Water Resources Research

RESEARCH ARTICLE
10.1029/2025WR041570

Data Assimilation in Integrated Subsurface Flow Models—Making Optimal Use of Cross-Compartmental Interactions

Bastian Waldowski[1], **Harrie-Jan Hendricks-Franssen**[2,3], **and Insa Neuweiler**[1]

[1]Institute of Fluid Mechanics and Environmental Physics in Civil Engineering, Leibniz University Hannover, Hannover, Germany, [2]Agrosphere (IBG-3), Forschungszentrum Jülich GmbH, Jülich, Germany, [3]Centre for High-Performance Scientific Computing in Terrestrial Systems, Jülich, Germany

Key Points:

- We investigate optimal use of soil moisture and groundwater observations to characterize subsurface flow variables during data assimilation (DA)
- Including an area above the water table in the DA avoids balancing flows and yields a better water table characterization
- For characterizing root zone soil moisture, knowledge of soil layering matters, and multivariate assimilation is beneficial

Correspondence to:
B. Waldowski,
bastian.waldowski@web.de

Citation:
Waldowski, B., Hendricks-Franssen, H.-J., & Neuweiler, I. (2025). Data assimilation in integrated subsurface flow models—Making optimal use of cross-compartmental interactions. *Water Resources Research*, *61*, e2025WR041570. https://doi.org/10.1029/2025WR041570

Received 8 JUL 2025
Accepted 18 NOV 2025

Abstract The ensemble Kalman filter is often used for data assimilation (DA) to forecast states and fluxes in integrated subsurface flow problems. Integrated subsurface flow models include two compartments: the vadose zone and groundwater aquifers. In DA with such models, observations from both compartments can be used with the goal of improving forecasts in the whole system. To study the potential risks and benefits of cross-compartmental DA in this context, we apply different DA methods, drawing observations from multiple heterogeneous virtual realities. Updating soil moisture with point observations can consistently improve spatially averaged soil moisture predictions at the land surface but frequently deteriorates estimates of the groundwater table height (worse for models with non-resolved layers). Bias correction and vertical localization can mitigate these deteriorations. Groundwater table assimilation that is limited to updating aquifer states can improve the estimation of spatially averaged groundwater tables (RMSEs on average reduced by 29% in our test examples), but rigorously taking out the partly saturated parts introduces balancing fluxes that impair the benefits of the groundwater assimilation. Additionally updating pressure heads in an area above the groundwater table can, on average, reduce RMSEs of groundwater table estimates by more than double (59%). Multivariate assimilation of both soil moisture and groundwater tables leads to the best results for predicting root zone soil moisture. Groundwater recharge predictions could often be improved in a subsequent forecast without DA if groundwater tables had been updated before.

1. Introduction

Reliable estimates of soil moisture are essential for many applications that include evaluating water availability for plants (e.g., Enenkel et al., 2016; Sohrabi et al., 2015), weather forecasting (e.g., De Rosnay et al., 2013; van den Hurk et al., 2012), flood prediction (e.g., Komma et al., 2008; Wanders et al., 2014), and landslide prediction (e.g., Brocca et al., 2012; Wicki et al., 2020). The necessity to estimate groundwater levels is obvious since groundwater is freshwater that is directly accessible to humans, clearly motivating its management and preservation. It follows that both subsurface compartments (vadose zone and groundwater) are monitored. Measurement availability, however, is limited, and numerical subsurface flow models are often used to interpret measurements and thus get additional information. There are different approaches for numerically simulating subsurface flow in the two compartments. One is to consider the vadose zone and groundwater as mostly *separated* and neglect two-way feedbacks. An example of this would be using a 1D vertical flow model for the vadose zone (e.g., HYDRUS-1D by Simunek et al. (2005)) with a constant head boundary condition at the bottom that allows for exchange flows with the groundwater, but neglects groundwater feedbacks like groundwater table movements. Another philosophy is to consider the two compartments as *integrated*, which means simulating them simultaneously and capturing two-way feedbacks.

In general, the inputs of subsurface flow models (subsurface flow parameters and boundary conditions) are unknown and uncertain, so the models rely on measurements of their states (outputs) to condition the model in an inverse manner. Sequential data assimilation (DA) is one way to use measurement information to reduce the uncertainty of the states and parameters of such models, with the goal of improving predictions. The DA method that is used in this work is the widely used ensemble Kalman filter (EnKF), which was introduced by Evensen (1994). It is a sequential method that consists of a *forecast* step (i.e., propagating an ensemble of model realizations forward in time) and an *update* step (i.e., conditioning the ensemble based on measurements). The update aims to find the most likely model states and parameters, given the measurements, reducing the state and parameter uncertainty of the model. In an integrated model, it theoretically makes sense to utilize all available

measurement information (measurements in the vadose zone, here soil moisture, and groundwater table measurements) to update states and parameters within the whole domain. This can be referred to as *multivariate* DA (its opposite is *univariate* DA). In practice, however, multivariate DA was not always found to lead to improved predictions, as will be elaborated on in the following.

Shi et al. (2015) used the EnKF for parameter estimation in a 1D variably-saturated domain with synthetic observations. They tested different kinds of observations (point scale pressure head and soil moisture, average soil moisture at the land surface, and groundwater level) and drew the conclusion that assimilating more types of data is helpful to improve the estimates of soil hydraulic parameters. However, it was also found that some observations improve soil parameters only in a limited area of influence (e.g., surface soil moisture data had almost no influence on the parameter estimates in the deeper layers).

D. Zhang et al. (2016) used the Ensemble Transform Kalman Filter and the coupled subsurface flow model MIKE SHE to jointly assimilate soil moisture and groundwater head with both a synthetic catchment and real observations from a catchment in Denmark. They used an ensemble of 60 members and corrected the bias between observations and simulations by subtracting the average offset from the observations. They found that DA improves the characterization of the variable that is assimilated but not other variables. No clear incentive for multivariate DA could be derived from their study.

Botto et al. (2018) used the EnKF on the fully integrated surface/subsurface flow model CATHY to assimilate pressure head, soil moisture, and integrated subsurface outflow in a two-layered hillslope with real measurements. They found that multivariate DA can lead to the degradation of model predictions for variables that are otherwise well-represented. However, they partly attributed this to their pressure head measurements being of relatively poor quality with disturbances caused by temperature fluctuations.

Hung et al. (2022) used the fully integrated surface/subsurface flow model TSMP-PDAF (ParFlow-CLM) in a 3D heterogeneous virtual catchment that is inspired by a real catchment in southwestern Germany. They found that multivariate assimilation of soil moisture and groundwater tables resulted in a worse soil moisture estimation than univariate soil moisture assimilation. They hypothesized that this might be related to the relatively small ensemble size of 64 members.

The most promising results with respect to multivariate DA were obtained by H. Zhang et al. (2018). They tested different updating schemes with TSMP-PDAF (with ParFlow-CLM) in an idealized setting with four (2×2) horizontal grid cells, vertically heterogeneous soil hydraulic parameters, and impermeable boundaries at the bottom and lateral sides of the model domain. They conducted DA experiments with virtual realities (VRs) that covered different climate, soil, and vegetation types. For each VR, they perturbed the saturated hydraulic conductivity to generate an ensemble of 128 members. They found that univariate assimilation of soil moisture near the land surface improves the soil moisture estimates in shallow depths but can worsen them in deeper layers. Univariate assimilation of groundwater tables was shown to improve soil moisture predictions in deeper layers but sometimes deteriorated them in upper layers. Multivariate assimilation was able to combine the best of the two and showed great potential with respect to improving predictions of root zone soil moisture.

In addition to the question *which* measurements should be used, the question of *how* these measurements should be used also arises. Measurements from one compartment can be used to update states and parameters in another one (referred to as *strongly coupled* in this work), or the update can be restricted to the respective compartment (referred to as *weakly coupled* in this work). Some of the previously cited studies also investigated weak and strong coupling.

D. Zhang et al. (2016) found that strongly coupled multivariate updates of the groundwater table and soil moisture without localization notably deteriorated predictions of groundwater tables, soil moisture at a deeper depth, and river discharge at the outlet. When they used distance localization (restricted updates to a certain area) and weak coupling (they referred to it as variable localization), they could fix the deteriorations. This means that weakly coupled experiments worked better than strongly coupled ones in their case.

H. Zhang et al. (2018) tested different strongly coupled update variants and compared them. They tested one variant where they assimilated groundwater table measurements to update pressure heads in both the saturated and the unsaturated zone and found this to not work well since pressure heads in the vadose zone were very negative at times, which skewed the ensemble distributions and led to unreasonable updates. In another strongly

coupled variant, assimilating groundwater table measurements to update pressure heads in the saturated zone and soil moisture in the vadose zone turned out to work much better, notably improving predictions of root zone soil moisture. Weakly coupled groundwater updates also improved the characterization of groundwater tables and root zone soil moisture, but to a lesser extent than the strongly coupled updating of soil moisture, and they converged slower towards the virtual truth.

The well-performing strongly coupled update variant introduced by H. Zhang et al. (2018) was then tested by Hung et al. (2022) for a more complex test case. Hung et al. (2022) found, however, weakly coupled updates of the groundwater table depth to work better than strongly coupled updates. In contrast to the strongly coupled updates, weakly coupled updates could even improve soil moisture predictions.

To sum up, in contrast to the expectation that fully integrated DA in the subsurface naturally brings advantages, if and how observations in the different compartments are beneficial for predictions in the coupled system is not so clear. Several authors observed that predictions of groundwater tables might deteriorate by assimilating soil moisture, compared to a simulation without DA (Botto et al., 2018; Hung et al., 2022). The impacts of updating groundwater tables on predictions of soil moisture are more diverse, with improvements, but also deteriorations being observed. Multivariate DA was sometimes found to be outperformed by univariate DA if the variable of interest was assimilated (Botto et al., 2018; Hung et al., 2022; D. Zhang et al., 2016). All in all, there are some apparently contradicting outcomes, and it is not clear if groundwater observations should be used if one is interested in soil moisture estimates and vice versa. It is also unclear *how* the observations are best utilized (e.g., weakly or strongly coupled) and when multivariate DA is beneficial.

It is obvious that the potential for an improved characterization by integrated DA does depend on catchment characteristics. A meaningful correlation between land surface processes and the groundwater table depth can only be expected up until a certain depth (Kollet & Maxwell, 2008). If the groundwater and the vadose zone are not meaningfully correlated, measurements in one compartment cannot improve predictions in the other one. However, direct deterioration of the modeled states in the other compartment by the updates of DA should, in theory, not occur and is usually linked to spurious correlations due to limited ensemble size (e.g., D. Zhang et al., 2016), or biases between the measurements and the numerical model (e.g., Botto et al., 2018). Ensemble size is always limited by computational power, and numerical models are always expected to give biased estimates of the true states due to reasons such as simplifying assumptions, lack of information on heterogeneous structures, or insufficient resolution, to name a few. These issues cannot be avoided, so it is important to test the impact of the deteriorations and work out strategies to deal with them. This is one topic of this work. These problems are expected to get worse as domain complexity increases. In studies for complex domains (Hung et al., 2022; D. Zhang et al., 2016) utilizing measurement information from the other compartment (strongly coupled DA) often appeared to do more harm than good. From an integrated modeling perspective, this is somewhat dissatisfying. There is an incentive to further investigate DA strategies that allow us to take the interrelation between the compartments into account. This is the second topic of this work.

In this study, we (a) investigate the occurrence of prediction deteriorations and test strategies to deal with these prediction deteriorations when doing cross-compartmental DA (referred to as *cross-compartmental deteriorations* hereinafter) and (b) test DA strategies for improving predictions by exploiting cross-compartmental interactions. This is done by using a numerical subsurface flow model on the plot scale. We investigate the role the model error plays in cross-compartmental deteriorations by approximating a heterogeneous reality with one homogeneously layered and one fully homogeneous ensemble. A bias-aware EnKF is used to directly tackle the errors due to the homogeneity of the soil hydraulic parameter field. This has been done successfully in the past by Erdal et al. (2014) for a 1D model of the vadose zone. Brandhorst and Neuweiler (2023) have investigated DA with a homogeneous ensemble for a 3D model, but also had a clear focus on estimates within the vadose zone and did not address the model structural errors. Using the bias-aware EnKF in the context of integrated modeling and cross-compartmental feedbacks poses additional challenges and has not been investigated before. To make optimal use of cross-compartmental interactions, different DA variants are tested. As an additional innovation presented in this work, we analyze the effects of subsurface flow DA on groundwater recharge predictions. This flux is crucial for qualitative and quantitative groundwater assessment. Since groundwater recharge occurs directly at the interface between the vadose zone and groundwater, it quantifies the interactions between the subsurface compartments.

In the upcoming section (Section 2), governing equations and methods for subsurface flow DA are presented. Section 3 shows the test case and the DA settings and variants. This is followed by the results and discussion section (Section 4), which leads to Section 5: Conclusions.

2. Governing Equations and Methods

We use the openly available iteratively coupled subsurface flow model by Brandhorst et al. (2021). It decreases the dimensionality of a 3D flow problem by coupling 1D variably saturated flow models with a 2D groundwater flow model. Two-way feedbacks between the subsurface compartments are taken into account through iteration. Brandhorst et al. (2021) showed that this coupled model could closely approximate estimates by the 3D model ParFlow for cases where no steep topography and groundwater tables are involved, using only a fraction of the compute time (required core hours were reduced by a factor of more than 12 in the cases of Brandhorst et al. (2021)). This computational efficiency allows elaborate testing with many ensemble runs. Section 2.1 presents the equations and methods for coupled subsurface flow and Section 2.2 for DA with the EnKF.

2.1. Coupled Subsurface Flow

We simulate horizontal groundwater flow in an unconfined aquifer by solving the depth-averaged groundwater flow equations based on the Dupuit assumption:

$$S_y \frac{\partial(h_{GW})}{\partial t} = \nabla \cdot \left(\overline{K_s}(h_{GW} - z_0)\nabla(h_{GW})\right) + R \tag{1}$$

with the specific yield $S_y\,[-]$, the groundwater table height $h_{GW}\,[L]$, the time $t\,[T]$, the depth averaged saturated hydraulic conductivity $\overline{K_s}\left[L\,T^{-1}\right]$, the bottom elevation of the aquifer $z_0\,[L]$, and groundwater recharge $R\left[L\,T^{-1}\right]$. The groundwater table height h_{GW} is deemed equal to the piezometric head $\psi_{GW} + z$, with the groundwater pressure head $\psi_{GW}\,[L]$ and the geodetic head $z\,[L]$.

In the vadose zone, we assume the flow to be exclusively vertical and solve the mixed form of the 1D Richards' equation (Richards, 1931):

$$S_s \frac{\theta(\psi)}{\phi}\frac{\partial \psi}{\partial t} + \frac{\partial \theta(\psi)}{\partial t} = \frac{\partial}{\partial z}\left[K(\psi)\left(\frac{\partial \psi}{\partial z} + 1\right)\right] + Q_s \tag{2}$$

with the specific storage $S_s\,\left[L^{-1}\right]$, water content $\theta\,[-]$, pressure head $\psi\,[L]$, porosity $\phi\,[-]$ (equal to maximum water content in this work), hydraulic conductivity $K\,\left[L\,T^{-1}\right]$ and the source/sink term $Q_s\,\left[T^{-1}\right]$. Please note that the terms water content and soil moisture are treated as synonyms within this work. Equations 1 and 2 are subject to initial and boundary conditions, which will be addressed in Section 3. We describe the dependence of the water content θ and hydraulic conductivity K on the pressure head ψ by the van Genuchten-Mualem model (van Genuchten, 1980):

$$\theta(\psi) = \begin{cases} \theta_r + \dfrac{(\phi - \theta_r)}{\left[1 + (\alpha\,|\psi|)^n\right]^m} & \text{if } \psi < 0 \\ \phi & \text{otherwise} \end{cases} \tag{3}$$

$$K(\psi) = \begin{cases} K_s \dfrac{\left[1 - \frac{(\alpha\,|\psi|)^{(n-1)}}{(1 + (\alpha\,|\psi|)^n)^m)}\right]^2}{\left[1 + (\alpha\,|\psi|)^n\right]^{(m/2)}} & \text{if } \psi < 0 \\ K_s & \text{otherwise} \end{cases} \tag{4}$$

with the residual water content $\theta_r\,[-]$, the saturated hydraulic conductivity $K_s\,\left[L\,T^{-1}\right]$ and the van Genuchten parameters $\alpha\,\left[L^{-1}\right]$, $n\,[-]$, and $m = 1 - (1/n)$.

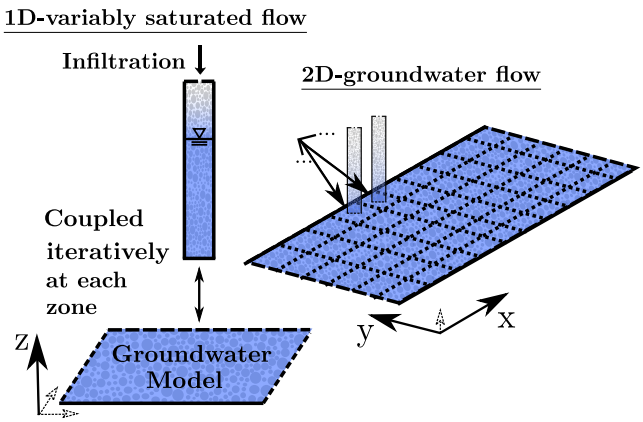

Figure 1. Sketch of the coupled subsurface flow model.

The model by Brandhorst et al. (2021) couples multiple 1D variably saturated flow models with a 2D groundwater flow model (see Figure 1) and solves Equations 1 and 2 based on the Finite Volume Method with an implicit Euler scheme for time integration. In the following, the coupling of the model is briefly described. For more details, please see Brandhorst et al. (2021).

Flow within the vadose zone is approximated with multiple vertical 1D columns that reach the bottom of the aquifer, where they are closed. They are only open at the top (like a bucket). Since there is no lateral flow possible in these columns, all groundwater recharge R can be estimated based on groundwater table fluctuations in the unsaturated zone model.

The recharge calculated over one coupling time step goes into the water balance of the connected groundwater model cell. The water balances in the unsaturated zone column and the groundwater cell are iterated until the correct value for the specific yield is found and a consistent water table estimate in both models is achieved. For technical details, please see (Brandhorst et al., 2021).

In comparison to fully integrated 3D models like ParFlow or HydroGeoSphere, the coupled model lacks a representation of overland flow and does not work for domains with steep slopes, as lateral flows within the vadose zone are not represented. However, for problems where these aspects are not relevant, findings with this coupled model are highly transferable to the fully integrated 3D models. Since groundwater tables are contained within the 1D columns and nonlinear feedbacks between the groundwater and the vadose zone are handled through iteration, this model considers the interactions between the zones in an integrated manner, well comparable to a 3D model.

2.2. Ensemble Kalman Filter for Data Assimilation

The EnKF was introduced by Evensen (1994), and it is a common choice when applying DA to subsurface flow problems. Its objective is to improve predicted model states with measurements and reduce state uncertainty. This is done by calculating the most likely model state, given a measurement. The following section is a brief summary of the equations related to the EnKF.

The output of a numerical model is the *forecast* (denoted with an upper index f) from time $k-1$ to time k, which is estimated according to Equation 5. $\vec{x}_k^f$ is the augmented *state vector*, which contains states, parameters, and possibly bias terms of the forecast at the time k. $\mathcal{F}$ is the forecast operator, which is the numerical model presented in Equations 1 and 2 for the states and identity (I) for parameters and bias terms. $\vec{x}_{k-1}^a$ is the initial condition for the forecast, which, for $k>1$, is the state vector estimated by DA in the previous update step.

$$\vec{x}_k^f = \underset{k-1\to k}{\mathcal{F}}(\vec{x}_{k-1}^a) \tag{5}$$

The state vector $\vec{x}^f$ can be transformed to the observation space by the operator $\mathcal{M}$, which links measurements and model states (see Equation 6).

$$\vec{y}^f = \mathcal{M}(\vec{x}^f) \tag{6}$$

$\vec{y}^f$ are the *simulated observations*. An ensemble of model realizations is simulated (number of ensemble members $= N_{ens}$). The ensemble reproduces the distribution of the model output. This results in a state vector for each ensemble member ($\vec{x}_1^f$ to $\vec{x}_{N_{ens}}^f$), which can be combined in a matrix X^f. $\overline{X}^f$ is a matrix with the ensemble mean of the states in each column, and Y^f and $\overline{Y}^f$ are the equivalent matrices for the simulated observations $\vec{y}^f$. With these matrices, the error covariance matrices of the model states $C_{xy,k}$ and $C_{yy,k}$ can be approximated at each time k (see Equations 7 and 8). The accuracy of this approximation is tied to the ensemble size.

$$C_{xy,k} \approx \frac{1}{N_{ens}-1}\left(X_k^f - \overline{X}_k^f\right)\left(Y_k^f - \overline{Y}_k^f\right)^T \tag{7}$$

$$C_{yy,k} \approx \frac{1}{N_{ens}-1}\left(Y_k^f - \overline{Y}_k^f\right)\left(Y_k^f - \overline{Y}_k^f\right)^T \tag{8}$$

$C_{xy,k}$ describes error cross-covariances between the model states x and the simulated observations y and $C^f_{yy,k}$ describes error covariances of the simulated observations y. The Kalman gain K_k can be computed according to Equation 9, taking into account the observation error covariance matrix $C_{\varepsilon\varepsilon}$, which is calculated based on the predefined observation error ε.

$$K_k = C_{xy,k}\left(C_{yy,k} + C_{\varepsilon\varepsilon}\right)^{(-1)} \tag{9}$$

Y_k^{obs} is a matrix that contains the *observations* perturbed by white Gaussian noise, the extent of which is determined by the variance of the observation error ε. In the *analysis* step, the augmented state matrix is updated according to Equation 10, with X_k^a as the matrix of the updated ensemble states, which is then used for the next forecast step (see Equation 5). β is a damping factor between 0 and 1, used to confine the update (Hendricks Franssen & Kinzelbach, 2008).

$$X_k^a = X_k^f + \beta K_k\left(Y_k^{obs} - Y_k^f\right) \tag{10}$$

In some simulations, localization is applied to mitigate the effects of spurious covariances, which occur due to the limited representation of the distribution by a finite and typically too-small ensemble. Within this work, the term localization refers to covariance localization, as it was described amongst others by Houtekamer and Mitchell (2001). In theory, the Schur (element-wise) product (denoted as $\odot$) of a *localization matrix* ρ_l and the state error covariance matrix $C_{xx,k}$ are calculated. ρ_l may hold values between 0 (update not applied) and 1 (update fully applied) but is chosen binary within this work to keep it simple. In the EnKF, $C_{xx,k}$ is not computed, so the localization is directly applied to $C_{xy,k}$ and $C_{yy,k}$ using the variants of the localization matrix $\rho_{l,xy}$ and $\rho_{l,yy}$, which are in line with the dimensions of the respective covariance matrices (see Equation 11).

$$K_k = C_{xy,k} \odot \rho_{l,xy}\left(C_{yy,k} \odot \rho_{l,yy} + R\right)^{(-1)} \tag{11}$$

2.3. Bias Correction

Model errors in the EnKF are assumed to follow an unbiased distribution. Biases between modeled states and real states can be accounted for with a bias-aware EnKF (simply referred to as *bias correction* in the following). Biases between the simulated and the observed variables are compensated by bias terms, which are included in the augmented state vector and are thus updated in the analysis step (Equation 10). In this work, bias correction is performed based on the colored noise filter approach introduced by Madsen and Canizares (1999). Bias terms $\vec{e}$ are initialized as spatially correlated noise terms for the states of every ensemble member. After the forecast and before the analysis step, a *corrected forecast* (index *cf*) is conducted for every ensemble member, resulting in a corrected state vector. In this correction, the bias terms are added to the forecast of the numerical model (see Equation 12). $\vec{x}^f_{s,k}$ is a subvector of the augmented state vector $\vec{x}^f_k$ that only contains the states and $\vec{x}^{cf}_{s,k}$ is the corrected state vector.

$$\vec{x}^{cf}_{s,k} = \vec{x}^f_{s,k} + \vec{e}^f_k. \tag{12}$$

In the analysis step (Equation 10), X_k^f consists of the corrected state vector $\vec{x}^{cf}_{s,k}$, the soil hydraulic parameters, and the bias terms $\vec{e}^f_k$ for each ensemble member. If the bias correction is successful, the updates condition the ensemble toward meaningful bias terms that correct for the biases between the numerical model and observations. A process noise ω needs to be added to the bias terms before the correction step (Equation 12): $\vec{e}^f_{k+1} = \vec{e}^a_k + \omega$.

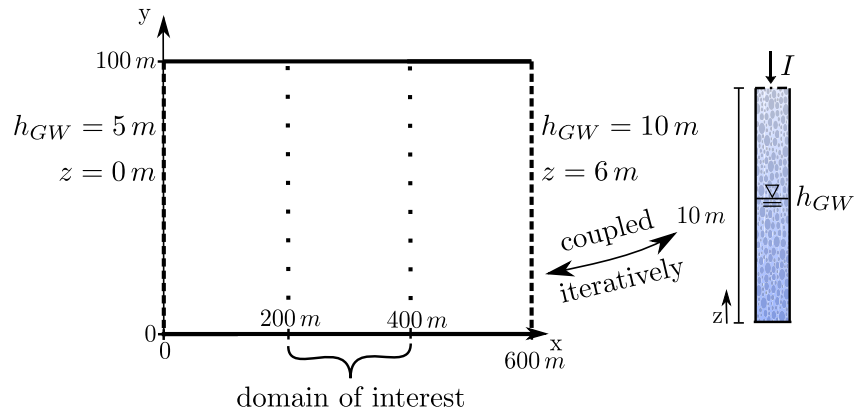

Figure 2. Domain of the Test case. The left side shows the groundwater model and the right side shows one vadose zone pillar, with the infiltration I applied on top. Solid lines are no-flow boundaries, dashed lines are Dirichlet boundaries, and the dotted lines indicate the domain of interest. z is the geodetic head at the bottom of the domain, and h_{GW} is the groundwater table height.

3. Test Case to Study Data Assimilation in Coupled Subsurface Systems

As this work is about investigating the interactions between the vadose zone and groundwater during DA, the test case that will be presented in the following is set up with this goal in mind. It is meant to be simple enough to clearly see the feedbacks without too many additional processes obscuring them and complex enough such that it can build a bridge between idealized studies like H. Zhang et al. (2018) and complex ones like Hung et al. (2022). It is spatially heterogeneous, and groundwater flow in the horizontal direction is captured. However, the effects of rivers, surface runoff, and other land surface processes are not considered. It exhibits shallow groundwater table depths for meaningful correlations between states of the saturated and unsaturated zones but no full saturation at the land surface to avoid modeling overland flow.

The domain is schematically shown in Figure 2. The length scale is chosen to be comparably small to examine heterogeneity at a meter scale. Simulations start at the beginning of the year and run for 200 d, capturing conditions during winter and summer. No-flow boundaries are defined at $y = 0$ and $y = 100$ m and Dirichlet (constant head) boundaries at $x = 0$ and $x = 600$ m. The full size of the domain is 600 m $\times$ 100 m $\times$ 10 m, but only the inner third in the x-direction is considered as the domain of interest to acquire more dynamic pressure distributions for the saturated zone and reduce effects of the Dirichlet boundary conditions. A constant slope of the land surface of 0.01 is applied in the x-direction.

The domain consists of variably saturated vertical soil columns that reach the bottom of the aquifer, which are coupled to a 2D-groundwater-flow model with a horizontal discretization of $\Delta x = \Delta y = 2$ m. The vertical columns are discretized into 172 cells with thickness decreasing from the bottom (12.5 cm) to the top (0.25 cm). The coupling time step is equal to the time step of the groundwater model ($\Delta t = 1\,d$). For the vadose zone model, Δt is chosen adaptively between 1 s and 1 d. The spatial and temporal resolutions are chosen according to a convergence study, and we have found that they give a good compromise between computational demand and accuracy, with relative errors in the range of 10^{-3}.

The soil parameters represent a sandy loam/sandy clay loam, which is common in Central Europe and other regions. Vertically, the domain can be divided into three distinct soil layers with sharp interfaces (*top layer*: land surface until 44 cm depth, *middle layer*: until 2.05 m depth, and *bottom layer*). Such distinct layering is common in real soils, and it complicates finding meaningful correlations over the depth of the soil, as flow behavior might notably differ between the different layers.

Observations of states are generated with model runs, called virtual reality (VR). Using VRs allows us to reflect distinctly upon the problems of DA itself in a controlled environment. To avoid results that are too case-specific, we conduct DA experiments with measurements from four different virtual realities that differ with respect to soil hydraulic parameters and boundary conditions. We have found these four VRs to differ enough to obtain a

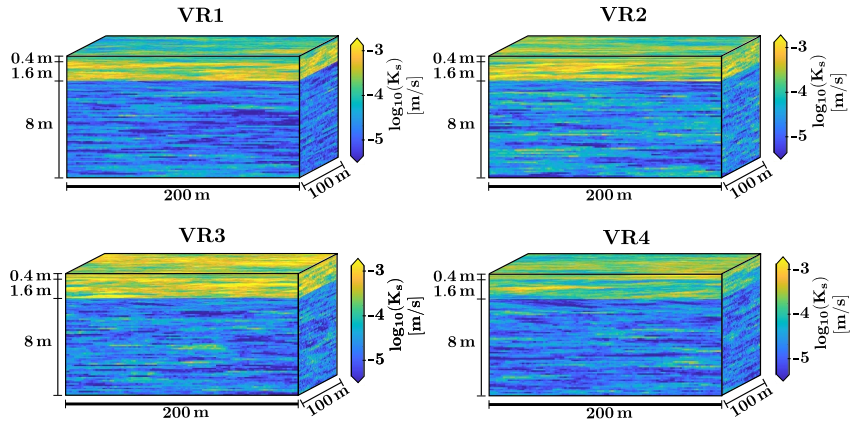

Figure 3. 3D distribution of the logarithmic hydraulic conductivity of the virtual realities 1 to 4. The z-axis is scaled for better visibility.

coherent picture and work out general findings and differences due to special cases. In the following, these four VRs are presented.

3.1. Virtual Realities

The VRs consist of 202 1D columns of different cross sections to simulate vertical flow in the vadose zone coupled to the 2D horizontal groundwater flow model. This way, horizontal heterogeneous patterns can be captured well, but the simulation is not too compute-intensive. 200 columns have identical cross-section sizes and are equally distributed over the domain of interest. Two columns are placed outside the domain of interest and cover the remaining area.

The constant pressure heads for the Dirichlet boundaries of each VR are sampled from a continuous uniform distribution in a range of ± 0.1 m, with average values of $h_{GW}(x = 0) = 5$ m and $h_{GW}(x = 600\,\text{m}) = 10$ m.

Within each soil layer, the average soil texture is sampled from a β-distribution, and then heterogeneous fields are generated. The heterogeneous soil texture parameters are represented by multivariate conditional random fields generated with GCOSIM3D (Gómez-Hernández & Journel, 1993) with variable correlation lengths within the ranges of $r_x = 110\,\text{m} \pm 25\,\text{m}$, $r_y = 40\,\text{m} \pm 12.5\,\text{m}$, and $rz = 18\,\text{cm} \pm 5\,\text{cm}$ that are sampled from uniform distributions. The parameters of K_s and n are calculated for each sample of the heterogeneous soil texture with the pedotransfer functions by Cosby et al. (1984) and Tóth et al. (2015). Since the pedotransfer function by Cosby et al. (1984) resulted in a low spatial variance of K_s, the spread of the spatial distribution of K_s was artificially increased while maintaining the patterns and spatial mean. The heterogeneous distributions of K_s for each VR are shown in Figure 3.

The heterogeneous spatial distributions of the parameter α were generated from the K_s-fields using Miller-scaling (Miller & Miller, 1956): $\alpha = A\sqrt{K_s}$. The scaling factor A was calculated such that the spatial average of α matches the spatial average that would have been obtained using the pedotransfer function of Tóth et al. (2015). Since it is known that pedotransfer functions can only give a rough approximation of the soil hydraulic parameters, their estimates are also considered uncertain. Spatial averages of K_s, α, and n are adjusted by β-distributed noise. Spatial means and standard deviations of the soil hydraulic parameters for each layer and each VR are shown in Table 1.

At the top of each vadose zone model, a Neumann boundary condition is applied to model infiltration at a daily resolution. For each model spinup (to reach a dynamic equilibrium), measurements of infiltration from a weather station in Gdansk, Poland (Beegum et al., 2018) were repeatedly applied. After the spinup, a net infiltration is applied with a different intensity for each VR (see Figure 4). The infiltration time series is derived from an ensemble of a larger model (ParFlow-CLM), which used conditional precipitation fields generated by the method of Haese et al. (2017) to estimate the perturbed infiltration and evapotranspiration patterns. Spatial variation of infiltration is neglected. Overland flow is not considered, and spatial variation of precipitation is negligible due to the small spatial extent of the domain.

Table 1
Spatial Means and Standard Deviations (in Brackets) of Soil Hydraulic Parameters for the Four Different Virtual Realities

		$\log_{10}(K_s)$ [m/s]	ϕ [−]	α [m^{-1}]	n [−]
VR1	Top layer	−3.93 (0.39)	0.40 (0.03)	2.95 (1.39)	1.33 (0.01)
	Middle layer	−3.38 (0.42)	0.50 (0.03)	22.63 (11.65)	1.61 (0.01)
	Bottom layer	−4.87 (0.41)	0.31 (0.03)	0.57 (0.28)	1.36 (0.01)
VR2	Top layer	−3.79 (0.37)	0.32 (0.03)	2.39 (1.08)	1.43 (0.01)
	Middle layer	−3.50 (0.44)	0.47 (0.03)	7.25 (4.03)	1.58 (0.01)
	Bottom layer	−4.62 (0.43)	0.38 (0.03)	4.33 (2.28)	1.30 (0.01)
VR3	Top layer	−3.31 (0.37)	0.39 (0.03)	24.59 (11.21)	1.30 (0.01)
	Middle layer	−3.42 (0.41)	0.53 (0.03)	25.42 (13.27)	1.61 (0.01)
	Bottom layer	−4.75 (0.43)	0.34 (0.03)	0.83 (0.45)	1.26 (0.01)
VR4	Top layer	−3.59 (0.37)	0.32 (0.03)	5.50 (2.37)	1.49 (0.01)
	Middle layer	−3.95 (0.40)	0.54 (0.03)	15.83 (7.88)	1.36 (0.01)
	Bottom layer	−4.77 (0.41)	0.36 (0.03)	0.66 (0.33)	1.29 (0.01)

3.2. Data Assimilation Settings

Data assimilation is performed with an ensemble of 100 members. Larger ensemble sizes were tested for some experiments, but they did not yield notably different results. As observations, we take measurements from a virtual reality (VR) at the locations shown in Figure 5. For soil moisture, measurement depths are 11 cm, 31 cm, and 71 cm. This means that the first two measurements are within the top layer, and the third is within the middle layer. A Gaussian error with a zero mean and a standard deviation of 0.01 is added to the water content observations. For the groundwater observations, we used an error of 0.01 m. To reduce the computational demand of the DA simulations, the amount of 1D columns is reduced from 202 to 52 for each ensemble member, meaning that the ensemble represents the vadose zone at a lower horizontal resolution than the VR. The ensemble does not exhibit horizontal soil heterogeneity and infiltration is assumed spatially invariant, so we found that the high resolution that is used within the VR is not needed, as the model structural error introduced by this coarsening is comparatively small. The spinup and sampling of the boundary conditions for the ensemble follow the same approaches used for the VRs.

The spatially distributed soil hydraulic parameters of the ensemble members are sampled from the same pdfs as used for the VRs (ensemble means and standard deviations shown in Table 2). This means we assume prior knowledge regarding the soil texture and its uncertainty, which we consider a realistic assumption given the availability of soil maps. Detailed information about heterogeneous patterns of the soil hydraulic parameters is usually not available, so the heterogeneous distribution is not represented by the ensemble. The case that depths of the soil layers are known is not rare, so one ensemble, which we will refer to as *homogeneously layered*, deals with this case. In this ensemble, the soil hydraulic parameters within each of the three soil layers are homogeneously distributed but differ among the layers. The case that one has no information about the soil layering should be regarded as a more extreme case, which is covered by the *fully homogeneous* ensemble. The two ensembles are not only representative of challenges one might face in real settings, but they also allow evaluation of the effects of the model error on DA with two different conceptual soil models. In prior experiments, we found that DA without any model structural error (with a homogeneous ensemble and VR) works very well, also across subsurface compartments. Here we present more realistic scenarios, as a necessary step forward from synthetic toward real-world applications, where such model structural errors are inevitable.

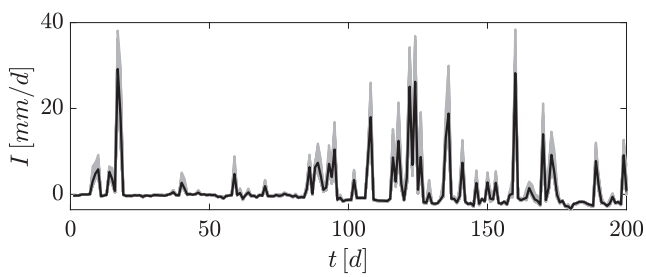

Figure 4. Net infiltration I over time. The gray lines show values for each model realization, and the black line shows their average.

Soil hydraulic parameters α, K_s, n, and ϕ are updated on a weekly basis. Soil hydraulic parameters make up a large part of prediction uncertainty by the forward model. Including them in the update for subsurface flow DA was found to be beneficial compared to only updating the states (e.g., Brandhorst & Neuweiler, 2023; Gebler et al., 2019). Soil moisture is updated daily, and

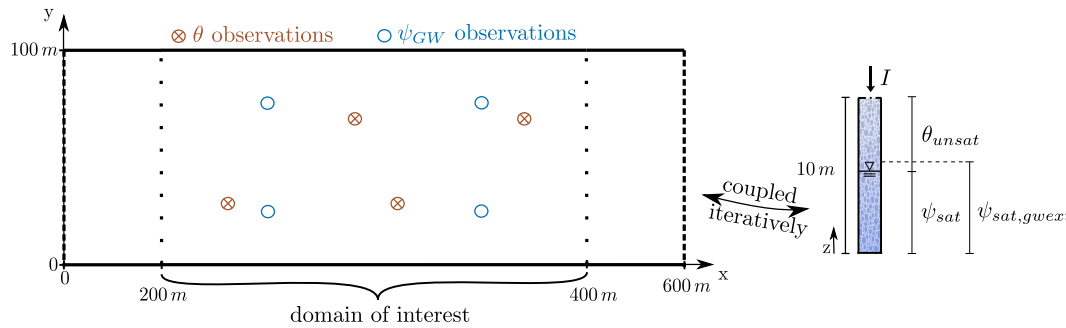

Figure 5. Measurement locations and variables relevant for the data assimilation (DA) experiments. The left side shows the groundwater model and measurement locations for DA, with θ indicating soil moisture and ψ_{GW} groundwater pressure head. The right side shows one vadose zone pillar and θ_{unsat}, ψ_{sat}, and $\psi_{sat,gwext}$ refer to point estimates within the respective ranges. θ_{unsat} is the soil moisture in the vadose zone, ψ_{sat} is the pressure head in the fully saturated zone, and $\psi_{sat,gwext}$ is the pressure head in the fully saturated zone and a variably saturated section above.

groundwater levels are updated weekly. We chose different update frequencies for soil moisture and groundwater since groundwater states change more slowly, and too frequent updating should be avoided as measurement errors are assumed to be temporally uncorrelated (Evensen et al., 2022). After 150 days, DA is turned off, and we proceed with an open-loop phase subsequent to DA with the initial states and parameters determined by the last update of the DA. We found the computational demand of one full ensemble simulation to vary between 120 and 528 core hours, allowing us to run many simulations on a comparably small compute cluster, using 24 cores in parallel.

3.3. Data Assimilation Variants

To investigate cross-compartmental interactions during DA, the uni- and multivariate DA schemes for soil moisture and groundwater shown in Table 3 are applied. The schemes include univariate (SM, GW) and multivariate ($SMGW$) DA. Groundwater table observations are used for weakly coupled GW_{wc} (strongly coupled GW_{sc}) DA, where the soil moisture within the vadose zone θ_{unsat} is not (is) updated along with the groundwater states.

A variation of the weakly coupled DA is tested ($GW_{wc,gwext}$), where additionally to the groundwater pressure head ψ_{GW} (2D groundwater flow model) the pressure heads up until 10 cells above the groundwater table of the 1D variably saturated flow models (see $\psi_{sat,gwext}$ in Figure 5) are also updated. As Δz is variable in the considered test case and decreases from bottom to top, the average vertical length associated with the 10 cells approximately varies between 25 cm and 50 cm in the conducted experiments. The decrease of this updated vertical length from bottom to top is intentional since pressure heads near the land surface are best avoided in the state vector because they may yield highly negative values. H. Zhang et al. (2018) found such values to be problematic for DA, as they skew the prior distribution. The length between 25 cm and 50 cm was found to work well for this test case, but it is expected to be test case specific, and a different number of cells or a fixed distance could also have been chosen. The essential point of the variant $GW_{wc,gwext}$ is that a section above the groundwater table is updated along with the water table of the aquifer. This could be considered a compromise between weakly coupled and strongly coupled update variants. The choice of pressure head instead of soil moisture for the state updates in the vadose zone is intentional. Updating pressure heads for the $GW_{wc,gwext}$ scheme avoids complications due to the nonlinear relationship between soil moisture and pressure head. Also, soil moisture values near the groundwater table are expected to be less informative about the flow conditions than pressure heads because they are near the volume of full saturation.

Table 2
Ensemble Means and Standard Deviations (in Brackets) of the Soil Hydraulic Parameters in the Three Soil Layers

	$\log_{10}(K_s)$ [m/s]	ϕ [−]	α [m^{-1}]	n [−]
Top layer	−3.76 (0.28)	0.42 (0.04)	8.30 (8.36)	1.4 (0.08)
Middle layer	−3.3 (0.3)	0.41 (0.05)	17.36 (14.16)	1.56 (0.09)
Bottom layer	−4.73 (0.32)	0.42 (0.04)	1.44 (2.53)	1.34 (0.10)

The multivariate DA (SMGW) is performed asynchronously, assimilating SM in the first step and $GW_{wc,gwext}$ in the second step. A full simulation of the ensemble without DA is referred to as an open-loop (OL) run. If bias correction is applied, this is labeled as *bias*. The bias terms are applied to the pressure heads in the 1D variably saturated models. Applying the bias terms

Table 3
Data Assimilation Variants

DA variant	Observations	State vector
SM	θ	$[\theta_{unsat}, P]$
SM_{bias}	θ	$[\theta_{unsat}, P, e]$
GW_{wc}	ψ_{GW}	$[\psi_{GW}, \psi_{sat}, P]$
GW_{sc}	ψ_{GW}	$(\psi_{GW}, \psi_{sat}, \theta_{unsat}, P)$
$GW_{wc,gwext}$	ψ_{GW}	$[\psi_{GW}, \psi_{sat,gwext}, P]$
$SMGW$	$\theta \mid \psi_{GW}$	$[\theta_{unsat}, P] \mid [\psi_{GW}, \psi_{sat,gwext}, P]$

Note. $P = [\log(K_s), \log(\alpha), n, \phi]$ are all soil hydraulic parameters. e are the bias terms. ψ_{GW} is the groundwater pressure head and refers to the groundwater model. θ is the water content and refers to the variably saturated model. ψ_{sat}, $\psi_{sat,gwext}$ and θ_{unsat} refer to the variably saturated model. θ_{unsat} is the water content, and it is only considered above the groundwater table. ψ_{sat} is the pressure head in the parts below the groundwater table. $\psi_{sat,gwext}$ is the pressure head considered below and a small stretch above the groundwater table. See also Figure 5. | indicates asynchronous updates: First, the measurement left from the vertical line is assimilated to update the left state vector, and then the right measurement is assimilated to update the right state vector.

to the soil moisture was tested, but it led to skewed pdfs. In the context of this work, bias terms can be considered source/sink terms that repeatedly correct for lacking heterogeneity of the soil hydraulic parameters. The label *loc* indicates that vertical covariance localization is applied to constrain the update to the respective soil layer (see Equation 11). The damping factors used in Equation 10 are $\beta = 0.1$ for SM, $\beta = 0.5$ for GW, $\beta = 0.1$ for parameters and $\beta = 1$ for the bias terms. These factors gave stable results in previous experiments, avoiding issues such as filter inbreeding. For the multivariate DA, parameters are updated twice as often as in the univariate experiments because of the two update steps. For this reason, the damping factor is decreased by a factor of two in the multivariate DA experiments.

3.4. Evaluation of Model Results

To quantify the forecast quality of the DA experiments, the root mean square error (RMSE) of a variable x is used:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} \left(\overline{x_i} - x_i^{\mathrm{VR}}\right)^2}{N}} \tag{13}$$

with N as the number of timesteps, $\overline{x_i}$ as the ensemble average of the variable, and x_i^{VR} as the true variable resulting from the simulation of the respective VR. The RMSE is calculated both for the OL run (labeled as $\mathrm{RMSE}_{OL\,run}$) and DA experiments (labeled as RMSE_{DA}). Their ratio $r = \frac{\mathrm{RMSE}_{DA}}{\mathrm{RMSE}_{OLrun}}$ is used as a measure of how much DA improves ($r < 100\%$) or deteriorates ($r > 100\%$) the model forecast.

Since the spatial extent of our domain is comparably small, we mostly concentrate on spatially averaged estimates, as we deem a correct forecast of them as more important for water management applications than accurately depicting local values. We also expect conclusions drawn from spatial averages to be better transferable to catchment scale models. These averages are shown with an overline throughout this work, to clearly distinguish them from point values. The variables shown are relevant with respect to different applications and representative of different soil depths and compartments. We investigate the land surface soil moisture $\overline{\theta}_{10cm}$ (known to be linked to evaporation), which we estimated as the average volumetric soil moisture for the upper 10 cm of the soil, the root zone soil moisture $\overline{\theta}_{2m}$ (known to be linked to evapotranspiration), which is the average volumetric soil moisture for the upper $2\,m$ of the soil, the spatially averaged groundwater table height $\overline{h}_{GW}$, and the spatially averaged groundwater recharge $\overline{R}$. Only for Section 4.1, we consider a local estimate: The soil moisture at the observation locations θ_{DAlocs}. While predicting this variable is not the main focus of this work, it illustrates how the bias correction locally emulates heterogeneity at the observation locations.

4. Results and Discussion

In this section, the test case is used to examine the two goals outlined in the introduction: testing DA strategies for (a) dealing with deteriorations and (b) making better use of the interactions between the vadose zone and groundwater compartments. We found the issue of deteriorations across the subsurface compartments to be particularly relevant for the soil moisture DA experiments, which are the focus of Section 4.1. Due to the connection between the deteriorations and model structural errors, DA experiments are conducted for both ensembles (fully homogeneous and homogeneously layered). Bias correction and localization are tested as measures to counteract the deteriorations.

Goal (b) is addressed using groundwater table DA and multivariate DA of both soil moisture and groundwater together. Section 4.2 deals with the question of *how* measurements of groundwater tables could be used to optimally improve states in the coupled system. The different degrees of coupling strength for groundwater table updates that were presented in Section 3.2 are tested with the homogeneously layered ensemble. By analyzing estimates of groundwater recharge, differences between the different DA coupling methods are quantified right at the interface between the two compartments (the groundwater table). In Section 4.3, univariate and multivariate

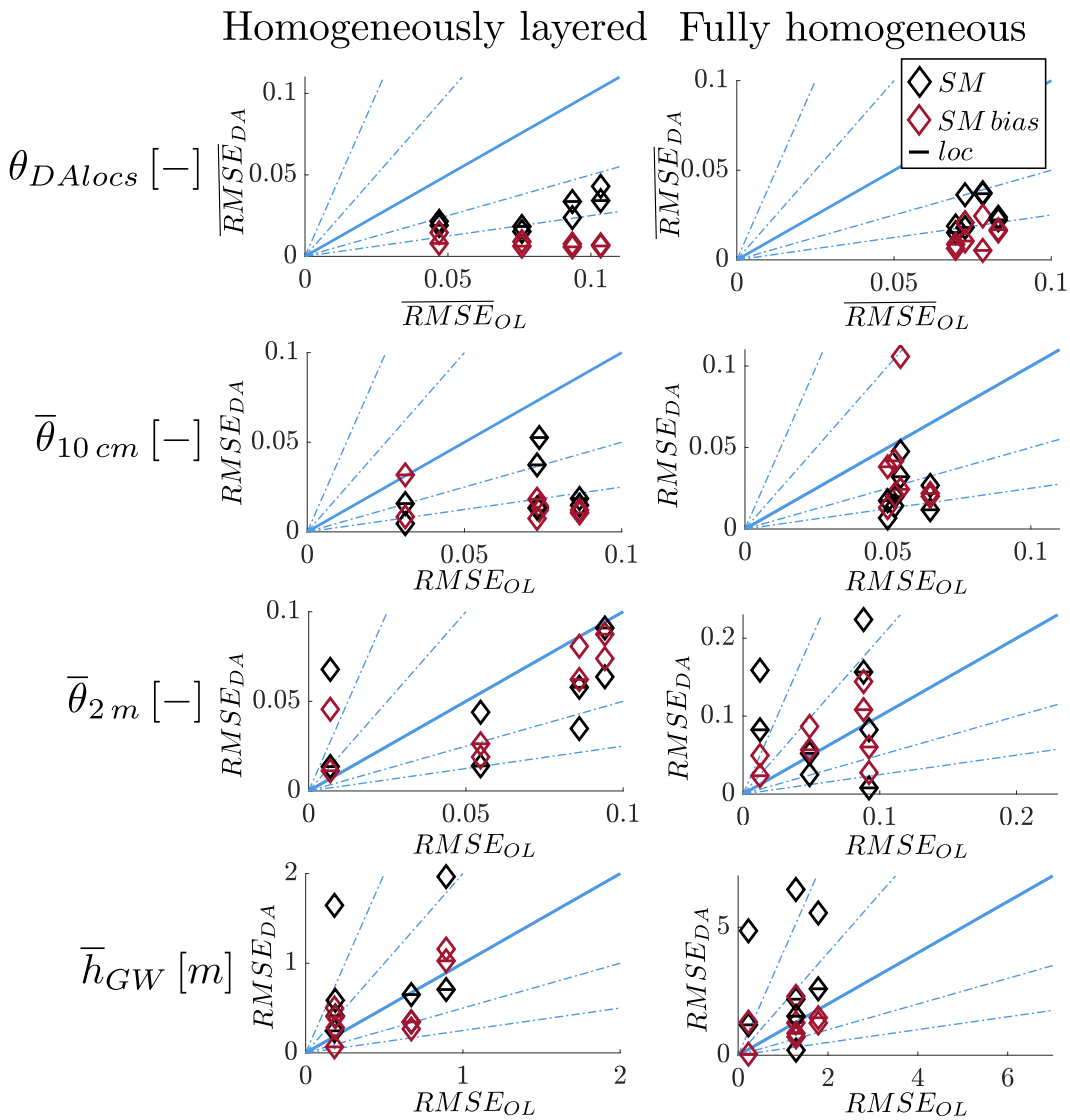

Figure 6. Root mean square errors (RMSEs) during the open-loop phase subsequent to data assimilation (DA) ($t > 150\,d$). The error on the *y*-axis compares the ensemble mean with DA to the virtual reality (VR), and the error on the *x*-axis compares the mean of the open-loop (OL) run (no DA) to the VR. RMSEs of the averages of land surface soil moisture, root zone soil moisture, and the groundwater table height are shown. For soil moisture at the observation locations, the spatial averages of the RMSEs (not the RMSE of the spatial averages) are shown to capture local errors. Soil moisture DA experiments (*SM*) are shown in black and soil moisture DA experiments with bias correction (*SM bias*) in crimson, both with and without vertical localization. Localization is indicated by a horizontal line. Below the straight blue line, DA improves compared to the OL run. The dashed lines indicate a quadrupling ($r = 400\%$), doubling ($r = 200\%$), halving ($r = 50\%$) and quartering ($r = 25\%$) of the RMSEs from the DA experiments.

DA experiments are compared. It deals with the question of *which* measurements should be used if certain variables are of interest.

If it is stated that DA improves or worsens a prediction, this always means with respect to the open-loop run, where no DA is carried out. This deviation is also quantified with a RMSE. For the RMSEs, we look at the timeframe $t > 150\,d$ (OL phase) because by then, the biggest differences between the experiments have emerged.

4.1. Soil Moisture Data Assimilation and Cross-Compartmental Deteriorations

Figure 6 compares the RMSEs for the DA and the OL runs for the soil moisture DA experiments (upper two rows in Table 3). Values below the straight line indicate an improvement compared to the OL run. The different VRs

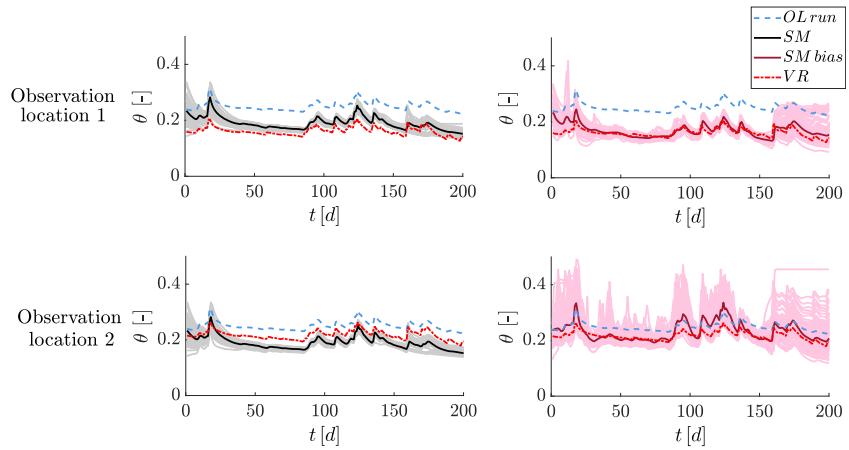

Figure 7. Soil moisture θ over time at two soil moisture observation locations at 31 cm depth. Data assimilation (DA) is conducted with the homogeneous ensemble with measurements from VR2, with the DA variants *SM* (left, mean in black and ensemble in gray) and *SM bias* (right, mean in crimson and ensemble in light pink).

are not labeled for the sake of clarity but can be distinguished by considering that all experiments conducted with the same VR share the same location on the x-axis since they have the same value of $\mathrm{RMSE}_{OL\ run}$. Figure 6 shows that most improvements are achieved closest to the observation locations (top), and deteriorations are strongest the furthest away from them (bottom). At the observation locations (θ_{DAlocs}), soil moisture predictions can be improved considerably by DA in all considered cases despite the ensembles being homogeneous approximations of the VRs. Localization does not play a major role here. With bias correction, the RMSEs can be reduced further since local heterogeneities of the VRs can be emulated at the observation locations. This is exemplified in Figure 7, which compares the temporal evolution of the estimates with the DA variants *SM* and *SM bias* for the homogeneous ensemble with observations from VR2 at two different locations. Due to the homogeneity of the ensemble, the two locations are highly correlated, and the updates by DA have to look very similar for the two locations. Local differences present in the VR can thus not be fully reproduced with the DA variant *SM*, seen in the left part of Figure 7. At the observation locations, the bias terms in the experiments with *SM bias* can compensate for such local differences as shown in the right part of Figure 7. However, bias correction can also notably inflate the ensemble, as seen at observation location 2. While this can counteract filter inbreeding, it increases uncertainty and can skew the pdfs if soil moisture constraints are reached.

Figure 6 further shows that the estimate of the land surface soil moisture $\overline{\theta}_{10cm}$ is almost always improved by the soil moisture assimilation, with localized simulations improving a bit less than those without localization. The land surface soil moisture only covers the top layer, and it appears that measurements from the middle layer actually contain valuable information for estimating $\overline{\theta}_{10cm}$, despite the sharp interface between the two layers. For the homogeneously layered ensemble, the estimate of the root zone soil moisture $\overline{\theta}_{2m}$ is improved in the majority of simulations. For the fully homogeneous ensemble, a deterioration of the estimation of $\overline{\theta}_{2m}$ is more often found than an improvement. We conclude that knowledge regarding soil layering matters a lot for this variable. The deteriorations can generally be reduced by bias correction and localization. Deteriorations of $\overline{\theta}_{2m}$ predictions often coincide with deteriorations of predictions of the groundwater table height $\overline{h}_{GW}$, but are often less pronounced in comparison. It is probable that the deteriorations of the estimates of both these variables ($\overline{\theta}_{2m}$ and $\overline{h}_{GW}$) are tied to the model error due to the assumption of homogeneous soils. This interpretation is supported by the fact that the deteriorations are much more pronounced and common for the experiments with the fully homogeneous ensemble. Bias correction and vertical localization are measures we used to address the model errors due to the assumption of homogeneity. For $\overline{h}_{GW}$, the amount of simulations where $r > 200\%$ (the RMSEs are doubled by DA) reduces from 75% to 25% when using localization or bias correction. A combination of localization and bias correction further reduces this amount to 12.5% and shifts the majority of experiments from severe deterioration ($r > 200\%$) to improvement ($r < 100\%$). It can thus be concluded that both bias correction and vertical localization are effective measures against cross-compartmental deteriorations. Even though the error values vary quite a bit between the different VR realizations, these findings apply to all VRs.

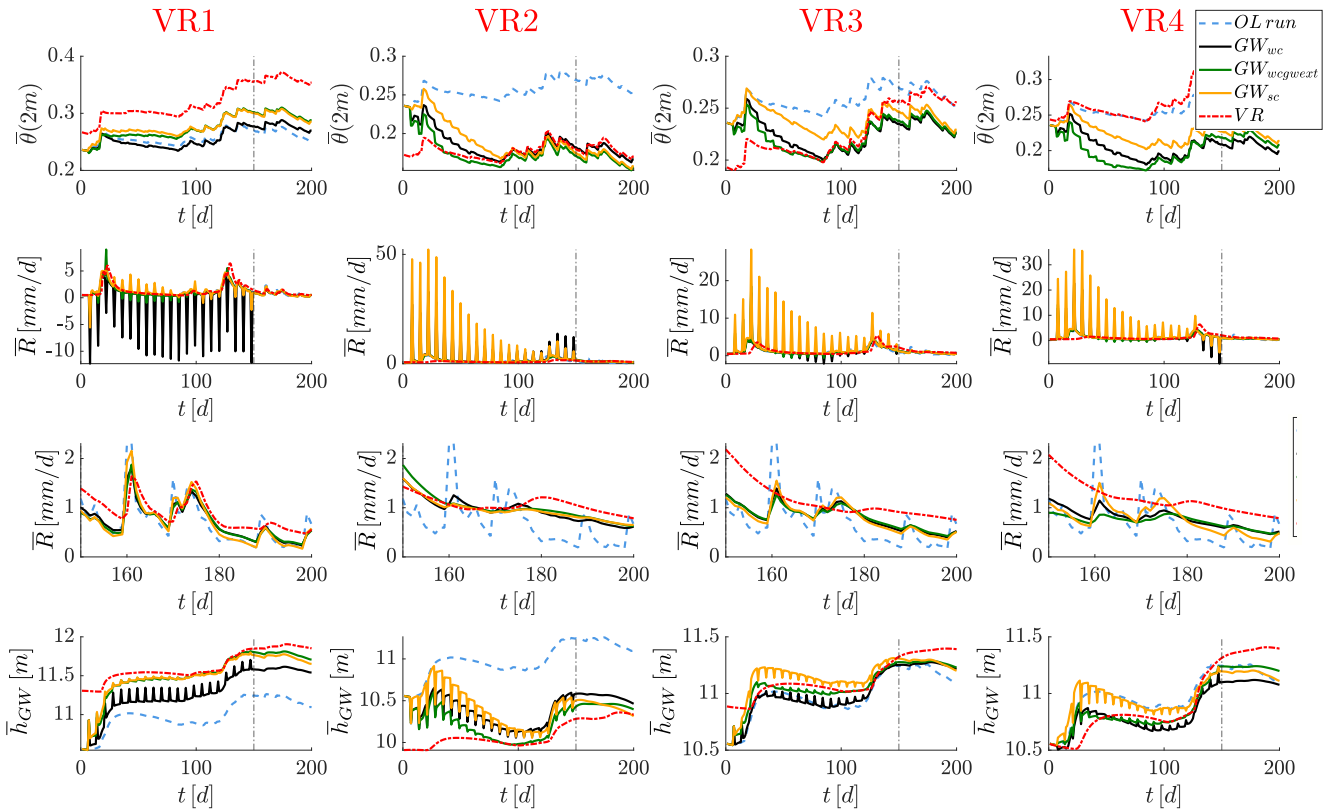

Figure 8. Spatially averaged root zone soil moisture $\overline{\theta}_{2m}$, groundwater recharge R, and groundwater table height $\overline{h}_{GW}$ for four different virtual realities and three different groundwater table update variants. The third row shows R during the open-loop phase subsequent to data assimilation ($t > 150\,d$). The average height of the land surface is $13\,m$.

We can generally summarize that groundwater table predictions were mostly deteriorated by soil moisture DA. So if one is only interested in groundwater table estimates, univariate soil moisture DA cannot be recommended based on these findings. If one is also interested in soil moisture estimates, one can mitigate the deteriorations with bias correction and/or localization.

4.2. Updating Across Subsurface Compartments With Groundwater Table Data Assimilation

Figure 8 shows the temporal evolution of the estimates of the spatially averaged root zone soil moisture $\overline{\theta}_{2m}$, groundwater recharge $\overline{R}$ and groundwater table height $\overline{h}_{GW}$ for the DA experiments with the homogeneously layered ensemble obtained with the groundwater table update variants GW_{wc}, $GW_{wc,gwext}$, and GW_{sc} (see Table 3). It can be seen that the weakly coupled groundwater updates (GW_{wc}) improve predictions of $\overline{h}_{GW}$ (RMSEs during the open-loop phase subsequent to DA on average reduced by 29% compared to the OL run without prior DA). However, $\overline{h}_{GW}$ values drift slightly back to those of the OL run after the update. We attribute this to the exchange fluxes with the vadose zone, illustrated in the second row of Figure 8 as groundwater recharge.

It can clearly be seen that ensemble averages of groundwater recharge estimates *during the DA time period* ($t < 150\,d$) are not representative of the true value at all. Mass balance violations by DA cause balancing fluxes across the compartments after each update. For GW_{wc}, such fluxes can be particularly high because cells below full saturation are not updated, leading to kinks in the pressure profile, which cause high gradients. For example, for VR1, the groundwater table height is increased with the DA method GW_{wc} at each update, but the vadose zone remains dry, as it is not affected by the updates. The capillary rise from groundwater into the vadose zone then impairs the update. Weakly coupled groundwater updates with extended groundwater ($GW_{wc,gwext}$) can deal best with the balancing fluxes (lowest peaks in $\overline{R}$) and thus also perform best with respect to predictions of $\overline{h}_{GW}$. The RMSEs of h_{GW} during the open-loop phase subsequent to DA are, on average, reduced by 58% compared to the OL run without prior DA, which is more than for the update variants GW_{wc} (29%) and GW_{sc} (49%). GW_{sc}

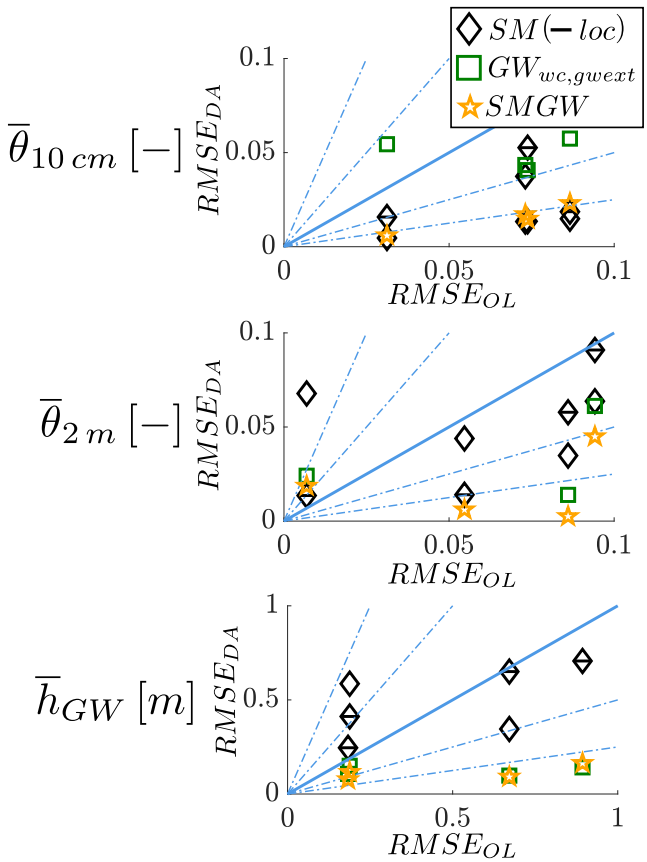

Figure 9. Root mean square errors (RMSEs) during the open-loop phase subsequent to data assimilation (DA) ($t > 150\,d$). The error on the *y*-axis compares the ensemble mean with DA to the virtual reality (VR), and the error on the *x*-axis compares the mean of the open-loop (OL) run (no DA) to the VR. RMSEs of the averages of land surface soil moisture, root zone soil moisture, and the groundwater table height are shown. Soil moisture DA experiments (*SM*) are shown in black, Groundwater table DA experiments ($GW_{wc,gwext}$) are shown in green, and multivariate DA experiments (SMGW) are shown in yellow. Localization is indicated by a horizontal line. Below the straight blue line, DA improves compared to the OL run. The dashed lines indicate a quadrupling ($r = 400\%$), doubling ($r = 200\%$), halving ($r = 50\%$) and quartering ($r = 25\%$) of the RMSEs from the DA experiments.

performs very well for VR1 and not so well for the other VRs. In the GW_{sc} variant, the states in the whole vadose zone are being updated and the cross-covariances between the groundwater table height and soil moisture in the vadose zone add water to the vadose zone that percolates and impairs the groundwater table updates. The balancing recharge flows in experiments with $GW_{wc,gwext}$ are much lower compared to the other two variants. Weakly coupled groundwater table updates (GW_{wc}) are a common choice in recent studies, so the finding that $GW_{wc,gwext}$ appears to give better groundwater table estimates is important.

During the open-loop phase subsequent to DA ($t > 150\,d$, the third row of Figure 8), groundwater recharge estimates can be improved by DA. However, if there is a bias in the temporal average recharge, DA is not able to compensate for that (see VRs 3 and 4, where the ensembles systematically underestimate groundwater recharge). The variability between the VRs and also in one VR over time does not allow us to evaluate which of the three *GW* DA variants performs better or worse in estimating recharge subsequent to DA.

Estimates of the root zone soil moisture $\overline{\theta}_{2m}$ are notably improved by groundwater updates for VR2 and notably deteriorated for VR4. In VR1, no DA strategy is able to retrieve the true root zone soil moisture. However, the estimates with the update variants $GW_{wc,gwext}$ and GW_{sc} get closer to the truth than the estimate with the weakly coupled variant GW_{wc}. Looking at all four VRs, it appears that groundwater table assimilation is often better able to decrease the values of $\overline{\theta}_{2m}$ than to increase them.

This section has shown that updating parts of the vadose zone with groundwater table observations is worthwhile, when interested in estimating groundwater tables. While a full strongly coupled update where the whole vadose zone is updated (GW_{sc}) is often inadvisable, as it is prone to spurious correlations, updating a section above the groundwater table $GW_{wc,gwext}$ can notably improve estimates over the commonly used weakly coupled updating approach (GW_{wc}). For the root zone soil moisture, the benefits of updating across the subsurface compartments can also be seen commonly, but not always. The positive findings regarding the update variant $GW_{wc,gwext}$ clearly contribute toward exploiting cross-compartmental interactions to improve predictions.

4.3. Synergetic Effects During Multivariate Data Assimilation

Figure 9 compares RMSEs from the univariate assimilation of soil moisture (with and without vertical localization) with the best-performing groundwater assimilation variant $GW_{wc,gwext}$ and the multivariate assimilation of soil moisture and groundwater SMGW. Multivariate DA experiments (SMGW) show a similarly good representation of the land surface soil moisture $\overline{\theta}_{10cm}$ as the univariate soil moisture assimilation (*SM*). This means that if $\overline{\theta}_{10cm}$ is the only variable of interest, assimilating soil moisture is sufficient, but additionally assimilating groundwater table observations does no harm. However, often, one is also interested in root zone soil moisture $\overline{\theta}_{2m}$ due to its relation to evapotranspiration. For the characterization of $\overline{\theta}_{2m}$, the multivariate DA experiments clearly show the best results. For predicting $\overline{h}_{GW}$, multivariate DA and univariate assimilation of groundwater work similarly well, which leads to the same conclusion as for $\overline{\theta}_{10cm}$.

Figure 10 gives further insights by showing the temporal development of $\overline{\theta}_{2m}$, $\overline{h}_{GW}$, and $\overline{R}$ and allows for a more in-depth analysis of the different cases. The temporal developments of $\overline{\theta}_{2m}$ for VR1 and 4 are especially interesting. The developments resulting from the multivariate DA (*SMGW*) are no mere superpositions of the developments from the univariate assimilation of soil moisture and groundwater tables. For VR1, it can be seen that, while the

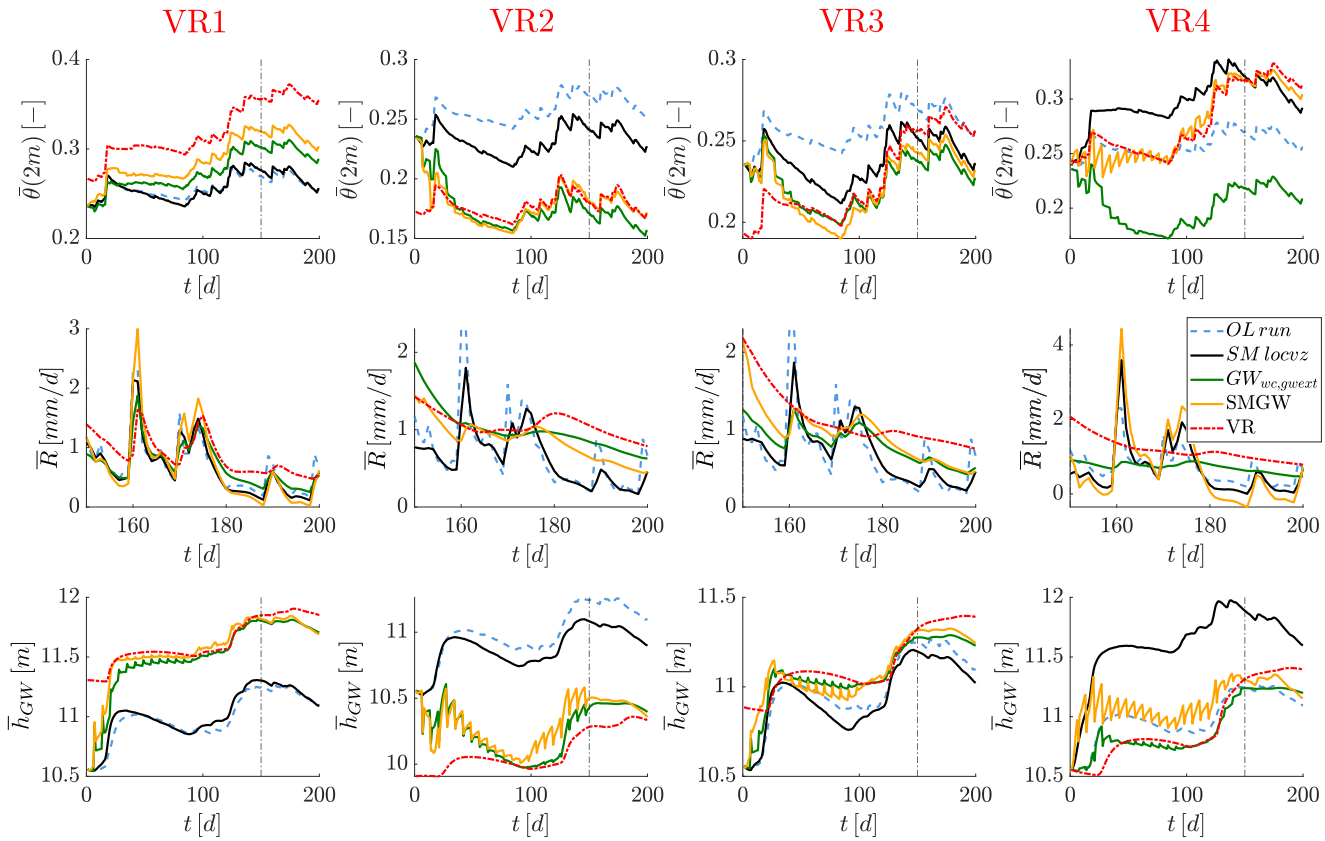

Figure 10. Spatially averaged root zone soil moisture $\overline{\theta}_{2m}$, groundwater recharge R (during the open-loop phase subsequent to data assimilation) and groundwater table height $\overline{h}_{GW}$. The average height of the land surface is 13 m.

DA variant *SM* does not improve predictions compared to the OL run, the variant *SMGW* performs better than the variant $GW_{wc,gwext}$, which means the soil moisture observations are only beneficial in combination with groundwater table observations for estimating root zone soil moisture. This synergy effect is presumably caused by the groundwater table assimilation putting the ensemble in a better position to utilize the soil moisture observations in the experiments with VR1. For VR4, predictions of $\overline{\theta}_{2m}$ are notably deteriorated in the experiments with the univariate DA variants *SM* and $GW_{wc,gwext}$ and notably improved by the multivariate DA variant *SMGW*. Still, the overestimation of $\overline{h}_{GW}$ by univariate soil moisture assimilation is also reflected in the multivariate DA experiment, which yields worse estimates of $\overline{h}_{GW}$ than univariate groundwater table DA for VR4. Basically, Figure 10 shows a high variability between the VRs, but a general benefit from the multivariate assimilation *SMGW* can be concluded, which contributes toward exploiting cross-compartmental interactions to improve predictions.

Forecasts of groundwater recharge $\overline{R}$ are mostly tied to prior updating the groundwater table. Univariate soil moisture assimilation generally does not improve the characterization of groundwater recharge, and multivariate DA also generally performs worse than univariate groundwater table assimilation in predicting groundwater recharge. This is particularly noticeable for multivariate DA with VR4. While predictions of $\overline{\theta}_{2m}$ are improved a lot and of $\overline{h}_{GW}$ a bit, the $\overline{R}$ estimate is poor. We generally found that groundwater recharge estimates are closely tied to the soil hydraulic parameters.

Figure 11 gives further insights into the poor groundwater recharge estimate by *SMGW* for VR4 by showing the spatially (and ensemble) averaged parameters over time for the DA experiments. For VR4 (last row), it can be seen that estimates of the parameter α deviate strongly from the averaged value of the VR in the multivariate DA experiment (*SMGW*) and the true value is severely underestimated at the end of the DA time period. In other cases (exemplified in Figure 11 for the top layer of VR1 and the middle layer of VR2), the multivariate DA experiments

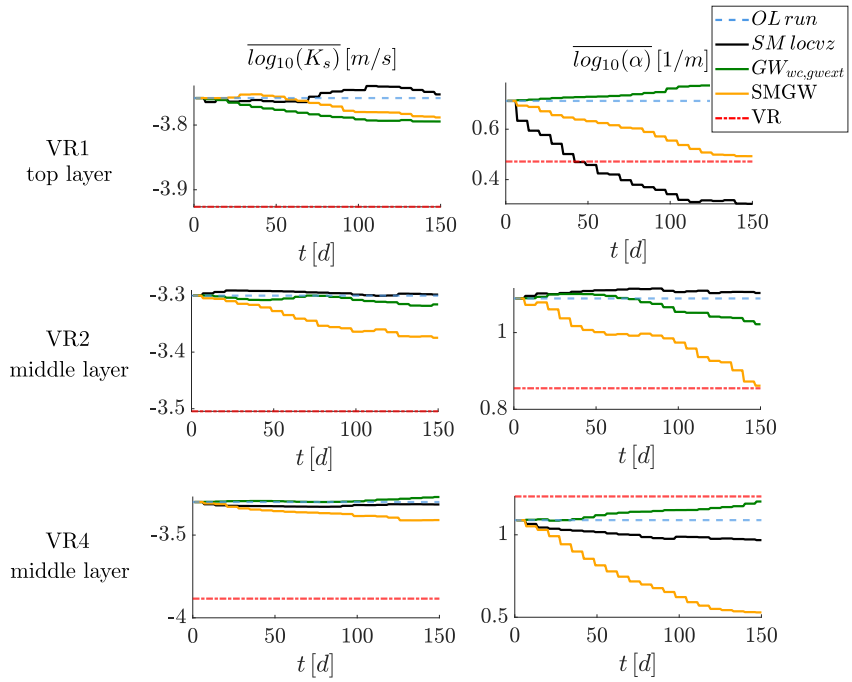

Figure 11. Parameter updates for the data assimilation (DA) experiments *SM loc* (shown in black), $GW_{wc,gwext}$ (shown in green), and *SMGW* (shown in yellow) with the virtual realities (VRs) 1, 2, and 4. Shown are spatial averages in the top and middle layers. The blue line indicates the open-loop run (where the parameters of all DA experiments start) and the red line indicates the VR (the true spatially averaged parameter values).

appear to be particularly suitable for improving estimates of the soil hydraulic parameters. In general, we find that DA experiments can improve the soil hydraulic parameters but sometimes also lead to a different set of parameters that can reproduce the states well but not the fluxes. Such deteriorations in the parameters during DA are found difficult to anticipate.

5. Conclusions

Based on the current state of the art of DA with integrated subsurface flow models (modeling both groundwater and the vadose zone), there is a clear incentive to get a better understanding of how the representation of both compartments in one DA system can be beneficial. We tackled this fundamental problem in two regards.

First, we dealt with deteriorations of predictions from assimilating a vadose zone variable on the groundwater characterization. Second, we investigated capabilities to improve vadose zone and groundwater characterization with both soil moisture and groundwater level measurements. To investigate how model errors influence the results, the ensembles for the forward model use homogenized parameters, whereas the VR from which the observations are taken is heterogeneous. The analyzed variables were the spatial average of the soil moisture near the land surface (linked to evaporation), the root zone soil moisture (linked to evapotranspiration), the groundwater table height, and groundwater recharge. We have assimilated soil moisture (point measurements), groundwater table height, and both combined (multivariate DA).

We found that DA with homogeneous models is well able to predict spatially averaged estimates of land surface soil moisture for heterogeneous realities, without the need for bias correction. The bias due to homogeneous models for a heterogeneous VR only had a notable impact on the soil moisture DA when extrapolating (forecasting far away from the observations). The prediction of the groundwater table height was often notably deteriorated if soil moisture was assimilated. These deteriorations were stronger for bigger model errors (in our case a more homogeneous model). Using bias correction and vertical localization, deteriorations across the subsurface compartments could be notably mitigated, which is one core finding of this work. For future studies, this implies that such structural errors should be addressed if soil moisture DA is conducted and groundwater table heights are of interest.

We observed that weak coupling of groundwater table updates impairs potential improvements of groundwater table height predictions due to balancing flows between the (not updated) vadose zone and groundwater. This is in line with the slower convergence of weakly coupled groundwater updates that H. Zhang et al. (2018) observed. The balancing flows are an artifact of DA, evoked by the updates. We have found that these flows can be dampened by updating a section above the groundwater table along with the fully saturated parts, which resulted in a notably improved forecast of the groundwater table height without any additional measurement information being necessary. We expect that this finding is transferable to more complex models. In terms of complexity, our model primarily lacks processes near the land surface, and we do not expect these to play a big role with respect to this finding. The idea behind this update variant is quite simple, which makes it easy to implement and probably appealing for future studies.

For predictions of the land surface soil moisture and groundwater table height, multivariate DA performed similarly well as univariate DA of the respective variable. While this means that multivariate DA has no negative impact on estimating these variables, it also means that there is no incentive to use soil moisture measurements if only groundwater tables are of interest and no incentive to use groundwater table measurements if land surface soil moisture is of interest. This coincides with the findings of many authors (Botto et al., 2018; Hung et al., 2022; D. Zhang et al., 2016). As land surface soil moisture and groundwater table height are spatially distant variables, it is consistent that the influence of one on the other is weak. For quantities of interest located between these two boundaries, this is different.

Multivariate DA was, in our study, the best choice to improve the root zone soil moisture characterization. This was consistent with all virtual realities we examined. It agrees with the findings by H. Zhang et al. (2018). It is a bit in contrast to the findings of Hung et al. (2022), who found that univariate assimilation works better for improving soil moisture predictions at validation locations than multivariate. However, Hung et al. (2022) examined point estimates of soil moisture and groundwater tables in their catchment are not everywhere as shallow as in our domain. For deep groundwater tables, meaningful correlations to the soil moisture would not be expected, and if groundwater assimilation does not deteriorate soil moisture characterization, this is already a good result. At locations where groundwater tables are shallow, we see big potential for multivariate DA with respect to estimating root zone soil moisture. Multivariate DA can thus yield improved characterization of subsurface states compared to univariate DA in the depths between the measurements and usually performs similarly well as univariate DA at the measurement locations.

For predictions of the root zone soil moisture, we found that information about the layering can improve predictions substantially. When assimilating groundwater table observations, we found that predictions of the root zone soil moisture improved stronger (less) if the DA ensemble was wetter (drier) than the VR, which means that the groundwater table was lowered (increased) by DA. It could thus be preferable to generate an ensemble that slightly overestimates the true groundwater table heights instead of one that slightly underestimates them.

For improvements in forecasting groundwater recharge, we found groundwater table updates to be most important. The update variant did not play a big role. Two limitations should be kept in mind with respect to forecasting groundwater recharge in the context of subsurface flow DA. First, groundwater recharge may represent artificial fluxes induced by local soil moisture and groundwater table changes caused by DA, which leads to poor forecasts. Second, estimates of groundwater recharge were sometimes deteriorated even after DA was turned off if soil hydraulic parameters differed from the true values. Soil hydraulic parameters are ambiguous, meaning that one state could be represented by different sets of parameters, and as Brandhorst et al. (2017) pointed out, it is usually not possible to retrieve the true parameters from subsurface flow DA experiments given a realistic measurement density. Still, in the majority of cases, groundwater recharge predictions could be improved by applying the parameters estimated by DA in forward runs without DA.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

Simulation results (raw outputs and postprocessed data) are archived in the Zenodo repository (Waldowski, 2022a, 2022b, 2022c). The forward model is available at http://doi.org/10.5281/zenodo.4422532.

Acknowledgments
This research was performed in the Research Unit FOR2131 "Data Assimilation for Improved Characterization of Fluxes across Compartmental Interfaces," funded by Deutsche Forschungsgemeinschaft (Grants: NE 824/12-2 and HE 6239/4-2). It was supported by the LUH compute cluster, which is funded by the Leibniz University Hannover, Germany. The authors sincerely thank Natascha Brandhorst for providing the numerical model and DA framework. Open Access funding enabled and organized by Projekt DEAL.

References

Beegum, S., Šimnek, J., Szymkiewicz, A., Sudheer, K., & Nambi, I. M. (2018). Updating the coupling algorithm between hydrus and modflow in the hydrus package for modflow. *Vadose Zone Journal*, *17*(1), 1–8. https://doi.org/10.2136/vzj2018.02.0034

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

Brandhorst, N., Erdal, D., & Neuweiler, I. (2017). Soil moisture prediction with the ensemble Kalman filter: Handling uncertainty of soil hydraulic parameters. *Advances in Water Resources*, *110*, 360–370. https://doi.org/10.1016/j.advwatres.2017.10.022

Brandhorst, N., Erdal, D., & Neuweiler, I. (2021). Coupling saturated and unsaturated flow: Comparing the iterative and the non-iterative approach. *Hydrology and Earth System Sciences*, *25*(7), 4041–4059. https://doi.org/10.5194/hess-25-4041-2021

Brandhorst, N., & Neuweiler, I. (2023). Impact of parameter updates on soil moisture assimilation in a 3D heterogeneous hillslope model. *Hydrology and Earth System Sciences*, *27*(6), 1301–1323. https://doi.org/10.5194/hess-27-1301-2023

Brocca, L., Ponziani, F., Moramarco, T., Melone, F., Berni, N., & Wagner, W. (2012). Improving landslide forecasting using ASCAT-derived soil moisture data: A case study of the Torgiovannetto landslide in central Italy. *Remote Sensing*, *4*(5), 1232–1244. https://doi.org/10.3390/rs4051232

Cosby, B., Hornberger, G., Clapp, R., & Ginn, T. (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

De Rosnay, P., Drusch, M., Vasiljevic, D., Balsamo, G., Albergel, C., & Isaksen, L. (2013). A simplified extended Kalman filter for the global operational soil moisture analysis at ECMWF. *Quarterly Journal of the Royal Meteorological Society*, *139*(674), 1199–1213. https://doi.org/10.1002/qj.2023

Enenkel, M., Steiner, C., Mistelbauer, T., Dorigo, W., Wagner, W., See, L., et al. (2016). A combined satellite-derived drought indicator to support humanitarian aid organizations. *Remote Sensing*, *8*(4), 340. https://doi.org/10.3390/rs8040340

Erdal, D., Neuweiler, I., & Wollschläger, U. (2014). Using a bias aware EnKf to account for unresolved structure in an unsaturated zone model. *Water Resources Research*, *50*(1), 132–147. https://doi.org/10.1002/2012wr013443

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

Evensen, G., Vossepoel, F. C., & Van Leeuwen, P. J. (2022). *Data assimilation fundamentals: A unified formulation of the state and parameter estimation problem*. Springer Nature.

Gebler, S., Kurtz, W., Pauwels, V., 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

Gómez-Hernández, J. J., & Journel, A. G. (1993). Joint sequential simulation of multigaussian fields. In *Geostatistics tróia'92*. (Vol. 1, pp. 85–94). Springer. https://doi.org/10.1007/978-94-011-1739-5_8

Haese, B., Hörning, S., Chwala, C., Bárdossy, A., Schalge, B., & Kunstmann, H. (2017). Stochastic reconstruction and interpolation of precipitation fields using combined information of commercial microwave links and rain gauges. *Water Resources Research*, *53*(12), 10740–10756. https://doi.org/10.1002/2017wr021015

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). 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), 123–137. https://doi.org/10.1175/1520-0493(2001)129<0123:asekff>2.0.co;2

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*(6), e2021WR031549. https://doi.org/10.1029/2021wr031549

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). https://doi.org/10.1029/2007wr006004

Komma, J., Blöschl, G., & Reszler, C. (2008). Soil moisture updating by ensemble Kalman filtering in real-time flood forecasting. *Journal of Hydrology*, *357*(3–4), 228–242. https://doi.org/10.1016/j.jhydrol.2008.05.020

Madsen, H., & Canizares, R. (1999). Comparison of extended and ensemble Kalman filters for data assimilation in coastal area modelling. *International Journal for Numerical Methods in Fluids*, *31*(6), 961–981. https://doi.org/10.1002/(sici)1097-0363(19991130)31:6<961::aid-fld907>3.0.co;2-0

Miller, E., & Miller, R. (1956). Physical theory for capillary flow phenomena. *Journal of Applied Physics*, *27*(4), 324–332. https://doi.org/10.1063/1.1722370

Richards, L. A. (1931). Capillary conduction of liquids through porous mediums. *Physics*, *1*(5), 318–333. https://doi.org/10.1063/1.1745010

Shi, L., Song, X., Tong, J., Zhu, Y., & Zhang, Q. (2015). Impacts of different types of measurements on estimating unsaturated flow parameters. *Journal of Hydrology*, *524*, 549–561. https://doi.org/10.1016/j.jhydrol.2015.01.078

Simunek, J., Van Genuchten, M. T., & Sejna, M. (2005). *The hydrus-1D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media* (Vol. 3, pp. 1–240). University of California-Riverside Research Reports.

Sohrabi, M. M., Ryu, J. H., Abatzoglou, J., & Tracy, J. (2015). Development of soil moisture drought index to characterize droughts. *Journal of Hydrologic Engineering*, *20*(11), 04015025. https://doi.org/10.1061/(asce)he.1943-5584.0001213

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

van den Hurk, B., Doblas-Reyes, F., Balsamo, G., Koster, R. D., Seneviratne, S. I., & Camargo, H. (2012). Soil moisture effects on seasonal temperature and precipitation forecast scores in Europe. *Climate Dynamics*, *38*(1–2), 349–362. https://doi.org/10.1007/s00382-010-0956-2

van Genuchten, 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

Waldowski, B. (2022a). Data assimilation in integrated subsurface flow models (1/3). https://doi.org/10.5281/zenodo.14637081

Waldowski, B. (2022b). Data assimilation in integrated subsurface flow models (2/3). https://doi.org/10.5281/zenodo.14638715
Waldowski, B. (2022c). Data assimilation in integrated subsurface flow models (3/3). https://doi.org/10.5281/zenodo.14639161
Wanders, N., Karssenberg, D., De Roo, A., De Jong, S., & Bierkens, M. (2014). The suitability of remotely sensed soil moisture for improving operational flood forecasting. *Hydrology and Earth System Sciences*, *18*(6), 2343–2357. https://doi.org/10.5194/hess-18-2343-2014
Wicki, A., Lehmann, P., Hauck, C., Seneviratne, S. I., Waldner, P., & Stähli, M. (2020). Assessing the potential of soil moisture measurements for regional landslide early warning. *Landslides*, *17*(8), 1881–1896. https://doi.org/10.1007/s10346-020-01400-y
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, 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