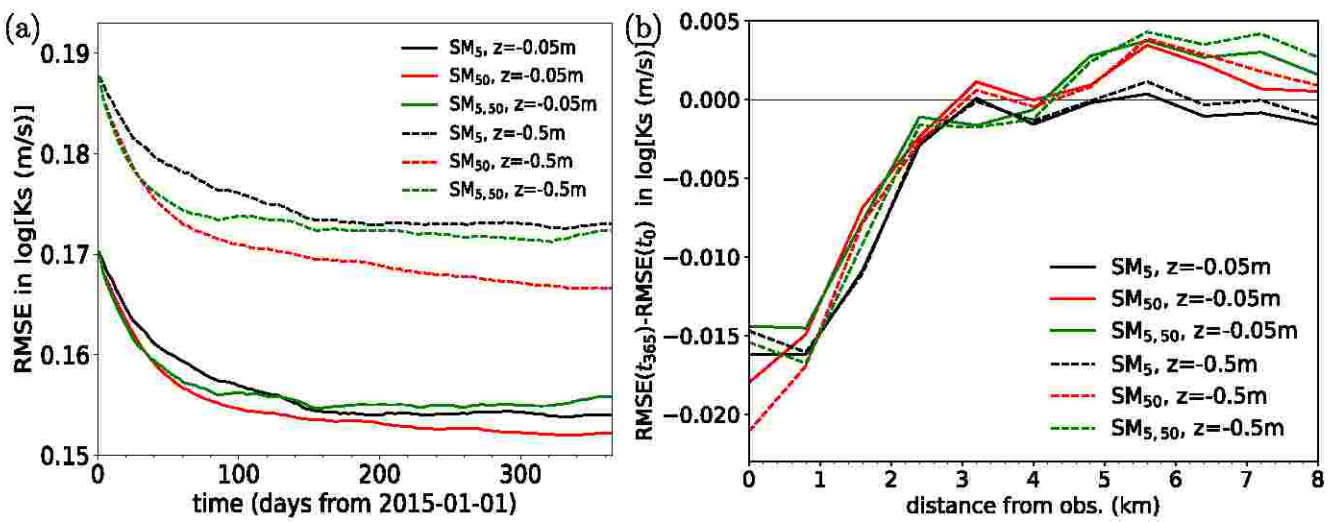

Figure 18. (a) Root mean square error (RMSE) of $\log_{10}K_s$, average over the observation locations at depth 5 cm (solid lines) and 50 cm (dashed lines) respectively. Results of 3 scenarios are shown, using observations at depth 5 cm (black lines), 50 cm (red lines), and both depths (green lines.) (b) Difference of RMSE of $\log_{10}K_s$ between the end and start of the year 2015, over samples of verification points at increasing horizontal ranges from the observation locations. Again results for depth 5 cm (solid lines) and 50 cm (broken lines) are displayed, in the cases soil moisture data at depths 5 cm (black), 50 cm (red) and both (green) (Experiments 6, 7, 8 resp.) are assimilated.

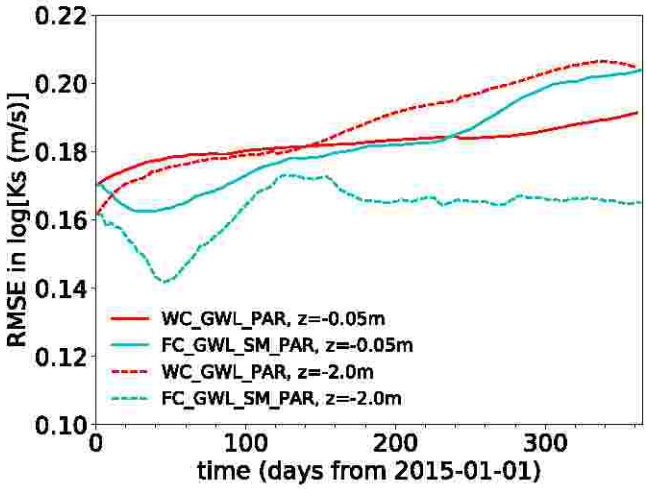

Figure 19. Root mean square error of $\log_{10}K_s$, averaged over the observation locations at depth 0.5 m (solid lines) and 2.0 m (dashed lines) respectively. Results of 2 scenarios are shown, WC_GWL_PAR (red lines) and FC_GWL_SM_PAR (cyan lines).

is improved with a RMSE-reduction in the range of 40%–60% for the soil moisture profile at the observation points. The improvement is smaller in the horizontal direction moving away from the observation points, namely 13% at 800 m distance and 3% at 8 km distance. Assimilation of near surface soil moisture data (5 cm) results in better soil moisture estimates near the surface. Assimilation of soil moisture measurements taken at 50 cm improves surface soil moisture characterization nearly as much as measurements from 5 cm, and improves the deeper soil moisture profile characterization more. Joint assimilation of soil moisture from 5 to 50 cm depth gives therefore the best improvement. The ET characterization also improves slightly by soil moisture DA, with nearly 1% RMSE reduction for the year 2015. Estimating hydraulic conductivity together with model states, brought the hydraulic conductivity locally closer to the VR in a vicinity of 4 km around the measurement locations.

In this work, we also assimilated for the first time GWL data in a coupled land surface-subsurface model over a large domain like the Neckar catchment and surroundings, and with a complex topography including many small valleys and a fully heterogeneous 3D subsurface. We found an improved forecast of the GWL at the observation locations, the estimation error is reduced by 76%–88%, but the benefits reduce when moving away from the observations. The weakly coupled assimilation gives a better performance with an improvement at least for a range of 8 km from the observations, while for the fully coupled DA improvements are only found up to 4 km distance from the observation locations. For the weakly coupled cases, assimilating GWL data alone can also slightly improve the soil moisture characterization, reducing the RMSE by 1%–6%, which is not the case for the fully coupled case. Even with additional assimilation of surface soil moisture data, the fully coupled scenarios can only improve the estimate of soil moisture at the observation locations. Overall, the weakly coupled scenarios perform better for assimilation of GWL data, in terms of better characterization of subsurface states between the measurement locations. The results show the difficulties in improving the characterization of subsurface states of a realistic land surface-groundwater model, with a complex topography involving many small valleys and hill ridges. On the other hand, it could also be shown that GWL assimilation is able to improve GWL estimation over large parts of the domain between the measurement locations. The river discharge at the selected locations is also investigated for the DA experiment. It is found that without DA, the difference between the values for the open loop simulation are very close to the reference model VR, leaving small room for improvements and therefore the benefit of DA is very small with less than 0.1% reduction in the estimation errors compared with the open loop simulation. Results show the challenge to improve the characterization of states of complex land surface-subsurface models by DA, and point to the need to explore the assimilation of multiple data types in those models, and increase computation efficiency to be able to use larger ensemble sizes.

Appendix A: Variograms of Atmospheric Variables Derived From Fully Coupled Ensemble Simulations

As clarified in Section 3, we conducted ensemble simulations (16 members) with Terrestrial Systems Modeling Platform including all three component models. These simulations had perturbed lateral atmospheric boundary conditions according the COSMO setting. Variograms and cross-variograms were determined for four atmospheric variables (precipitation, shortwave radiation, longwave radiation and air temperature). These variograms are calculated on the basis of the data from the year 2015, for days with precipitation and during daytime (so that the correlation between precipitation and shortwave radiation can be evaluated). The yearly averages and standard deviations of the sill, spatial and temporal correlation ranges are given. We assumed isotropy in space for the different variables. In the following, the parameters of the variograms are expressed in matrices and the first to forth row is respectively for precipitation, shortwave radiation, longwave radiation and air temperature. We only give the upper triangular entries as the matrices are symmetric. The sills are given by:

$$\begin{pmatrix} 0.8 \pm 0.1 & -0.13 \pm 0.04 & -0.01 \pm 0.02 & -0.21 \pm 0.04 \\ * & 0.93 \pm 0.05 & -0.6 \pm 0.1 & 0.09 \pm 0.06 \\ * & * & 0.85 \pm 0.04 & 0.19 \pm 0.06 \\ * & * & * & 0.83 \pm 0.04 \end{pmatrix}, \tag{A1}$$

the ranges in space (km) are:

$$\begin{pmatrix} 70 \pm 20 & 110 \pm 50 & 10 \pm 20 & 120 \pm 30 \\ * & 120 \pm 30 & 120 \pm 30 & 100 \pm 50 \\ * & * & 100 \pm 30 & 80 \pm 30 \\ * & * & * & 120 \pm 30 \end{pmatrix}, \tag{A2}$$

and the temporal correlations (in hours) are:

$$\begin{pmatrix} 4.1 \pm 0.2 & 4.3 \pm 0.9 & 1 \pm 2 & 4.1 \pm 0.2 \\ * & 4.3 \pm 0.9 & 4.3 \pm 0.9 & 3 \pm 2 \\ * & * & 4.3 \pm 0.9 & 4.3 \pm 0.9 \\ * & * & * & 4.3 \pm 0.9 \end{pmatrix}. \tag{A3}$$

Data Availability Statement

Computing time has been provided by the Gauss Centre for Supercomputing (http://www.gauss-centre.eu/gauss-centre/EN/Home/home_node.html) and the Facilities are operated by the Juelich Supercomputing Centre (JSC) (http://www.fz-juelich.de/ias/jsc/EN/Home/home_node.html).

Acknowledgments
This research is funded by the Deutsche Forschungsgemeinschaft (DFG, FOR2131: "Data Assimilation for Improved Characterization of Fluxes across Compartmental Interfaces"). Majority of simulations are performed on the supercomputer JUWELS at JSC. Harrie-Jan Hendricks Franssen kindly acknowledges support from the Earth System Modeling project (ESM), funded by German Helmholtz Society (HGF).

References

Anderson, J. L. (2001). An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, *129*(12), 2884–2903. https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2

Anyah, R. O., Weaver, C. P., Miguez-Macho, G., Fan, Y., & Robock, A. (2008). Incorporating water table dynamics in climate modeling: 3. Simulated groundwater influence on coupled land-atmosphere variability. *Journal of Geophysical Research*, *113*(D7), D07103. https://doi.org/10.1029/2007JD009087

Ashby, S. F., & Falgout, R. D. (1996). A parallel multigrid preconditioned conjugate gradient algorithm for groundwater flow simulations. *Nuclear Science & Engineering*, *124*(1), 145–159. https://doi.org/10.13182/NSE96-A24230

Baatz, R., Bogena, H., Franssen], H.-J. H., Huisman, J., Qu, W., Montzka, C., & Vereecken, H. (2014). Calibration of a catchment scale cosmic-ray probe network: A comparison of three parameterization methods. *Journal of Hydrology*, *516*, 231–244. (Determination of soil moisture: Measurements and theoretical approaches). https://doi.org/10.1016/j.jhydrol.2014.02.026

Baroni, G., Zink, M., Kumar, R., Samaniego, L., & Attinger, S. (2017). Effects of uncertainty in soil properties on simulated hydrological states and fluxes at different spatio-temporal scales. *Hydrology and Earth System Sciences*, *21*(5), 2301–2320. https://doi.org/10.5194/hess-21-2301-2017

Bogena, H. R., Herbst, M., Huisman, J., Rosenbaum, U., Weuthen, A., & Vereecken, H. (2010). Potential of wireless sensor networks for measuring soil water content variability. *Vadose Zone Journal*, *9*(4), 1002–1013. https://doi.org/10.2136/vzj2009.0173

Bogena, H. R., Huisman, J. A., Baatz, R., Hendricks Franssen, H.-J., & Vereecken, H. (2013). Accuracy of the cosmic-ray soil water content probe in humid forest ecosystems: The worst case scenario. *Water Resources Research*, *49*(9), 5778–5791. https://doi.org/10.1002/wrcr.20463

Botto, A., Belluco, E., & Camporese, M. (2018). Multi-source data assimilation for physically based hydrological modeling of an experimental hillslope. *Hydrology and Earth System Sciences*, *22*(8), 4251–4266. https://doi.org/10.5194/hess-22-4251-2018

Burgers, G., Jan van Leeuwen, P., & Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, *126*(6), 1719. https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2

Butts, M., Drews, M., Larsen, M. A., Lerer, S., Rasmussen, S. H., Grooss, J., et al. (2014). Embedding complex hydrology in the regional climate system – Dynamic coupling across different modelling domains. *Advances in Water Resources*, *74*, 166–184. https://doi.org/10.1016/j.advwatres.2014.09.004

Chaudhuri, A., Franssen, H.-J. H., & Sekhar, M. (2018). Iterative filter based estimation of fully 3D heterogeneous fields of permeability and Mualem-van Genuchten parameters. *Advances in Water Resources*, *122*, 340–354. https://doi.org/10.1016/j.advwatres.2018.10.023

Chen, X., & Hu, Q. (2004). Groundwater influences on soil moisture and surface evaporation. *Journal of Hydrology*, *297*(1), 285–300. https://doi.org/10.1016/j.jhydrol.2004.04.019

Chen, Y., & Zhang, D. (2006). Data assimilation for transient flow in geologic formations via ensemble Kalman filter. *Advances in Water Resources*, *29*(8), 1107–1122. https://doi.org/10.1016/j.advwatres.2005.09.007

Cosby, B. J., Hornberger, G. M., Clapp, R. B., & Ginn, T. R. (1984). A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. *Water Resources Research*, *20*(6), 682–690. https://doi.org/10.1029/WR020i006p00682

Craig, A., Valcke, S., & Coquart, L. (2017). Development and performance of a new version of the OASIS coupler, OASIS3-MCT_3.0. *Geoscientific Model Development*, *10*(9), 3297–3308. https://doi.org/10.5194/gmd-10-3297-2017

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

De Lannoy, G. J. M., Houser, P. R., Pauwels, V. R. N., & Verhoest, N. E. C. (2006). Assessment of model uncertainty for soil moisture through ensemble verification. *Journal of Geophysical Research*, *111*(D10), D10101. https://doi.org/10.1029/2005JD006367

Deutsch, C. V., & Journel, A. G. (1992). *Geostatistical software library and user's guide* (Vol. 119).

Dickinson, E., Henderson-Sellers, A., & Kennedy, J. (1993). *Biosphere-atmosphere transfer scheme (bats) version 1E as coupled to the NCAR community climate model*. https://doi.org/10.5065/D67W6959

Dickinson, R. E. (1986). *Biosphere/atmosphere transfer scheme (bats) for the NCAR community climate model. Technical report*. NCAR. https://doi.org/10.5065/D6668B58

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., et al. (2010). The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE*, *98*(5), 704–716. https://doi.org/10.1109/jproc.2010.2043918

Evensen, G. (1992). Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *Journal of Geophysical Research*, *97*(C11), 17905–17924. https://doi.org/10.1029/92JC01972

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, *99*(C5), 10143–10162. https://doi.org/10.1029/94JC00572

Franz, T. E., Wang, T., Avery, W., Finkenbiner, C., & Brocca, L. (2015). Combined analysis of soil moisture measurements from roving and fixed cosmic ray neutron probes for multiscale real-time monitoring. *Geophysical Research Letters*, *42*(9), 3389–3396. https://doi.org/10.1002/2015GL063963

Gasper, F., Goergen, K., Shrestha, P., Sulis, M., Rihani, J., Geimer, M., & Kollet, S. (2014). Implementation and scaling of the fully coupled Terrestrial Systems Modeling Platform (TerrSysMP v1.0) in a massively parallel supercomputing environment – A case study on JUQUEEN (IBM Blue Gene/Q). *Geoscientific Model Development*, *7*(5), 2531–2543. https://doi.org/10.5194/gmd-7-2531-2014

Gebler, S., Kurtz, W., Pauwels, V. R. N., Kollet, S., Vereecken, H., & Hendricks Franssen, H.-J. (2019). Assimilation of high-resolution soil moisture data into an integrated terrestrial model for a small-scale head-water catchment. *Water Resources Research*, *55*(12), 10358–10385. https://doi.org/10.1029/2018WR024658

Gerten, D., Schaphoff, S., Haberlandt, U., Lucht, W., & Sitch, S. (2004). Terrestrial vegetation and water balance—Hydrological evaluation of a dynamic global vegetation model. *Journal of Hydrology*, *286*(1), 249–270. https://doi.org/10.1016/j.jhydrol.2003.09.029

Gómez-Hernández, J. J., & Journel, A. G. (1993). Joint sequential simulation of multigaussian fields. In *Geostatistics troia'92* (pp. 85–94). Springer.

Gregory, D., Morcrette, J.-J., Jakob, C., Beljaars, A. C. M., & Stockdale, T. (2000). Revision of convection, radiation and cloud schemes in the ECMWF integrated forecasting system. *Quarterly Journal of the Royal Meteorological Society*, *126*(566), 1685–1710. https://doi.org/10.1002/qj.49712656607

Hamill, T. M., Whitaker, J. S., & Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, *129*(11), 2776–2790. https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2

Harlim, J., & Hunt, B. R. (2007). Four-dimensional local ensemble transform Kalman filter: Numerical experiments with a global circulation model. *Tellus A: Dynamic Meteorology and Oceanography*, *59*(5), 731–748. https://doi.org/10.1111/j.1600-0870.2007.00255.x

He, X., Lucatero, D., Ridler, M.-E., Madsen, H., Kidmose, J., Øyvind, H., et al. (2019). Real-time simulation of surface water and groundwater with data assimilation. *Advances in Water Resources*, *127*, 13–25. https://doi.org/10.1016/j.advwatres.2019.03.004

Hendricks-Franssen, H. J., & Kinzelbach, W. (2008). Real-time groundwater flow modeling with the ensemble Kalman filter: Joint estimation of states and parameters and the filter inbreeding problem. *Water Resources Research*, *44*(9), W09408. https://doi.org/10.1029/2007WR006505

Houtekamer, P. L., & Mitchell, H. L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, *129*(1), 123. https://doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2

Hunt, B., Kostelich, E., & Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, *230*(1–2), 112–126. https://doi.org/10.1016/j.physd.2006.11.008

Isaaks, E. H., & Srivastava, R. M. (1994). An introduction to applied geostatistics. *Geographical Analysis*, *26*(3), 282–283. https://doi.org/10.1111/j.1538-4632.1994.tb00325.x

Jiang, X., Niu, G.-Y., & Yang, Z.-L. (2009). Impacts of vegetation and groundwater dynamics on warm season precipitation over the central United States. *Journal of Geophysical Research*, *114*(D6), D06109. https://doi.org/10.1029/2008JD010756

Jones, J. E., & Woodward, C. S. (2001). Newton–Krylov-multigrid solvers for large-scale, highly heterogeneous, variably saturated flow problems. *Advances in Water Resources*, *24*(7), 763–774. https://doi.org/10.1016/S0309-1708(00)00075-0

Keller, J., Hendricks Franssen, H.-J., & Marquart, G. (2018). Comparing seven variants of the ensemble Kalman filter: How many synthetic experiments are needed? *Water Resources Research*, *54*(9), 6299–6318. https://doi.org/10.1029/2018WR023374

Kollet, S. J., & Maxwell, R. M. (2006). Integrated surface–groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model. *Advances in Water Resources*, *29*(7), 945–958. https://doi.org/10.1016/j.advwatres.2005.08.006

Kollet, S. J., & Maxwell, R. M. (2008). Capturing the influence of groundwater dynamics on land surface processes using an integrated, distributed watershed model. *Water Resources Research*, *44*(2), W02402. https://doi.org/10.1029/2007WR006004

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., et al. (2004). Regions of strong coupling between soil moisture and precipitation. *Science*, *305*(5687), 1138–1140. https://doi.org/10.1126/science.1100217

Kurtz, W., He, G., Kollet, S. J., Maxwell, R. M., Vereecken, H., & Hendricks Franssen, H.-J. (2016). TerrSysMP-PDAF (version 1.0): A modular high-performance data assimilation framework for an integrated land surface-subsurface model. *Geoscientific Model Development*, *9*(4), 1341–1360. https://doi.org/10.5194/gmd-9-1341-2016

Larsen, M. A. D., Refsgaard, J. C., Drews, M., Butts, M. B., Jensen, K. H., Christensen, J. H., & Christensen, O. B. (2014). Results from a full coupling of the HIRHAM regional climate model and the MIKE SHE hydrological model for a Danish catchment. *Hydrology and Earth System Sciences*, *18*(11), 4733–4749. https://doi.org/10.5194/hess-18-4733-2014

Majewski, D., Liermann, D., Prohl, P., Ritter, B., Buchhold, M., Hanisch, T., et al. (2002). The operational global icosahedral–hexagonal gridpoint model GME: Description and high-resolution tests. *Monthly Weather Review*, *130*(2), 319–338. https://doi.org/10.1175/1520-0493(2002)130<0319:TOGIHG>2.0.CO;2

Manabe, S. (1969). Climate and the ocean circulation. *Monthly Weather Review*, *97*(11), 739–774. https://doi.org/10.1175/1520-0493(1969)097<0739:CATOC>2.3.CO;2

Maxwell, R. M. (2013). A terrain-following grid transform and preconditioner for parallel, large-scale, integrated hydrologic modeling. *Advances in Water Resources*, *53*, 109–117. https://doi.org/10.1016/j.advwatres.2012.10.001

Maxwell, R. M., Chow, F. K., & Kollet, S. J. (2007). The groundwater–land-surface–atmosphere connection: Soil moisture effects on the atmospheric boundary layer in fully-coupled simulations. *Advances in Water Resources*, *30*(12), 2447–2466. https://doi.org/10.1016/j.advwatres.2007.05.018

Maxwell, R. M., Lundquist, J. K., Mirocha, J. D., Smith, S. G., Woodward, C. S., & Tompson, A. F. B. (2011). Development of a coupled groundwater–atmosphere model. *Monthly Weather Review*, *139*(1), 96–116. https://doi.org/10.1175/2010MWR3392.1

McCreight, J., FitzGerald, K., Cabell, R., Fersch, B., donaldwj, Dugger, A., et al. (2021). *NCAR/wrf_hydro_nwm_public: WRF-Hydro® v5.2.0*. Zenodo. https://doi.org/10.5281/zenodo.4479912

McLaughlin, D. (2002). An integrated approach to hydrologic data assimilation: Interpolation, smoothing, and filtering. *Advances in Water Resources*, *25*(8), 1275–1286. https://doi.org/10.1016/S0309-1708(02)00055-6

Mizuta, R., Oouchi, K., Yoshimura, H., Noda, A., Katayama, K., Yukimoto, S., et al. (2006). 20-km-mesh global climate simulations using JMA-GSM model –mean climate states. *Journal of the Meteorological Society of Japan. Ser. II*, *84*(1), 165–185. https://doi.org/10.2151/jmsj.84.165

Mölders, N., & Rühaak, W. (2002). On the impact of explicitly predicted runoff on the simulated atmospheric response to small-scale land-use changes—An integrated modeling approach. *Atmospheric Research*, *63*(1), 3–38. https://doi.org/10.1016/S0169-8095(02)00002-9

Moradkhani, H., Nearing, G. S., Abbaszadeh, P., & Pathiraja, S. (2019). Fundamentals of data assimilation and theoretical advances. In Q. Duan, F. Pappenberger, A. Wood, H. L. Cloke, & J. C. Schaake (Eds.), *Handbook of hydrometeorological ensemble forecasting* (pp. 675–699). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-39925-1_30

Morita, M., & Yen, B. C. (2002). Modeling of conjunctive two-dimensional surface-three-dimensional subsurface flows. *Journal of Hydraulic Engineering*, *128*(2), 184–200. https://doi.org/10.1061/(ASCE)0733-9429(2002)128:2(184)

Myneni, R., Hoffman, S., Knyazikhin, Y., Privette, J., Glassy, J., Tian, Y., et al. (2002). Global products of vegetation leaf area and fraction absorbed par from year one of modis data. *Remote Sensing of Environment*, *83*(1), 214–231. (The Moderate Resolution Imaging Spectroradiometer (MODIS): A new generation of Land Surface Monitoring). https://doi.org/10.1016/S0034-4257(02)00074-3

Nerger, L., & Gregg, W. W. (2007). Assimilation of SeaWiFS data into a global ocean-biogeochemical model using a local SEIK filter. *Journal of Marine Systems*, *68*(1), 237–254. https://doi.org/10.1016/j.jmarsys.2006.11.009

Nerger, L., & Hiller, W. (2013). Software for ensemble-based data assimilation systems—Implementation strategies and scalability. *Computers & Geosciences*, *55*, 110–118. (Ensemble Kalman filter for data assimilation). https://doi.org/10.1016/j.cageo.2012.03.026

Oleson, K., Dai, Y., Bonan, B., Bosilovichm, M., Dickinson, R., Dirmeyer, P., et al. (2004). Technical description of the community land model (clm).

Oleson, K., Niu, G.-Y., Yang, Z.-L., Lawrence, D., Thornton, P., Lawrence, P., et al. (2008). Improvements to the Community Land Model and their impact on the hydrological cycle. *Journal of Geophysical Research*, *113*(G1), G01021. https://doi.org/10.1029/2007jg000563

Panday, S., & Huyakorn, P. S. (2004). A fully coupled physically-based spatially-distributed model for evaluating surface/subsurface flow. *Advances in Water Resources*, *27*(4), 361–382. (A Tribute to George F. Pinder). https://doi.org/10.1016/j.advwatres.2004.02.016

Qu, Y., & Duffy, C. J. (2007). A semidiscrete finite volume formulation for multiprocess watershed simulation. *Water Resources Research*, *43*(8), W08419. https://doi.org/10.1029/2006WR005752

Rasmussen, J., Madsen, H., Jensen, K. H., & Refsgaard, J. C. (2015). Data assimilation in integrated hydrological modeling using ensemble Kalman filtering: Evaluating the effect of ensemble size and localization on filter performance. *Hydrology and Earth System Sciences*, *19*(7), 2999–3013. https://doi.org/10.5194/hess-19-2999-2015

Rasmussen, J., Madsen, H., Jensen, K. H., & Refsgaard, J. C. (2016). Data assimilation in integrated hydrological modelling in the presence of observation bias. *Hydrology and Earth System Sciences*, *20*(5), 2103–2118. https://doi.org/10.5194/hess-20-2103-2016

Reichle, R. H., Koster, R. D., Liu, P., Mahanama, S. P. P., Njoku, E. G., & Owe, M. (2007). Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the scanning multichannel microwave radiometer (SMMR). *Journal of Geophysical Research*, *112*(D9), D09108. https://doi.org/10.1029/2006JD008033

Ridler, M.-E., Zhang, D., Madsen, H., Kidmose, J., Refsgaard, J. C., & Jensen, K. H. (2017). Bias-aware data assimilation in integrated hydrological modelling. *Hydrology Research*, *49*(4), 989–1004. https://doi.org/10.2166/nh.2017.117

Rosenbaum, U., Huisman, J., Weuthen, A., Vereecken, H., & Bogena, H. (2010). Sensor-to-sensor variability of the ECH_2O EC-5, TE, and 5TE sensors in dielectric liquids. *Vadose Zone Journal*, *9*(1), 181–186. https://doi.org/10.2136/vzj2009.0036

Schalge, B., Baroni, G., Haese, B., Erdal, D., Geppert, G., Saavedra, P., et al. (2021). Presentation and discussion of the high-resolution atmosphere–land-surface–subsurface simulation dataset of the simulated Neckar catchment for the period 2007–2015. *Earth System Science Data*, *13*(9), 4437–4464. https://doi.org/10.5194/essd-13-4437-2021

Schrön, M., Köhli, M., Scheiffele, L., Iwema, J., Bogena, H. R., Lv, L., et al. (2017). Improving calibration and validation of cosmic-ray neutron sensors in the light of spatial sensitivity. *Hydrology and Earth System Sciences*, *21*(10), 5009–5030. https://doi.org/10.5194/hess-21-5009-2017

Sellers, P. J., Mintz, Y., Sud, Y. C., & Dalcher, A. (1986). A simple biosphere model (SiB) for use within general circulation models. *Journal of the Atmospheric Sciences*, *43*(6), 505–531. https://doi.org/10.1175/1520-0469(1986)043<0505:ASBMFU>2.0.CO;2

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., et al. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, *99*(3), 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004

Seneviratne, S. I., Lüthi, D., Litschi, M., & Schär, C. (2006). Land–atmosphere coupling and climate change in Europe. *Nature*, *443*(7108), 205–209. https://doi.org/10.1038/nature05095

Seuffert, G., Gross, P., Simmer, C., & Wood, E. F. (2002). The influence of hydrologic modeling on the predicted local weather: Two-way coupling of a mesoscale weather prediction model and a land surface hydrologic model. *Journal of Hydrometeorology*, *3*(5), 505–523. https://doi.org/10.1175/1525-7541(2002)003<0505:TIOHMO>2.0.CO;2

Shi, Y., Davis, K. J., Zhang, F., Duffy, C. J., & Yu, X. (2014). Parameter estimation of a physically based land surface hydrologic model using the ensemble Kalman filter: A synthetic experiment. *Water Resources Research*, *50*(1), 706–724. https://doi.org/10.1002/2013WR014070

Shi, Y., Davis, K. J., Zhang, F., Duffy, C. J., & Yu, X. (2015). Parameter estimation of a physically-based land surface hydrologic model using an ensemble Kalman filter: A multivariate real-data experiment. *Advances in Water Resources*, *83*, 421–427. https://doi.org/10.1016/j.advwatres.2015.06.009

Shrestha, P., Sulis, M., Masbou, M., Kollet, S., & Simmer, C. (2014). A scale-consistent terrestrial systems modeling platform based on COSMO, CLM, and ParFlow. *Monthly Weather Review*, *142*(9), 3466–3483. https://doi.org/10.1175/mwr-d-14-00029.1

Tian, W., Li, X., Cheng, G.-D., Wang, X.-S., & Hu, B. X. (2012). Coupling a groundwater model with a land surface model to improve water and energy cycle simulation. *Hydrology and Earth System Sciences*, *16*(12), 4707–4723. https://doi.org/10.5194/hess-16-4707-2012

Tóth, B., Weynants, M., Nemes, A., Makó, A., Bilas, G., & Tóth, G. (2015). New generation of hydraulic pedotransfer functions for Europe. *European Journal of Soil Science*, *66*(1), 226–238. https://doi.org/10.1111/ejss.12192

Valcke, S. (2013). The OASIS3 coupler: A European climate modelling community software. *Geoscientific Model Development*, *6*(2), 373–388. https://doi.org/10.5194/gmd-6-373-2013

VanderKwaak, J. E., & Loague, K. (2001). Hydrologic-response simulations for the R-5 catchment with a comprehensive physics-based model. *Water Resources Research*, *37*(4), 999–1013. https://doi.org/10.1029/2000WR900272

vanGenuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal*, *44*(5), 892–898. https://doi.org/10.2136/sssaj1980.03615995004400050002x

White, G., Yang, F., & Tallapragada, V. (2018). *The development and success ofncep's global forecast system*. National Centers for Environmental Prediction. Retrieved from https://ams.confex.com/ams/2019Annual/webprogram/Manuscript/Paper350196/technotegfsh.pdf

York, J. P., Person, M., Gutowski, W. J., & Winter, T. C. (2002). Putting aquifers into atmospheric simulation models: An example from the mill creek watershed, northeastern Kansas. *Advances in Water Resources*, *25*(2), 221–238. https://doi.org/10.1016/S0309-1708(01)00021-5

Zhang, D., Madsen, H., Ridler, M. E., Kidmose, J., Jensen, K. H., & Refsgaard, J. C. (2016). Multivariate hydrological data assimilation of soil moisture and groundwater head. *Hydrology and Earth System Sciences*, *20*(10), 4341–4357. https://doi.org/10.5194/hess-20-4341-2016

Zhang, D., Madsen, H., Ridler, M. E., Refsgaard, J. C., & Jensen, K. H. (2015). Impact of uncertainty description on assimilating hydraulic head in the MIKE SHE distributed hydrological model. *Advances in Water Resources*, *86*, 400–413. (Data assimilation for improved predictions of integrated terrestrial systems). https://doi.org/10.1016/j.advwatres.2015.07.018

Zhang, H., Kurtz, W., Kollet, S., Vereecken, H., & Franssen, H.-J. H. (2018). Comparison of different assimilation methodologies of groundwater levels to improve predictions of root zone soil moisture with an integrated terrestrial system model. *Advances in Water Resources*, *111*, 224–238. https://doi.org/10.1016/j.advwatres.2017.11.003

Zhou, H., Gómez-Hernández, J. J., Franssen], H.-J. H., & Li, L. (2011). An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering. *Advances in Water Resources*, *34*(7), 844–864. https://doi.org/10.1016/j.advwatres.2011.04.014

Zreda, M., Desilets, D., Ferré, T. P. A., & Scott, R. L. (2008). Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons. *Geophysical Research Letters*, *35*(21), L21402. https://doi.org/10.1029/2008GL035655