Using a bias aware EnKF to account for unresolved structure in an unsaturated zone model

D. Erdal,[1] I. Neuweiler,[1] and U. Wollschläger[2]

[1] When predicting flow in the unsaturated zone, any method for modeling the flow will have to define how, and to what level, the subsurface structure is resolved. In this paper, we use the Ensemble Kalman Filter to assimilate local soil water content observations from both a synthetic layered lysimeter and a real field experiment in layered soil in an unsaturated water flow model. We investigate the use of colored noise bias corrections to account for unresolved subsurface layering in a homogeneous model and compare this approach with a fully resolved model. In both models, we use a simplified model parameterization in the Ensemble Kalman Filter. The results show that the use of bias corrections can increase the predictive capability of a simplified homogeneous flow model if the bias corrections are applied to the model states. If correct knowledge of the layering structure is available, the fully resolved model performs best. However, if no, or erroneous, layering is used in the model, the use of a homogeneous model with bias corrections can be the better choice for modeling the behavior of the system.

1. Introduction

[2] The unsaturated zone, with its placement between the land surface and the groundwater table, plays an important role in the terrestrial water cycle. It is an important part of any simulation of surface-land-atmosphere systems. It is a complex heterogeneous system that is poorly accessible for observations. A crucial point in modeling processes in the unsaturated zone is therefore to make use of as many observations of the system as possible. It has been shown by many [e.g., *Papafotiou et al.*, 2008; *Mertens et al.*, 2005; *Ines and Mohanty*, 2008b; *Kumar et al.*, 2010] that model parameters for one spatial scale may not represent the system at another scale. Hence, it is important to obtain parameters for a model at the scale it is to be used at. However, the available observations have often observation volumes that do not match the scale of the model. When, for example, using observations from time domain reflectometry (TDR) probes, one might face a situation where parameters for a model with grid cells that exceed the typical volume of the observations that are used for calibration and validation have to be found.

[3] The usual approach in subsurface modeling is to calibrate a model using observations from one or several measurement campaigns, where information is only available for a restricted time period (a review of methods is given, for example, by *Vrugt et al.* [2008]). If boundary conditions do not change significantly compared to the calibration conditions, the calibrated model can be used for predictions, also for time spans that do not immediately follow the measurement period. When continuous observations are available and the model is used to make predictions for limited time periods, the situation is different. Here, an estimation of states as initial condition for a prediction can be used. Methods for such situations that are gaining increasing attention also in unsaturated zone modeling are the data assimilation methods [see e.g., *Reichle*, 2008]. Of these, the Ensemble Kalman Filter (EnKF) is a commonly used sequential method. In sequential data assimilation, the states that control the model at a given moment, and possibly the model parameters, are updated sequentially over time at time steps where observations are available, instead of just updating parameters once for all times. These methods are suited for situations, where continuous observation series are available and the model is used to make predictions for limited time periods. If parameter updates are carried out, data assimilation can be considered also a sequential calibration method. Also in data assimilation with and without parameter updates, the problem of mismatch between length scales of model and observation might occur.

[4] The Ensemble Kalman Filter (in the following denoted as EnKF) was presented by *Evensen* [1994] as an ensemble estimation of the original Kalman filter and tested for an ocean model. Since then, the EnKF has been used for updating states, or states and parameters, in reservoir modeling [e.g., *Oliver et al.*, 2010; *Vogt et al.*, 2012], atmospheric modeling [e.g., *Houtekamer and Mitchell*, 1998], groundwater modeling [e.g., *Hendricks Franssen*

[1]Institute of Fluid Mechanics and Environmental Physics in Civil Engineering, Leibniz University Hannover, Hannover, Germany.

[2]UFZ—Helmholtz Centre for Environmental Research, Department Monitoring and Exploration Technologies, Leipzig, Germany.

Corresponding author: D. Erdal, Institute of Fluid Mechanics and Environmental Physics in Civil Engineering, Leibniz University Hannover, DE-30167 Hannover, Germany. (erdal@hydromech.uni-hannover.de)

and Kinzelbach, 2008; *Kollat et al.*, 2011; *Hendricks Franssen et al.*, 2011], and in coupled modeling of surface-subsurface systems [e.g., *Kumar et al.*, 2008; *Pasetto et al.*, 2012]. The EnKF has also been used to assimilate soil moisture observations, both from satellite or other near-surface observations [e.g., *Walker et al.*, 2001; *Reichle et al.*, 2002; *Crow and Van Loon*, 2006; *Crow and van den Berg*, 2010] and using local soil water content observations at distinct depths [e.g., *De Lannoy et al.*, 2007a; *De Lannoy et al.*, 2007b; *Monsivais-Huertero et al.*, 2010; *Han et al.*, 2012]. The joint estimation of states and parameters by an EnKF for problems in the unsaturated zone has also been given an increasing attention lately [*Monsivais-Huertero et al.*, 2010; *Li and Ren*, 2011; *Wu and Margulis*, 2011]. *Monsivais-Huertero et al.* [2010] used a bucket-type model for the water balance in the unsaturated zone to update both soil moisture states and model parameters for both real and synthetic observations, while both *Wu and Margulis* [2011] and *Li and Ren* [2011] used models based on the Richards equation for water flow in the unsaturated zone for the same purpose. *Wu and Margulis* [2011] assimilated soil moisture observations to successfully estimate hydraulic parameters for both a homogeneous and a layered synthetic test case. *Li and Ren* [2011], on the other hand, used pressure head observations to estimate parameters for a wide range of synthetic soils, and showed that the EnKF can be a potentially useful method, but also that certain setups provided erroneous results. Despite using different observations, both *Li and Ren* [2011] and *Wu and Margulis* [2011] used models that have primary variables that are identical to the observed states.

[5] An important decision connected with any type of modeling and the observations taken is to what extent the model should capture details in both the medium's structure and the processes in the system and to what extent they can be simplified. Naturally, the more details of the real system are captured the better it can model reality, but lack of information and use of simplified models will always lead to nonperfect results when modeling a real system. This is known as model structure error, or model structural adequacy, as suggested by *Gupta et al.* [2012] in their recent review of model adequacy. In unsaturated zone modeling, it has been shown that using insufficient spatial structure (e.g., not resolving soil layers) can create models that cannot mimic the true system [e.g., *Erdal et al.*, 2012]. Yet, a model without resolved structure might be adequate to predict the average behavior of the system, such as the total outflow. It might also be that a nonstructured model is required for reasons of computation time or because the structure is unknown and cannot be obtained from observations. In this case, as outlined above, the observations used to control the model might have length scales that are smaller than a volume representative for the mean behavior of the system. For example, a TDR probe might be located in one layer and does not give information about the water content averaged over several layers. If the layered structure is not resolved in the model, this results in a structure error that can be regarded as a consistent mismatch between observations and modeling results.

[6] *Erdal et al.* [2012] suggested to include error models to correct for bias during a batch calibration process for such a scenario. In the EnKF community, bias correction is common (though not aiming at nonresolved medium

structure) and discussed, for example, by *Dee and Da Silva* [1998]. It has been used previously in both groundwater [e.g., *Drécourt et al.*, 2006; *Kollat et al.*, 2011] and ocean modeling [e.g., *Deng et al.*, 2011], but also to correct soil moisture simulations [e.g., *De Lannoy et al.*, 2007a, 2007b]. There are two common ways for treating a bias correction within a filter [*Drécourt et al.*, 2006]. Either the bias is treated with a separated filter, as is for example done by *Friedland* [1969] or *De Lannoy et al.* [2007a], or it is treated as a colored noise within the current filter. The latter idea was introduced by *Madsen and Canizares* [1999] and a comparison of the two methods has been performed by *Drécourt et al.* [2006]. The authors found that both treatments improved a groundwater model when different biases were added.

[7] In this paper, we investigate how to deal with model inadequacy in unsaturated zone data assimilation. We consider in particular model inadequacy due to layered soil structure, which is difficult to recover as only few local observations are available (which is a typical situation in unsaturated flow modeling). For this purpose, we use the concept of a colored noise filter in an EnKF with a homogeneous Richards model to describe water flow in the unsaturated zone with layered structure, where the bias is supposed to correct for the nonresolved soil structure. We update model parameters, model states (here pressure heads) and bias terms. In difference to the previous works mentioned above, we apply the EnKF using a nonlinear relation between model states (pressure heads) and model observations (soil water content), where the shape of the relation function is controlled by the simultaneously estimated parameters. As continuous observations are usually not easily available in the subsurface, the question of predictions for time spans without observations is quite important. This question is strongly related to the parameter estimates in the model. We discuss the performance of different model approaches for predictions during longer time spans without observations and the related parameter updates in these models. To test different modeling approaches, we consider both artificial and real test cases in layered media. We assume that time series of only few observations are available, which have an observation volume that is significantly smaller than the system. In summary, the paper addresses the following questions:

[8] Can a simplified model for flow in the unsaturated zone be improved with bias corrections within the EnKF to predict the local and the average behavior, such as the mass of water in the system, of a more complicated system? The simplification concerns both nonresolved structure and simplified parameterization of the flow model.

[9] 1. Can the resulting bias correction parameter be interpreted in a meaningful way? Can it be used to get information about nonresolved structure?

[10] 2. How severe are uncertainties in the prior information about the medium's structure?

2. Methods

[11] As outlined in section 1, we test an EnKF to make predictions of variables in the unsaturated zone using measurements with an observation volume that is considerably smaller than the system and using a model that simplifies reality. Simplifications are due to a simple parameterization

of the models and nonresolved subsurface layering structure. Two data sets are used in this paper. First, observations from two virtual (synthetic) lysimeters are used to perform the development and testing of the data assimilation methods using bias correction. These data are explained in section 2.4. Here, we predict the total mass of the lysimeter as well as water content at the observation locations, using water content data taken from TDR probes. Second, the method is validated using a real data set from four TDR probes in a layered soil that has previously been used and calibrated with good results by *Wollschläger et al.* [2009]. This data set is explained in section 2.5, and here both observations and predictions are local water content measurements from TDR probes.

2.1. Ensemble Kalman Filter

[12] The Ensemble Kalman Filter [*Evensen*, 1994, 2003] can be described as a sequential data assimilation scheme, in which a state vector is sequentially updated based on a finite ensemble Monte Carlo approximation of the covariance between model states and model observations. An ensemble of N_{ens} state vectors (containing for example the primary variables) is initiated. At time t_n, the state vector x of ensemble member j can be propagated forward from the previous assimilation time t_{n-1}:

$$
\begin{aligned}
x_n^{(j),f} &= f(x_{n-1}^{(j),a}) + q_n^{(j)} \\
y_n^{(j)} &= g(x_n^{(j),f}) + r_n^{(j)}
\end{aligned}
\tag{1}
$$

where f is the model (here the Richards equation, equation (8)), q is the process noise, y is the observation vector, meaning here the vector of modeled observations, g is the observation model (here equation (10)), and r is the observation noise. The indices a and f stand for analysis and forecast, respectively, and equation (1) is therefore known as the forecast step. The analysis step consists of updating the states according to a comparison of modeled and observed variables:

$$
x_n^{(j),a} = x_n^{(j),f} + K_n(y_n^{(obs)} - y_n^{(j)})
\tag{2}
$$

where K is the Kalman gain and $y^{(obs)}$ is the real (true) observation vector, while $y^{(j)}$ is the modeled (simulated) observation vector. The Kalman gain is calculated as:

$$
K = \frac{P_{xy}}{P_{yy} + R}
\tag{3}
$$

where P_{xy} is the estimated error cross-covariance matrix between the states and the observations, P_{yy} is the estimated covariance of the observations, and R is a matrix with the variance of the observation error on the main diagonal. Both covariance matrices are estimated from the ensemble of realizations:

$$
\begin{aligned}
P_{xy} &= \frac{1}{N_{ens}-1} E_x(E_y)^T \\
P_{yy} &= \frac{1}{N_{ens}-1} E_y(E_y)^T
\end{aligned}
\tag{4}
$$

where N_{ens} is the size of the ensemble, $E_x = [x^{(1),f} - \overline{x}^f, \ldots, x^{(N_{ens}),f} - \overline{x}^f]$ is the ensemble error matrix, where $\overline{x}^f$ is the ensemble mean, and $E_y = [y^{(1)} - \overline{y}, \ldots, y^{(N_{ens})} - \overline{y}]$ is the observation error matrix. From the build up of the EnKF, it is obvious that the quality of the estimates depends on the ensemble size.

2.1.1. State and Parameter Estimation

[13] When considering a solution to the Richards equation, also a choice of flow model parameter values must be selected (see later section 2.3 and equations (9) and (10)). Since these parameters are difficult to select without prior information, they are also required to be estimated by the EnKF.

[14] As outlined in section 1, the estimation of parameters is an important point for predictions during time periods without observations. The most straight forward way of updating parameters is to expand the state vector (x) to include both parameters (ϕ) and primary model variables (here pressure heads (h)), so that $x^{(j)} = [\phi^{(j)}, h^{(j)}]^T$. This is commonly known as a joint parameter and state estimation and the approach has, for example, been used successfully by *Hendricks Franssen and Kinzelbach* [2008], *Monsivais-Huertero et al.* [2010], and *Li and Ren* [2011]. The initial sampling of parameters in this work was done uniformly from a range that is limited to physically reasonable values using the Latin Hypercube method [*McKay et al.*, 1979].

2.1.2. Filter Tuning

[15] To limit the impact of very large state updates, a so-called dampening factor is applied to the Kalman gain. The dampening factor is a number between 0 and 1 that is multiplied with the Kalman gain. In this work, a dampening of 0.3 is applied to the part of the Kalman gain that controls the parameter updates. This is done to ensure a more stable parameter update that does not create jumps across the full parameter space in just one (potentially erroneous) update. This application of the dampening factor is similar to the dampening successfully used by *Hendricks Franssen and Kinzelbach* [2008].

2.2. Bias Correction

[16] It has been shown many times that it is problematic to fit a homogeneous model to a heterogeneous reality [e.g., *Durner et al.*, 2007; *Ines and Mohanty*, 2008a]. Different approaches on how to deal with this problem have been suggested in the literature. One approach to deal with missing spatial structure is to estimate a fully layered model from the given data [e.g., *Shin et al.*, 2012], as done in our first and second models explained in section 2.4.3. Another approach is to introduce a bias term to a homogeneous model. In this paper, we use the concept of the colored noise filter [*Madsen and Canizares*, 1999], in which a bias term is applied in the filter and estimated as part of the state vector. This creates a state vector consisting of parameters (ϕ), pressure heads (h), and bias terms (b) that are updated in the EnKF.

$$
x = [\phi, h, b]^T
\tag{5}
$$

[17] When implementing the bias corrections for a solution to the Richards equation taking into account a nonlinear relation between the states and the observations, there are three ways one can follow. The first way is to add the bias directly to the observations and perform no bias

correction on the states. The second is to apply the bias directly to the states and the third is to apply the bias to the observations with a feedback to the state. The first way is not expected to work, since the bias that acts only on the observations has no direct contact with the states, and hence there is no correlation between the bias corrections and the states. Therefore, this bias correction would only have an impact on the system at the few observation locations [*De Lannoy et al.*, 2007b] and hardly on the whole system. In this work, the second way is used. The application of a bias correction acting directly on the states alters the original forecast equation (equation (1)) into:

$$
\begin{aligned}
b_n &= b_{n-1} + w_n \\
h_n &= f([\phi_{n-1}, h_{n-1}]^T) + b_n + q_n \\
y_n &= g([\phi_{n-1}, h_n]^T) + r_n
\end{aligned}
\tag{6}
$$

where w_n is the bias noise. This means that the number of bias parameters needed is the number of states in the model. For the purpose of better understanding the bias corrections, the third of the error models is also presented here:

$$
\begin{aligned}
h_{int} &= f([\phi_{n-1}, h_{n-1}]^T) + q_n \\
b_n &= b_{n-1} + w_n \\
y_{int} &= g([\phi_{n-1}, h_{int}]^T) + b_n \\
h_n &= g^{-1}(y_{int}) \\
y_n &= y_{int}(z_{obs}) + r_n
\end{aligned}
\tag{7}
$$

where y_{int} is the intermediate observation vector of the same size as the pressure head vector, g^{-1} is the inverse of the observation model (here equation (10)), and z_{obs} is the index of the observation nodes. Although the last method (equation (7)) is not followed in this paper, it has been tested and generally leads to similar results as the first method (equation (6)).

[18] The physical interpretation of the bias corrections is that it can remove or add water in each cell of the numerical model as a means to compensate for the unresolved structure. Hence, the bias corrections can be seen as a, potentially time dependent, sink or source term. However, in difference to the numerical model of the Richards equation, no conservation of mass is required from the bias corrections, meaning that the total mass of the system can be altered when the bias corrections are applied. This potential breaking of the mass conservation of the numerical model is, however, not unique for the bias corrections. It is, in fact, an inherent feature of EnKF state updates, where no mass is guaranteed to be conserved, even if the bias correction is not used. If the bias corrections are considered as source/sink terms, it is logical to apply them also during predictions, as the problem of the unresolved structure is as present during the predictions (no observations) as during the filtering (with observations). Here, an addition/subtraction of mass resulting from the bias corrections could lead to systematic overprediction/underprediction of the total mass if the periods between the observations are long and the bias corrections are poorly defined. This problem is further discussed in section 4.

[19] The noise of the system is controlled by the three noise variables known as the process noise q, the observation error r, and the bias noise w. As is commonly done, the observation error is assumed to be Gaussian white noise [*Doherty and Welter*, 2010], in all cases used here with a standard deviation of 10^{-2}, since this corresponds to the noise added to the perfect virtual reality measurements. Also the process noise, that is supposed to cover for errors in the model setup, such as in the cases shown here, when the model parameterization is simplified, is assumed to be a Gaussian white noise. As described above, the bias can be viewed as a colored noise, in which b is the nonzero mean and w is a random white noise. From the derivation of the EnKF, it is clear that the filter cannot update any states that have no correlation with the observations. Since the bias corrections are added after the simulation model (f in equation (6)) is run, the bias corrections can be split into two categories. The bias corrections that are on the same location as the observations will be strongly correlated with the observations, whilst the correlation between all other bias corrections and the observations would be extremely weak, for practical use considered as nonexisting. This makes it impossible for the filter to update any other bias terms than the ones located exactly at an observation location [*De Lannoy et al.*, 2007b]. This creates a problem, as no bias correction would occur in most of the medium and only additional noise (the bias noise) would be added to the states. The problem is circumvented here by imposing a strong spatial correlation when generating the bias noise (w).

[20] The setup of the correlation assumes a Gaussian spatial correlation function with a correlation length of 75% of the length of the medium. The correlation is imposed during the generation of the noise, such that the noise on each of the bias correction locations correlates with the noise at the other bias correction locations. Hence, when calculating the correlations (equation (4)), any bias correction would have a correlation with one of the four bias corrections that are at a measurement location and, subsequently, any bias correction correlates with an observation. The use of a strong spatial correlation, such as the one used here, has been shown by *Drécourt et al.* [2006] to be advantageous when dealing with spatial errors. A different implementation of the correlation structure, in which a layering structure is considered, is also presented in Appendix A.

2.3. Flow Model

[21] Flow in the unsaturated zone is modeled with the Richards equation,

$$
\frac{\partial \Theta(h)}{\partial t} - \nabla \cdot (K_u(h)(\nabla h + \mathbf{1})) = 0
\tag{8}
$$

where Θ (m³/m³) is the water content, K_u (m/s) is the unsaturated hydraulic conductivity, h (m) is the water pressure head (negative for unsaturated conditions), and $\mathbf{1}$ is the unit vector in the z-direction positive upward. The relations between the water content and the pressure head and between the unsaturated hydraulic conductivity have to be modeled explicitly. In this paper, we consider two formulations. In the equations below, all positive pressure heads

are implicitly assumed to result in full water saturation. The first parameterization model is the Mualem-van Genuchten formulation [*van Genuchten*, 1980; *Mualem*, 1976]

$$S_e = \frac{1}{[1+(\alpha_{VG}|h|)^n]^m}$$

$$K_u = K_s[1-(1-S_e^{1/m})^m]^2 S_e^{0.5}$$

(9)

where S_e (–) is the saturation, which relates to water content as $\Theta = n_f S_e$, n_f being the porosity of the soil, α_{VG} (m^{-1}), n (–), and m (–) are the parameters, with $m = 1-1/n$, and K_s (m/s) is the saturated hydraulic conductivity. Residual saturations are here neglected for simplicity.

[22] The second parameterization model is the so-called Gardner-Russo model [*Gardner*, 1958; *Russo*, 1988],

$$S_e = [e^{-0.5\alpha_{GR}|h|}(1+0.5\alpha_{GR}|h|)]^{2/2.5}$$

$$K_u = K_s e^{\alpha_{GR}|h|}$$

(10)

where α_{GR} (m^{-1}) is a shape parameter. As can be seen from the smaller number of parameters in the Gardner-Russo model, this function is a strong simplification of the more complex van Genuchten function. To highlight the difference between the two parameterization models, an example is shown in Figure 1. More detailed comparisons between the two parameterization models can be found for example in *Zhu and Mohanty* [2003] and *Zhu et al*. [2007].

[23] It is, here, in relation to the filtering problem, important to point out that since any positive pressure head in the Richards equation represents a fully saturated soil ($S_e = 1$), there can be a nonunique relation between pressure and saturation. This is, for example, the case when an infiltration is applied that is larger than the saturated hydraulic conductivity of the soil.

[24] The flow equation (8) is solved using a finite volume approach on a nonuniform grid. An implicit Euler scheme is used for the time discretization and the nonlinearities are treated with a Newton-Raphson scheme with line search.

2.4. Virtual Reality Data

[25] To explore the capabilities of different EnKF models in unsaturated zone modeling, two virtual reality (VR) cases were set up. The conceptualization of the VR models are shown in Figure 2 and the total mass (water weight plus 3000 kg soil matrix) of the VR lysimeters are shown in
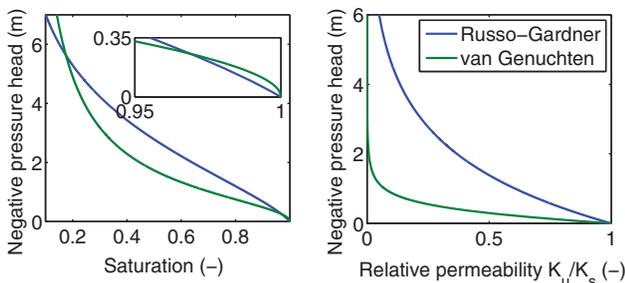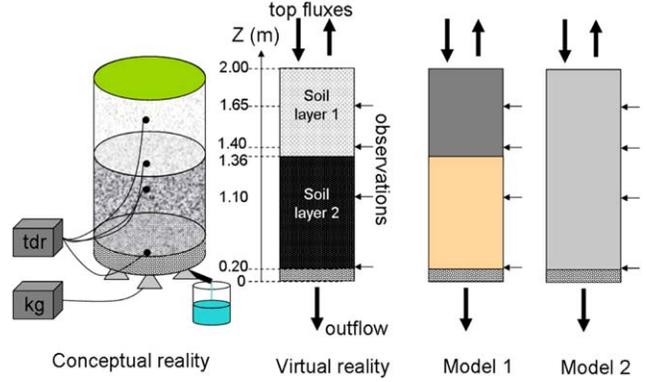
Figure 2. Visualization of the conceptual lysimeter, the virtual reality, and the two models used in the filter problems.

Figure 3. The VR are set up in accordance to real lysimeters with layering structure [e.g., *Meissner et al*., 2010; *Fank and von Unold*, 2005]. We used here the lysimeter placed in Wagna (Austria) that is described in *Fank and von Unold* [2005]. The TDR observations from this lysimeter were, however, not used, as evapotranspiration fluxes on the vegetated surface were not controlled and the numerical modeling of the bottom boundary condition was unclear. Instead, we use the real precipitation data in combination with a simplified evapotranspiration flux and synthetic observations for the lysimeter masses as described below.

[26] The lysimeters have a height of 2 m and both virtual reality data sets (referenced as VR-A and VR-B) are set up to mimic the real lysimeter with two soil layers and a coarse gravel layer at the bottom. The first virtual reality, which is most similar to the real lysimeter, consists of two different sands. The second reality, created to have more contrasts between the water content in the two soil layers, consists of a sand and a loam. All simulation parameters for the VR data sets use the van Genuchten parameterization (equation (9)) and the values can be found in Table 1.

[27] The VR models are calculated with the numerical model explained above, using a grid with 245 cells with a spatial discretization varying from 0.1 (top) to 1 (bottom) cm. The VR models are run for 1 year and the available data are divided into three periods, shown in Figure 3. Here, the EnKF is used from day 10 to day 280 (filter period) and the time from day 281 to day 363 is used for the prediction (where the filter is turned off). The first 10 days are the burn-in period.

2.4.1. Boundary Conditions and Forcings

[28] The upper boundary condition is simulated as a flux using a real time series of daily precipitation from the Wagna data set (see Figure 3) combined with a fixed (low) evaporation rate (10^{-10} m/s). The bottom is simulated using an open boundary. This particular bottom boundary condition is implemented in the numerical flow model in such a way that the flow from the coarse gravel layer is either set to gravity outflow, if the bottom cell is fully saturated, or otherwise is switched off (no flow).

2.4.2. Observations

[29] Observations of water content, the same information as would be available from TDR probes in a real data set,

Figure 1. Example of the two parameterization models shown in equations (9) and (10). The example uses the parameter values $\alpha_{VG} = \alpha_{GR} = 1$ m^{-1} and $n = 2$.
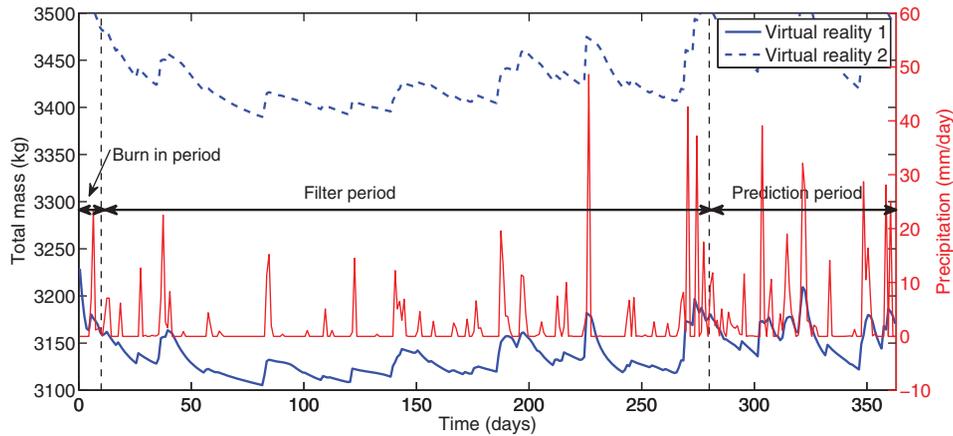
Figure 3. Total mass over time of the two virtual reality simulations together with the precipitation used as top flow boundary.

are taken at four depths (see Figure 2) once per day. The observation volume of the TDR probes is 1 cm. A measurement error consisting of white noise with the standard deviation of 0.01 is added to all water content measurements to mimic a measurement device error. Since we want to predict the average behavior of the lysimeter, the total weight (Figure 3) is the validation observation of primary interest. To avoid problems arising from a simplified parameterization of the bottom gravel layer, which is here only considered as a boundary with fixed parameters, all masses are calculated ignoring any water present in this layer.

2.4.3. Simulation Models

[30] For the EnKF data assimilation and prediction simulations, we compare two different flow model setups. The first setup is a two-layered model, like the true model (Figure 2), but with a simplified parameterization (equation (10) instead of (9)). This model is close to the reality it is supposed to predict. It is assumed that the structure of the soil, meaning the depths of the layer boundaries, is known exactly and only the parameterization is unclear. The parameterization chosen for data assimilation does not capture all features of the virtual reality model. Note that if we had chosen the same parameterization for the data assimilation model, the predictions should meet the virtual observations and parameters exactly and it is only a question of ensemble size of the EnKF and sensitivity of the observations on the parameters to get there. The different parameterization models create a true model error that is clearly different from Gaussian errors.

[31] As the knowledge of structure in reality is very difficult to obtain, we use a second model for data assimilation and prediction, which is identical to the first model

described above. Only the interface is wrongly assumed to be at 0.45 m below the soil surface whilst in the virtual reality the interface is in fact at 0.65 m below soil surface.

[32] The third simulation model setup consists of only one homogeneous layer and also uses the simplified parameterization. This model is an example of the case described in section 1, where the detailed structure of the soil should or could not be included in the simulation model. Such models are used, for example, to predict the average behavior of the system. This model has a severe structural model error and naturally one would expect that this is problematic, since observations are taken only at four locations. The ratio between observations and model states (pressure heads) is hence very low. This, however, is not unlikely for such a system and the full observation setup is identical to the one used for the real lysimeter.

[33] To mitigate this error, we explore the use of the bias correction in the fourth simulation model. The model is set up as the third model, but a bias correction is added that is meant to correct for the mismatch due to the nonresolved structure.

[34] All simulation models are, apart from the parameterization and the internal structure, set up identically to the VR simulations. This means that, for example, the model forcings are perfect in the simulation models. For the parameter estimation, all three parameters in the simulation models are estimated: saturated hydraulic conductivity (K_{sat}), α_{GR}, and the porosity of one or two materials, respectively.

2.5. Real Data

[35] For the real data simulations, we use a 1 year time series of TDR measurements obtained from a layered field soil at the Grenzhof Test Site, SW Germany. The site and data are described in detail in *Wollschläger et al.* [2009]. Hence, only a brief summary is provided here.

[36] The soil profile at the Grenzhof Test Site consists of four layers (0–0.28, 0.28–0.82, 0.82–1.10, and 1.10–1.54 m below surface) which all can be classified as sandy loam [*Soil Survey Division Staff*, 1993]. They are underlain by fluvial gravels and sands starting at 1.54 m depth. The soil surface at the top of the profile was covered with grass which was regularly cut to a height of a few centimeters.

Table 1. Setup of the Virtual Realities

	K_{sat} (m/s)	α (1/m)	n (–)	Θ_{sat} (m³/m³)
VR-A				
Top	1.73×10^{-5}	10	2.1	0.36
Bottom	5.8×10^{-4}	36	1.7	0.22
VR-B				
Top	1.73×10^{-5}	10	2.1	0.36
Bottom	3.6×10^{-6}	3.6	1.5	0.43

2.5.1. Boundary Conditions and Forcings

[37] The atmospheric upper boundary condition was measured onsite by an automatic weather station. Reference evapotranspiration was calculated using the Food and Agriculture Organization (FAO) Penman-Monteith equation [*Allen et al.*, 1998]. The lower-boundary condition is represented by a hypothetical groundwater table located at 4 m depth which is far below the TDR measurements of interest.

2.5.2. Observations

[38] Volumetric soil water content was measured with TDR at four depths (0.13, 0.63, 0.92, and 1.16 m) where one sensor was placed in each of the uppermost four soil layers. Volumetric water contents at each depth were estimated from the measured dielectric permittivities and temperature corrected according to *Roth et al.* [1990] using soil temperature measurements from a nearby profile.

2.5.3. Simulation Models

[39] The model was used to simulate the daily average soil water contents at the Grenzhof Test Site. It follows the setup described in *Wollschläger et al.* [2009]. The numerical model was set up with a uniform grid consisting of 400 cells of which the top 154 were considered for the assimilation and the bottom 246, which represent the gravel and sand layers and where no TDR measurements were available, were kept without updating. Initial and boundary conditions were the same as described in *Wollschläger et al.* [2009] with the simplification that the evapotranspiration flux was already scaled by the factor of 0.61 [see *Wollschläger et al.*, 2009] and the distributed root water uptake was neglected. Instead, the complete flux across the upper boundary was exchanged via the uppermost cell of the model grid. Instead of resolving the layering, we use the homogeneous model only (third and fourth model setups in section 2.4.3). A calibration for the four-layered model has been carried out by *Wollschläger et al.* [2009].

[40] Two different setups of the homogeneous models with and without bias corrections were tested. The first uses the simpler parameter model of Gardner and Russo (equation (3)) and the second setup uses the van Genuchten model (equation (9)). As was done in *Wollschläger et al.* [2009], we do not estimate the porosity, but set it to the measured values, leaving the two setups with 3 and 2 estimation parameters, respectively.

2.6. EnKF Simulations and Evaluation Setup

[41] For the evaluation of the data assimilation with the different models introduced in section 2.4.3, data assimilation results are compared to the VR and real data observations. Two different measurement types, taken in the prediction period (cf. Figure 3), are used in this paper. First, the total mass of the system is compared. Second, the water content at the point of the measurements is compared. For the real data, since no total mass is available, only the predictions of the water content are used for the evaluation. The performance of the EnKF simulation is quantified using the root mean square error (RMSE) of the ensemble mean of the interested quantity (mass or water content) during the prediction period:

$$RMSE = \sqrt{\frac{1}{N_{pred}} \sum_{n=t_{T-X}}^{t_T} (\overline{Q}_n^{(s)} - Q_n^{(t)})^2} \qquad (11)$$

where $N_{pred} = t_T - t_{T-X}$ is the length of the prediction period (cf. Figure 3), $\overline{Q}_n^{(s)}$ is the ensemble mean of the interested quantity at time n, and $Q_n^{(t)}$ represents the true quantity of the system at time n. There is no unique measure to quantify the performance of the simulations and one could also consider the deviations between modeled and actual mass of the system over the whole time period. We decided to use only the prediction time period, as we see the prediction in periods without observations as an important task for a model. In subsurface systems, the situation that observations are continuously available over very long time periods are rare.

[42] When using any ensemble-based method, a crucial filter parameter to decide on is the size of the ensemble. In principle, the ensemble size should be larger than the effective number of degrees of freedom. In general, the larger the ensemble, the better the approximation of the covariances becomes and the more appropriate the update is. A smaller ensemble, on the other hand, is often needed due to computational restrictions. In this work, an ensemble sizes of 1000 ensemble members and is considered large enough to give a consistent result when the assimilation is repeated. To assure that the ensemble is large enough, all simulations are repeated 10 times, though (since they are similar) only one example is shown in the results. The topic of ensemble size is further discussed in Appendix A.

[43] When using the EnKF to estimate states and parameters for the strongly nonlinear Richards equation, certain problems have been noted by the authors. The combination of low conductivity and an evaporation top boundary condition leads in the numerical flow model to large negative pressure heads. During a parameter estimation in the EnKF, these large negative pressure heads are at risk to be combined with rather small pressure heads. The mean of the ensemble of states will then strongly be influenced by the few strongly negative heads. Since the updates of the EnKF are based around the ensemble mean, this can result in large state updates that create enormous positive (unphysical) pressure heads. This, in turn, cannot be detected as poor prediction by the observations, since any positive pressure head only results in full saturation ($S_e = 1$), independent of its value. This effect may lead to severe problems in the filtering and larger ensemble sizes cannot solve this issue. It is a real problem of the nonlinearity of the model.

[44] The issue is interesting and will require more investigation, but in this paper, the unwanted effect is kept negligible by limiting the initial sampling (please note: only the initial sampling) of the hydraulic conductivity to a minimum of $K_s = 10^{-6}$ m/s and by limiting the negative pressure heads arising from evaporation to $h = -10$ m. Although the limiting value may seem very low, it did not change the general result presented here, but only served to make the result consistent when repeating large ensemble simulations.

[45] As outlined above, a large selection of different data assimilation setups are used in this paper. All tested setups are summarized in Table 2.

Table 2. Setup of Used Simulation Scenarios[a]

	Data	Simulation Layers	Parameter Model	Bias
Case 1	VR-A	1	GR	No
Case 2	VR-A	2	GR	No
Case 3	VR-A	2-incorr.	GR	No
Case 4	VR-A	1	GR	Yes
Case 5	VR-B	1	GR	No
Case 6	VR-B	2	GR	No
Case 7	VR-B	1	GR	Yes
Case 8	RD	1	GR	No
Case 9	RD	1	GR	Yes
Case 10	RD	1	VG	Yes

[a]RD is the real data, simulation layers signifies the number of soil layers in the numerical flow model, 2-incorr means the two-layered simulation model with the incorrectly placed interface, GR is the Gardner-Russo parameter model, and VG is the van Genuchten parameter model.

3. Results

3.1. One-Layered and Two-Layered Flow Models

[46] The first assimilations are performed using a simulation model with one layer (no bias correction), using a two-layered model with a correctly placed interface and using a two-layered model with a misplaced interface. This corresponds to the first, second, and third setups outlined in section 2.4.3 and to case numbers 1, 2, and 3 in Table 2. These three cases are compared to estimate the performance of a classical EnKF with the parameter update using a model where structure is not resolved, a model with the perfect structure and a model where it is assumed that the structure is known, but is wrong. The result from these so called original simulations is shown in Figure 4. It shows the mass of the lysimeter over time, for a full year, indicating the mass of each ensemble member together with the ensemble mean mass and the true mass. The resulting RMSE-values for each of the cases are shown in Table 3.

[47] From Figure 4 and Table 3, it is clear that the predicted masses in the one-layered model have very little in common with the mass of the true system (RMSE: 124.5). This is expected, as the model is a strong simplification of the true system. The model inadequacy due to the simplified parameterization is not severe, but the inadequacy due to the nonresolved structure leads to very poor performance. The correctly set up two-layered model, on the other hand, can well approximate the true mass of the system (RMSE: 6.8). The incorrectly set up two-layered model shows, in difference to the one set up correctly, a very poor performance (RMSE: 345.7). So the model inadequacy due to the wrong structure is here as severe as that one caused by ignoring the structure in the first place. This exemplifies how important it is to know the details of the modeled system correctly, here where the layer interface is placed, when using a more complex model.

3.2. Use of Bias Corrections

[48] It is now tested, whether and to what extent bias correction can compensate for the model inadequacy when a model without spatial structure is used. The bias correction acting directly on the model states, described in equation (6), is considered. This corresponds to the fourth setup described in section 2.4.3 and case 4 in Table 2. The result is shown in Figure 5 and Table 3. As can be seen when
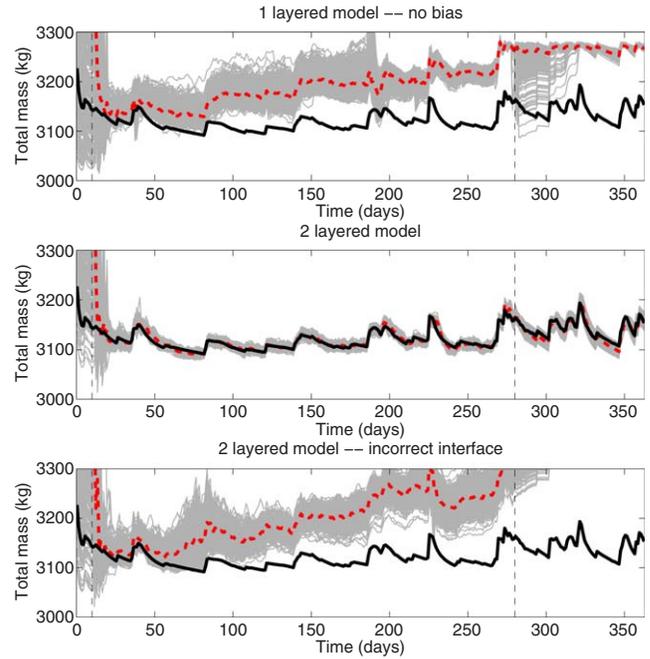
Figure 4. EnKF simulation result of the total masses using a one-layered model without bias correction (case 1), a two-layered model with correct interfaces (case 2), and a two-layered model with incorrect interface (case 3). Gray areas show the spread of the ensemble and red lines are the ensemble mean. The black lines mark the true mass and the vertical lines mark the start and end of the filtering period.

comparing Figures 4 and 5, the performance of the flow model is improving when using the bias correction. It should be noted from Figure 5 that even though the spread of the ensemble is rather large during the prediction period, the mean of the ensemble is still a reasonable prediction of the total mass (RMSE: 34.3), far better than a one-layered model without the bias corrections, though naturally not as good as the perfect two-layered model. It is also important that the dynamic behavior of the system is now reproduced well, despite the offset seen in Figure 5.

[49] In Figure 5, the masses of the lysimeter are compared, in order to address the question if the averaged behavior of the system can be predicted from the local

Table 3. Resulting RMSE of Used Simulation Scenarios[a]

	Total Mass (kg)	Water Content (−)
Case 1	124.5	0.045
Case 2	6.8	0.008
Case 3	345.7	0.238
Case 4	34.3	0.017
Case 5	126.3	0.136
Case 6	17.8	0.007
Case 7	24.7	0.018
Case 8	–	0.048
Case 9	–	0.008
Case 10	–	0.009

[a]Shown is the RMSE value (equation (11)) for the prediction period. For the water content, the value is taken as an average over the four measurements. For comparison, the RMSE values of the true signal to its temporal mean is 21 and 0.016 (mass and water content, respectively) for VR-A, 33 and 0.020 for VR-B, and 0.026 for the real data.
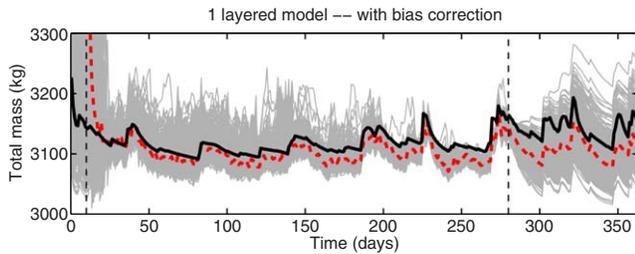
Figure 5. EnKF simulation result of the total masses using a one-layered model with bias correction (case 4). Gray areas show the spread of the ensemble, and red lines are the ensemble mean. The black lines mark the true mass and the vertical lines mark the start and end of the filtering period.

observations in a heterogeneous setting. However, depending on the problem, one might also want to predict the local states that are measured. One would expect that this prediction is less problematic than the mean behavior of the system. Figure 6 shows the water content at each of the measurement locations for the case of using a correct two-layered model and the one-layered model with and without the bias corrections (cases 2, 4, and 1, Table 2). It becomes clear from the figure that the water content is estimated similarly well by all simulation models during the filtering period, which simply means that the filter controls the water content with the state updates. As long as observations are available, the model mismatches are not so severe. In the prediction period, predictions made with the two-layered model are stable and well fitting the true water
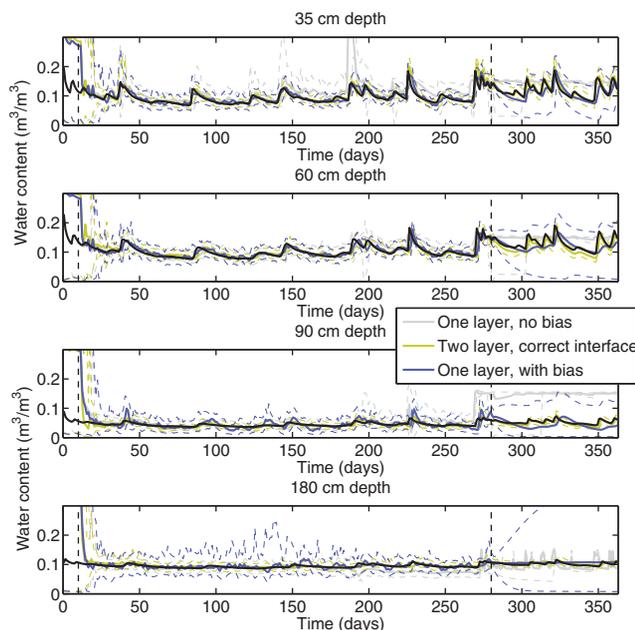
Figure 6. Water content as a function of time at each of the four measurement depths. The dashed lines mark the envelope of the ensemble, the solid lines the mean of the ensemble, and the black lines mark the true water content. The vertical lines mark the start and end of the filtering period.

content for all measurements. This is expected, since the simulation model is very similar to the virtual reality. The one-layered model without the bias, however, is a poor simulation model for the water content in the prediction period. The one-layered model with the bias correction leads to a good prediction of the water content at the two uppermost measurement depths, whilst the prediction of the lower measurements comes with a larger spread of the ensemble. The good performance on the uppermost measurements is showing the strength of the bias correction method, whilst the bottom measurement (180 cm) is showing a large spread of the ensemble due to the uncertain value of the bias correction at the bottom of the domain (see further the discussion about the resulting bias corrections below). To be noted is that even though the spread of the ensemble is large in the prediction when using the bias corrections, which can indicate that the bias corrections are far from optimal, the mean of the ensemble is always close to the true water content.

[50] It is interesting to note that even though the one-layered model without a bias correction has a reasonable value of the water content during the filter period (Figure 6), it does not resolve well the total mass (Figure 4). This is most likely related to the poor performance of the estimated parameters, which is further discussed in section 3.4. The misfit in the masses shows how important it is to have other measures of performance for a filter method then just the measurements that have been used for assimilation when robustness is an issue.

[51] It is clear from the previous section that the two-layered model can perform very well when the layer is correctly placed but when the layer is misplaced the simulation result is not satisfactorily. The poor performance of the two-layered incorrect model is by no means surprising, since a misplacement of 1/4 of the observations is a rather severe mistake. It is noticeable that the result of the one-layered model with bias (Figure 5) is clearly better performing than the two-layered incorrect model (Figure 4). This suggests that, rather than guessing the placement of the layer interface, a simpler model with a bias correction can be used with better result. One could interpret the results in the way that the perfect structure model has the smallest errors, the structured model with the wrong structure has the largest errors and the models that ignore structure and compensate this with a bias correction are in-between (Table 3).

3.3. Virtual Reality B

[52] The results discussed above have been obtained for one setup of a layered model. In order to test if they can be reproduced in a different setting, we compare the same measures as discussed above for the virtual reality model B (VR-B). In this model, the material properties are different from VR-A. The VR-B has a stronger contrast between the water contents of the two materials, and it should therefore be more difficult for the filter with the one-layer model to make good predictions of the mass of the system. In addition to the increased contrast, the order of the layers is reversed, so that the VR-B has the finer material at the bottom. The VR-B is also used to analyze the information contained in the bias term, however, this aspect will be discussed in section 3.6.
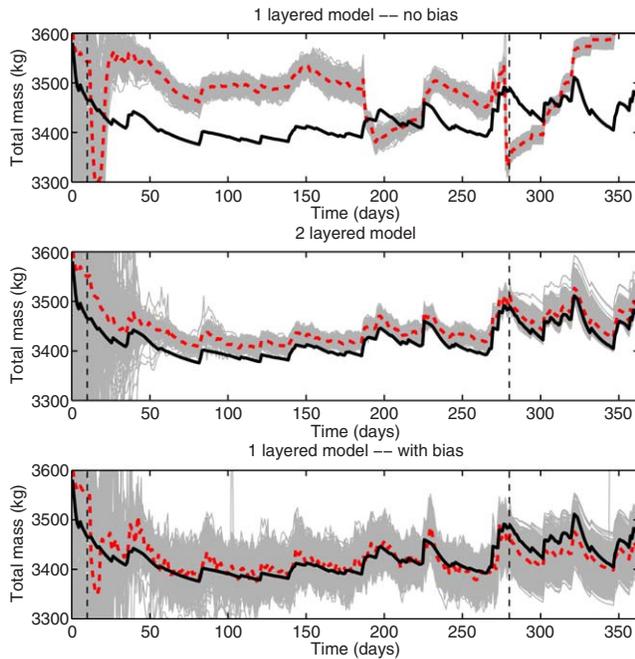
Figure 7. EnKF simulation result of the total masses for VR-B using a one-layered model without bias correction (case 5), a two-layered model (case 6), and a one-layered model with bias correction (case 7). Gray areas show the spread of the ensemble and red lines are the ensemble mean. The black lines mark the true mass and the vertical lines mark the start and end of the filtering period.

[53] As the performance using this virtual reality is similar to the previous one, we only show here the comparison of the total mass of the system in Figure 7, where the results of cases 5–7 in Table 2 are shown. The one-layered model without a bias is, as expected, not a good simulation model. The two-layered model has a rather good performance, though not as good as in the VR-A (case 2), which most likely is due to the stronger contrasts in the water contents that makes the simplified parameterization less suitable to model the more complex reality. The one-layered model with the bias corrections applied has, in difference to the one-layered model without bias correction, a large spread of the ensemble but a prediction of the ensemble mean mass that is similar to the true mass.

3.4. Parameter Estimations

[54] As described in section 2, not only the pressure head states are updated during the filtering process but also the flow model parameters. An important aspect of the parameter filtering is whether or not the parameters stabilize in time around a reasonable final value. If this is the case, the model with the parameters can be trusted more to capture the physical processes and to be useful for predictions where observations are not available. In the case that parameters do not stabilize and change with time over large spans of values, the model is more data driven and cannot be expected to make good predictions when no observations are available. We found in the test cases discussed above that for all runs that produce a good result (low RMSE, see equation (11)), the parameters are also far more
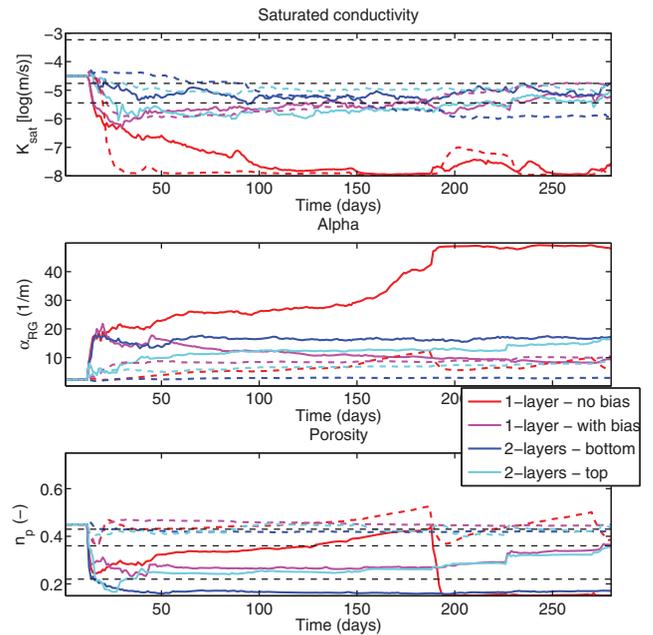
Figure 8. Development of the ensemble mean of the estimated parameters shown for VR-A (solid lines) and VR-B (dashed lines). Dashed black lines mark the true parameter values (Table 1).

stable than for the cases in which the performance was poor. In the latter, the parameters showed sudden jumps in time, or converged to parameter values on the edge of the allowed ranges. These aspects are shown in Figure 8, where the development of the estimated parameters for the mean of the ensembles are plotted against the assimilation time for cases 1, 2, 4, 5, 6, and 7 (Table 2).

[55] In the lowest plot, the porosity has values very close to the true porosity values (dashed black lines) for the good cases, whilst the one-layered model without bias correction shows an unstable performance. The porosity is straight forward to compare, since it has the same physical meaning and definition in all parameterization models. The picture gets more complicated when interpreting the other two parameters (alpha parameter and saturated hydraulic conductivity). Since the parameterization is different between the virtual realities and the EnKF simulations, the alpha parameters of the two models are not comparable. It is, however, positive to note that the parameter alpha in the EnKF simulations is very constant throughout the good simulations.

[56] The saturated hydraulic conductivity is, as can be seen from Figure 8, always predicted in the low range. It can be argued that this is not good, since the saturated conductivity is a material property that has a clear physical meaning and should therefore have nothing to do with the parameterization of the unsaturated hydraulic conductivity function. This, however, is only true at full saturation, which rarely occurs in the models tested in this paper, and especially rarely occurs at the observation locations. Hence, the model becomes insensitive to the hydraulic conductivity at full saturation. The model is, however, not insensitive to the total (unsaturated) hydraulic conductivity and therefore the model most likely compensates parts of

the missing parameterization flexibility (using a Gardner-Russo model for assimilation, which has a simpler parameterization than the van Genuchten model used to create the virtual realities) by reducing the saturated hydraulic conductivity. Indeed, this can be seen when comparing the unsaturated hydraulic conductivity curves (not shown) of the virtual reality model and the EnKF simulation model; with the saturated hydraulic conductivity resulting from the EnKF simulation model, the conductivity curve of the EnKF simulation model is more similar to the conductivity curve of the VR than if the true hydraulic conductivity would be used for the EnKF simulation model.

[57] The saturated hydraulic conductivity values obtained with the EnKF are consistent with the setups of the virtual realities. The bottom layer in case 2 (EnKF simulation model for VR-A) is more permeable than the top layer, and the other way around in case 6 (EnKF simulation model for VR-B). This reflects the relations in the two virtual realities (Table 1). Further, the hydraulic conductivity of the bottom layer in the EnKF simulation model for VR-B (case 6) is by far the lowest conductivity estimated in a well-performing simulation. This is positive, since this corresponds well to the VR setup in which the bottom of the VR-B has the lowest hydraulic conductivity (Table 1).

[58] The overall consistency, in combination with the correct parameter relations and reasonable model performances speaks for the fact that the EnKF simulations have performed well and are capable of filtering both parameters, states and bias corrections.

3.5. Real Data

[59] In difference to the results shown with the virtual reality test cases, the real data contains no validation data other than the water content observation themselves. The performance of the model can therefore be assessed best with the observations during the prediction period, where the filter is turned off. To look more into the prediction period, the filter problem is set up to have a long assimilation period, followed by a clear break (the prediction period), which in the end is followed by yet another assimilation. The final assimilation period is used to test the errors of the predictions during the crossover from the prediction to the time span where observations are again available. The result of the data assimilation is shown, for each of the four depths where TDR measurements of water content are available, in Figure 9. The resulting parameters are shown in Figure 10. As can be seen from the figures, the prediction of the water content using an assimilation without the bias corrections is, as expected, rather poor. Similarly, the resulting parameters are close to the allowed boundaries (which are the same as the extent of the vertical axes in Figure 10) and show sudden, large, jumps. When the bias corrections are used, however, the prediction of the water content improves and the mean of the ensemble is similar to the measured water content. Though, similar to the VR cases, the spread of the ensemble is rather large in the lower layers. Just after the start of the second assimilation period, the models using the bias corrections directly fit well to the observations, while the model without a bias correction takes several days to get back to reasonable values. The result presented here supports the use of bias
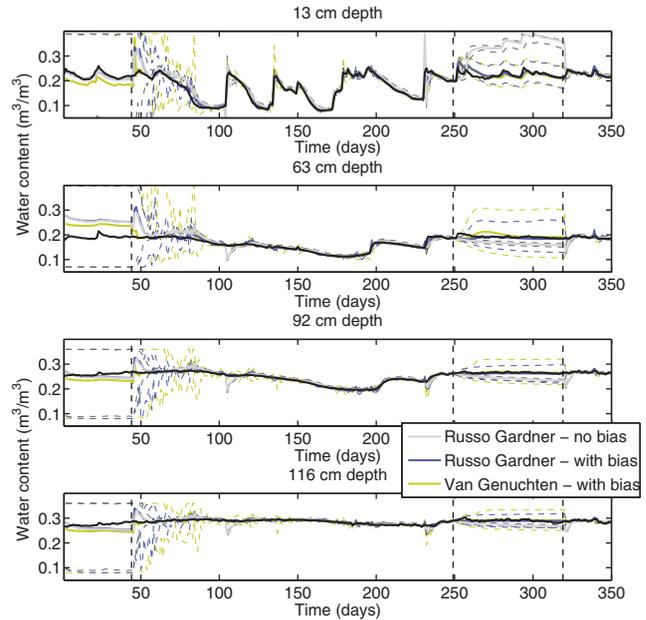
Figure 9. Water content as a function of time at each of the four measurement depths. The dashed lines mark the envelope of the ensemble, the solid lines the mean of the ensemble, and the black lines mark the true water content. The vertical lines mark the start and end of the filtering period.

corrections to account for nonresolved medium structure, also for real field measured data.

[60] When looking at the parameters, the one-layered Gardner-Russo model without bias is unstable and shows
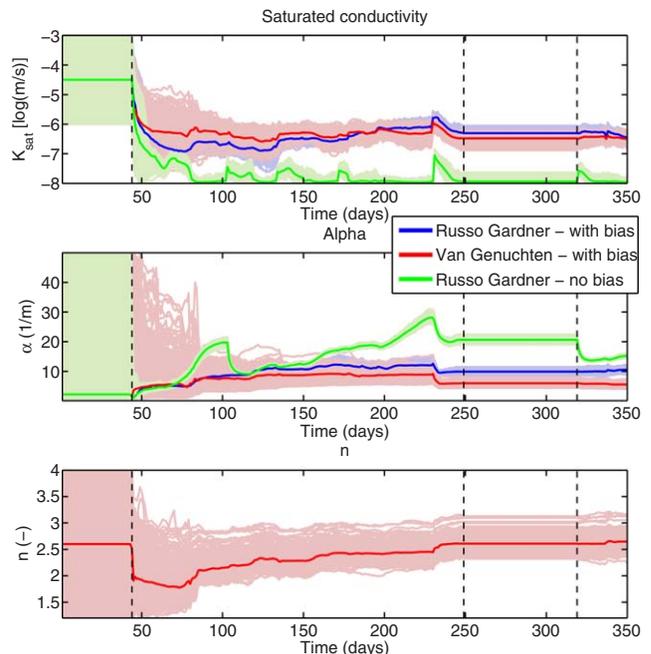
Figure 10. Parameter values plotted over time for the real data simulations (cases 8–10). Dark colors mark the mean of the ensemble and light colors the respective ensemble spreads. Please note that the middle plot shows both α_{RG} and α_{VG} though these cannot one to one be compared.

parameters that are close to the allowed boundaries. This is the same as in the VR cases and does not improve if the van Genuchten parameter model is used. The parameters resulting from the assimilations using the bias corrections are a lot more reasonable. When looking at the alpha parameter plot (Figure 10), it is important to keep in mind that the two parameter models are distinctly different and that the alpha parameter values therefore cannot be compared one to one. When comparing the behavior of the van Genuchten model with the simpler Gardner-Russo model, both with bias correction, all parameters are well estimated in both parameterization models and the prediction performance (Figure 9) is very similar. This suggests that for the assimilation problem with the bias correction, the use of a more complex parameterization model has no positive effect, and the simpler Gardner-Russo model could be used.

3.6. Analysis of the Bias Correction

[61] A relevant question to ask when using the bias correction methods to compensate for missing subsurface structure is if the suggested bias corrections can be meaningfully interpreted. It could be useful, if the bias terms would allow to infer the variability of, for example, the water content in the soil. Three examples of the bias parameter profiles are shown in Figure 11 for the one-layered model with the bias correction for VR-A, VR-B, and the real data (cases 4, 7, and 9 in Table 2). Since the bias corrections used in this paper turned out to be rather stationary, only the resulting values at the end of the assimilation are shown.

[62] When comparing the two virtual realities, the resulting bias for the VR-B has higher values, and is mirrored in comparison with the VR-A bias corrections. This reflects the structure of the soil, given that the differences between the two VR models are the contrast of water content between the two soil layers and the order of these layers. This is a clear strength of the bias correction as a methodology to capture nonresolved substructure in a model, when small-scale observations are used. The effect of the bias

correction is strongest around the observation that is just below the true interface. In the VR-A, the bias subtracts water just below the nonexistent interface and in the VR-B water is added. This, in regard to what a homogeneous model on its own would produce (cf. Figure 4), suggests that the bias corrections resulting from the EnKF simulations can be interpreted as really correcting for the missing layering structure.

[63] When it comes to the real data, the interpretation is, as is often the case with the real data, less clear. For a start, the magnitude of the bias terms is smaller than in the VR data cases. This can be understood by the far less obvious contrasts in water content that exists in the real data compared with the VR data. In the paper of *Wollschläger et al.* [2009], four different layers were used for parameter estimation (one measurement per layer). When comparing their resulting parameters between the four layers [*Wollschläger et al.*, 2009, Table 3], it is clear that the two bottom layers are more similar to each other than the two top ones. Further, it is difficult to say which of the two top most layers is coarser. When looking at the bias corrections for the real data in Figure 11, the same trend can be seen. The bottom is clearly showing a uniform bias correction trend, whilst the top is different from the bottom and no clear trend can be seen at the top. This, again, supports what has been seen in the VR data that the bias corrections can correct for missing spatial structure.

4. Summary and Discussion

[64] In this paper, we have investigated possibilities to fit two models with different level of details to two virtual reality lysimeters and a real field experiment using the Ensemble Kalman Filter. These models were considered for predictions in an unsaturated system, where observations are available for long time periods. We focus on the case where the observations have a small observation volume compared to the size of the system. We used observations of local water content from TDR probes to make predictions of either the water mass of the whole system or
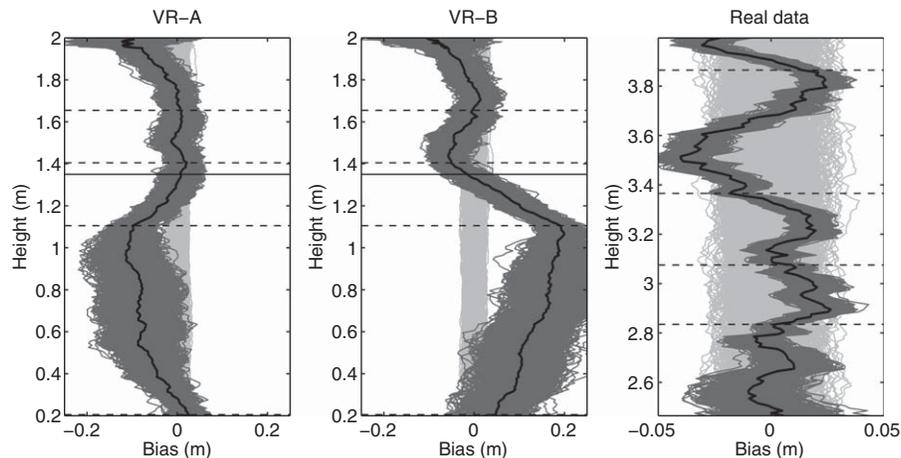
Figure 11. Bias corrections, taken at the end of the assimilation period for (left) case 4, (middle) case 7, and (right) case 9. Seen in light gray is the initial bias correction, in dark gray the final bias correction, and in black the mean of the final bias correction. The dashed lines mark the positions of the measurements and the solid lines mark the layer interface in the virtual realities.

the water content at the measurement locations. The tested models are two two-layered flow models with the simplified parameterizations and a one-layered flow model, also with the simplified parameterization. The results for the virtual realities show that the EnKF with the two-layered model with correct subsurface structure can well predict the behavior of the true system, while both the two-layered model with incorrect layering and the one-layered model on its own are very poor representations of the true system. When the bias corrections are introduced in the EnKF, a clear increase in the predictive capability of the one-layered model is shown. The resulting bias correction contains information about where the largest errors are found and thus also some information about the unresolved subsurface structure.

[65] It should be stressed that the use of an EnKF the way it is done here is only useful if long time series of observations are available (possibly with gaps). Only in such settings is it reasonable to update states as initial condition for the prediction over a given (short) time span, until a new observation is available. If observations are only available over a short time span, the model parameters are the most important factors that need to be estimated as good as possible, while errors in initial conditions are not controllable. For such cases, a batch calibration (using all observations for one optimization) will be the better approach to estimate these parameters and the advantage of the sequential EnKF would only be in its efficiency. For example, in case of a setup similar to the ones discussed here, it may take some 10,000s of model evaluations for a successful estimation of the wanted parameters in a Markov Chain Monte Carlo parameter estimation. In contrast, we here use 1000 model evaluations for the EnKF. This is, no doubt, an advantage of the EnKF, but also clearly a warning. The reduction of model evaluation is not only beneficial, but may also create large estimation problems if the estimation problem is not well set up (see further Appendix A).

[66] Also when states are updated, it has been found to be an important goal of the EnKF parameter estimation to find stable parameters. This was demonstrated in the discussion of the results (Figures 8 and 10). The stable parameters achieved with the bias corrections are, however, only representative for the system under similar conditions as in the assimilation time. Since the data used in this paper contains rather slow flow processes (observations only once per day), the estimated retention functions are probably well matched to the real retention behavior. If the data would, on the other hand, show strong dynamic behavior, such as during heavy rainfall after a dry period with some observations per hour, the effective conductivity behavior would not be captured by the effective parameters, and the result would probably look different. An example of a misfit between retention behavior and conductivity behavior can be seen in the example in Figure 1, where the parameter functions for the Gardner-Russo model and the van Genuchten model are compared. The retention functions of the two models are reasonably similar, whilst the hydraulic conductivity functions are clearly different. The difference of the relative permeability shown in the figure can be compensated by adjusting the saturated conductivity, such that the unsaturated hydraulic conductivity functions (the product of saturated conductivity and relative permeability) of the two models become more similar. This is most likely what is being seen in the EnKF simulations discussed above, where K_{sat} is always lower than the true values. If the model would be used to predict a fast flow event under wet conditions, which is sensitive to the saturated hydraulic conductivity, it is to be expected that such a prediction would be poor. Similarly, if other scenarios are considered, the formulation of the filter and the bias corrections cannot be expected to work without first being adopted carefully to the new scenario. An example is contaminant transport were the impact of the layering on the transport may rather result in a time delay than the offset discussed in this work, and hence the bias corrections would need to compensate differently.

[67] We have demonstrated with the virtual reality cases and the real data that nonresolved subsurface structure in the simulation model can be compensated by bias correction in the EnKF. Such a bias correction could also be used in a batch calibration, where model parameters are estimated from the full observation data set. One could also, here, be interested in calibrating a homogeneous model to observations in a heterogeneous reality (which means the goal is to find the effective model parameters). It has been shown by many authors that this is a difficult task and bias corrections could, here, also be helpful. For example, *Erdal et al*. [2012] used a flow model with bias correction to calibrate a homogeneous unsaturated flow model to observations in a heterogeneous system in a Markov Chain Monte Carlo (MCMC) framework. In a calibration there can be no feedback from the bias to the model, as the observations are not treated sequentially. There are fundamental differences between using a bias correction in a calibration (with the aim to find model parameters) and in an EnKF (with the aim to find model parameters and states for predictions of the next time step). One important difference is the use of the bias correction in a prediction.

[68] The bias corrections used in this paper are all applied both during the assimilation (equivalent to the calibration) and during the prediction period, while in the model calibration of *Erdal et al*. [2012], they are only applied during the calibration. The missing soil structure is, hence, in the EnKF explicitly taken into account with the bias in every forward modeling step, also during the prediction. This implies that the longer the prediction period, hence the longer the time without any update of the EnKF, the larger the spread of the resulting ensemble becomes. This may eventually lead to the deterioration of the prediction of the ensemble mean. If, however, the bias corrections would be removed during the prediction period, the initial condition for that period would be very wrong and lead to artifacts such as infiltration fronts stemming from the "release" of a heterogeneous water distribution in a homogeneous model. This is not problematic for the calibration of a model as, once the model is calibrated, the states are not controlled anymore. The initial condition for a prediction would therefore be chosen according to the homogeneous model setup and would not be influenced by spatially heterogeneous states.

[69] The different use of bias correction in data assimilation and model calibration shows that it is important to know the purpose of the modeling to judge about
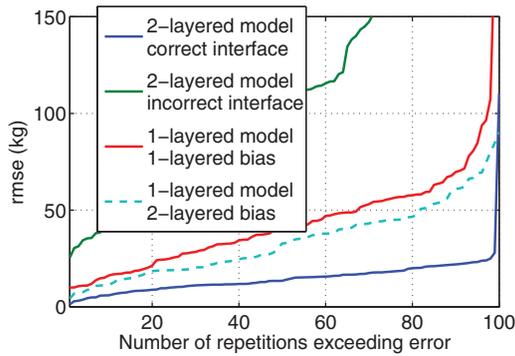
Figure 12. Error in the predictions for 100 repetitions of the EnKF using 20 ensemble members, sorted in order of increasing error.

appropriate bias estimate models. If the purpose is prediction of fully unknown scenarios, the calibration with a bias model used by *Erdal et al.* [2012] would be the better choice. If the purpose is prediction for shorter time frames and with continuous assignment of data, the EnKF with bias corrections explored in this paper would be a good choice.

[70] The resulting bias corrections shown in this work does well resolve the structure of the true system. How to extract the true soil structure from the bias corrections is, though, not obvious, especially for more complicated systems. What can be said is rather which of the observations are placed in soils that are more similar to each other, since these observations would result in similar bias corrections. In this work, for example, the two upper bias correction values in the VR cases show similar values, which corresponds well with them being placed in the same layer. It would, however, not be possible to detect from the bias corrections where the layer boundary is, only that the two top observations are different to the two lower ones.

[71] A final remark regarding the resulting bias corrections is that though, in the cases discussed above, they stabilize in time (leading to pretty stationary values), this is by no means something that is required by the model. The system investigated in this paper is intentionally set up without reaching any extremes, such as storms, droughts, etc. This means that the bias corrections required by the filter are kept on a rather constant level. If there would be more extreme conditions included, such as a strong and more natural transpiration in the summer cycle, the stationary feature of the bias corrections may be replaced by for example a yearly cycle, such as the leakage parameter in the groundwater filtering problem of *Hendricks Franssen et al.* [2011]. This, however, is for further research to investigate.

Appendix A: Reduced Ensemble Size

[72] The results shown in this paper are generated using an ensemble size in the EnKF that is large enough to ensure stable results (1000 members). When using a large ensemble size, a resimulation of the same setup is expected to generate the same result. When the ensemble size decreases, the system becomes more sensitive to, for example, initial sampling and the random noises added during the simulation. For any practical purpose, the use of large

ensembles is often difficult and it might be necessary in a specific application to use ensemble sizes as small as possible. To demonstrate the impact of the ensemble size on the results shown and, further, to demonstrate how the EnKF can be improved when small ensemble sizes are being used, four test cases are here presented. The cases shown are: the two two-layered models using a correctly and an incorrectly placed interface as well as the one-layered model using the bias correction (small ensemble versions of cases 2, 3, and 4 in Table 2). All cases are assimilated using an ensemble size of 20 members and the result is showing the RMSE values for 100 repetitions of the exact same problem using different random initializations, plotted in order of increasing RMSE. As can be seen from Figure 12, when looking at the two-layered model with correct interface, the result is very stable and very good (comparable with the results in Table 3), while the two-layered model with incorrect interface performs poor. The one-layered models, on the other hand, may have good results, but the RMSE curve is clearly increasing toward the right end of the axis in Figure 12. This shows that if the model is a very good approximation of reality, a small ensemble is sufficient, whilst if the model gets more complicated (as with the bias corrections that introduce more parameters), a larger ensemble size is needed to provide a stable result.

[73] A step in the direction of lowering the ensemble size is to aid the correlations in the EnKF. This could, for example, be done by using localization covariance matrices, which are applied to the computed covariances and reduce the correlation where no correlation should be [see e.g., *Houtekamer and Mitchell*, 1998; *Chen and Oliver*, 2009]. Since the focus of this work is the use of bias corrections, we demonstrate here the use of a prescribed correlation structure for the bias noise (w in equation (6)) in which the strongly correlated noise used for the bias corrections is divided into two separate layers, such that the correlation is only present between points within the same layer. The correlation layer interface is here placed at the position of the true layer interface. As can be seen from Figure 12, both bias correction models perform similarly if the simulation is good (low RMSE values are comparable), but the two-layered bias correlation structure model has clearly lower errors over all of the repetitions than the one-layered bias model, suggesting that the former has a more stable result. This suggests that a better prior knowledge of the structure of the bias can help to reduce the required ensemble size to get reasonable predictions with the EnKF.

References

Allen, R. G., L. S. Pereira, D. Raes, and M. Smith (1998), Crop-evapotranspiration (guidelines for computing crop water requirements), *FAO Irrig. and Drain. Pap. 56*, Food and Agric. Org. of the U. N., Rome.

Chen, Y., and D. S. Oliver (2009), Cross-covariances and localization for EnKF in multiphase flow data assimilation, *Comput. Geosci.*, *14*(4), 579–601, doi:10.1007/s10596-009-9174-6.

Crow, W. T., and M. J. van den Berg (2010), An improved approach for estimating observation and model error parameters in soil moisture data assimilation, *Water Resour. Res.*, *46*, W12519, doi:10.1029/2010 WR009402.

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

Dee, D., and A. Da Silva (1998), Data assimilation in the presence of forecast bias, *Q. J. R. Meteorol. Soc.*, *124*(545), 269–295, doi:10.1002/qj.49712454512.

De Lannoy, G. J. M., P. R. Houser, V. R. N. Pauwels, and N. E. C. Verhoest (2007a), State and bias estimation for soil moisture profiles by an ensemble Kalman filter: Effect of assimilation depth and frequency, *Water Resour. Res.*, *43*, W06401, doi:10.1029/2006WR005100.

De Lannoy, G. J. M., R. H. Reichle, P. R. Houser, V. R. N. Pauwels, and N. E. C. Verhoest (2007b), Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter, *Water Resour. Res.*, *43*, W09410, doi:10.1029/2006WR005449.

Deng, Z., Y. Tang, and H. J. Freeland (2011), Evaluation of several model error schemes in the EnKF assimilation: Applied to Argo profiles in the Pacific Ocean, *J. Geophys. Res.*, *116*, C09027, doi:10.1029/2011 JC006942.

Doherty, J., and D. Welter (2010), A short exploration of structural noise, *Water Resour. Res.*, *46*, W05525, doi:10.1029/2009WR008377.

Drécourt, J.-P., H. Madsen, and D. Rosbjerg (2006), Bias aware Kalman filters: Comparison and improvements, *Adv. Water Resour.*, *29*, 707–718, doi:10.1016/j.advwatres.2005.07.006.

Durner, W., U. Jansen, and S. C. Iden (2007), Effective hydraulic properties of layered soils at the lysimeter scale determined by inverse modelling, *Eur. J. Soil Sci.*, *59*, 114–124, doi:10.1111/j.1365-2389.2007.00972.x.

Erdal, D., I. Neuweiler, and J. A. Huisman (2012), Estimating effective model parameters for heterogeneous unsaturated flow using error models for bias correction, *Water Resour. Res.*, *48*, W06530, doi:10.1029/2011WR011062.

Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*(10), 143–162.

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

Fank, J., and G. von Unold (2005), *Wägbare Monolithische Lysimeter Unter Maschineller Freilandbewirtschaftung (Wagna, Austria)*, pp. 55–60, Höhere Bundeslehr- und Forschungsanst. fur Landwirt. Raumberg-Gumpenstein, Irdning, Austria.

Friedland, B. (1969), Treatment of bias in recursive filtering, *IEEE Trans. Autom. Control*, *14*(4), 359–367.

Gardner, W. (1958), Some steady state solutions of unsaturated moisture flow equations with applications to evaporation from a water table, *Soil Sci.*, *85*, 228–232.

Gupta, H. V., M. P. Clark, J. A. Vrugt, G. Abramowitz, and M. Ye (2012), Towards a comprehensive assessment of model structural adequacy, *Water Resour. Res.*, *48*, W08301, doi:10.1029/2011WR011044.

Han, E., V. Merwade, and G. C. Heathman (2012), Application of data assimilation with the root zone water quality model for soil moisture profile estimation in the upper Cedar Creek, Indiana, *Hydrol. Processes*, *26*(11), 1707–1719, doi:10.1002/hyp.8292.

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

Hendricks Franssen, H. J., H. P. Kaiser, U. Kuhlmann, G. Bauser, F. Stauffer, R. Müller, and W. Kinzelbach (2011), Operational real-time modeling with ensemble Kalman filter of variably saturated subsurface flow including stream-aquifer interaction and parameter updating, *Water Resour. Res.*, *47*, W02532, doi:10.1029/2010WR009480.

Houtekamer, P., and H. Mitchell (1998), Data assimilation using an ensemble Kalman filter technique, *Mon. Weather Rev.*, *126*, 796–811.

Ines, A. V., and B. P. Mohanty (2008a), Near-surface soil moisture assimilation for quantifying effective soil hydraulic properties under different hydroclimatic conditions, *Vadose Zone J.*, *7*(1), 39, doi:10.2136/vzj2007.0048.

Ines, A. V., and B. P. Mohanty (2008b), Parameter conditioning with a noisy Monte Carlo genetic algorithm for estimating effective soil hydraulic

properties from space, *Water Resour. Res.*, *44*, W08441, doi:10.1029/2007WR006125.

Kollat, J. B., P. M. Reed, and R. M. Maxwell (2011), Many-objective groundwater monitoring network design using bias-aware ensemble Kalman filtering, evolutionary optimization, and visual analytics, *Water Resour. Res.*, *47*, W02529, doi:10.1029/2010WR009194.

Kumar, S., M. Sekhar, D. V. Reddy, and M. S. Mohan Kumar (2010), Estimation of soil hydraulic properties and their uncertainty: Comparison between laboratory and field experiment, *Hydrol. Processes*, *24*(23), 3426–3435, doi:10.1002/hyp.7775.

Kumar, S. V., R. H. Reichle, C. D. Peters-Lidard, R. D. Koster, X. Zhan, W. T. Crow, J. B. Eylander, and P. R. Houser (2008), A land surface data assimilation framework using the land information system: Description and applications, *Adv. Water Resour.*, *31*(11), 1419–1432, doi:10.1016/j.advwatres.2008.01.013.

Li, C., and L. Ren (2011), Estimation of unsaturated soil hydraulic parameters using the ensemble Kalman filter, *Vadose Zone J.*, *10*(4), 1205–1227, doi:10.2136/vzj2010.0159.

Madsen, H., and R. Canizares (1999), Comparison of extended and ensemble Kalman filters for data assimilation in coastal area modelling, *Int. J. Numer. Methods Fluids*, *31*, 961–981.

McKay, M., R. Beckman, and W. Conover (1979), Comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, *21*(2), 239–245.

Meissner, R., H. Rupp, J. Seeger, G. Ollesch, and G. W. Gee (2010), A comparison of water flux measurements: Passive wick-samplers versus drainage lysimeters, *Eur. J. Soil Sci.*, *61*(4), 609–621, doi:10.1111/j.1365-2389.2010.01255.x.

Mertens, J., H. Madsen, M. Kristensen, D. Jacques, and J. Feyen (2005), Sensitivity of soil parameters in unsaturated zone modelling and the relation between effective, laboratory and in situ estimates, *Hydrol. Processes*, *19*(8), 1611–1633, doi:10.1002/hyp.5591.

Monsivais-Huertero, A., W. D. Graham, J. Judge, and D. Agrawal (2010), Effect of simultaneous state-parameter estimation and forcing uncertainties on root-zone soil moisture for dynamic vegetation using EnKF, *Adv. Water Resour.*, *33*(4), 468–484, doi:10.1016/j.advwatres.2010.01.011.

Mualem, Y. (1976), A new model for predicting the hydraulic conductivity of unsaturated porous media, *Water Resour. Res.*, *12*(12), 513–522.

Oliver, D. S., Y. Chen, and G. Nævdal (2010), Updating Markov chain models using the ensemble Kalman filter, *Comput. Geosci.*, *15*(2), 325–344, doi:10.1007/s10596-010-9220-4.

Papafotiou, A., et al. (2008), From the pore scale to the lab scale: 3-D lab experiment and numerical simulation of drainage in heterogeneous porous media, *Adv. Water Resour.*, *31*(9), 1253–1268, doi:10.1016/j.advwatres.2007.09.006.

Pasetto, D., M. Camporese, and M. Putti (2012), Ensemble Kalman filter versus particle filter for a physically-based coupled surface-subsurface model, *Adv. Water Resour.*, *47*, 1–13, doi:10.1016/j.advwatres.2012.06.009.

Reichle, R. H. (2008), Data assimilation methods in the Earth sciences, *Adv. Water Resour.*, *31*(11), 1411–1418, doi:10.1016/j.advwatres.2008.01.001.

Reichle, R. H., D. B. McLaughlin, and D. Entekhabi (2002), Hydrologic data assimilation with the ensemble Kalman filter, *Mon. Weather Rev.*, *130*(1), 103–114.

Roth, K., R. Schulin, H. Flühler, and W. Attinger (1990), Calibration of time domain reectrometry for water content measurement using a composite dielectric approach, *Water Resour. Res.*, *26*(10), 2267–2273, doi:10.1029/90WR01238.

Russo, D. (1988), Determining soil hydraulic properties by parameter estimation: On the selection of a model for the hydraulic properties, *Water Resour. Res.*, *24*(3), 453–459.

Shin, Y., B. P. Mohanty, and A. V. M. Ines (2012), Soil hydraulic properties in one-dimensional layered soil profile using layer-specific soil moisture assimilation scheme, *Water Resour. Res.*, *48*, W06529, doi:10.1029/2010WR009581.

Soil Survey Division Staff (1993), *Soil Survey Manual*, U.S. Dep. of Agric. [Available at http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_050993.pdf.]

van Genuchten, M. (1980), A closed-form equation for predicting the hydraulic conductivity of unsaturated soils, *Soil Sci. Soc. Am. J.*, *44*, 892–898.

Vogt, C., G. Marquart, C. Kosack, A. Wolf, and C. Clauser (2012), Estimating the permeability distribution and its uncertainty at the EGS demonstration reservoir Soultz-sous-Forêts using the ensemble Kalman filter, *Water Resour. Res.*, *48*, W08517, doi:10.1029/2011WR011673.

Vrugt, J. A., P. H. Stauffer, T. Wöhling, B. A. Robinson, and V. V. Vesselinov (2008), Inverse modeling of subsurface flow and transport properties: A review with new developments, *Vadose Zone J.*, *7*(2), 843–864, doi:10.2136/vzj2007.0078.

Walker, J., G. Willgoose, and J. Kalma (2001), One-dimensional soil moisture profile retrieval by assimilation of near-surface measurements: A simplified soil moisture model and field application, *J. Hydrometeorol.*, *2*, 356–373.

Wollschläger, U., T. Pfaff, and K. Roth (2009), Field-scale apparent hydraulic parameterisation obtained from TDR time series and inverse modelling, *Hydrol. Earth Syst. Sci.*, *13*, 1953–1966.

Wu, C.-C., and S. A. Margulis (2011), Feasibility of real-time soil state and flux characterization for wastewater reuse using an embedded sensor network data assimilation approach, *J. Hydrol.*, *399*(3–4), 313–325, doi:10.1016/j.jhydrol.2011.01.011.

Zhu, J., and B. P. Mohanty (2003), Effective hydraulic parameters for steady state vertical flow in heterogeneous soils, *Water Resour. Res.*, *39*(8), 1227, doi:10.1029/2002WR001831.

Zhu, J., M. H. Young, and M. T. van Genuchten (2007), Upscaling schemes and relationships for the Gardner and van Genuchten hydraulic functions for heterogeneous soils, *Vadose Zone J.*, *6*(1), 186–195, doi:10.2136/vzj2006.0041.