Challenges with effective representations of heterogeneity in soil hydrology based on local water content measurements

Hannes H. Bauser\textsuperscript{1,2} | Lukas Riedel\textsuperscript{2,3,5} | Daniel Berg\textsuperscript{2,3} | Peter A. Troch\textsuperscript{1,4} | Kurt Roth\textsuperscript{2,5}

\textsuperscript{1} Univ. of Arizona, Biosphere 2, 845 N Park Ave., Tucson, AZ 85721, USA
\textsuperscript{2} Heidelberg Univ., Institute of Environmental Physics, Im Neuenheimer Feld 229, Heidelberg 69120, Germany
\textsuperscript{3} Heidelberg Univ., HGS MathComp, Im Neuenheimer Feld 205, Heidelberg 69120, Germany
\textsuperscript{4} Dep. of Hydrology and Atmospheric Sciences, Univ. of Arizona, 1133 E. James E. Rogers Way, Tucson, AZ 85721, USA
\textsuperscript{5} Heidelberg Univ., Interdisciplinary Center for Scientific Computing, Im Neuenheimer Feld 205, Heidelberg 69120, Germany

Correspondence
Hannes H. Bauser, Univ. of Arizona, Biosphere 2, 845 N Park Ave., Tucson, AZ 85721, USA.
Email: hbauser@email.arizona.edu

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: BA 6635/1-1; Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg, Grant/Award Number: 33-7533.-30-20/6/2

Abstract
The description of soil water movement at the field scale requires soil hydraulic material properties. These cannot be measured directly and exhibit an intrinsic multiscale heterogeneity. The estimation of soil hydraulic properties through inverse modeling or data assimilation requires measurements of the state of the system. However, these measurements are typically scarce and do not resolve soil heterogeneity leading to effective material properties with limited predictive capabilities. In a synthetic study at the scale of soil profiles, we explore the possibility to estimate an effective reduced one-dimensional representation based on only four local water content measurements in a vertical profile located in one- and two-dimensional heterogeneous media. We allow the effective material properties to deviate locally to consider the impact of the small-scale heterogeneity on the water content measurements. In the synthetic experiments, the estimated reduced one-dimensional representations predict soil water content and fluxes sufficiently well but can fail to predict the accurate timing and magnitude of infiltration fronts. Cumulative fluxes below the water content measurements were predicted correctly even on short time scales, if evaporation was estimated accurately. However, this proved to be challenging if the local evaporation fluxes above the measurement profile deviated strongly from the average evaporation fluxes of the heterogeneous medium. The one-dimensional estimation leads to soil hydraulic parameters describing effective material properties that differ from the true material properties to compensate the missing representation of heterogeneity and two-dimensional flow.

Abbreviations: AHFR, average horizontal flux ratio; EnKF, ensemble Kalman filter; NSE, Nash–Sutcliffe efficiency; 1D, one-dimensional truth and one-dimensional data assimilation; TDR, time domain reflectometry; 2D, two-dimensional truth and one-dimensional data assimilation.

1 | INTRODUCTION

The quantitative description of soil water movement on scales from meters to entire landscapes remains an open challenge. A critical reason is the heterogeneity of the...
porous soil structure at all scales, which strongly influences the water flow.

On the smaller scale of soil profiles, which is considered to be the critical scale for a reliable representation of soil hydrology (Vogel, 2019), soil water movement is described by the Richards equation, which requires hydraulic material properties of the porous structure. These properties are typically parameterized, commonly by the Mualem–van Genuchten parameterization. The required soil hydraulic parameters cannot be measured directly with sufficient accuracy or transferred reliably from the laboratory to the field. The unknown parameters are a priori often the source for the largest uncertainties in soil hydrologic modeling.

Therefore, soil hydraulic parameters have been estimated inversely (Hopmans, Šimůnek, Romano, & Durner, 2002; Vrugt, Stauffer, Wöhling, Robinson, & Vesselinov, 2008) or more recently through data assimilation (Botto, Belluco, & Camporese, 2018; Li & Ren, 2011; Shi, Song, Tong, Zhu, & Zhang, 2015; Wu & Margulis, 2011, 2013; Yu et al., 2019) based on measurements of the temporal development of the system's state. These studies resolve heterogeneity up to layers with discrete parameter sets but do not directly consider heterogeneity within these layers.

The issue of unresolved heterogeneity is well known in soil hydrology and has been addressed by describing the heterogeneous soil through effective material properties. For example, Wildenschild and Jensen (1999) performed a laboratory experiment with two-dimensional flow in a soil with small-scale heterogeneity. Based on detailed information about their experimental setup, they were able to find effective homogeneous material properties that could describe their specific experiment adequately. In contrast, an estimation based on local measurements of potentials through inverse modeling yielded dissatisfying results. However, effective material properties not only depend on the heterogeneous distribution but also on boundary conditions, and due to the nonlinear dynamics, simple averaging or first-order upscaling is unable to fully reproduce upscaled fluxes (Vereecken, Kasteel, Vanderborght, & Harter, 2007).

One possibility to address the challenge of heterogeneity is to fully resolve the heterogeneity through distributed measurements. This has been demonstrated for large-scale heterogeneity with extension of tens of meters in a synthetic study on the field scale by Chaudhuri, Franssen, and Sekhar (2018), who used data assimilation to estimate three-dimensional parameter fields. However, fully resolving heterogeneity through measurements is already challenging for large-scale heterogeneities where only a few measurement infrastructures exist that allow the necessary measurement resolution through distributed measurement profiles, but it is typically unfeasible in practical field applications to resolve smaller submeter heterogeneity for the entire field scale. Alternatively, prior structural information about the soil hydrological properties needs to be added to the inverse estimation (Vereecken et al., 2007), which is again difficult to provide for small-scale heterogeneity on the field scale.

On the scale of soil profiles, Song, Shi, Ye, Yang, and Navon (2014) estimated a two-dimensional field of the saturated hydraulic conductivity based on two measurement profiles close to each other in a synthetic study. In many cases, however, only one-dimensional heterogeneity can be considered in parameter estimation based on information from local measurement profiles. For example, Zha, Zhu, Zhang, Mao, and Shi (2019) directly estimated a one-dimensional parameter field, and Brandhorst, Erdal, and Neuweiler (2017), Erdal, Neuweiler, and Wollschläger (2014), and Zhang et al. (2019) estimated an additional bias to the state to consider model structural error, which can also stem from unresolved small-scale heterogeneity.

Heterogeneity within soil layers that are typically considered homogeneous has been reported and addressed in real-world applications as well. For example, Bauser, Jaumann, Berg, and Roth (2016) found relevant heterogeneity within individual soil layers based on water content measurements in a soil profile and estimated a simplified one-dimensional heterogeneity by employing Miller scaling. In artificial hillslopes at the Landscape Evolution Observatory (LEO) at Biosphere 2, which were constructed from a homogeneous material, Gevaert et al. (2014) reported that water content decreased with depth during a constant infiltration experiment and considered soil heterogeneity to be one possible cause. Subsequent studies represented the heterogeneity vertically by introducing soil layers in the model (Scudeler et al., 2016). Nevertheless, even this research infrastructure does not allow to resolve three-dimensional heterogeneity on a submeter scale.

To the best of our knowledge, it has not been demonstrated to what extent an effective representation of small-scale heterogeneity can describe soil water movement, when estimated on the basis of few sparse local measurements without additional information about the

---

**Core ideas**

- We estimated effective material properties based on local water content measurements.
- Estimated effective soil hydraulic material properties differ from true material properties.
- Accurate description of evaporation is crucial for flux prediction.
structure of the heterogeneity. In this study, we quantitatively explore this question with synthetic experiments by estimating simplified one-dimensional representations of heterogeneous soils based on only four local water content measurements in a vertical profile. The simplified representation of the heterogeneity is based on an effective description of the hydraulic material properties but allows local deviations at the measurement locations to consider the local deviations of the measurements there. For a set of scenarios, we generate the respective two-dimensional heterogeneous synthetic truth using Miller-similar media. We assimilate the local water content measurements with an ensemble Kalman filter (EnKF) to estimate the effective one-dimensional representation of the soil hydraulic material properties. We only estimate the heterogeneity at measurement locations and interpolate in between. To differentiate between the effect of this interpolation and two-dimensional flow, we additionally generate heterogeneous one-dimensional truths and apply the EnKF to them as well. Based on these synthetic experiments, we investigate the estimated material properties and predictive capabilities for water content and flux of a one-dimensional representation of two-dimensional heterogeneous water flow.

2 MATERIALS AND METHODS

2.1 Soil water movement

Under the assumption of local equilibrium and single phase flow, the Richards equation describes water movement in a porous medium as

$$\frac{\partial \theta}{\partial t} - \nabla \cdot \{K(\theta) \left[ \nabla h_m(\theta) - e_z \right] \} = 0$$  \hspace{1cm} (1)

with volumetric soil water content \( \theta \) [-], time \( t \) [T], unit vector in direction of gravity \( e_z \), and soil hydraulic material properties, which relate the isotropic hydraulic conductivity \( K(\theta) \) [L T\(^{-1}\)] and the matric head \( h_m(\theta) \) [L] to the water content. These properties are most commonly parameterized with the Mualem–van Genuchten parameterization:

$$h_m(\theta) = h_m^s \left( \frac{1}{\xi(x)} \right)$$  \hspace{1cm} (5)

$$K(\theta) = K^s \left( \frac{\theta}{\theta_s} \right)^{2/3}$$  \hspace{1cm} (6)

where * denotes the reference material properties and \( \xi [-] \) is the Miller scaling parameter. This way a heterogeneous field can be generated based on a scalar field of Miller scaling parameters.

Miller-similar media reproduce an essential feature of heterogeneous soils. At small negative matric heads, locations with coarse texture (larger scaling factors) feature a higher hydraulic conductivity than locations with finer texture (smaller scaling factors). This relation becomes reversed for more negative matric heads. Then, locations with finer texture feature a higher hydraulic conductivity than locations with coarse texture. This behavior leads to crossing points between different realizations of the hydraulic conductivity function \( K(h_m) \) scaled with different Miller scaling parameters. A limitation of Miller scaling is, that it does not scale the porosity, which is consequently assumed to be the porosity of the reference material at all locations.

2.2 Ensemble Kalman filter

The EnKF (Burgers, Jan van Leeuwen, & Evensen, 1998; Evensen, 1994) is a data assimilation method to combine model and measurement information into an improved estimate of the model’s state based on the assumption of Gaussian error distributions. It is the Monte Carlo extension of the Kalman filter (Kalman, 1960) for nonlinear models. To not only estimate states but also parameters, an augmented state can be defined, which then contains the model states as well as parameters. For the estimation, the filter sequentially iterates between a forecast and an analysis step.
In the forecast step, an ensemble of augmented states \( \mathbf{u}^n \) is propagated from time \( k-1 \) to time \( k \) when a measurement is available:

\[
\mathbf{u}^n_{f,k} = M \left( \mathbf{u}^n_{a,k-1} \right) + \eta^m_{k-1}
\]

where \( n \) denotes the ensemble member, \( M \) is the forward model, and \( \eta \) is the process noise, which can be used to describe unrepresented model errors. The subscripts “\( f \)” and “\( a \)” denote forecast and analysis, respectively. In this paper, the augmented state consists of a vector of water contents \( \theta \) representing the water content state, soil hydraulic parameters, and Miller scaling factors, and it has the length \( N_u = 208 \), where the water content state component accounts for 200 dimensions. The forward model for the water content is the Richards equation, whereas for the parameters it is a constant model.

Assuming Gaussian distributions, the uncertainty of the ensemble is characterized by the \( N_u \times N_u \) forecast error covariance matrix:

\[
\mathbf{P}_{f,k} = \left[ \mathbf{u}^n_{f,k} - \hat{\mathbf{u}}^n_{f,k} \right] \left[ \mathbf{u}^n_{f,k} - \hat{\mathbf{u}}^n_{f,k} \right]^T
\]

where the overline denotes the expectation value and \( \hat{\mathbf{u}}^n_{f,k} \) is the ensemble mean. In the practical application of the EnKF, expectation values are approximated by the ensemble.

In the analysis step, the model information is combined with measurement data \( \mathbf{d} \) based on their respective uncertainty. In this paper, the length of \( \mathbf{d} \) is the number of measurements \( N_m \), with \( N_m = 4 \). The measurement uncertainty can be represented through an ensemble of measurements \( \mathbf{d}^n \):

\[
\mathbf{d}^n = \mathbf{d} + \varepsilon^n
\]

where \( \varepsilon \) is the measurement error. The \( N_m \times N_m \) measurement error covariance matrix is then

\[
\mathbf{R} = \left[ \varepsilon^n \right] \left[ \varepsilon^n \right]^T
\]

The measurements are linked to the augmented state through the linear \( N_m \times N_u \) measurement operator \( \mathbf{H} \), which maps from the state space to the measurement space. A measurement at time step \( k \) of the postulated true system state \( \mathbf{u}_{true,k} \) is given by

\[
\mathbf{d}_k = \mathbf{H}_k \mathbf{u}_{true,k} + \varepsilon_k
\]

The forecast ensemble is then updated with the measurements to the analysis ensemble:

\[
\mathbf{u}^n_{a,k} = \mathbf{u}^n_{f,k} + \mathbf{K}_k \left( \mathbf{d}^n - \mathbf{H}_k \mathbf{u}^n_{f,k} \right)
\]

where the \( N_u \times N_m \) Kalman gain \( \mathbf{K} \) weights the forecast error covariance matrix with the sum of forecast error covariance matrix and measurement error covariance matrix and maps from the measurement space back to the state space:

\[
\mathbf{K}_k = \mathbf{P}_{f,k} \mathbf{H}_k^T \left[ \mathbf{H}_k \mathbf{P}_{f,k} \mathbf{H}_k^T + \mathbf{R} \right]^{-1}
\]

The augmented state is iteratively updated by alternating between forecast and analysis throughout the available measurement data.

We use several extensions of the EnKF to improve its performance in soil hydrological application. Through spurious correlations, as well as non-Gaussian distributions, the uncertainties described by the error covariance matrix of the analysis ensemble can become too small. This leads to filter inbreeding and can ultimately give rise to filter divergence. The issue is intensified if the process noise describing the model error is unknown. To reduce this effect, a damping factor \( \gamma \in (0,1) \) can be multiplied to the correction vector in the analysis update in Equation 12 (Hendricks Franssen & Kinzelbach, 2008). This lowers the reduction of the uncertainty in each analysis step. Extending the damping factor to a vector \( \mathbf{\gamma} \) (with an entry-wise multiplication) allows to treat each augmented state component differently (Wu & Margulis, 2011), which is also done in the present work.

Multiplicative inflation is an additional possibility to counteract filter inbreeding. To increase the model uncertainty, the forecast uncertainty can be raised by enhancing the distance between each ensemble member and the ensemble mean. Anderson and Anderson (1999) multiplied the distance with a constant inflation factor \( \lambda \geq 1 \):

\[
\mathbf{u}^n_{f,inf} = \sqrt{\lambda} \left( \mathbf{u}^n_{f} - \hat{\mathbf{u}}^n_{f} \right) + \hat{\mathbf{u}}^n_{f}
\]

However, the inflation factor is specific for each problem. By comparing the distance between forecast and measurements with their uncertainties, the factor can be estimated adaptively both in time (Anderson, 2007; Li, Kalnay, & Miyoshi, 2009; Wang & Bishop, 2003) and space for the augmented state (Anderson, 2009). In this paper, we use the spatiotemporally adaptive inflation method.
proposed by Bauser, Berg, Klein, and Roth (2018), which uses a Kalman filter within the EnKF to estimate the inflation factors. The method requires setting an uncertainty for the inflation factor, which determines how fast the factor can be adjusted.

Due to a limited ensemble size, the signal/noise ratio worsens for decreasing correlations in the error covariance matrix. However, correlations typically decrease with increasing physical distance between two state locations (Hamill, Whitaker, & Snyder, 2001). The resulting spurious correlations can be counteracted through localization (Houtekamer & Mitchell, 1998), which reduces covariances with distance through a correlation function (Houtekamer & Mitchell, 2001). The localization alters the calculation of the Kalman gain to

\[
K_k = \left( \rho \circ P_{f,k} \right) H_k^T \left( \rho \circ P_{f,k} H_k^T + R_k \right)^{-1}
\]

where \( \rho \) is a correlation function and \( \circ \) denotes the element-wise multiplication with each entry of the covariance matrix based on the distance between the positions in the augmented state. We use the typically employed fifth-order piecewise rational function defined by Gaspari and Cohn (1999) as a correlation function. The function is similar to a Gaussian function but with local support, which introduces a cutoff length.

### 2.3 Ensemble Kalman filter settings

Throughout this paper, we chose an ensemble size of \( N = 100 \) and set the process noise to \( \eta = 0 \). The damping factor was chosen analogously to, for example, Erdal et al. (2014); we set a factor of \( \gamma = 0.3 \) for the parameter component of the augmented state and applied no damping (\( \gamma = 1.0 \)) to the water content component. For the inflation methods, we followed the settings used by Bauser et al. (2018) for the estimation of the inflation factor with a separate Kalman filter: we set a value of 1.0 for the uncertainty of the inflation factor and applied the same damping factors as in the EnKF. For the localization, we chose a cutoff length scale of the correlation function of 1.5 m.

### 3 SYNTHETIC EXPERIMENTS

To investigate the effect of unrepresented heterogeneity when estimating soil hydraulic properties based on local measurements, we performed a total of 80 synthetic experiments, which are specified below and summarized in Table 1. The synthetic experiments are designed to include the unrepresented heterogeneity as the only model structural error. The local water content measurements were placed in the middle of a two-dimensional heterogeneous soil at depths of 9.5, 19.5, 39.5, and 79.5 cm. For the heterogeneous soil, we chose a spatial extension of \( 2 \times 2 \) m, which we expect to be sufficient to describe the impact of the small-scale heterogeneity on the local measurements.

#### 3.1 Material properties

To generate heterogeneous material fields, we used Miller scaling, which requires a scalar field of scaling factors for the entire domain. We followed Roth (1995) to create lognormally distributed Miller scaling fields. In a first step, we generated realizations of scalar Gaussian random fields with expectation 0 and variance \( \sigma^2 \) using a generator based on circulant embedding (Klein, 2016). We used Gaussian autocovariance functions with axes parallel to

| Scenario     | Miller scaling field | Soil type | Boundary condition |
|--------------|----------------------|-----------|-------------------|
| Default      | \( \sigma^2 = 0.1 \), \( \ell_h = 10 \), \( \ell_v = 10 \) | sandy loam | standard          |
| MedVar       | \( \sigma^2 = 0.2 \), \( \ell_h = 10 \), \( \ell_v = 10 \) | sandy loam | standard          |
| HighVar      | \( \sigma^2 = 0.5 \), \( \ell_h = 10 \), \( \ell_v = 10 \) | sandy loam | standard          |
| LowFlux      | \( \sigma^2 = 0.1 \), \( \ell_h = 10 \), \( \ell_v = 10 \) | sandy loam | low flux          |
| Horizontal   | \( \sigma^2 = 0.1 \), \( \ell_h = 30 \), \( \ell_v = 10 \) | sandy loam | standard          |
| Vertical     | \( \sigma^2 = 0.1 \), \( \ell_h = 10 \), \( \ell_v = 30 \) | sandy loam | standard          |
| Silt         | \( \sigma^2 = 0.1 \), \( \ell_h = 10 \), \( \ell_v = 10 \) | silt loam  | standard          |
| SiltLowFlux  | \( \sigma^2 = 0.1 \), \( \ell_h = 10 \), \( \ell_v = 10 \) | silt loam  | low flux          |

*\( \sigma^2 \), variance; \( \ell_h \), horizontal correlation length; \( \ell_v \), vertical correlation length.
FIGURE 1  Example of Miller scaling fields. The two-dimensional field in Panel a was generated with vertical correlation length \(\ell_v = 10\) cm and horizontal correlation length \(\ell_h = 10\) cm and variance \(\sigma^2_\xi = 0.1\) for the underlying Gaussian random field. The one-dimensional Miller field in Panel b was generated by using a vertical slice of the 2D field at width 100.5 cm. Note that the color map scales exponentially the anisotropy defined by the vertical correlation length \(\ell_v\) and horizontal correlation length \(\ell_h\). The spatial resolution of the field was set to 1 cm in all cases. The Miller scaling factors are then calculated at each location from the random number \(r\) at that location with

\[
\xi = \exp (r - \sigma^2_\xi) \tag{16}
\]

This way, the expectation of the hydraulic conductivity function remains the same as for the reference material.

We created five different realizations of the random field through different seeds for each set of \(\ell_v, \ell_h,\) and \(\sigma^2_\xi\) listed in Table 1. We restricted the number of different realizations to five to limit the computational demand. An example for a Miller field with \(\ell_v = \ell_h = 10\) cm and \(\sigma^2_\xi = 0.1\) is shown in Figure 1a. As a control for each two-dimensional heterogeneous Miller field, we also created an effectively one-dimensional heterogeneous Miller field by using a vertical slice at horizontal position 100.5 cm of the two-dimensional heterogeneous Miller field (Figure 1b). This slice allows to explore the effect of the reduced representation of heterogeneity alone without the impact of two-dimensional flow.

As the reference material for the Miller media, we used the Mualem–van Genuchten parameters reported by Carsel and Parrish (1988) for sandy loam \((\theta_s = 0.41, \theta_r = 0.065, \alpha = -7.5\) m\(^{-1}\), \(n = 1.89, K_0 = 1.23 \times 10^{-5}\) m s\(^{-1}\), \(\tau = 0.5\)) and silt loam \((\theta_s = 0.45, \theta_r = 0.067, \alpha = -2.0\) m\(^{-1}\), \(n = 1.41, K_0 = 1.25 \times 10^{-6}\) m s\(^{-1}\), \(\tau = 0.5\)). The reference material was homogeneous for the entire domain. We did not include layers to focus on the small-scale heterogeneity as present within layers. However, they could be added explicitly through different sets of reference materials.

3.2  | Initial condition

To create the initial condition for the synthetic experiments, we used a spin-up phase of 30 d. The initial condition for this spin-up phase was hydraulic equilibrium with a groundwater table at the bottom of the domain. The lower boundary condition was kept constant with a matric head of 0 m throughout the entire time. The upper boundary condition was an infiltration flux of \(1.0 \times 10^{-7}\) m s\(^{-1}\) (0.36 mm h\(^{-1}\)) for the first 10 d, followed by no flux for the remaining 20 d. Figures 2a and 2b show the initial conditions for the two- and the one-dimensional Miller field given in Figure 1 \((\ell_v = \ell_h = 10\) cm and \(\sigma^2_\xi = 0.1\)) with sandy loam as the reference material. Figure 2c compares the vertical water content distribution of the one-dimensional case with the two-dimensional water content distribution at 100.5 cm. Although the Miller scaling factors are exactly the same, the water contents differ due to the influence of the two-dimensional flux even after 20 d of no flux as the upper boundary condition. In this example, this can be seen at a depth of around 100 cm.
3.3 Boundary conditions

The boundary condition sequence applied to the upper boundary is shown in Figure 3. The total simulation time is 52 d. There are five rain events with duration of 4 d, each followed by 4 d of evaporation. The sixth rain event lasts for 2 d, followed by 10 d of evaporation. The first four irrigation events have a flux of $1.0 \times 10^{-7}$ m s$^{-1}$ (0.36 mm h$^{-1}$), the fifth event has a lower flux of $0.25 \times 10^{-7}$ m s$^{-1}$ (0.09 mm h$^{-1}$), and the sixth rain event has an increased flux of $4.0 \times 10^{-7}$ m s$^{-1}$ (1.44 mm h$^{-1}$). The evaporation is realized through a matric potential that starts at 0 m at the end of the previous rain event and is linearly decreased by 1 m d$^{-1}$. If this potential would lead to a flux into the domain, a no-flux boundary condition is applied instead. At the lower boundary, we keep a constant matric head of 0 m, corresponding to a fixed water table, and we apply no-flux boundary conditions at the lateral boundaries.

We separated the entire time into a data assimilation phase during which the EnKF was used to estimate water content states, soil hydraulic parameters, and Miller heterogeneity, and a verification phase used to investigate the predictive capabilities. The assimilation phase comprises the first three rain and evaporation events and ends after 24 d. The rain events are large enough to cause a response at all measurement locations. The following verification phase includes one rain event identical to the one used in the data assimilation phase, and rain events with lower and higher intensity. This allows us to investigate the prediction accuracy within and beyond the calibrated dynamics.

To alter the impact of the Miller scaling on the dynamics, we created a second boundary condition sequence with lower fluxes, for which the material properties differ more strongly for the same scaling field. We created this boundary condition by dividing the flux of each rain event by 4. The rate of change of the matric potential for the evaporation was decreased to only 0.25 m d$^{-1}$. To ensure that the dynamics reaches similar depths as before, all times were extended by a factor of 4. This increased the total time.
to 208 d. Hereafter, we refer to these longer cases as the low-flux boundary condition, whereas the previous boundary condition of 52 d length is referred to as the standard boundary condition.

### 3.4 Data assimilation

Synthetic measurements of the water content were generated every 2 h (8 h for the low-flux boundary condition) at horizontal position 100.5 cm at depths of 9.5, 19.5, 39.5, and 79.5 cm. The true values were disturbed by the measurement uncertainty drawn from an unbiased normal distribution with a standard deviation of 0.007, which can be assumed for time domain reflectometry (TDR) probes (Jaumann & Roth, 2017).

In the data assimilation phase, we used the EnKF to estimate a one-dimensional representation based on the synthetic water content measurements. However, other methods for the inverse estimation could be used as well. The augmented state in the EnKF comprised the water content state, the Mualem–van Genuchten parameters α, n, and K₀, the Miller scaling factors at the four measurement locations, and additionally the Miller scaling factor at the cell closest to the surface (0.5-cm depth). For K₀ and the Miller scaling factors, we chose to include the logarithm instead of the direct quantity into the augmented state. The parameters θᵢ, θᵣ, and τ were set to the respective true values.

As the initial guess of the water content state for the data assimilation, we linearly interpolated the water content measurements at time 0 d and the saturation water content at the lower boundary. The value between the upper boundary and the topmost measurement was set constant to that of the measurement. To represent the uncertainty of measurements and interpolation, we used a normal distribution with a standard deviation of 0.01 and a spatial covariance given by the fifth-order piecewise rational function by Gaspari and Cohn (1999) with a cutoff length scale of 20 cm.

For the initial guess of the Mualem–van Genuchten parameters we used univariate normal distributions, whose mean was set to the respective true value. By choosing the true values for the mean, the parameter estimation becomes easier for the EnKF. This is intended to reduce the impact of the filter performance on the different synthetic experiments, as it is not the goal of this paper to investigate the performance of the EnKF itself. The standard deviations for the Mualem–van Genuchten parameters were set to σₐ = 2.0 m⁻¹, and σₙ log₁₀(K₀) = 1.0. For n, we chose different uncertainties for the sandy loam and the silt loam cases to avoid too small of values, which create difficulties for numeric solvers due to dramatically increasing sensitivity of the hydraulic conductivity to small changes in the water content close to saturation (Vogel, van Genuchten, & Cislerova, 2000). For the sandy loam, we set σₐ = 0.25, whereas for the silt loam we set σₐ = 0.1.

The data assimilation was only performed on one-dimensional domains with a reduced Miller scaling field representation explained in the subsection below. Synthetic measurements, on the other hand, were conducted on one- and two-dimensional domains with a fully resolved Miller field. These two cases (two-dimensional truth and one-dimensional data assimilation, and one-dimensional truth and one-dimensional data assimilation) are in the following only referred to as “2D” and “1D,” respectively. Hence, there are two sources of model uncertainty: the reduced representation of heterogeneity in both 2D and 1D cases, and the influence of horizontal fluxes in the 2D cases. The resulting soil hydraulic parameters and Miller scaling field are not expected to match the true values because the data assimilation process estimates an effective representation that predicts the measured dynamics within the limited frame of the hydraulic model used in it.

### 3.5 Miller scaling representation

We chose to not include the entire Miller scaling field into the augmented state, because local water content measurements typically convey this information insufficiently, if they cannot resolve the field. Therefore, we only estimated five Miller scaling factors, one at each measurement position and one at the surface, and aim for an effective description where the measurements cannot resolve the heterogeneity. We constructed this reduced representation for the Miller scaling factor at depth z as

\[
\xi(z) = 1 + \varphi(z)^T \mathbf{A}^{-1} [\xi - 1]
\]

(17)

where \(\xi\) is the vector containing the five Miller scaling factors \(\xi_i\). The vector \(\varphi(z)\) consists of Gaussian kernel functions \(\varphi_i(z)\) whose means are the positions of the corresponding Miller scaling factors. These functions are not normalized, leading to a maximum value of 1.0 at the location of the respective Miller scaling factor. For the standard deviation of the kernels, we chose 10 cm. This way the difference to no scaling (scaling factor of 1) of each Miller scaling factor is weighted with the kernel function depending on the distance to its respective location. The matrix \(\mathbf{A}\) accounts for possible correlations of Miller scaling factors close to each other to avoid values exceeding the individual Miller scaling factors. This is achieved by setting the diagonal entries to \(A_{ii} = 1\) and the off-diagonal entries to \(A_{ij} = \varphi_i(z)\varphi_j(z)\).
The logarithm of each of the five scaling factors was included in the augmented state. Their initial guess was drawn from normal distributions with standard deviation \( \sigma_{\log(\xi)} = 0.2 \). We chose the respective distribution mean to be the logarithm of the true value of the Miller scaling field at the corresponding location.

### 3.6 Implementation

We solved the Richards equation numerically with the DORiE software package (Riedel, Ospinar De Los Rios, Klein, Häfner, & Rixinger, 2019), which is based on the Distributed and Unified Numerics Environment (DUNE) (Blatt et al., 2016) and DUNE-PDELab (Bastian, Heimann, & Marnach, 2010).

For the spatial discretization, we chose a vertical grid resolution of 1 cm in all simulations. In the one-dimensional simulations, one grid cell spanned the entire domain width of 2 m. In two-dimensional simulations, we chose a horizontal resolution of 1 cm as well, which resulted in a square grid of 200 \( \times \) 200 cells. The spin-up phase simulations were conducted with the symmetric weighted interior penalty variant of the discontinuous Galerkin method. During the synthetic experiments, the ensemble states were propagated by a finite volume solver with upwinding. The simpler solver and the relatively low resolution were chosen to reduce the computational effort, which enabled a larger ensemble size. For the temporal discretization, the diagonally implicit second order method proposed by Alexander (1977) was used. We used an adaptive time step scheme that varies simulation time steps between 1 s and 1 h depending on the solution's convergence rate.

For computing the state propagation of a single ensemble member, we calculate a matric head state from the water content state used in the EnKF via the hydraulic parameterization. To that end, the Miller scaling field is reconstructed from the Miller scaling factors stored in the augmented state and their respective positions following Equation 17. The resulting field in the solver has one constant scaling value per grid cell. DORiE then propagates the matric head state by using the current boundary conditions. Before returning the new state back to the EnKF, it is transferred into water content again.

### 3.7 Evaluation

After the data assimilation phase, the propagation of the ensemble was continued without any state or parameter update during the verification phase. To evaluate and compare the performance of the different cases during the verification phase, we calculated summarizing measures.

For the water content we calculated the RMSE as

$$\text{RMSE}_w = \sqrt{\frac{1}{N_t N_m} \sum_{t=1}^{N_t} \sum_{m=1}^{N_m} (\theta_{\text{true},t,m} - \theta_{\text{true},t,m})^2}$$

where \( \theta_{\text{true},t,m} \) and \( \theta_{\text{true},t,m} \) are the synthetic true water content and mean water content for measurement location \( m \) at time step \( t \), respectively. The total number of measurement locations is \( N_m = 4 \), and the total number of 2-h time steps (8-h time steps for the low-flux boundary condition) during the verification phase is \( N_t = 336 \).

For the fluxes, we calculated the ensemble mean of the predicted vertical flux at three different locations—(a) at the surface at 0 cm, (b) in between the water content measurements at a depth of 50 cm, and (c) below the depth covered by the water content measurements at 100 cm—and compare it with the true value. Note that the synthetic truth for the 1D cases, as well as the predictions for all cases, are one-dimensional with a horizontally homogeneous flux, whereas the synthetic truth for the 2D cases is two-dimensional and the flux direction and magnitude varies horizontally. Therefore, we use the horizontal average of the vertical flux component at the different depths as the true flux.

For the depth of 100 cm, we calculated the relative cumulative flux error as

$$\Delta J = \frac{\sum_{t=1}^{N_t} (j_{\text{true},t} - j_{\text{true,0}})}{\sum_{t=1}^{N_t} J_{\text{true,0}}}$$

where \( j_{\text{true},t} \) and \( j_{\text{true,0}} \) are the true and mean predicted vertical flux at the depth of 100 cm at time step \( t \). The total number of time steps is again \( N_t = 336 \).

Additionally, we calculated the Nash–Sutcliffe efficiency (NSE),

$$\text{NSE}_j = \frac{\sum_{t=1}^{N_t} (j_{\text{true},t} - j_{\text{true,0}})^2}{\sum_{t=1}^{N_t} (j_{\text{true},t} - \bar{j})^2}$$

where \( j_{\text{true}} \) is the temporal average of the true fluxes at 100-cm depth.

To quantify the relevance of the two-dimensional flux, we calculated the average horizontal flux ratio (AHFR) for the synthetic truth at horizontal position 100.5 cm where the measurements are located:

$$\text{AHFR} = \frac{1}{N_t} \sum_{t=1}^{N_t} \frac{<j_{\text{true},t}>}{<j_{\text{true},t}>}$$

where \( <j_{\text{true},t}> \) is the spatial average of the absolute horizontal flux in the top 100 cm and \( <j_{\text{true},t}> \) is the spatial average
FIGURE 4 Water content predictions for the one-dimensional truth and one-dimensional data assimilation (1D) and two-dimensional truth and one-dimensional data assimilation (2D) case of the Default scenario with the Miller field shown in Figure 1. The top panels show the water contents at the four different observation depths. The dashed line marks the synthetic truth, and the dots the noisy measurements. The solid line is the mean of the ensemble prediction. The standard deviation of the ensemble is hardly visible. The middle panels show the absolute (abs.) residual between synthetic truth and ensemble mean. The bottom panel shows the relative (rel.) residual calculated by dividing the absolute residual by the ensemble standard deviation. The horizontal black dashed lines indicate one and three standard deviations of the absolute vertical flux in the top 100 cm at each time step t.

4 RESULTS AND DISCUSSION

In this section, we first discuss the results for the water content and flux predictions for a single example and summarize the results for all synthetic experiments. We then explore the effect of surface heterogeneity for the particular example with the strongest impact, as well as the effect on the prediction of evaporation fluxes. The last subsection presents the details of the example with the largest AHFR value to highlight the impact of two-dimensional flow.

4.1 Water content and flux prediction

Figure 4 shows the water content predictions during the verification phase for the 1D and 2D case of the Default scenario for the Miller random field shown in Figure 1. The water content predictions at all four measurement locations agree very well with the synthetic measurements and truth. The uncertainty predicted by the ensemble is small. The residuals between synthetic truth and predicted mean water contents are low as well. However, there is an increase from the 1D to the 2D case. In the 2D case, the most significant errors are spikes in the residual between 40 and 44 d, due to an inaccurate prediction of the arrival time of the infiltration front.

The RMSE over all four measurement locations is small with RMSEθ = 0.0025 for the 1D case and a slight increase to RMSEθ = 0.0035 for the 2D case. This finding is confirmed by the average RMSE over all five seeds for the Default scenario. It increases from an average RMSEθ = 0.0022 for the 1D cases to RMSEθ = 0.0040 in the 2D cases (Table 2).

The small errors demonstrate that in this scenario, the EnKF was able to estimate a reduced one-dimensional representation of the heterogeneous soil, based on only four local water content measurements. This reduced representation was sufficient to predict the water contents at the measurement locations with high accuracy for (a) a truth generated with a more detailed one-dimensional
TABLE 2 Summary measures of the one-dimensional truth and one-dimensional data assimilation (1D) and two-dimensional truth and one-dimensional data assimilation (2D) cases for all scenarios (see Table 1), where the overline denotes the average over the five different seeds

| Scenario       | AHFR 2D | RMSE<sub>θ</sub> 1D  | RMSE<sub>θ</sub> 2D | NSE<sub>θ</sub> 1D | NSE<sub>θ</sub> 2D | ΔJ 1D | ΔJ 2D |
|----------------|---------|------------------|------------------|-----------------|-----------------|------|------|
| Default        | 0.14    | 0.0022           | 0.0040           | 0.02            | 0.10            | 0.97 | 0.84 |
| MedVar         | 0.18    | 0.0029           | 0.0049           | 0.03            | 0.13            | 0.92 | 0.73 |
| HighVar        | 0.24    | 0.0051           | 0.0072           | 0.05            | 0.15            | 0.78 | 0.36 |
| LowFlux        | 0.17    | 0.0022           | 0.0035           | 0.03            | 0.08            | 0.94 | 0.84 |
| Horizontal     | 0.09    | 0.0021           | 0.0022           | 0.02            | 0.05            | 0.98 | 0.97 |
| Vertical       | 0.09    | 0.0019           | 0.0041           | 0.01            | 0.06            | 0.97 | 0.92 |
| Silt           | 0.04    | 0.0023           | 0.0026           | 0.08            | 0.06            | 0.94 | 0.96 |
| SiltLowFlux    | 0.04    | 0.0015           | 0.0014           | 0.05            | 0.07            | 0.98 | 0.99 |

Note. AHFR, average horizontal flux ratio; RMSE<sub>θ</sub>, RMSE of the water content; NSE<sub>θ</sub>, Nash-Sutcliffe efficiency; ΔJ, relative cumulative flux error. The NSE<sub>θ</sub> and ΔJ are calculated at a depth of 100 cm. Note that we use the absolute value of the relative cumulative fluxes to calculate ΔJ. The AHFR of all 1D cases are 0.

heterogeneity as well as (b) a truth generated with a two-dimensional heterogeneous Miller-similar medium. However, the results also show that the errors increase in case of unrepresented two-dimensional flux.

To see how well the error is predicted by the ensemble spread, we divided the absolute residual by the standard deviation of the ensemble. This relative residual is displayed in the bottom panels in Figure 4. They show that the uncertainty is slightly underestimated, with a stronger underestimation in the 2D case. The ensemble can only predict uncertainties represented in the augmented state. Errors due to the two-dimensional heterogeneity in the 2D case and the reduced representation of the heterogeneity in the 1D and 2D case cannot be fully represented. This can lead to the underestimated uncertainties. Compared with the absolute residuals, the structure of the relative residuals changes as well. The peaks due to the inaccurate prediction of the arrival time become smaller in relation to the other errors. This shows that the ensemble, at least in part, is able to predict the uncertainty in the arrival time.

To investigate the relation of the performance of water content predictions and heterogeneity over all scenarios and cases, Figure 5 shows the RMSE<sub>θ</sub> over the AHFR. The AHFR is calculated from the synthetic truth and indicates the strength of two-dimensional flux in the profile. Per definition, AHFR values of the 1D cases are 0. We therefore use the AHFR of the 2D cases for the corresponding 1D cases as well, as indicated in Table 2.

Figure 5 shows that RMSE<sub>θ</sub> generally increases with the AHFR. We find that RMSE<sub>θ</sub> is below 0.012 for all cases, and even below 0.005 for a majority of them. This means that most cases show a smaller RMSE<sub>θ</sub> than the measurement uncertainty of TDR probes of around 0.007 (Jauermann & Roth, 2017). This confirms that the water content predictions have a high accuracy, as indicated by the example case in Figure 4. Most 2D cases have a larger RMSE<sub>θ</sub> than their 1D counterparts due to the unrepresented two-dimensional flux.

The variance of the underlying Miller field rises through scenarios Default, MedVar, and HighVar. This increases both AHFR and RMSE<sub>θ</sub>, following the overall trend in Figure 5. The same is true when comparing scenarios Default and Silt. Scenario Default uses sandy loam as reference material, whereas scenario Silt uses silt loam. The material properties of silt loam with the same Miller scaling field applied differ less for the occurring fluxes than for sandy loam, since the system is closer to the crossing points of the hydraulic conductivity functions K(h<sub>m</sub>). Therefore, the flow field heterogeneity is lower and AHFR and RMSE<sub>θ</sub> become smaller.

Comparing scenarios Default and LowFlux, AHFR and RMSE<sub>θ</sub> behave oppositely: the AHFR increases, but the RMSE<sub>θ</sub> becomes smaller. The reason for this is that Default uses the standard boundary condition, whereas LowFlux uses the low flux boundary condition. The material properties differ stronger for the lower fluxes, since the system is farther away from the crossing points of the hydraulic conductivity functions K(h<sub>m</sub>), leading to a larger AHFR. The lower fluxes also cause smaller changes in the water content, which in combination with an increasing slope of the
FIGURE 6  Flux predictions for the 1D and 2D case of the Default scenario with the Miller field shown in Figure 1. The top panels show the flux at three different depths. The dashed line marks the synthetic truth that is calculated by averaging horizontally over the entire domain. The solid line is the mean of the ensemble predictions and the pale area around shows the standard deviation of the ensemble. The bottom panels show the residuals between synthetic truth and ensemble mean.

hydraulic conductivity function $K(\theta)$ towards lower water contents then lead to smaller errors and a reduced RMSE$\theta$. The same explanation is true for the scenarios with silt loam, Silt and SiltLowFlux, where the RMSE$\theta$ is smaller for SiltLowFlux.

Increasing the horizontal or vertical correlation length of the Miller scaling field in scenarios Horizontal and Vertical promotes vertical flow paths, manifesting in a lower AHFR compared with the isotropic setting. In the former, this is due to the scaling field becoming more similar to the respective 1D case. Indeed, the difference of RMSE$\theta$ between 1D and 2D case becomes nearly negligible. The impact on the RMSE$\theta$ of the Vertical scenario is low, the differences between 1D and 2D cases persist and the RMSE$\theta$ values remain similar to the ones of the Default scenario.

Synthetic experiments offer the advantage that we can analyze the flux predictions as well. Figure 6 shows the flux predictions during the verification phase for the 1D and 2D case of the Default scenario for the Miller random field shown in Figure 1. The fluxes are predicted well for both the 1D and 2D case. Again, the residuals are larger for the 2D case than for the 1D case. At the depths of 50 and 100 cm, they are mainly due to the inexact prediction of the arrival time of the infiltration fronts. At the upper boundary, at a depth of 0 cm, the flux during the rain events is given through the boundary condition and the residuals are 0. The evaporation flux is determined through a matric potential leading to larger residuals, especially in the 2D case, due to the inexact onset time of the evaporation and due to the maximum evaporation flux after the onset.

The NSE$_j$ at a depth of 100 cm is 0.95 for the 1D case and 0.94 for the 2D case in Figure 6. The relative cumulative flux error $\Delta J$ at 100 cm is 0.01 for the 1D case and 0.07 for the 2D case. In this example, the fluxes are predicted well, but again better for the 1D case than for the 2D case. This is also true for the average over all five seeds for NSE$_j$ and $\Delta J$. The 1D cases yield $\overline{NSE}_j = 0.97$ and $|\overline{\Delta J}| = 0.02$, whereas the 2D cases result in $\overline{NSE}_j = 0.84$ and $|\overline{\Delta J}| = 0.10$ (see Table 2).

To investigate the fluxes in more detail, Figure 7 shows the NSE$_j$ over the AHFR for all scenarios and cases. Most NSE$_j$ are close to 1 and larger than 0.8. Even for large values of the AHFR, corresponding to a stronger heterogeneous flux, NSE$_j$ values close to 1 are possible. However, the spread of NSE$_j$ increases with AHFR and makes lower NSE$_j$ more likely. Although a 1D case can have a lower NSE$_j$ than the corresponding 2D case for individual realizations, most 2D cases have smaller NSE$_j$ than the 1D cases for all Scenarios with sandy loam. Only for scenarios Silt and SiltLowFlux do the 1D cases have smaller NSE$_j$ than the 2D cases. However, these differences are small with NSE$_j$ values close to 1 (see Table 2). We also note that there is one particular case that has even a negative NSE$_j = -0.57$. This is a realization of scenario HighVar, which features the largest variance of the Miller scaling field.
The relative cumulative flux error $\Delta J$ at a depth of 100 cm over the AHFR is shown in Figure 8. Again, no clear tendency with the AHFR can be seen. However, the 2D cases tend to have larger errors than the 1D cases (see also Table 2). In particular, we note a group of five different 2D cases with $\Delta J < -0.15$. The worst performance is again part of scenario *HighVar*, which has a relative cumulative flux error of $\Delta J < -0.49$.

### 4.2 Effects of surface heterogeneity

We investigate the outlier noted in Figure 8 in more detail. This case also corresponds to the single case with the negative NSE$_J$ in Figure 7 and the largest RMSE$_\theta$ in Figure 5. The water content predictions of this case are shown in Figure 9a, and the flux predictions are shown in Figure 10a.

The water content predictions show strong deviations from the true values. The RMSE$_\theta$ over all four measurement locations is 0.012. However, the deviations are stronger closer to the surface. The RMSE at the measurement depth of 9.5 cm is 0.017, where the largest deviations occur during the infiltration events. The flux predictions show large deviations at all depths. The evaporation at a depth of 0 cm is overestimated strongly, whereas the infiltration flux at 50- and 100-cm depth is underestimated. The cumulative flux at 100 cm is underestimated by 49%.

The reason for this poor performance is in the specific realization of the Miller field. In this particular case, the scaling factor at the surface in the middle of the domain (where the measurements are located) is very small with $\xi = 0.09$. This corresponds to smaller pore sizes and consequently higher evaporation fluxes at more negative potentials compared with larger scaling factors.

The data assimilation is able to follow the trend of the true Miller field correctly and estimates a small Miller scaling factor of $\xi = 0.11$ at the surface. Note, however, that the Miller scaling factors cannot be compared entirely because the relevant material properties at the surface are also influenced by the estimated reference material. Although the low scaling factor in the estimated 1D field influences the evaporation flux over the entire width of the domain, it is a localized deviation in the 2D synthetic truth with little influence on the horizontally averaged fluxes. This leads to incorrect estimation of the evaporation, which influences the subsequent infiltration, causing lower fluxes at greater depths.

To investigate the impact of the Miller scaling at the surface in more detail, we modified the data assimilation setup and removed the Miller scaling factor at the surface from the EnKF estimation. The chosen interpolation to create the reduced representation of the Miller field approaches scaling factors of 1 in sufficiently large distances from the locations of the Miller scaling factors represented in the augmented state (see Equation 17). In
**FIGURE 9** Water content predictions for (a) the two-dimensional truth and one-dimensional data assimilation (2D) case with the worst performance with respect to the RMSE of water content ($\text{RMSE}_\theta$), Nash–Sutcliffe efficiency ($\text{NSE}_j$), and relative cumulative flux error ($\Delta J$) and (b) the same case, but with modified data assimilation setup, where the Miller scaling factor at the surface is removed from the augmented state, which relaxes the scaling field towards values of 1 there. The top panels show the water contents at the four different observation depths. The dashed line marks the synthetic truth, and the dots the noisy measurements. The solid line is the mean of the ensemble prediction, and the pale area around it shows the standard deviation of the ensemble. The bottom panels show the absolute (abs.) residual between synthetic truth and ensemble mean.

**FIGURE 10** Flux predictions for (a) the two-dimensional truth and one-dimensional data assimilation (2D) case with the worst performance with respect to RMSE of water content ($\text{RMSE}_\theta$), Nash–Sutcliffe efficiency ($\text{NSE}_j$), and relative cumulative flux error ($\Delta J$) and (b) the same case, but with modified data assimilation setup, where the Miller scaling factor at the surface is removed from the augmented state, which relaxes the scaling field towards values of 1 there. The top panels show the flux at three different depths. The dashed line marks the synthetic truth, which is calculated by averaging horizontally over the entire domain. The solid line is the mean of the ensemble predictions, and the pale area around it shows the standard deviation of the ensemble. The bottom panels show the residuals between synthetic truth and ensemble mean.
this modified example, this resulted in a scaling factor of $\xi = 0.58$ at the surface.

Figure 9b shows that the water content predictions of this modified estimation changed structurally. Now the largest errors occur during the evaporation phases and extend over longer time periods. The $\text{RMSE}_\theta$ increases to 0.014 because we reduced the degrees of freedom in the augmented state of the EnKF. The now larger Miller scaling factor at the surface deviates stronger from the true scaling, which worsens the water content predictions directly below. However, the larger errors also lead to an increased uncertainty in the predictions.

Figure 10b shows that the flux predictions of the modified estimation improved strongly. The residuals for the flux at the surface are very small. The larger Miller scaling factor at the surface keeps the material properties closer to the reference material properties, which are more representative for the average evaporation of the heterogeneous field. As a consequence, the fluxes at the depths of 50 and 100 cm improve strongly as well. Nevertheless, relevant errors remain. The $\text{NSE}_j$ for the flux at 100 cm only increases to 0.64, and the cumulative flux is still underestimated by 14%. The errors mainly stem from the inexact predictions of arrival time and maximum flux of the infiltration fronts.

We also achieve a similar predictive performance, if we use the true flux across the upper boundary condition of the synthetic truth and use this flux as the boundary condition in the data assimilation phase, as well as the verification phase. In this case, the description of the water content also worsens to an $\text{RMSE}_\theta$ of 0.014, whereas the flux predictions improve to an $\text{NSE}_j$ of 0.59 and a cumulative flux error of 12% for the depth of 100 cm.

These examples demonstrate that it is imperative to describe the evaporation accurately. Then, the one-dimensional representation of the two-dimensional synthetic truth can predict cumulative fluxes adequately. However, this may lead to a worse description of local water content measurements and therefore requires additional information. Contrary to a real-world experiment, we could safely omit the estimation of the Miller scaling factor at the surface because we knew that the reference material properties do not change. Alternatively, we were able to use the true evaporation values of the synthetic truth, which is also difficult in real-world cases. More importantly, this causes unphysical situations, when the soil is locally unable to provide the average evaporation flux.

In Figure 11, we take a closer look at the effect of the correct prediction of the evaporation. It shows the cumulative flux error at a depth of 100 cm (not the relative cumulative flux error $\Delta J$) over the cumulative flux error at the surface for all scenarios and cases. Note that the error at the surface only stems from the evaporation, since the true flux is used during the infiltration. The cumulative flux errors show an almost perfect identity relation. Due to the mass balance a relation between the cumulative flux errors at different depths is not surprising. However, it holds in these examples even on the rather short timescales and with a more negative cumulative flux at 100 cm than at the surface (on average over all 80 experiments, $\sim$19 mm). This emphasizes the relevance of a correct estimation of the evaporation. It also shows that the average change of water content in the two-dimensional reality could be predicted well by the one-dimensional representations. This indicates that the EnKF was able to find an effective description of the soil water characteristics of the two-dimensional reality based on information from sparse local measurements only.

### 4.3 Effects of two-dimensional flow

We further investigate the effect of two-dimensional flow on the predictions with another example. We choose the case with the largest occurring AHFR value of 0.33 because we expect to see the strongest effects there. The 1D case yields $\text{RMSE}_\theta = 0.0058$ over all measurement locations, a relative cumulative flux error $\Delta J = 0.03$, and $\text{NSE}_j = 0.93$ at the depth of 100 cm, whereas the 2D case has $\text{RMSE}_\theta = 0.0078$, $\Delta J = 0.02$, and $\text{NSE}_j = 0.27$. The corresponding predictions of water content and flux are shown in Figures 12 and 13.
FIGURE 12  Water content predictions for the one-dimensional truth and one-dimensional data assimilation (1D) and two-dimensional truth and one-dimensional data assimilation (2D) case of the HighVar scenario with the largest average horizontal flux ratio (AHFR). The top panels show the water contents at the four different observation depths. The dashed line marks the synthetic truth, and the dots the noisy measurements. The solid line is the mean of the ensemble predictions, and the pale area around it shows the standard deviation of the ensemble. The bottom panels show the absolute residual between synthetic truth and ensemble mean.

FIGURE 13  Flux predictions for the one-dimensional truth and one-dimensional data assimilation (1D) and two-dimensional truth and one-dimensional data assimilation (2D) case of the HighVar scenario with the largest average horizontal flux ratio (AHFR). The top panels show the flux at three different depths. The dashed line marks the synthetic truth, which is calculated by averaging horizontally over the entire domain. The solid line is the mean of the ensemble predictions, and the pale area around it shows the standard deviation of the ensemble. The bottom panels show the residuals between synthetic truth and ensemble mean.
Errors in both water content and flux predictions are larger than in the first example case but show the same overall behavior. As indicated by the summary measures, the errors in the 2D case are larger than in the 1D case. However, we particularly note the large error at the lowest measurement location in the 1D case, which stems from the reduced representation of heterogeneity. In the 2D case, the largest errors occur because the water contents during the first infiltration event are overestimated and the arrival time of the last event is predicted too early. It demonstrates that prediction errors depend on the specific configuration and that we cannot conclude general statements about expected errors. The flux errors in the 2D case are larger than in the 1D case, but the cumulative flux as well as the evaporation are predicted well in both cases.

We conclude again that the reduced one-dimensional representation predicts the soil water dynamics sufficiently well if exact arrival times and magnitudes of infiltration fronts are not relevant and if the evaporation is predicted accurately. However, it is not clear if this would hold for further processes such as solute transport, which are based on the description of the water dynamics. There, the actual flow paths may have a larger impact and the reduced description may become insufficient for predictions.

For the same example, we investigate the effect of the reduced one-dimensional representation on the estimated material properties. Figures 14a and 14b show the true and estimated soil water characteristics and hydraulic conductivity functions at the four measurement locations. To compensate the nonrepresented heterogeneity, these estimated effective material properties deviate considerably from the true material properties in both the 1D and 2D case. However, the deviations still remain smaller than the variations between all measurement locations. Although the deviations do not show a clear direction in the 1D case, the deviations in the 2D case show a consistent bias. For the same water content, matric potentials are more negative and the hydraulic conductivities are larger.

Figure 14c shows the true scaling field in the middle of the domain, where the measurements are located, together with the reduced representations based on the estimation of Miller scaling factors at the surface and the measurement locations only. As expected, the reduced representations are unable to reproduce the true field. Only close to the surface, where measurements are denser, is the structure better resolved. For lower measurement densities, the chosen interpolation relaxes towards scaling factors of 1. Note that the scaling factors cannot be compared directly, since the material properties are also altered by the reference material properties, which are estimated as well. Therefore, we compare these resulting material properties at the measurement locations only.

The differences between true material properties and estimated effective material properties, as well as the
5 | CONCLUSIONS

In this study, we quantitatively explored the effect of unrepresented small-scale heterogeneity on prediction uncertainties for soil water movement, when estimating effective soil hydraulic material properties based on sparse local water content measurements. For this, we generated a set of synthetic experiments, where the synthetic truths feature one- and two-dimensional small-scale heterogeneity described through Miller-similar media. The estimations of the effective hydraulic material properties are based on four local water content measurements in a single vertical profile only. The effective description uses a reduced representation of the vertical heterogeneity through a reference material but allows local deviations at the measurement locations through Miller scaling to consider the local deviations of the measurements. We used a data assimilation method, the EnKF, for the estimation and investigated the predictive capabilities with respect to soil water content, as well as fluxes of the reduced effective representation.

For the synthetic experiments in this study, the estimated reduced representation proved to be sufficient for water content and flux predictions in most cases. In detail, the uncertainty in the water content predictions rises with increasing heterogeneity but remains sufficiently low. Still, these uncertainties are underestimated with the EnKF. Flux errors are larger and more relevant. They mainly stem from inexact prediction of arrival times and magnitudes of infiltration events.

The limitation of the reduced description lies in the capability to predict evaporation accurately, which was not possible in few cases. This capability of the reduced representation depends strongly on the location of the measurements within the two-dimensional heterogeneous field. In the case of strong local heterogeneity at the surface above the measurement profile, the estimation based on local water content measurements becomes more difficult due to the highly nonlinear material properties. Resulting cumulative errors of the evaporation translated directly into errors of the cumulative flux below the soil profile, even on the short time scales of the synthetic examples in this study. We demonstrated on one example that in such a particular case, additional information about the evaporation or the structure of the heterogeneity is necessary to estimate effective material properties, which allow the description of the cumulative fluxes as well. Therefore, if the evaporation is represented accurately and if the exact timing of infiltration fronts and flux magnitudes are not important, the estimation and description of the full heterogeneity are not necessary and can be achieved based on few local measurements.

Even though the estimated effective soil hydraulic material properties are allowed to adjust locally to describe the water content measurements, they differ strongly from the true material properties to compensate the reduced representation of the heterogeneity. This highlights the relevance of in situ measurements for inverse estimation of material properties and demonstrates the limitations of direct determination of “accurate” material properties (e.g., based on soil samples).

As concluding remarks, we would like to point out that we limited the description of the synthetic truth to up to two dimensions and did not include a three-dimensional truth. We expect that this would only lead to a quantitative modification of our findings, not to a qualitative one. Finally, we would like to emphasize that our results only apply to the soil water dynamics. Driving other processes like solute transport or chemical and microbial transformations with the reduced soil water dynamics cannot be expected to yield a faithful representation.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS
HH Bauser designed, implemented, performed and analyzed the presented study with contributions from L Riedel. L Riedel and D Berg provided computational software. All authors participated in continuous discussions. HH Bauser and L Riedel prepared the manuscript with contributions from all authors.

ACKNOWLEDGMENTS
We thank editor Sjoerd Van Der Zee and two anonymous reviewers for their comments, which helped to improve this paper. H.H. Bauser was funded by the Deutsche Forschungsgemeinschaft (DFG) through Project BA 6635/1-1. L. Riedel was funded by the Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg (Az 33-7533.-30-20/6/2). The Biosphere 2 facility and the capital required to conceive and construct LEO were provided through a charitable donation from the Philecology
Foundation, and its founder, Mr. Edward Bass. We gratefully acknowledge that charitable donation. Additional funding support was provided by the Water, Environmental, and Energy Solutions (WEEES) initiative at the University of Arizona and by the Office of the Vice President of Research at the University of Arizona. These funding sources are gratefully acknowledged.

ORCID

Hannes H. Bauser https://orcid.org/0000-0002-6617-8114
Lukas Riedel https://orcid.org/0000-0002-4667-3652
Kurt Roth https://orcid.org/0000-0003-2634-8825

REFERENCES

Alexander, R. (1977). Diagonally implicit Runge-Kutta methods for stiff O.D.E.’s. SIAM Journal on Numerical Analysis, 14, 1006–1021. Retrieved from https://www.jstor.org/stable/2156678
Andersen, J. L. (2007). A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. Monthly Weather Review, 127, 2761–2758. https://doi.org/10.1175/1520-0493(1999)127(2761:ACIAF)2.0.CO;2
Andersen, J. L., & Anderson, S. L. (1999). A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. Monthly Weather Review, 127, 2741–2758. https://doi.org/10.1175/1520-0493(1999)127(2741:AMCIOT)2.0.CO;2
Bastian, P., Heimann, F., & Marnach, S. (2010). Generic implementation of finite element methods in the Distributed and Unified Numerics Environment (DUNE). Kybernetica, 46, 294–315. Retrieved from https://dml.cz/dmlcze/140745.
Bauser, H. H., Berg, D., Klein, O., & Roth, K. (2018). Inflation method for ensemble Kalman filter in soil hydrology. Hydrology and Earth System Sciences, 22, 4921–4934. https://doi.org/10.5194/hess-22-4921-2018
Bauser, H. H., Jaumann, S., Berg, D., & Roth, K. (2016). EnKF with closed-eye period—Towards a consistent aggregation of information in soil hydrology. Hydrology and Earth System Sciences, 20, 4999–5014. https://doi.org/10.5194/hess-20-4999-2016
Blatt, M., Burchardt, A., Dedner, A., Engwer, C., Fahlke, J., Flemisch, B., … Sander, O. (2016). The Distributed and Unified Numerics Environment, Version 2.4. Archive of Numerical Software, 4(100), 13–29. https://doi.org/10.11588/ans.2016.100.26526
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, 4251–4266. https://doi.org/10.5194/hess-22-4251-2018
Brandhorst, N., Erdal, D., & Neuwiler, 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.adwres.2017.10.022
Burgers, G., Jan van Leeuwen, P., & Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. Monthly Weather Review, 126, 1719–1724. https://doi.org/10.1175/1520-0493(1998)126d1719:ASITEKe2.0.CO;2
Carsel, R. F., & Parrish, R. S. (1988). Developing joint probability distributions of soil water retention characteristics. Water Resources Research, 24, 755–769. https://doi.org/10.1029/WR024i005p00755
Chaudhuri, A., Hendricks Franssen, H. J., & Sekhar, M. (2018). Iterative filter based estimation of fully 3D heterogeneous fields of permeability and Mualem-van Genuchten parameters. Advances in Water Resources, 122, 340–354. https://doi.org/10.1016/j.advwatres.2018.10.023
Erdal, D., Neuwiler, I., & Wollschläger, U. (2014). Using a bias aware EnKF to account for unresolved structure in an unsaturated zone model. Water Resources Research, 50, 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: Oceans, 99, 10143–10162. https://doi.org/10.1029/94JC00572
Hendricks Franssen, J. H., & 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
Gaspari, G., & Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. Quarterly Journal of the Royal Meteorological Society, 125, 723–757. https://doi.org/10.1002/qj.4971255417
Gevaert, A. I., Teuling, A. J., Uijlenhoet, R., DeLong, S. A., Huaxman, T. E., Pangle, L. A., … Zeng, X. (2014). Hillslope-scale experiment demonstrates the role of convergence during two-step saturation. Hydrology and Earth System Sciences, 18, 3681–3692. https://doi.org/10.5194/hess-18-3681-2014
Hamill, T. M., Whitaker, J. S., & Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. Monthly Weather Review, 129, 2776–2790. https://doi.org/10.1175/1520-0493(2001)129<2776:DDFBEV>2.0.CO;2
Hopmans, J. W., Šimůnek, J., Romano, N., & Durner, W. (2002). Inverse methods. In J. H. Dane & C. G. Topp (Eds.), Methods of soil analysis. Part 4. Physical methods (pp. 963–1008). Madison, WI: SSSA. https://doi.org/10.2136/sssabookser5.4.c40
Houtekamer, P. L., & Mitchell, H. L. (1998). Data assimilation using an ensemble Kalman filter technique. Monthly Weather Review, 126, 796–811. https://doi.org/10.1175/1520-0493(1998)126d0796:DAUAEK2.0.CO;2
Houtekamer, P. L., & Mitchell, H. L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. Monthly Weather Review, 129, 123–137. https://doi.org/10.1175/1520-0493(2001)129<123:aekf2d>AWEKE2.0.CO;2
Jaumann, S., & Roth, K. (2017). Effect of unrepresented model errors on estimated soil hydraulic material properties. Hydrology and Earth System Sciences, 21, 4301–4322. https://doi.org/10.5194/hess-21-4301-2017
Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. Journal of Basic Engineering, 82, 35–45.
Klein, O. (2016). DUNE-Randomfield (Version 2.6). GitLab repository. Retrieved from https://gitlab.dune-project.org/oklein/dune-randomfield
Li, C., & Ren, L. (2011). Estimation of unsaturated soil hydraulic parameters using the ensemble Kalman filter. Vadose Zone Journal, 10, 1205–1227. https://doi.org/10.2136/vzj2010.0159

Li, H., Kalnay, E., & Miyoshi, T. (2009). Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. Quarterly Journal of the Royal Meteorological Society, 135, 523–533. https://doi.org/10.1002/qj.371

Riedel, L., Ospinar De Los Rios, S., Klein, O., Häfner, D., & Riexinger, F. (2019). DORIE: The DUNE Operated Richards Equation Solving Environment. GitLab repository. Retrieved from https://ts-gitlab.iup.uni-heidelberg.de/dorie/dorie

Roth, K. (1995). Steady state flow in an unsaturated, two-dimensional, macroscopically homogeneous, Miller-similar medium. Water Resources Research, 31, 2127–2140. https://doi.org/10.1029/95WR00946

Scudeler, C., Pangle, L., Pasetto, D., Niu, G. Y., Volkmann, T., Paniconi, C., … Troch, P. (2016). Multiresponse modeling of variably saturated flow and isotope tracer transport for a hillslope experiment at the Landscape Evolution Observatory. Hydrology and Earth System Sciences, 20, 4061–4078. https://doi.org/10.5194/hess-20-4061-2016

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

Song, X., Shi, L., Ye, M., Yang, J., & Navon, I. M. (2014). Numerical comparison of iterative ensemble Kalman filters for unsaturated flow inverse modeling. Vadose Zone Journal, 13(2). https://doi.org/10.2136/vzj2013.05.0083

Vereecken, H., Kasteel, R., Vanderborght, J., & Harter, T. (2007). Upscaling hydraulic properties and soil water flow processes in heterogeneous soils. Vadose Zone Journal, 6, 1–28. https://doi.org/10.2136/vzj2006.0055

Vogel, H.-J. (2019). Scale issues in soil hydrology. Vadose Zone Journal, 18(1). https://doi.org/10.2136/vzj2019.01.0001

Vogel, T., van Genuchten, M. Th., & Cislerova, M. (2000). Effect of the shape of the soil hydraulic functions near saturation on variably-saturated flow predictions. Advances in Water Resources, 24, 133–144. https://doi.org/10.1016/S0309-1708(99)00037-3

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

Wang, X., & Bishop, C. H. (2003). A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. Journal of the Atmospheric Sciences, 60, 1140–1158. https://doi.org/10.1175/1520-0469(2003)060d1140:ACOBAno.2.0.CO;2

Wildenschild, D., & Jensen, K. H. (1999). Numerical modeling of observed effective flow behavior in unsaturated heterogeneous sands. Water Resources Research, 35, 29–42. https://doi.org/10.1029/98WR01959

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

Wu, C. C., & Margulis, S. A. (2013). Real-time soil moisture and salinity profile estimation using assimilation of embedded sensor data streams. Vadose Zone Journal, 12(1). https://doi.org/10.2136/vzj2011.0176

Yu, D., Yang, J., Shi, L., Zhang, Q., Huang, K., Fang, Y., & Zha, Y. (2019). On the uncertainty of initial condition and initialization approaches in variably saturated flow modeling. Hydrology and Earth System Sciences, 23, 2897–2914. https://doi.org/10.5194/hess-23-2897-2019

Zha, Y., Zhu, P., Zhang, Q., Mao, W., & Shi, L. (2019). Investigation of data assimilation methods for soil parameter estimation with different types of data. Vadose Zone Journal, 18(1). https://doi.org/10.2136/vzj2019.01.0013

Zhang, Q., Shi, L., Holzman, M., Ye, M., Wang, Y., Carmona, F., & Zha, Y. (2019). A dynamic data-driven method for dealing with model structural error in soil moisture data assimilation. Advances in Water Resources, 132. https://doi.org/10.1016/j.advwatres.2019.103407

How to cite this article: Bauser HH, Riedel L, Berg D, Troch PA, Roth K. Challenges with effective representations of heterogeneity in soil hydrology based on local water content measurements. Vadose Zone Journal. 2020;19:e20040. https://doi.org/10.1002/vzj2.20040