

Water Resources Research



RESEARCH ARTICLE

10.1029/2018WR024658

Key Points:

- Information from soil moisture sensor network was assimilated with EnKF to explore the limits of data assimilation with integrated terrestrial model
- Different assimilation scenarios including alternative geostatistical parameterizations were applied on real-world and synthetic experiments
- Performance differences between synthetic and real-world cases are due to model structural errors

Supporting Information:

- Figure S1
- Figure S2
- Figure S3

Correspondence to:

H.-J. Hendricks Franssen,
h.hendricks-franssen@fz-juelich.de

Citation:

Gebler, S., Kurtz, W., Pauwels, V. R. N., Kollet, S. J., Vereecken, H., & Hendricks Franssen, H.-J. (2019). Assimilation of High-Resolution Soil Moisture Data Into an Integrated Terrestrial Model for a Small-Scale Head-Water Catchment. *Water Resources Research*, 55, 10,358–10,385. <https://doi.org/10.1029/2018WR024658>

Received 24 DEC 2018

Accepted 12 OCT 2019

Accepted article online 6 NOV 2019

Published online 10 DEC 2019

Assimilation of High-Resolution Soil Moisture Data Into an Integrated Terrestrial Model for a Small-Scale Head-Water Catchment

S. Gebler^{1,2,3} , W. Kurtz^{1,2,4} , V. R. N. Pauwels⁵ , S.J. Kollet^{1,2} , H. Vereecken^{1,2} , and H.-J. Hendricks Franssen^{1,2}

¹Agrosphere Institute (IBG-3), Forschungszentrum Jülich, Jülich, Germany, ²Centre for High-Performance Scientific Computing in Terrestrial Systems: HPSC TerrSys, Forschungszentrum Jülich GmbH, Jülich, Germany, ³BASF SE, Agricultural Solutions-Global Environmental Fate Modelling, Limburgerhof, Germany, ⁴Leibniz Supercomputing Center of the Bavarian Academy of Sciences and Humanities, Garching, Germany, ⁵Department of Civil Engineering, Monash University, Clayton, Victoria, Australia

Abstract Land surface-subsurface modeling combined with data assimilation was applied on the Rollesbroich hillslope (Germany). Dense information from a soil moisture sensor network was assimilated with the ensemble Kalman filter applying different scenarios including the update of model states with or without updating of saturated soil hydraulic conductivity on an ensemble size of 128 (or 256) realizations with 3-D heterogeneous fields of Mualem-van Genuchten parameters. Simulations were also carried out with a synthetic test case mimicking the Rollesbroich site, to get more insight in the role of model structural errors. The combination of joint updating of model states and hydraulic conductivity was more efficient in updating the soil water content than state updating alone for the real-world case. On average, the root-mean-square error at the sensor locations was reduced by 14% if states and parameters were updated jointly, but discharge estimation was not improved significantly. Synthetic simulations showed much better results with an overall root-mean-square error reduction by 55% at independent verification locations in case of daily soil water content data assimilation including parameter estimation. Individual synthetic data assimilation scenarios with parameter estimation showed an increase of the Nash-Sutcliffe-Efficiency for discharge from −0.04 for the open loop run to 0.61. This shows that data assimilation in combination with high-resolution physically based models can strongly improve soil moisture and discharge estimation at the hillslope scale. Large performance differences between synthetic and real-world experiments indicated the limits of such an approach associated with model structural errors like errors in the prior geostatistical parameters.

1. Introduction

Modeling the soil water content (SWC) is of high interest for various geoscientific research fields. The SWC influences the water and energy cycles at the local, regional, and global scales (Vereecken et al., 2016). It controls, for example, the partitioning of net radiation into latent, sensible and ground heat fluxes, as well as the partitioning of precipitation into infiltration and runoff (Grayson et al., 1997; Robinson et al., 2008). Thus, a precise characterization and prediction of the SWC patterns is essential for understanding and quantifying the water and energy cycles for applications like weather prediction, flood prediction, and real-time irrigation scheduling.

Given this importance, recent modeling studies have applied integrated terrestrial model platforms (e.g., AquiferFlow-SiB2, Tian et al., 2012; CATHY, Bixio et al., 2002; Camporese et al., 2010; MikeShe, Abbott et al., 1986; Graham & Butts, 2005; ParFlow-CLM, Maxwell & Miller, 2005; Kollet & Maxwell, 2006; ParFlow-WRF, Maxwell et al., 2011; PIHM, Qu & Duffy, 2007; Kumar, 2009; TerrSysMP, Shrestha et al., 2014) for modeling two-way feedbacks between different terrestrial compartments. Coupled land surface-subsurface models are well suited for in-depth investigation of SWC spatial variability in combination with in situ SWC data at the hillslope scale. Examples are studies on the controls of SWC variability (e.g., Cornelissen et al., 2014; Fatichi et al., 2015; Herbst et al., 2006; Ivanov et al., 2010) as well as the relation between SWC distribution and the rainfall runoff response (e.g., Herbst & Diekkrüger, 2003; Sciuto & Diekkrüger, 2010). These studies demonstrated that fully coupled land surface-subsurface models can be a

©2019. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

valuable tool for a better understanding of the small-scale processes at the hillslope scale. However, due to the high computational needs of these models, studies were often conducted with uncalibrated parameters and without model uncertainty estimation. The uncertainty of modeled SWC is related to erroneous model forcings (e.g., precipitation), model parameters (e.g., soil hydraulic properties), and initial conditions, as well as model structural deficits.

With the help of data assimilation (DA) techniques, model predictions can be improved by merging (uncertain) observation data with uncertain model predictions (Burgers et al., 1998; Vrugt et al., 2005). Ensemble based sequential DA methods rely on a probabilistic framework where an ensemble of model realizations is propagated forward and updated each time observation data are available. One of the most commonly applied algorithms is the ensemble Kalman filter (EnKF; Evensen, 1994; Burgers et al., 1998). The EnKF is also used for updating model parameters, for example, by an augmented state vector approach (Chen & Zhang, 2006; Hendricks Franssen & Kinzelbach, 2008). This is especially important for subsurface applications, since parameter uncertainty is an important source of uncertainty for subsurface terrestrial system models. The computational costs for the EnKF are not excessive as the needed ensemble size is in general not too large (less than 1,000 ensemble members) and parallelization to speed up the simulations is trivial (Kurtz et al., 2016; Pauwels & De Lannoy, 2009). This is one of the reasons that the EnKF has been used in combination with different types of terrestrial system models (e.g., atmospheric models, hydrological models, and land surface models) and various kinds of observation data.

Only relatively few studies at the hillslope scale combine DA with integrated terrestrial system models. Nevertheless, some authors were able to show the high potential of the EnKF in combination with integrated hydrological models. For example, Camporese et al. (2009) found that streamflow predictions improved by assimilating different combinations of pressure head and streamflow data. They performed the DA experiments for a synthetic tilted v-catchment with a fully coupled surface water-groundwater flow (CATHY) model including 3-D subsurface flow and found that pressure head data always improved the characterization of streamflow, whereas standalone streamflow assimilation showed less improvement. This was also confirmed by Bailey and Baù (2012) who estimated the spatial distribution of hydraulic conductivity by the assimilation of streamflow and water table data in a similar test case. Shi et al. (2014) assimilated six data types (discharge, groundwater table depth, SWC, land-surface temperature, sensible heat, latent heat, and transpiration) with the EnKF and the land surface-subsurface model Flux-PIHM mimicking a small forested catchment. They estimated several, but spatially homogeneous, Mualem-van Genuchten soil hydraulic properties and found an improvement of the model simulations in association with a high sensitivity of the subsurface soil hydraulic properties (i.e., the Mualem-van Genuchten air entry and shape parameters). Shi et al. (2015) confirmed these findings also for experiments with real-world observations of the Shale Hills catchment in Pennsylvania, USA. Both studies by Shi et al. (2014, 2015) were conducted with a simplified representation of subsurface flow, namely, 2-D groundwater flow and 1-D flow in the unsaturated zone.

Some studies with integrated hydrologic models focused more on experimental design. For instance, in a synthetic experiment, which mimicked the Biosphere 2 Landscape Evolution Observatory hillslopes, Pasetto et al. (2015) estimated saturated hydraulic conductivity values and investigated the impact of the number and spatial distribution of assimilated SWC data for a hillslope on the characterization of the hydrologic response of the CATHY model with a 3-D synthetic subsurface. If the number of sensors fell below a certain threshold (100 instead of 496), the model was not able to reproduce the synthetic truth by SWC assimilation with the EnKF. Rasmussen et al. (2015) investigated the relationship between the number of ensemble members and the number of pressure head measurement data on the estimation of the discharge rate with help of the ensemble transform Kalman filter. They performed this study for the agriculture dominated Karup catchment in Denmark and concluded that less head observations required more ensemble members to reproduce the synthetic discharge and pressure head observations. Assimilating discharge observations in combination with parameter estimation required a larger ensemble size than the assimilation of groundwater table measurements. Zhang et al. (2015) found that the EnKF performance is very sensitive to the estimation of the initial model parameter uncertainty. The majority of these examples are limited to synthetic test cases or, in case of real-world case studies, strong simplifications of the subsurface compartment. In particular, complex subsurface structures (i.e., fully 3-D heterogeneous fields of subsurface hydraulic properties) and their impacts on the EnKF performance within a fully coupled land surface-subsurface model at high resolution have not yet been investigated.

DA studies in combination with highly nonlinear vadose zone flow models and heterogeneity of soil hydraulic parameters face additional challenges. The EnKF shows an optimal performance for Gaussian distributions, but strongly skewed pressure head distributions in the vadose soil zone can be expected under very dry conditions (Erdal et al., 2015; Zhang et al., 2018). Furthermore, it is important that the spatial structure of the heterogeneous soil hydraulic parameters (i.e., horizontal layers) is well captured. Otherwise, a calibration with the EnKF leads to poor results (Erdal et al., 2014). DA experiments for different soil types (e.g., Li & Ren, 2011; Montzka et al., 2011; Tran et al., 2014) also indicated that, besides stratification and heterogeneity, the capacity of DA to adequately characterize model states and parameters depends on the soil type and its related hydraulic properties. Within this context, Li and Ren (2011) indicated that the joint update of multiple hydraulic properties (i.e., Mualem-van Genuchten inverse air entry (α) and shape parameter (n) in combination with hydraulic conductivity (K_s) by the EnKF using a 1-D subsurface hydraulic model is more efficient than updating hydraulic conductivity alone. Other examples for the joint update of hydraulic conductivity, α , and n in combination with the EnKF and a 1-D subsurface model are given by Wu and Margulis (2011, 2013). Montzka et al. (2011, 2013) used the particle filter for updating these soil hydraulic parameters in a 1-D model. The vast majority of 2-D and 3-D studies where soil hydraulic properties are estimated only update saturated hydraulic conductivity, since the estimation of other soil hydraulic parameters is associated with numerical instability (Rasmussen et al., 2015; Vereecken et al., 2016). Exceptions are studies by Shi et al. (2014, 2015) and Pasetto et al. (2015). However, these studies made various simplifications. Shi et al. (2014, 2015) assumed spatially homogeneous zones within their subsurface setup. Pasetto et al. (2015) estimated 3-D spatially distributed hydraulic conductivity and porosity under assumption of perfect knowledge of the remaining hydraulic parameters. Chaudhuri et al. (2018) estimated 3-D spatially distributed hydraulic conductivity and Mualem-van Genuchten parameters with an iterative filter for a synthetic case. The impact of only updating saturated hydraulic conductivity on the performance of an integrated 3-D terrestrial model for a real-world case, which includes also other spatial heterogeneous and uncertain soil hydraulic parameters (e.g., Mualem-van Genuchten α , n), has not yet been explored.

This gap of applying the EnKF on a complex high-resolution integrated terrestrial model at the field scale with 3-D heterogeneous fields of Mualem-van Genuchten subsurface parameters using real-world data is covered in this study. More specifically, SWC observations were assimilated in the Terrestrial Systems Modeling Platform (TerrSysMP) developed by Shrestha et al. (2014), which was coupled to the Parallel Data Assimilation Framework (PDAF) DA framework (Nerger & Hiller, 2013) by Kurtz et al. (2016). The physically based model ParFlow (Ashby & Falgout, 1996; Jones & Woodward, 2001; Kollet & Maxwell, 2006) and the Community Land Model (Oleson et al., 2004; Oleson et al., 2008) of TerrSysMP were used. ParFlow and Community Land Model (CLM) are fully coupled via source/sink terms at root zone layer. This allows transient interaction between 3-D subsurface flow, overland flow, and land-surface processes. Data sets collected in the context of the infrastructure initiative Terrestrial Environmental Observatories (TERENO; Zacharias et al., 2011), and the Transregional Collaborative Research center (TR 32; Simmer et al., 2015) were used for assimilation and also for verification. In particular, SWC, evapotranspiration (ET), and discharge measurements were available for this study in the Rollesbroich headwater grassland catchment in the Eifel (Germany). The simulations were made with fully 3-D heterogeneous fields of soil hydraulic properties which results in a model system with 0.3 million unknowns. DA experiments were both made for the real-world case and a synthetic case, which mimics the real-world case, with the aim to unravel the role of model structural errors. In the DA-experiments different combinations of state (i.e., pressure head) and parameter (i.e., saturated hydraulic conductivity) updating were explored. The following research questions are addressed with the conducted experiments:

1. Can the characterization of the hydrology at the hillslope scale (SWC, ET, and discharge) be improved with a combination of physically based integrated hydrological models and a dense network of SWC data using DA?
2. Is the joint update of model states and saturated soil hydraulic conductivity sufficient to adjust the model simulations closer to the observations or is it necessary to include additional soil hydraulic properties (e.g., inverse air entry) within the parameter update to constrain this complex nonlinear coupled model system with spatially heterogeneous soil hydraulic properties?
3. Is there a systematic difference in the performance of DA between synthetic experiments and real-world experiments which points to processes which are not captured by the high-resolution integrated model?

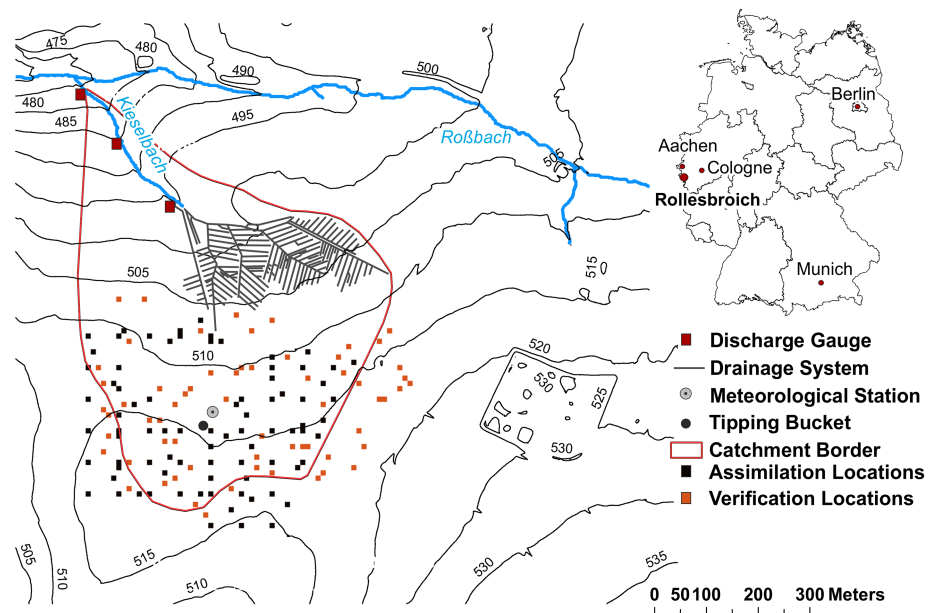


Figure 1. Rollesbroich overview map showing the locations of discharge gauges, the meteorological station and the drainage system. The locations for soil moisture assimilation (SoilNet) and verification are also indicated.

2. Materials and Methods

2.1. Study Site

The Rollesbroich study site ($50^{\circ} 37' 27''\text{N}$, $6^{\circ} 18' 17''\text{E}$) is located in the Eifel low mountain range (Germany). The Rollesbroich study site is a subcatchment of the river Rur and has an area of $\sim 0.38 \text{ km}^2$, an altitude varying between 474 and 518 m above sea level and slopes between 0% and 10%. As a part of the TERENO infrastructure, the Rollesbroich test site is an extensively managed grassland site dominated by smooth meadow grass (*Poa pratensis*) and perennial ryegrass (*Lolium perenne*). The prevalent soils are Cambisol (gleyic), Stagnosol and Cambisol-Stagnosol (Food and Agriculture Organization classification) with a depth between only 0.5 and 1.5 m (Korres et al., 2010; Koyama et al., 2010). The underlying bedrock is sandstone and siltstone with a heavily weathered top layer (saprolite) of ~ 0.1 - to 0.5-m thickness. Data from a meteorological station in approximately 4-km distance operated by the North Rhine-Westphalian State Environment Agency (LUA NRW) for the period 1981–2001 show that the study site has an annual mean temperature of 7.7°C and a mean yearly precipitation of 1,033 mm. A drainage system in the northwestern part of the study site prevents flooding induced by high groundwater tables. The outlet of the drainage system is located close to the Kieselbach source. The diameters of the ~ 80 -year-old clay pipes vary between 3 and 20 cm. However, the state of the pipes (possibly broken or blocked) and their detailed impact on the hydrological characteristics of the study site are uncertain. Figure 1 provides an overview map of the study site and the measurement equipment used in this study.

Precipitation measurements at the study site are performed with an interval of 10 min and a resolution of 0.1 mm using a standard Hellmann type tipping bucket balance (TB) rain gauge (ecoTech GmbH, Bonn, Germany). The device setup 1 m above ground is in agreement with the policy of the German weather service for areas with heavy snowfall and altitude > 500 m above sea level (World Meteorological Organization guidelines recommend 0.5 m). To avoid measurement errors due to instrument freezing, the device is temporally heated during the winter period.

Latent and sensible heat fluxes are measured by an eddy covariance station at the southern part of the study site ($50^{\circ} 37' 19''\text{N}$, $6^{\circ} 18' 15''\text{E}$, 514 m above sea level). At this location also temperature and air humidity are recorded by a HMP45C, Vaisala Inc., Helsinki, Finland, at 2.58 m above ground. Furthermore, a four-component net radiometer (CNR-1, Kipp and Zonen, Delft, The Netherlands) measuring incoming and outgoing longwave and shortwave radiation, a sonic anemometer (CSAT3, Campbell Scientific, Inc., Logan,

USA) recording wind speed and direction are installed there at 2.6 m above ground surface. At this location also a gas analyzer (LI7500, LI-COR Inc., Lincoln, NE, USA) for specific humidity and air pressure measurements is installed. These on-site data were used as hourly forcing for the CLM model simulations for 2011. For 2010 and for gap filling off-site meteorological data from the nearby LUA NRW station were applied instead. Furthermore, the data were used to calculate actual ET, which was used for verification in the project.

The SWC at the study site is measured with a wireless sensor network (SoilNet; Qu et al., 2013). SoilNet data were used in the DA and for model verification. At each of the 179 sensor locations two redundant SPADE ring oscillator sensors (Model 3.04, sceme.de GmbH i.G., Horn-Bad Meinberg, Germany) are vertically installed at 5-, 20-, and 500cm depths within a distance of ~10 cm to each other. This increases measurement volume and precision and helps to avoid data inconsistencies (e.g., contact issues with the soil matrix). Additional technical details can be found in Qu et al. (2013) and Qu et al. (2016). A long-term evaluation of the SWCs measured with the SPADE sensors at the study site indicated a drift in SWC values for individual SoilNet locations. The SWC data exhibited a gradual increase of maximum SWCs during saturated soil conditions for the period 2011 to 2013, which was not in line with the annual precipitation trend (Gebler et al., 2017). A linear SWC trend correction was applied which takes the maximum SWC of the wet periods in 2011 and 2012 into account. Trend correction was initially conducted for 82 SoilNet locations located in the southern part of the study site and separately for all three sensor depths (Gebler et al., 2017). This data set was further reduced to 61 sensor locations, since only gap-free SWC time series were used for DA. The sensors of the northern part of the study site were not included in our experiments as SWC values for the northern part of the study site were not available before 2012.

Discharge measurements for the Kieselbach, used for verification in this study, are conducted with a Venturi-Gauge weir close to the catchment outlet and two upstream Tomson gauges close to the headwaters of the Kieselbach (Qu et al., 2016).

2.2. TerrSysMP

For this study, ParFlow and CLM were applied in coupled fashion within the TerrSysMP (Shrestha et al., 2014) using the external coupler OASIS3-MCT (Valcke, 2013). While ParFlow simulates subsurface flow as well as overland flow, CLM calculates land-atmosphere exchange fluxes, snow, and vegetation processes. Both components were used in combination with the PDAF (Nerger & Hiller, 2013), which was implemented by Kurtz et al. (2016).

The subsurface model ParFlow simulates transient, variably saturated subsurface flow two-way coupled with overland flow at a high spatial resolution. On this account, ParFlow is designed to run in parallel on high-performance computing platforms (Kollet et al., 2010; Maxwell, 2013). Driven by external boundary conditions, ParFlow calculates the water pressure and saturation field as function of time. It solves the three-dimensional Richard's equation (Richards, 1931) using a cell-centered finite differences scheme in space:

$$S_s S_w(\psi) = \frac{\partial \psi}{\partial t} \varphi \frac{\partial S_w(\psi)}{\partial t} = \nabla \cdot \mathbf{q} + q_s + \frac{q_e}{m'} \quad (1)$$

$$\mathbf{q} = -\mathbf{K}_s k_r(\psi) \nabla(\psi - z) \quad (2)$$

where S_s is the specific storage coefficient (L^{-1}), S_w relative saturation ($-$), ψ pressure head (L), t time (T), φ porosity ($-$), $\mathbf{q}$ the volume specific Darcy flux (L/T), q_s a source/sink term (T^{-1}) (e.g., pumping and injection), q_e a general source/sink term that represents exchange fluxes between surface water and subsurface (L/T), m' an interfacial thickness (L), $\mathbf{K}_s$ saturated hydraulic conductivity tensor (L/T), k_r relative permeability ($-$), and z positive downward vertical coordinate (L).

Relative saturation and permeability in ParFlow are expressed by the Mualem-van Genuchten relationships (van Genuchten, 1980). The subsurface parametrization in ParFlow can be done with constant input or spatially distributed input files like the ones generated with geostatistical conditioning methods (i.e., turning bands algorithm and parallel Gaussian simulation). Groundwater and overland flow in ParFlow are coupled via a backward Euler scheme in time using a free-surface overland flow boundary condition (Kollet & Maxwell, 2006). Shallow overland flow is simulated by the kinematic wave model and the 2-D

momentum equation for streamflow is resolved without parameterized routing routines. Thus, the free surface flow is exclusively driven by gravity. The flow discharge relationships are described by Manning's equation (Kollet & Maxwell, 2008; Maxwell & Miller, 2005). For robust solutions of the subsurface and overland flow equations, the Newton-Krylov method in combination with multigrid preconditioning as described in Jones and Woodward (2001) is computed in a globally implicit manner. With an optional terrain-following grid (TFG) transformation which was introduced by Maxwell (2013), the computational efficiency of ParFlow can be improved, since TFG enables a reduction of model layers for study areas with large topographic gradients.

CLM v.3.5 is the upper boundary condition of ParFlow. Both models are dynamically coupled via source/sink terms at 10 overlapping subsurface horizons (root zone layer) of variable extent. This allows a transient feedback in two ways. Via the OASIS3-MCT coupler (Valcke, 2013), CLM provides ET to ParFlow and controls infiltrating precipitation. ParFlow passes SWC and pressure to the CLM subsurface. All water and heat fluxes in the CLM are calculated according the principles of mass and energy conservation. The CLM hydrologic cycle interacts with the land surface heat fluxes and accounts for snow, soil infiltration, and vegetation water uptake as well as for precipitation interception by plant foliage, throughfall, and stem flow. All these water fluxes can add (source) or extract (sink) water from the upper 10 subsurface horizons which are coupled with ParFlow. In the coupled ParFlow-CLM model, water flow in the subsurface, SWC content, and pressure heads is calculated by ParFlow. In the CLM, the Monin-Obukhov similarity theory according to Zeng et al. (1998) is used for calculating latent and sensible heat flux. The boundary layer resistance terms are implemented using the Biosphere-Atmosphere Transport Scheme, which was developed by Dickinson et al. (1993) and adapted by Zeng et al. (2005). The CLM distinguishes between five different land surface classes (glacier, lake, wetland, urban, and vegetation). The vegetated area can be further subdivided into plant functional types (Bonan et al., 2002) for which physiological plant parameters and other variables are defined. In the coupled mode, the TOPMODEL-based CLM runoff scheme and CLM variably saturated subsurface flow are replaced by ParFlow. Atmospheric forcing data (temperature, precipitation, solar radiation, humidity, and barometric pressure) are provided in a spatially distributed manner.

2.3. State and Parameter Estimation With the EnKF

In this study, the EnKF was applied to the ParFlow-CLM model of the Rollesbroich study site. The EnKF is an ensemble based sequential DA method to improve the estimation of model states (and possibly parameters) on the basis of optimally combining information from model simulations and observations.

The assimilation cycle of the EnKF is a two-step process consisting of a forecast and an analysis, which are repeated sequentially. In the forecast step, different model runs (in our case 128 or 256 which are also called stochastic realizations) with the ParFlow-CLM model (M) are propagated forward in time. The ParFlow-CLM model, as introduced in section 2.2, solves equations (1) and (2) for subsurface flow by the finite volume method, and in addition, it solves overland flow and the exchange of water and heat between the land and the atmosphere. The stochastic realizations differ because they use different initial conditions, saturated hydraulic conductivity, and precipitation as input. These are considered the main sources of uncertainty. The forecast step is given by the following:

$$\mathbf{x}_i^{f,t} = M(\mathbf{x}_i^{t-1}, \mathbf{p}_i, \mathbf{q}_i^t) \quad (3)$$

where $\mathbf{x}_i^{f,t}$ represents the predicted model state vector at time step t , $\mathbf{x}_i^{t-1}$ is the model state vector of the previous time step $t - 1$, $\mathbf{p}_i$ represents all the model parameters, and $\mathbf{q}_i^t$ are the model forcings for each ensemble realization i ($i = 1, 2, \dots, n_{\text{realiz}}$).

The prognostic variable in ParFlow is pressure. The model state vector $\mathbf{x}_i^{f,t}$ consists of the pressure head (ψ) simulated for the N different grid cells ($\mathbf{x}_i^{f,t} = \boldsymbol{\psi}_i^{f,t}$). In this study, SWC (θ) observations are assimilated:

$$\mathbf{y}_i^t = \boldsymbol{\theta}^t + \boldsymbol{\varepsilon}_{iy} \quad (4)$$

where $\mathbf{y}_i^t$ is the observation vector and $\boldsymbol{\varepsilon}_{iy}$ is a vector whose elements contain random values drawn from a normal distribution $\mathcal{N}(0, \sigma^2)$ with mean zero and standard deviation σ (Burgers et al., 1998). The dimension of all the vectors in equation (4) is equal to the number of observations M assimilated for a certain time step.

The pressure heads in the model state vector are updated by assimilating the SWC observations:

$$\mathbf{x}_i^{a,t} = \mathbf{x}_i^{f,t} + \mathbf{K}^t (\mathbf{y}_i^t - \mathbf{y}_i^{f,t}) \quad (5)$$

where $\mathbf{x}_i^a$ is the updated state vector, $\mathbf{K}$ the Kalman gain, $\mathbf{y}_i^t$ is the observation vector, and $\mathbf{y}_i^{f,t}$ are the forecasted observations according $\mathbf{y}_i^{f,t} = h(\mathbf{x}_i^{f,t})$ where h is the nonlinear measurement operator, which handles the mapping between (measured) SWC and (simulated) pressure via the Mualem-van Genuchten relationship according to Reichle et al. (2002).

The Kalman gain $\mathbf{K}$ is calculated according

$$\mathbf{K}^t = \text{Cov}(\mathbf{x}_i^{f,t}, \mathbf{y}_i^{f,t}) (\text{Cov}(\mathbf{y}_i^{f,t}, \mathbf{y}_i^{f,t}) + \mathbf{R})^{-1} \quad (6)$$

where **Cov** indicates covariances and $\mathbf{R}$ ($M \times M$) is the observation error covariance matrix, which contains on its diagonal the SWC measurement error variances, which are defined prior to the simulations in correspondence to equation (5).

The model covariance matrix is not completely calculated in equation (6), as only the covariances between the measurements and the model states are needed. The complete model covariance matrix can be determined from the ensemble of model forecasts, where $\bar{\mathbf{x}}^{f,t}$ contains the average model states calculated over all realizations n_{realz} at each time step:

$$\mathbf{P}^t = \frac{1}{n_{\text{realz}} - 1} \sum_{i=1}^{n_{\text{realz}}} (\mathbf{x}_i^{f,t} - \bar{\mathbf{x}}^{f,t}) (\mathbf{x}_i^{f,t} - \bar{\mathbf{x}}^{f,t})^T \quad (7)$$

The EnKF is used in this work to also update uncertain parameters (e.g., Chen et al., 2011; Chen & Zhang, 2006; Hendricks Franssen & Kinzelbach, 2008).

The augmented state vector for updating both states and parameters is as follows:

$$\mathbf{x}^{f,t} = \begin{pmatrix} \boldsymbol{\psi}^t \\ \mathbf{Y}^t \end{pmatrix} \quad (8)$$

where $\mathbf{x}$ is now the augmented state vector including pressure heads ($\boldsymbol{\psi}$) (L) and the logarithm of soil hydraulic conductivities ($\mathbf{Y} = \log_{10} \mathbf{K}_s$ (L/T)) at all model grid cells N . This implies that for the case of joint state-parameter updating the augmented state vector is of dimension $2N$.

In some of the scenarios where both states and parameters are updated, a damping factor (α) was used to reduce filter inbreeding (Hendricks Franssen & Kinzelbach, 2008). Filter inbreeding is the underestimation of the ensemble variance after applying the EnKF updating equation repeatedly. This underestimation of the variance is related to the low rank approximation of the covariance matrix by a limited number of ensemble members and the linearization of the relation between states and parameters. A damping factor reduces the impact of filter inbreeding because parameter updates are made in smaller steps, limiting the impact of non-linear relations between parameters and states, which are not captured by the covariances, and limiting the impact of spurious covariances. This results in the following updating equation for the joint state-parameter estimation problem:

$$\mathbf{x}_i^{a,t} = \mathbf{x}_i^{f,t} + \boldsymbol{\alpha}^T \mathbf{K} (\mathbf{y}_i^{f,t} - \mathbf{y}_i^{f,t}) \quad (9)$$

where $\mathbf{x}_i^a$ is the updated augmented state vector after DA and $\boldsymbol{\alpha}^T$ a vector with damping factors, which are 1 for updating states and between 0 and 1 for updating parameters.

2.4. The TerrSysMP-PDAF Setup

The PDAF (v1.16) module coupled to TerrSysMP was used for the assimilation of SWC observations from SoilNet infrastructure installed at the Rollesbroich study site. Figure 2 schematically illustrates the operation mode of PDAF with the EnKF when applied with ParFlow-CLM. States and parameters of the different

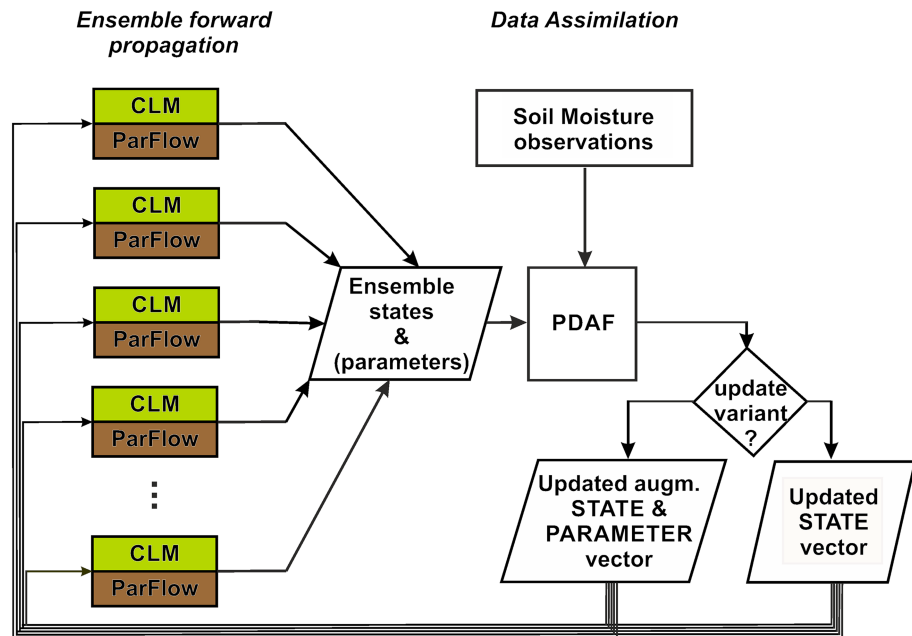


Figure 2. Schematic setup of the assimilation of soil water content with Parallel Data Assimilation Framework (PDAF) and ParFlow-CLM. Different possibilities of state and parameter update are indicated on the right.

ParFlow-CLM model runs are collected by PDAF after a predefined assimilation interval. By assimilating SWC observations either model states or states and parameters are updated and passed back to the ParFlow-CLM realizations which are then propagated further in time. For the different variants of the PDAF module within TerrSysMP as well as additional technical details see Kurtz et al. (2016).

3. Setup of the DA Experiments

The Rollesbroich model domain covers an area of $1,280 \times 1,120$ m with a lateral resolution of 10×10 m and a total depth of 3.2 m. This results in a model domain of $128 \times 112 \times 20$ (286,720) grid cells and includes the Rollesbroich catchment area of 38 ha as well as parts of the Rollesbroich residential area, streets, and other anthropogenic artificial structures. The slopes in x and y directions, which represent the topographic driven overland flow in the domain, were calculated with the help of a digital elevation model. Further details of the digital elevation model preparation can be found in Gebler et al. (2017). A Manning's roughness value of $0.001 \text{ hr/m}^{1/3}$ accounts for high (bank) vegetation density and the small stream channel bed of approximately 30 cm. The lower boundary was set impermeable at 3.2-m soil depth. The ParFlow subsurface domain was built with a variable vertical resolution making use of the TFG approach (Maxwell, 2013). From surface to bedrock the model resolution was gradually reduced with depth. The first layer with a vertical extent of 0.025 m is followed by 10 layers of 0.05-m vertical resolution and a twelfth layer with 0.1-m depth. The vertical resolution then further decreases from 0.2 m (Layer 13–17) to 0.5 m (Layer 18–19), and 0.575 m (Layer 20) for the deepest layer (bedrock).

The subsurface parameterization scheme mainly follows the heterogeneous subsurface setup by Gebler et al. (2017) with exception of the bedrock layers. For the assimilation experiments, the subsurface was separated into three soil horizons and two bedrock layers with spatially heterogeneous two-dimensional stochastic fields of Mualem-van Genuchten hydraulic properties. These stochastic fields include saturated hydraulic conductivity (K_s), inverse air entry (α), shape factor (n) as well as residual (θ_r) and saturated water content (θ_s) which also contain porosity information. A detailed description of the parameter sampling is given in section 3.2.

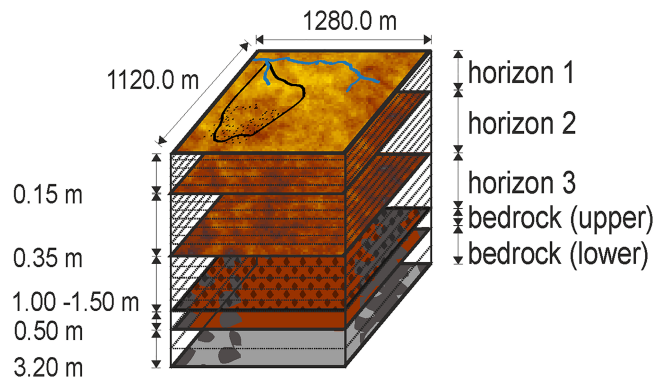


Figure 3. Schematic overview of the five heterogeneous horizons of soil hydraulic parameters within ParFlow. Study site borders (black line drawn inside Horizon 1), channels (blue), and locations of the SoilNet devices (black dots) are also indicated. The hatchings between the third and the upper bedrock horizon represent the variable bedrock depth at 1.0–1.5 m.

For the weathered siltstone and sandstone bedrock layers preferential flow through fractures is very likely. To mimic these characteristics, the bedrock was subdivided into two separate layers. The mean hydraulic conductivity ($\log_{10}K_s$) of the bedrock was increased to the magnitude of a sandy soil to enable an adequate drainage of the upper soil horizons and a high lateral flow component in the bedrock layer which also mimics the efficiency of the drainage system. Figure 3 illustrates the simulation setup with the different horizons. The default CLM subsurface setup with 10 exponentially increasing subsurface layers was adapted to match the before mentioned extent of 10 corresponding ParFlow model layers (0.475-m total depth). The CLM vegetation parameterization mainly relies on standard spatial uniform CLM C3-grass parameters with a leaf area index varying between 0.3 and 3.0 over the year. The CLM default root distribution parameters (roota and rootb) were adapted in accordance with the modifications made for the CLM subsurface extent (roota: 10.6; rootb: 6.0). This modification, which assigns 90% of grass roots within the upper 30-cm soil, is in line with literature values (e.g., Brown et al., 2010). No flow boundary conditions were imposed for the northern, eastern, southern, and western domain boundaries.

Precipitation, air pressure, wind speed, specific humidity, and incoming shortwave and longwave radiation from on-site measurements (2011) and the nearby LANUV station (2010) were used as input to CLM in hourly time steps (Gebler et al., 2017).

In total, a period of two years (2010–2011) was simulated with ParFlow-CLM.

3.1. PDAF Setup and Assimilation Scenarios

First, ParFlow-CLM open loop simulations (Table 1) for the entire ensemble and the entire simulation period (January 2010 to December 2011) were performed. This period includes the model spin-up, which was conducted for a period of 16 months (January 2010 to April 2011) beginning with an initial hydrostatic equilibrium condition and a groundwater table at 1.5-m depth. Second, several assimilation variants (Table 1) were tested with the EnKF assimilating SWC data taken from 61 locations in the southern part of the study site at 5-, 20-, and 50-cm depths. More specifically these variants are as follows:

1. State update on daily basis
2. Joint update of states and hydraulic conductivity on daily basis
3. Joint update of states and hydraulic conductivity where states are updated daily and hydraulic conductivity only each 5 days. In addition a damping factor ($\alpha = 0.1$) was applied for the parameter update.

In order to investigate potential model structural errors or methodical issues with the EnKF, the aforementioned variants were also tested with synthetic data mimicking the real-world data at identical measurement locations. This includes a scenario using deterministic Muelem-van Genuchten α and n using the true parameter set of 2-D heterogeneous layers for the entire ensemble. Furthermore, a scenario with biased saturated soil hydraulic conductivity ($\log_{10}K_s$: +0.5 in all soil layers) and an incorrect spatial correlation structure (increased variogram range by factor 2) was investigated. The impact of ensemble size was also explored. The selected synthetic reference truth was an ensemble member of the open loop ensemble run (i.e., SYN_OL), and synthetic measurement data were extracted from this reference run at 61 observation locations at 5-, 20-, and 50-cm depths. The synthetic observations were used in all synthetic experiments including experiments with biased K_s and erroneous spatial correlation structure. The assimilation period for real-world and synthetic experiments was from 1 May to 31 December 2011. The synthetic cases were further evaluated against 63 random locations (Figure 1) different from the assimilated locations in 5-, 20-, and 50-cm depths. In addition, two layers of the upper and lower bedrock in 130- and 190-cm depths were examined for verifying the performance of the DA runs.

The standard deviation of the measurement error was set to $0.04 \text{ cm}^3/\text{cm}^3$ and assumed to be spatially uncorrelated for synthetic and real-world scenarios. Surface cells with overland flow (i.e., with saturated conditions) were excluded from the update. This avoids numerical instability by an on/off switching of

Table 1
Simulation Scenarios for Different Synthetic and Real-World Data Assimilation Experiments

Scenario	Observation	Realizations	Parameter update frequency (d)	Parameter damping	Mualem-van Genuchten parametrization scheme (α , n)
REAL_OL	SoilNet	128	n/a	n/a	stochastic
REAL_ST	SoilNet	128	n/a	No	stochastic
REAL_PAR_1	SoilNet	128	1	No	stochastic
REAL_PAR_5d	SoilNet	128	5	Yes	stochastic
SYN_OL	synthetic	128	n/a	n/a	stochastic
SYN_Ks_OL	synthetic	128	n/a	n/a	stochastic
SYN_ST	synthetic	128	n/a	No	stochastic
SYN_PAR_1	synthetic	128	1	No	stochastic
SYN_PAR α n_1	synthetic	128	1	No	deterministic
SYN_PARKs_1	synthetic	128	1	Yes	stochastic
SYN_PAR_5d	synthetic	128	5	Yes	stochastic
SYN_OL_256	synthetic	256	n/a	n/a	stochastic
SYN_PAR α n_5d_256	synthetic	256	5	Yes	deterministic

Note. The open loop simulations refer to the synthetic reference (SYN_OL) or to the associated real-world case (REAL_OL). The “ST” refers to simulation with exclusive state update, while “PAR” includes simulation with state and parameter update. The letter “d” indicates scenarios with damping. The “Ks” specifies simulation scenarios with biased saturated conductivity and incorrect spatial correlation in all soil layers. Some parameters were not available in specific scenarios (“n/a”). The assimilation period for the data assimilation scenarios is May–December 2011.

the surface runoff generation which is caused by the perturbation of the pressure field after the state update.

3.2. Ensemble Generation

In order to account for input uncertainties, model forcings as well as soil hydraulic parameters of the ParFlow-CLM model were perturbed. This resulted in 128 (or 256) different stochastic model realizations for the synthetic and real-world experiments. Hourly precipitation, assumed homogeneous for the catchment, was perturbed according to a normal distribution by multiplying with a Gaussian random value ($\mathcal{N}(1, 0.15^2)$). As precipitation measurements were available in the catchment, it was assumed that the uncertainty with respect to this forcing was mainly related to measurement uncertainty and therefore relatively small which justifies the Gaussian assumption.

The creation of an ensemble of uncertain subsurface hydraulic properties departed from van Genuchten soil hydraulic properties estimated by HYDRUS-1D (Šimůnek et al., 2008) and the Shuffled Complex Evolution algorithm (SCE-UA; Duan et al. (1992) for 82 locations in the southern part of the study area. This approach was already demonstrated to capture the spatial variability at the Rollesbroich test site (Qu et al., 2014) and tested in combination with a 3-D ParFlow-CLM model (Gebler et al., 2017). On this account, Gebler et al. (2017) extended the initial approach by an uncertainty assessment of the estimated soil hydraulic properties (Figure 4), which was not provided by the original SCE-UA parameter estimation. Therefore, the discrepancy between the measurement support (at the optimization location) and the model discretization was considered on the basis of ordinary block kriging (Burgess & Webster, 1980) using the VESPER (Whelan et al., 2002) software in an unconditional manner at 10×10 -m model resolution for all soil hydraulic parameters. The estimated uncertainty was similar to the uncertainty that alternatively was derived with the help of measured texture data and the ROSETTA (ROS) pedotransfer function (Schaap et al., 2001; e.g., Mualem-van Genuchten inverse air entry α : $\pm 1.3 \text{ cm}^{-1}$ [SCE-UA]; $\pm 1.5 \text{ cm}^{-1}$ [ROS]) as clarified in Gebler et al. (2017). These data (optimized SCE-UA parameter with uncertainty estimation by block kriging) were the basis for the sampling of 128 (256) stochastic realizations with 3-D heterogeneous fields of soil hydraulic properties which were used in the DA approach.

The subsequent sampling of Mualem-van Genuchten soil hydraulic properties (Figure 4) consists of four major steps. First, for the upper three layers, samples were taken from the multivariate distribution of the soil hydraulic parameters for each model realization. This distribution is defined by the individual parameter mean values and (co)variances, which were given by the optimized 1-D parameter set and the available uncertainty at each spatial location and depth. For the underlying siltstone and sandstone bedrock (Horizons 4 and 5), the hydraulic properties were randomly sampled at each location on the basis of

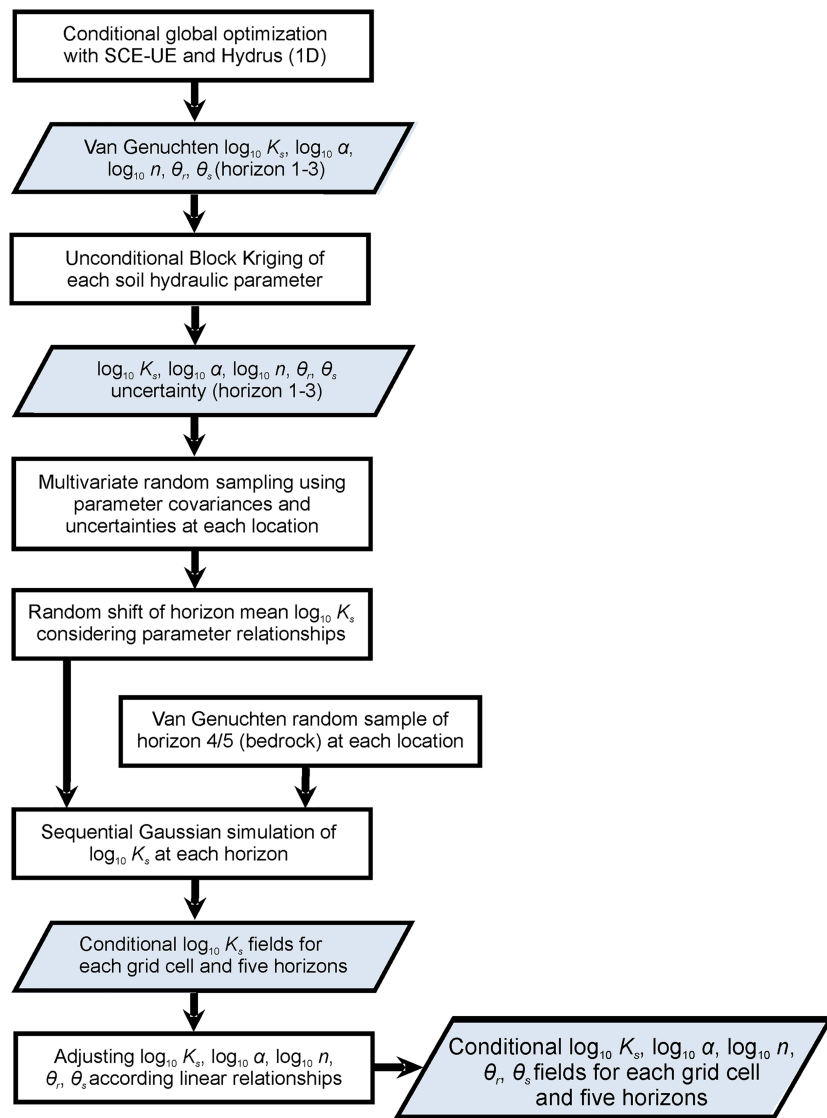


Figure 4. Flowchart of the subsurface parameter sampling creating a heterogeneous layered subsurface for each ParFlow-CLM realization (modified from Gebler et al., 2017).

parameter ranges provided by Boga (2003). The bedrock hydraulic properties were further artificially increased to the magnitude of a sandy soil with the aim to mimic the hydraulic characteristics of the saprolite and the on-site drainage system. The hydraulic conductivity of the lower bedrock horizon (1.2- to 1.5-m thickness) was more increased than the upper bedrock horizon (0.2- to 0.5-m thickness). This guarantees a realistic drainage of the lower soil horizon.

Second, the logarithmic hydraulic conductivity ($\log_{10}K_s$) at each location of the upper layers (5, 20, and 50 cm) was perturbed with a spatially homogeneous normal distributed Gaussian random value $\mathcal{N}(0, 0.25^2)$ to create more variable logarithmic hydraulic conductivity averages between different realizations for the individual soil horizons. With this procedure, additional model uncertainty was taken into account which may originate from an underestimation of the 1-D inversion uncertainty for hydraulic conductivity.

Third, a spatial heterogeneous field of $\log_{10}K_s$ was generated with sequential Gaussian simulation using GCOSIM3D (Gómez-Hernández & Journel, 1993) independently for each soil layer using the generated $\log_{10}K_s$ samples for the individual sensor locations. The required variogram parameters (range, sill, and nugget) were estimated by fitting the experimental semivariograms for the different soil layers to an exponential model for each soil layer (Horizons 1–3). For the upper and lower bedrock layer only very limited

information regarding the spatial dependence of the hydraulic properties was available. Therefore, the variogram parameters of Horizon 3 for the bedrock stochastic simulations were used.

In the fourth and last step, $\log_{10} \alpha$, $\log_{10} n$, θ_r , and θ_s were adapted according to their relations with $\log_{10} K_s$. The logarithmic transform was also used for α and n as the spatial variability of these parameters can in general not be described very well with the Gaussian approximation (e.g., Carsel & Parrish, 1988). The spatially heterogeneous $\log_{10} K_s$ fields, obtained in the third step, render $\log_{10} \alpha$, $\log_{10} n$ also spatially variable using the multivariate relationships of $\log_{10} K_s$ with the other variables. This results in stochastic realizations for the different soil and bedrock layers of $\log_{10} K_s$, $\log_{10} \alpha$, $\log_{10} n$, θ_r , and θ_s . For the different scenarios 128 (256) realizations were used in the DA procedure in ParFlow. In case of synthetic scenarios with deterministic α and n the true realization values were taken for the entire ensemble.

3.3. Performance Validation

The performance of the open loop and different DA scenarios was evaluated with measured or synthetic observations of daily SWC, daily ET and daily discharge. The Nash-Sutcliffe efficiency (NSE) index, the root-mean-square error (RMSE), and bias (BIAS) were calculated for the forecasted state variables of each simulation scenario.

The NSE was calculated according to the following:

$$\text{NSE} = 1 - \left(\frac{\sum_{t=1}^{n_{\text{tstep}}} \sum_{j=1}^{n_{\text{sample}}} (\bar{\mathbf{x}}_j^t - \mathbf{y}_j^t)^2}{\sum_{t=1}^{n_{\text{tstep}}} \sum_{j=1}^{n_{\text{sample}}} (\mathbf{y}_j^t - \bar{\mathbf{y}}_j)^2} \right) \quad (10)$$

where $\bar{\mathbf{x}}_j^t$ are the simulated ensemble mean values at the observation (or verification) locations for the j th sample at time t , $\mathbf{y}_j^t$ are the observed data (or synthetic observations from the synthetic reference), $\bar{\mathbf{y}}_j$ the average of the observed data over time, n_{sample} the number of samples (per layer), and n_{tstep} the total number of time steps. The NSE range is between $-\infty$ and 1.0. Negative values indicate unacceptable simulation performance whereas positive values suggest good performance (Nash & Sutcliffe, 1970; Moriasi et al., 2007).

The model bias (BIAS) and the relative model bias in percent (PBIAS) are given by the following:

$$\text{BIAS} = \frac{1}{(n_{\text{sample}} \times n_{\text{tstep}})} \sum_{t=1}^{n_{\text{tstep}}} \sum_{j=1}^{n_{\text{sample}}} (\bar{\mathbf{x}}_j^t - \mathbf{y}_j^t) \quad (11)$$

$$\text{PBIAS} = 100 \times \left(\frac{\sum_{t=1}^{n_{\text{tstep}}} \sum_{j=1}^{n_{\text{sample}}} (\bar{\mathbf{x}}_j^t - \mathbf{y}_j^t)}{\sum_{t=1}^{n_{\text{tstep}}} \sum_{j=1}^{n_{\text{sample}}} \mathbf{y}_j^t} \right) \quad (12)$$

The RMSE (equation (13)) was calculated to compare observed values with the ensemble mean from the simulations:

$$\text{RMSE} = \sqrt{\frac{1}{(n_{\text{sample}} \times n_{\text{tstep}})} \sum_{t=1}^{n_{\text{tstep}}} \sum_{j=1}^{n_{\text{sample}}} (\bar{\mathbf{x}}_j^t - \mathbf{y}_j^t)^2} \quad (13)$$

The RMSE was also calculated for individual locations. These RMSEs at individual locations were averaged for each individual soil layer, which was a further performance statistics to evaluate the model simulations.

4. Results

4.1. Synthetic Experiments

4.1.1. SWC

In this section results are discussed for the SWC modeling for the synthetic case. In section 4.2 results for the real-world experiments follow. The open loop simulations (SYN_OL) show a good performance for all soil layers (NSE: 0.480–0.759; Table 2). This is similar for SYN_Ks_OL at 5 and 20 cm (0.695–0.813), but poor performance is found at 50 cm (NSE: −10.368). After DA with state updating only (SYN_ST), and DA scenarios with joint state-parameter updating (SYN_PAR_1, SYN_PAR α n_1, and SYN_PAR_5d) the average SWC at 5- and 20-cm depths is almost perfectly matched (NSE $\geq$ 0.85). The RMSE decreased by 0.007–0.030 cm³/cm³

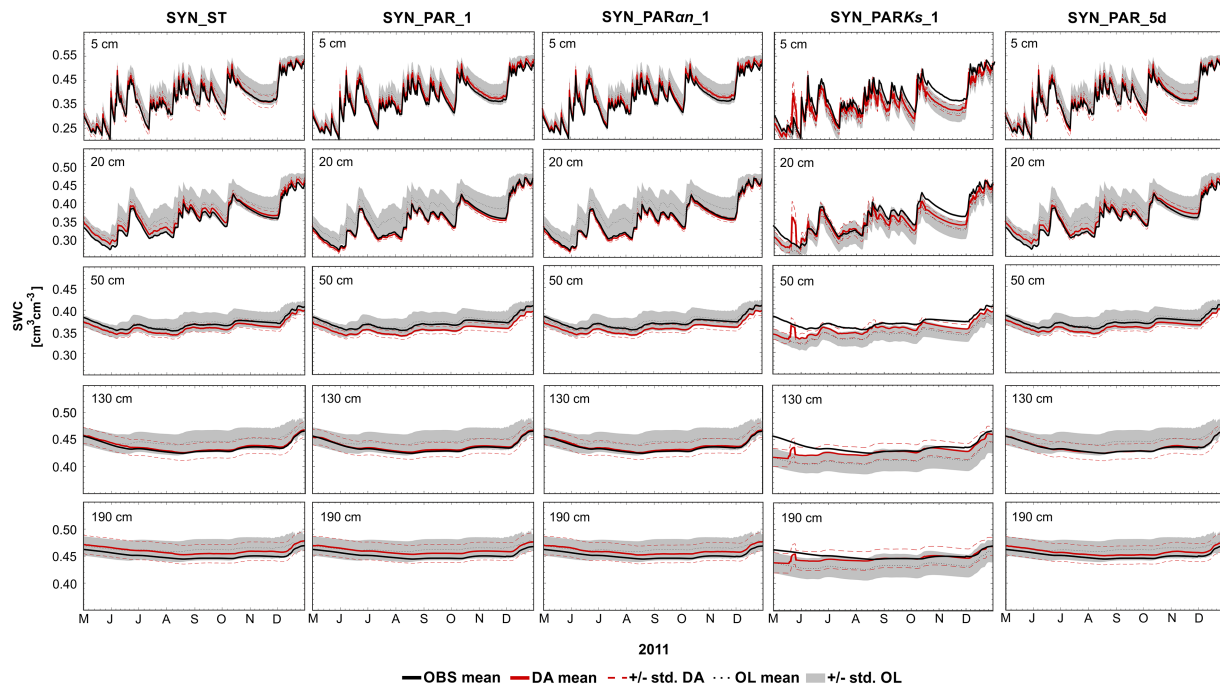


Figure 5. Mean soil water content (SWC) and standard deviation of SWC for the open loop run (OL) and the synthetic data assimilation experiments (DA) experiments at the verification locations. Mean synthetic reference observations are indicated as well.

(24–88%) compared to the open loop runs depending on the scenario and layer. At 50-cm depth, the simulations with biased K_s (SYN_PARKs_1) show a performance improvement indicated by 41% RMSE reduction compared to the corresponding open loop. In contrast, all other synthetic DA simulations at 50 cm show larger deviations from the synthetic truth than their open loop counterparts (NSE: -0.772 to -0.341 ; RMSE: 0.014 – 0.016 cm^3/cm^3). The synthetic simulations with 256 ensemble members (SYN_PARan_5d_256) show similar results to SYN_PARan_1, which indicates that the performance is not strongly influenced by ensemble size once the ensemble size is large enough.

Figure 5 shows the temporal evolution of the mean SWC for the open loop simulations and four different DA experiments for the verification locations in 2011. The differences between the individual DA scenarios are rather small. The different DA runs show larger differences in terms of SWC differences in May as well as in October and November 2011. Another notable difference is the reduced ensemble spread of SYN_PARan_1, which is indicated by a $\sim 75\%$ smaller standard deviation of mean SWC compared to other synthetic scenarios. SYN_OL shows a good performance (Table 2) for the upper soil layers (NSE: 0.571 – 0.793). In contrast, negative NSE is found at 130- and 190-cm depths (NSE: -6.761 to -2.333). With exception of the SWC in 20-cm depth (NSE: 0.631), SYN_Ks_OL shows an overall reduced performance compared to SYN_OL, particularly at 50, 130, and 190 cm (NSE: -14.738 to -9.486).

After DA, a performance improvement is found for the mean SWC at 5-, 20-, 130, and 190-cm depths indicated by an RMSE decrease of 0.002 – 0.020 cm^3/cm^3 (6–63%). Particularly, the mean SWC at 5, 20, and 130 cm shows a very good fit (NSE: 0.765 – 0.944) after DA, also for SYN_Ks_OL, which exhibits a large SWC fluctuation in all soil layers. For the bedrock layers the RMSE decrease ranges between 31% and 84%. In contrast, at 50-cm depth all assimilation scenarios show larger deviations from the reference. For this layer there is a decrease of performance in terms of mean SWC at the verification locations (NSE: -1.363 to -0.159 ; BIAS: -0.009 – 0.013 cm^3/cm^3). An exception is SYN_PARKs_1 (NSE: -2.846), which exhibits a strong performance improvement compared to SYN_Ks_OL (NSE: -9.484) and indicates that the biases in this simulation scenario could be partly corrected by DA. The performance of SYN_Ks_OL improves despite the strong fluctuation in SWC at the end of May.

Hereinafter, the RMSE at the individual assimilation locations is discussed (Figures 1 and 2 in the supporting information) for the different subsurface layers at the individual verification locations at 5, 20, 50, 130, and

Table 2
Performance Measures for the Reproduction of SWC for the Synthetic Experiments

		SYN_OL	SYN_Ks_OL	SYN_ST	SYN_PAR_1	SYN_PARan_1	SYN_PARKs_1	SYN_PAR_5d	SYN_PARan_5d (256)
5 cm	NSE (–)	0.751	0.813	0.978	0.971	0.926	0.892	0.973	0.934
	BIAS (cm ³ /cm ³)	0.035	–0.027	0.006	0.008	0.019	–0.006	0.007	0.017
	RMSE (cm ³ /cm ³)	0.036	0.029	0.011	0.012	0.019	0.022	0.012	0.018
20 cm	NSE (–)	0.480	0.695	0.936	0.931	0.993	0.849	0.924	0.990
	BIAS (cm ³ /cm ³)	0.031	–0.020	0.011	0.010	0.003	–0.002	0.012	–0.004
	RMSE (cm ³ /cm ³)	0.034	0.022	0.012	0.012	0.004	0.016	0.013	0.005
50 cm	NSE (–)	0.759	–10.368	–0.341	–0.527	–0.772	–3.214	–0.609	–1.112
	BIAS (cm ³ /cm ³)	–0.003	–0.031	–0.013	–0.014	–0.015	–0.018	–0.015	–0.017
	RMSE (cm ³ /m ³)	0.006	0.032	0.014	0.014	0.016	0.019	0.015	0.017

Note. Shown are performance statistics (Nash-Sutcliffe efficiency, NSE; root-mean-square error, RMSE; and absolute BIAS) for the open loop run (SYN_OL) and the data assimilation experiments for the assimilation locations.

190 cm. For the open loop simulation the layer-averaged RMSE (SYN_OL: 0.0311–0.0492 cm³/cm³; SYN_Ks_OL: 0.0384–0.0635) at the verification locations is 8–40% smaller than the RMSE at the assimilation locations. The mean RMSE for simulations without biased Ks over all verification locations decreases between 40% (SYN_ST) and 55% (SYN_PARan_1). For the individual layers, the mean RMSE averaged over the locations decreases between 20% and 66% compared to the open loop run depending on scenario and layer. The scenarios with parameter estimation show, for almost all layers, a better performance than SYN_ST with state update only. SYN_PARan_1 performs best in terms of layer-averaged RMSE for the individual locations over all five soil layers showing the strongest performance increase. The RMSE reduction at 130- and 190-cm depths is smaller than in the upper soil layers. Particularly, the RMSE reduction at 190 cm is relatively small (20–34%). These layers also show more locations (58–63%) with decreasing performance. In contrast, fewer locations with decreasing performance are found at the upper soil layers particularly for SYN_PARan_1 (0.0–1.5%). For SYN_PARKs_1 (Figure S2) the overall RMSE decrease (13%) as well as the RMSE decrease of the individual layers is smaller (9–18%) than for the other simulation scenarios. The best performance is found at 50 cm, where 83% of the individual locations show an improvement.

4.1.2. Parameter Estimation

Figure 6 illustrates the temporal evolution of the hydraulic conductivity ($\log_{10}K_s$) estimates for the synthetic experiments. In the synthetic scenario, the bedrock layer $\log_{10}K_s$ is reproduced well, despite relative large differences between initial average $\log_{10}K_s$ and the synthetic truth. Results for the other soil horizons for the synthetic case vary among the different DA experiments. While SYN_PAR_1 (Figure 6a) overestimates the hydraulic conductivity for Horizon 2 (20 cm) and Horizon 3 (50 cm), SYN_PAR_5d and SYN_PARan_1 overestimate the conductivity for Horizon 1 (5 cm). However, in most of these cases the deviations are very small ($\leq 0.05 \log_{10}K_s$). Only the SYN_PARan_1 Horizon 3 shows larger deviations ($\sim 0.10 \log_{10}K_s$). The parameter estimation for this scenario is also affected by temporal instability for Horizons 1 and 4 during the first month of the estimation, due to an insufficient ensemble spread. Higher differences between open loop and DA simulation are found for SYN_PARKs_1 (Figure 6b). While $\log_{10}K_s$ is slightly overestimated in Horizons 1 and 2, for the deeper horizons $\log_{10}K_s$ is strongly adjusted to the synthetic true reference. In this context, Horizon 4 shows an almost perfect fit although all horizons are affected by temporal instability during the first month of assimilation. The parameter updates (Figure S3) for the synthetic scenarios with biased Ks are very similar to each other. The parameter updates for SYN_PAR_1 and SYN_PARan_1 are only slightly stronger than for SYN_PAR_5d. SYN_PAR_1 and SYN_PARan_1 exhibit higher $\log_{10}K_s$ contrasts between the sensor locations than SYN_PAR_5d. On the contrary, high contrasts are found for SYN_PARKs_1 particularly for 50, 130, and 190 cm. This includes also differences in the spatial $\log_{10}K_s$ distribution compared to the other DA scenarios.

Figure 7 shows the spatial distribution of the differences between the ensemble hydraulic conductivity estimated for different DA scenarios and the reference hydraulic conductivity. This gives further insight in the change of the spatial patterns of the hydraulic conductivity during the assimilation period. For the scenarios SYN_PAR_1, SYN_PARan_1, SYN_PARKs_1, and SYN_PAR_5d in general a strong reduction of the RMSE

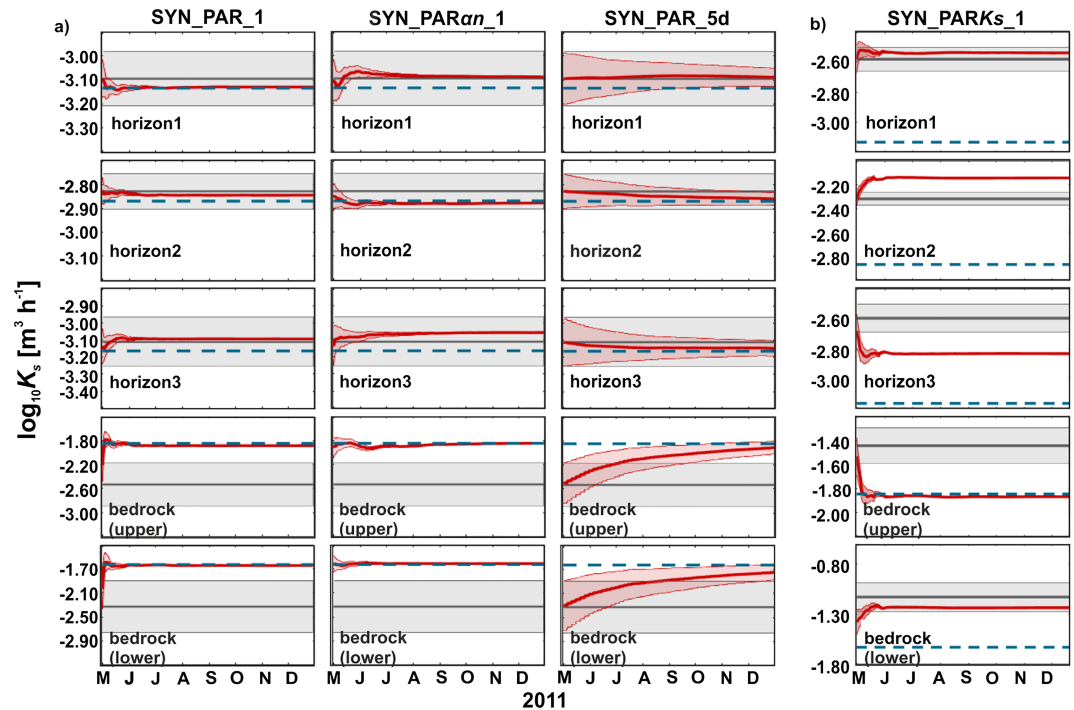


Figure 6. Ensemble mean soil hydraulic conductivity and standard deviation for synthetic experiments (red) against the open loop parameter set (gray) for different subsurface horizons of the Rollesbroich catchment. This includes scenarios without (a) and with biased hydraulic conductivity (b). The dashed line in blue indicates the synthetic truth. Notice that the used scales on the vertical axis differ for the different horizons.

is found. For SYN_Ks_OL and SYN_PARKs_1 a higher level of RMSE is found due to the introduction of the biased K_s . A large improvement is found at the sensor locations of Horizons 1–3 and also within the bedrock layers where no measurements were available. The spatial distribution of the RMSE reduction differs between SYN_PAR_1, SYN_PAR_5d, and the DA scenarios with biased K_s and spatial correlation range. Horizon 1 of SYN_PAR_1 is on average closer to the reference than its SYN_PAR_5d counterpart but shows some locations with relatively high RMSE (0.4 – $0.5 \log_{10} K_s$) in the center and the west of the study site. This is similar for Horizon 3. In contrast, the RMSE reduction is smoother and less spiky for the SYN_PAR_5d scenario. The spatial pattern of RMSE reduction for SYN_PARan_1 follows closely the spatial pattern of SYN_PAR_1. The spatial RMSE distribution for the bedrock layers also differs among the simulation scenarios. Local maxima of RMSE are found mainly in the north and center of the study site. SYN_PARKs_1 shows a significant RMSE reduction for Horizons 3–5 but less for Horizons 1 and 2. The RMSE patterns in SYN_Ks_OL and SYN_PARKs_1 reflect also the larger correlation range of K_s in the simulation setup. From these findings it can be concluded that the different spatial patterns of the estimated hydraulic conductivity are strongly associated with the update strategy resp. the model scenarios.

4.1.3. Discharge

SYN_ST shows strong discharge fluctuations until the high flow period associated with a high ensemble spread. In some realizations the Kieselbach fell dry during assimilation, in particular during the period May–June 2011 (Figure S4). This leads to an underestimation of discharge which is also indicated by a negative relative bias (-22.94% ; Table 3). The NSE (0.44) is higher and the RMSE smaller ($8.88 \text{ m}^3/\text{hr}$) compared to SYN_OL. This is probably related to the strong reduction of discharge peaks for which these measures are most sensitive. In contrast, SYN_PAR_1 shows limited discharge fluctuations and low ensemble spread. The performance for SYN_PAR_1 is much better than for the open loop run (SYN_OL) or state updating only (NSE: 0.61; PBIAS: 12.16%; RMSE: $7.34 \text{ m}^3/\text{hr}$), indicating a good adaptation of the simulations to the synthetic reference. The discharge SYN_PARan_1 has similar characteristics, whereas the discharge of SYN_PAR_5d can be characterized as intermediate between SYN_PAR_1 and SYN_ST. With an

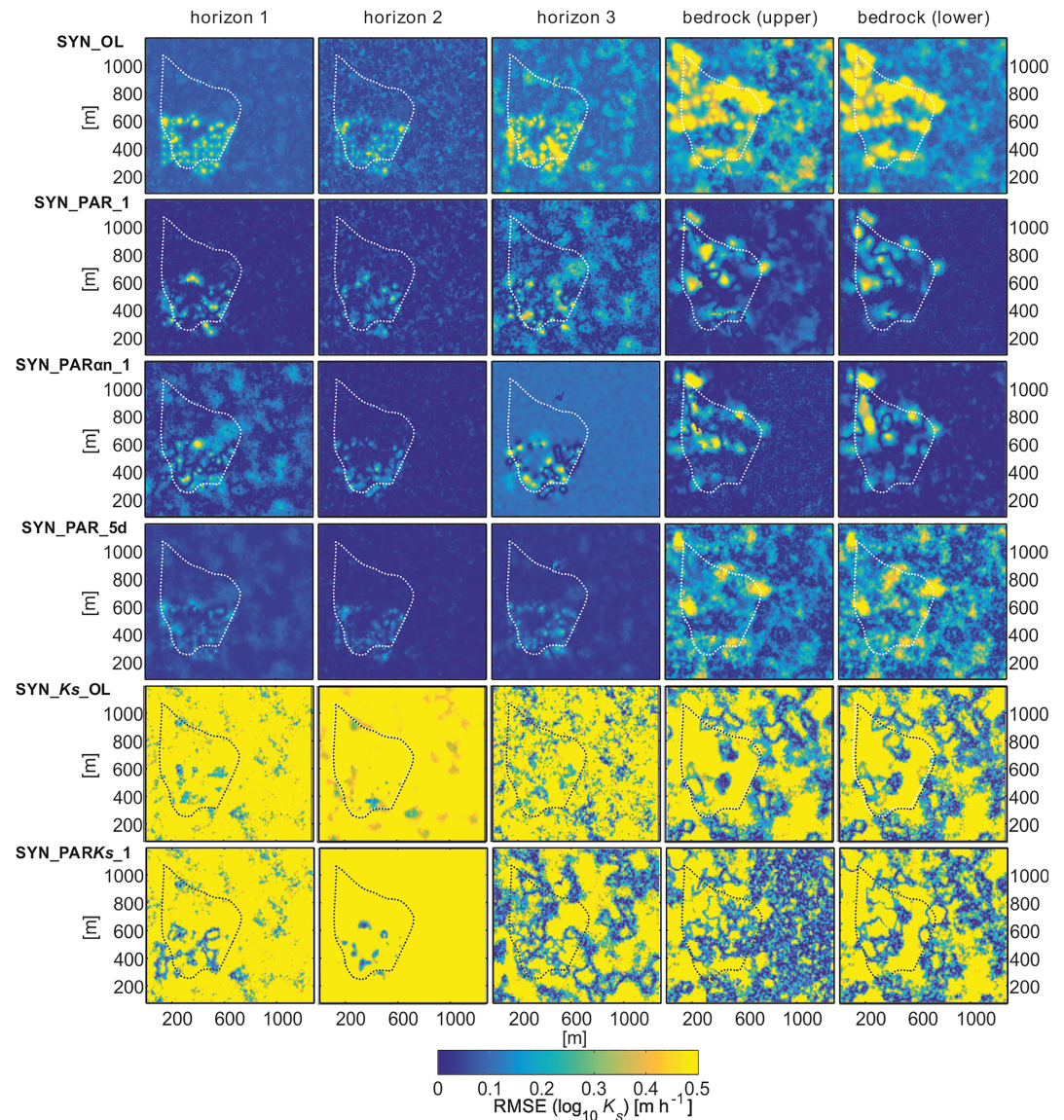


Figure 7. Root-mean-square error (RMSE) of the ensemble average hydraulic conductivity compared to the reference hydraulic conductivity ($\log_{10}K_s$) for the open loop and data assimilation experiments and for different soil horizons. The study site is indicated by a white or black dashed line.

increasing number of parameter updates during assimilation the discharge fluctuations and ensemble spread decrease, with a better adjustment of the discharge to the synthetic reference, an exception from the before mentioned findings SYN_PARKs_1. Compared to the associated open loop simulations (SYN_Ks_OL), SYN_PARKs_1 exhibits large bias (−39.28%) and a decrease of NSE and RMSE in combination with strong fluctuations for the majority of the assimilation period.

4.1.4. ET and Water Balance

In the synthetic experiments, the reproduction of ET is already almost perfect for the open loop run (without biased K_s) exhibiting very small bias and RMSE and very high NSE (0.99). In contrast, SYN_Ks_OL and SYN_PARKs_1 show a poor performance (NSE: 0.43–0.46) due to the K_s induced SWC bias in the upper soil layers. The DA runs with exception of SYN_PARan_1 yield an additional small improvement.

The water balance gaps for the synthetic experiments are mainly related to errors in the simulated discharge (Table 4) for simulations without K_s bias. The discharge of the open loop simulations is higher than the reference truth in the assimilation period (+24%). This gap is reduced by all DA scenarios. All DA scenarios are

Table 3

Discharge Model Performance at the Catchment Outlet for Different Update Scenarios With Synthetic Observations for 2011

Scenario	NSE (–)	PBIAS (%)	RMSE (m ³ /hr)
SYN_OL	–0.04	78.42	14.03
SYN_Ks_OL	0.56	–8.53	8.5
SYN_ST	0.44	–22.94	8.88
SYN_PAR_1	0.61	12.16	7.34
SYN_PARcan_1	0.46	12.38	8.72
SYN_PARKs_1	–0.91	–39.28	17.66
SYN_PAR_5d	0.59	–5.92	7.6

Note. The model performance is indicated by the mean Nash-Sutcliffe efficiency (NSE), percent bias, and the root-mean-square error (RMSE).

closer to the reference truth but show a negative bias (–17% to –9%) which is 62–68% of the total open loop discharge. The EnKF does not preserve the water balance and can add/extract water in correspondence with the measurements. Hence, large differences up to 134 mm can be found for the DA experiments in the assimilation period particularly for experiments with larger impact of the state update (e.g., SYN_ST and SYN_PAR_5d). For SYN_Ks_OL and SYN_PARKs_1 the water balance gap is related to both ET and discharge. Particularly, SYN_PARKs_1 shows a high deficit (26%), which is related to extraction of water during DA.

4.2. Real-World Experiments

4.2.1. SWC

In this section simulation results are discussed for SWC characterization with the help of the assimilation of data for the real-world case. Figure 8 shows the temporal evolution of the mean SWC for the open loop simulations and three different DA experiments for the assimilated locations in 2011. The measured mean SoilNet SWC at the study site for this period is high during winter and periods with intensive rainfall or thunderstorm events. The smallest SWC for 2011 is reached in May–June. This basic seasonality is captured well by the open loop and DA simulations. The comparison with the measured average SWC (Table 5) shows that the open loop simulation (REAL_OL) is close to the measurements at 5-cm depth (NSE: 0.894; RMSE: 0.026 cm³/cm³), but not at 20- and 50-cm depths with negative NSE (20 cm: –0.269; 50 cm: –7.711), high RMSE (20 cm: 0.048 cm³/cm³; 50 cm: 0.059 cm³/cm³) and significant bias (20 cm: 0.046 cm³/cm³; 50 cm: 0.057 cm³/cm³). The DA scenarios REAL_ST (state updating only) and REAL_PAR_1 (daily joint state-parameter updating) showed a slightly improved SWC average compared to the open loop run for May and June 2011 for the 5-cm layer, but these scenarios diverge from the observations afterward until the end of November 2011. Also, for the REAL_PAR_5d scenario with daily updates of states and each 5-day updates of parameters the scenario diverges from the observations during the (late) summer and autumn 2011. The average SWC at 20- and 50-cm depths is better characterized than in the open loop run for the three DA scenarios. For the 50-cm layer the high initial model bias gradually decreased over the entire assimilation period. Whereas the three DA scenarios for 5- and 20-cm depths give very similar results (October and November), at 50-cm depth REAL_PAR_5d outperforms REAL_ST and REAL_PAR_1, which shows the smallest performance improvement compared to the open loop. The bias for SWC-characterization at 50-cm depth decreases from 0.057 cm³/cm³ (REAL_OL) to 0.017 cm³/cm³ (REAL_PAR_5d).

Figure 9 displays the RMSE for individual SoilNet observations at 5-, 20-, and 50-cm depths. The real-world experiments REAL_ST and REAL_PAR_1 show decreasing simulation performance (larger RMSE than in open loop) for 62–67% of the sensor locations at 5-cm depth. Only less frequent assimilation in combination

Table 4

Observed and Simulated Yearly Water Balance Components: Precipitation (P), Evapotranspiration (ET), and Surface Runoff (Q) of the Rollesbroich Catchment for Different Synthetic Simulation Scenarios

Scenario	P (mm)	ET (mm)			Q (mm)			P-ET-Q (mm)	
	Sum	Sum	Max	Min	Sum	Max	Min	Mean	%
SYN_OL	953	488	513	437	546	1118	165	–81	–8.5
SYN_Ks_OL	953	459	486	409	371	797	78	–123	–12.9
Reference Truth	953	487	n/a	n/a	406	n/a	n/a	60	6.3
SYN_ST	953	483	496	462	336	625	160	134	14.1
SYN_PAR_1	953	484	496	463	371	594	205	98	10.3
SYN_PARcan_1	953	484	492	470	371	587	206	113	11.9
SYN_PARKs_1	953	364	378	338	339	388	292	–250	26.2
SYN_PAR_5d	953	484	496	462	353	599	170	116	12.2

Note. n/a = not available.

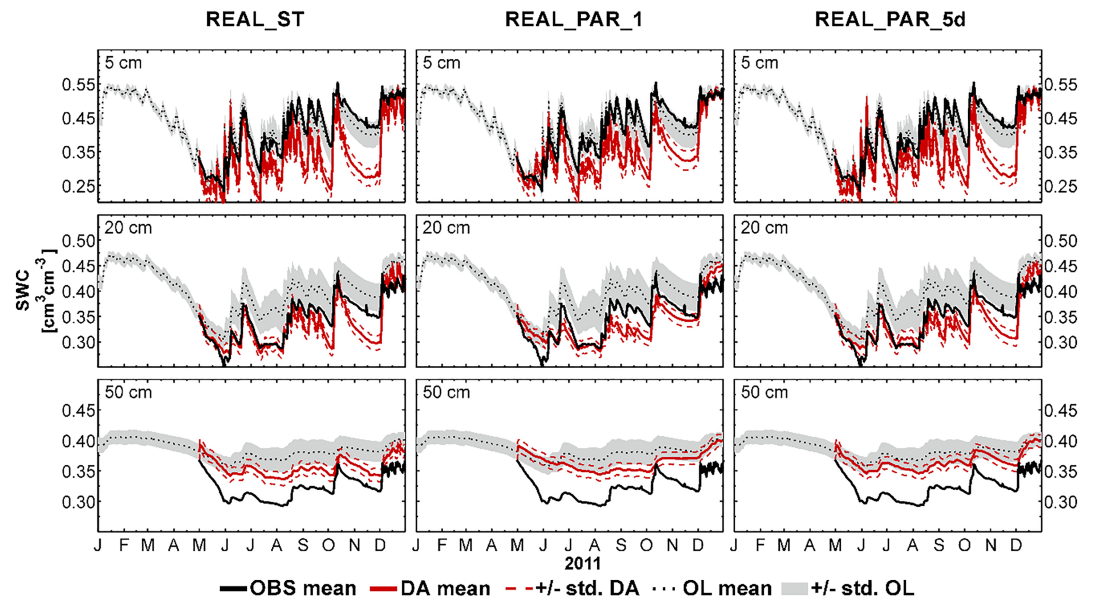


Figure 8. Mean soil water content (SWC) and standard deviation of SWC for the open loop and the real-world data assimilation (DA) experiments for the assimilation locations. Mean SoilNet observations are indicated as well.

with joint state-parameter estimation results in a small decrease in layer-averaged RMSE of the individual sensors (REAL_OL: $0.0936 \text{ cm}^3/\text{cm}^3$; REAL_PAR_5d: $0.0858 \text{ cm}^3/\text{cm}^3$) at this depth. However, an improvement was found for SWC model estimation at 20- and 50-cm depths for all real-world DA experiments (16–28% RMSE-reduction at 20 cm and 8–18% RMSE reduction at 50 cm). Nevertheless, at 50-cm depth the initial bias is only partly removed in the different DA scenarios. REAL_PAR_5d reproduces SWC profiles the best over all three layers showing a reduction by 14% for the RMSE averaged over all sensor locations. For REAL_ST the RMSE decreases by 6%, whereas REAL_PAR_1 shows only a 2% RMSE decrease. RMSE decreases or increases thereby do not show vertical correlations. Spatial RMSE structures are not very similar for the different simulation scenarios. In particular, the spatial RMSE structure for REAL_PAR_5d differs from REAL_ST and REAL_PAR_1.

Compared to the open loop simulation for the synthetic case, the open loop simulation for the real-world case (RMSE: $0.0825\text{--}0.094 \text{ cm}^3/\text{cm}^3$) has a larger layer-averaged RMSE over the observation locations ($0.044\text{--}0.056 \text{ cm}^3/\text{cm}^3$). These numbers suggest that 40% of the layer-averaged RMSE over the observation locations for the real-world case could be related to model structural errors (processes which are not

Table 5
Performance Measures for the Reproduction of Average SWC for the Real-World Experiments

		REAL_OL	REAL_ST	REAL_PAR_1	REAL_PAR_5d
5 cm	NSE (–)	0.894	–0.429	0.007	0.436
	BIAS (cm^3/cm^3)	0.002	–0.079	–0.064	–0.041
	RMSE (cm^3/cm^3)	0.026	0.095	0.079	0.053
20 cm	NSE (–)	–0.269	0.599	0.596	0.576
	BIAS (cm^3/cm^3)	0.046	–0.013	–0.009	–0.025
	RMSE (cm^3/cm^3)	0.048	0.027	0.027	0.030
50 cm	NSE (–)	–7.711	–2.000	–3.935	–1.372
	BIAS (cm^3/cm^3)	0.057	0.033	0.042	–0.017
	RMSE (cm^3/cm^3)	0.059	0.034	0.044	0.017

Note. Performance statistics (Nash-Sutcliffe efficiency, NSE; root-mean-square error, RMSE; and absolute BIAS) for the open loop run (REAL_OL) and the DA experiments for May–December 2011 are shown.

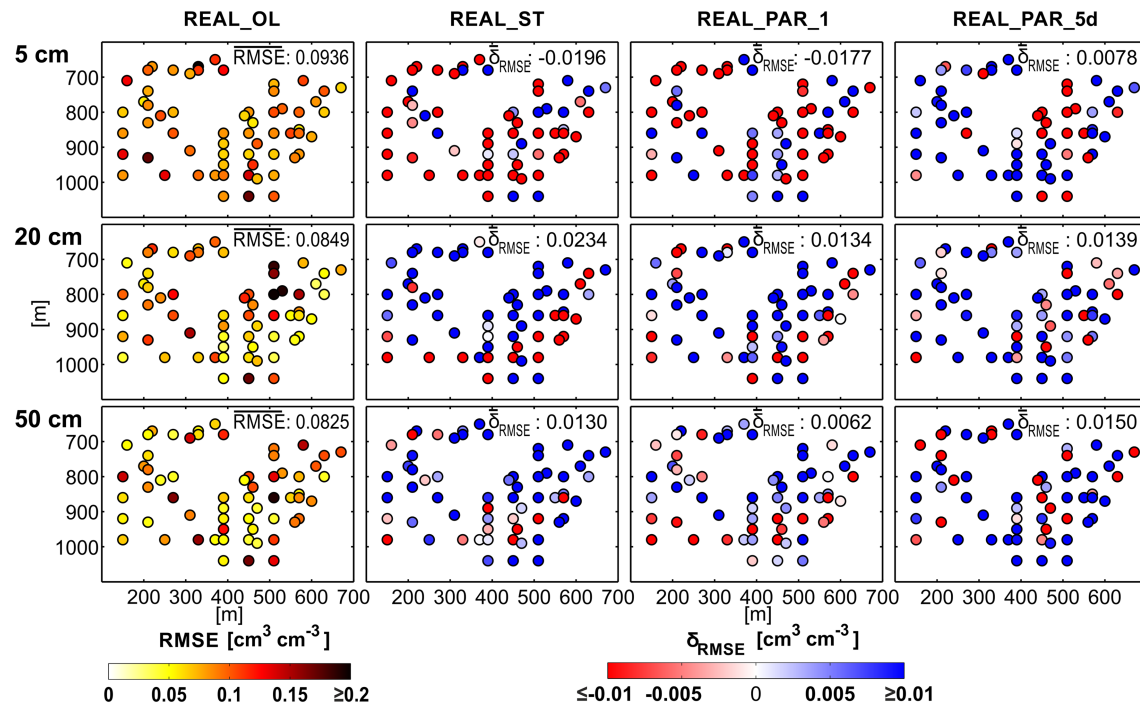


Figure 9. Root-mean-square error (RMSE) of soil water content at individual locations for the open loop runs (left column) and changes in RMSE (increase implies improvement and decrease implies impoverishment) for three data assimilation scenarios (three columns on the right) of the real-world case for 2011.

captured by the model) or systematic biases in the model input (systematic errors in model forcings or values for soil hydraulic parameters).

4.2.2. Parameter Estimation

Figure 10 illustrates the temporal evolution of the hydraulic conductivity ($\log_{10}K_s$) estimates for the real-world experiments. For REAL_PAR_1 the ensemble spread rapidly decreases within the first weeks of assimilation. In contrast, the ensemble spread of REAL_PAR_5d with less frequent parameter updating (each five days only) still exhibits ~25% of the initial ensemble spread at the end of the assimilation period. The different DA scenarios show large spatial variations in $\log_{10}K_s$ estimates for the real-world case. In particular, $\log_{10}K_s$ of REAL_PAR_1 varies between the individual horizons and the final parameter estimates show major differences in comparison with REAL_PAR_5d. The temporal evolution of $\log_{10}K_s$ for REAL_PAR_1 shows also some instability. The $\log_{10}K_s$ of REAL_PAR_1 for the second and third layer first rapidly decreases resp. increases until the end of May. Afterward, it increases resp. decreases again until it converges to a final parameter estimate which is reached at the end of July. The final estimates differ up to $0.6 \log_{10}K_s$ (m/hr) between REAL_PAR_1 and REAL_PAR_5d.

Figure 11 shows the differences of the ensemble average hydraulic conductivity fields, as updated in the DA scenarios, compared to the ensemble average hydraulic conductivity of the open loop run, for the real-world case. The figure shows that for REAL_PAR_1 parameter updates are stronger than for REAL_PAR_5d. Similar to the synthetic scenarios, REAL_PAR_1 exhibits high $\log_{10}K_s$ contrasts between the sensor locations and also for the bedrock layer between different areas of the northern study site. Also REAL_PAR_5d shows positive and negative changes in $\log_{10}K_s$ which are more limited spatially.

4.2.3. Discharge

Figure 12 shows the discharge dynamics for the real-world experiments for 2011. In general, the observations at the study site show high flow during winter rain and snowmelt, and low flow from May–November. This seasonality is captured by the open loop simulations, although discharge is overestimated during low flow conditions and for individual winter peaks. These findings are also reflected by the performance measures for the open loop simulations in Table 6. In comparison with the on-site measurements, the open loop simulations show a high relative bias (64.98%) and high RMSE ($15.77 \text{ m}^3/\text{hr}$). The DA simulations

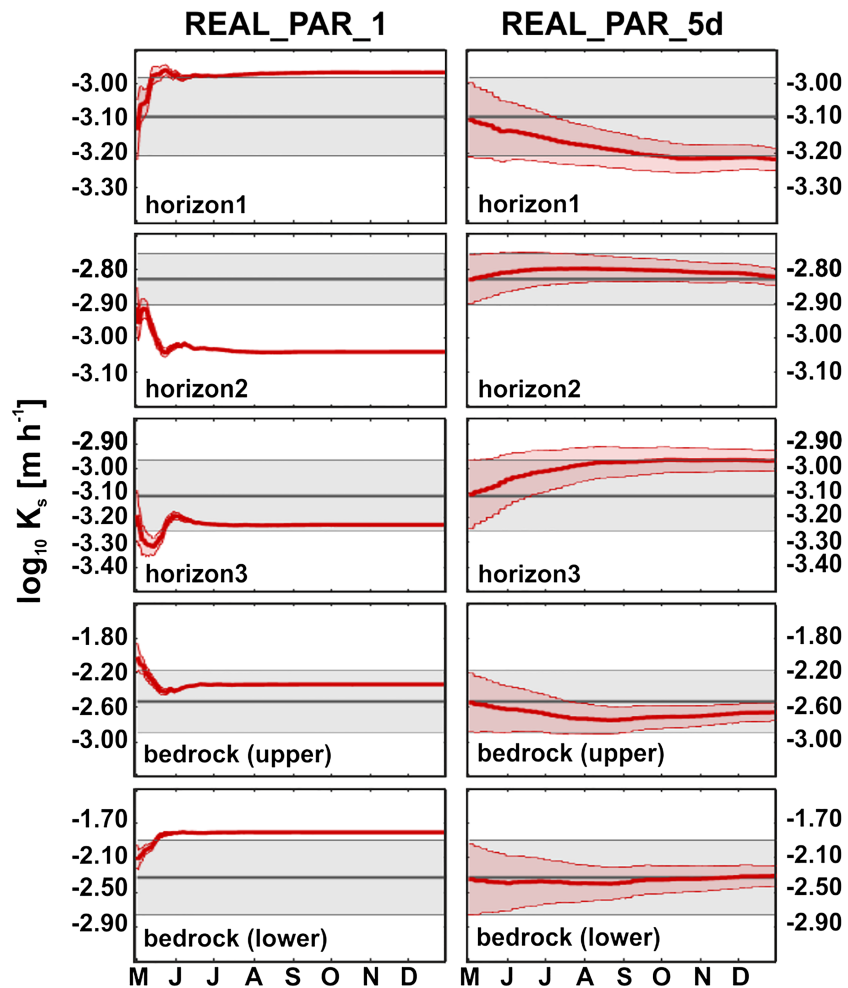


Figure 10. Mean and standard deviation of the soil hydraulic conductivity for the real-world experiments (red) against the open loop parameter set (gray) for different subsurface horizons of the Rollesbroich catchment. Please note that the used scales on the vertical axis differ for the different horizons.

exhibit a very different behavior. With the more intensive adjustment of SWC in the upper horizons, also discharge is affected by the assimilation cycles. This leads to absence of streamflow in the REAL_ST and the REAL_PAR_5d in November 2011. As a consequence, both scenarios exhibit a high relative bias (REAL_ST: -88.84% ; REAL_PAR_5d: -75.00%) indicating a discharge underestimation. The performance of REAL_ST (NSE: 0.05; RMSE: $20.43 \text{ m}^3/\text{hr}$) is even lower than for the open loop run (NSE: 0.49; PBIAS: 64.98% ; RMSE: $15.77 \text{ m}^3/\text{hr}$). Parameter estimation (REAL_PAR_1) results in an improved modeling of discharge (NSE: 0.67; PBIAS: -23.55% ; RMSE: $12.08 \text{ m}^3/\text{hr}$). In particular the bias during low flow periods is strongly reduced for these simulations. However, the peak discharge of REAL_PAR_1 is still overestimated during the entire assimilation period.

4.2.4. Water Balance

The water balance gaps for the real-world experiments are mainly related to discharge differences. Table 7 gives an overview of the different measured and modeled water balance components at the study site for 2011, where all model simulations overestimate discharge. Large differences up to 224 mm can be found for the DA experiments, which indicate that the EnKF extracts a significant amount of water in the assimilation period through the adaptation of the pressure heads. Similar to the synthetic experiments, higher balance gaps are found for the real-world experiments with the strong adaptation of the pressure heads within the simulations (REAL_ST and REAL_PAR_5d).

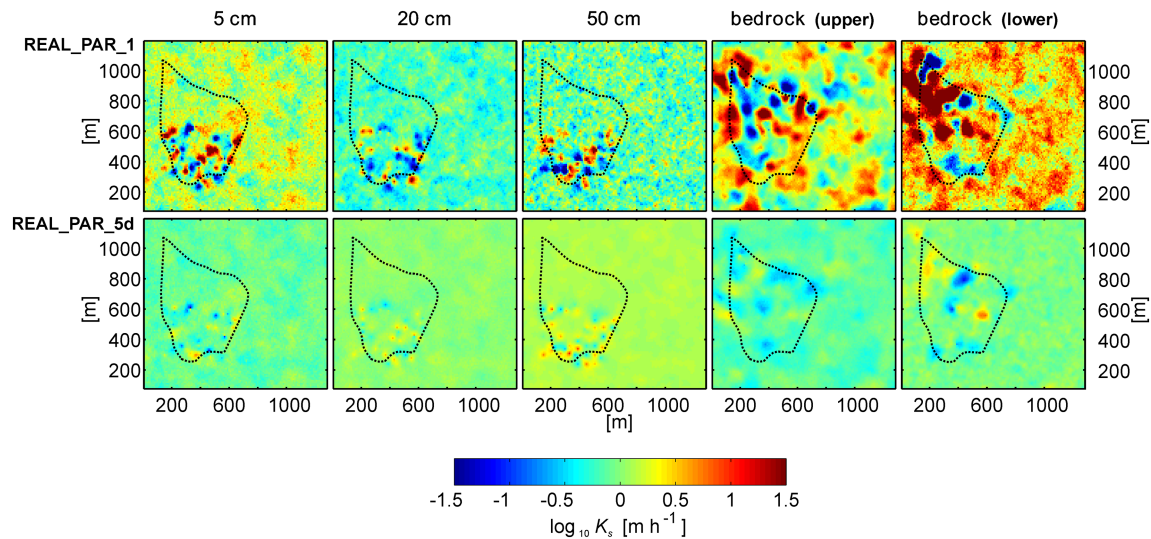


Figure 11. Ensemble averaged differences of hydraulic conductivity fields ($\log_{10}K_s$) between open loop and real-world scenarios at the end of the data assimilation period (December 2011). Indicated are results for the soil layers at 5-, 20-, and 50-cm depths as well as the upper and lower bedrock layer. The dashed line marks the test site borders within the model domain.

Measured ET was not available for 2011. However, given the range of lysimeter observations for 2012 (467–523 mm with some gap filling according Gebler et al., 2015) and a possible reduction of actual ET in 2011 due to water limitation in May–June, the simulated ET (467–488 mm) is in a reasonable range.

5. Discussion

Our findings indicate that the EnKF in combination with fully coupled land surface-subsurface models and SWC data from a dense observation network can potentially improve the characterization of the hydrology at the hillslope scale. For the real-world case joint updating of model states and hydraulic conductivity is

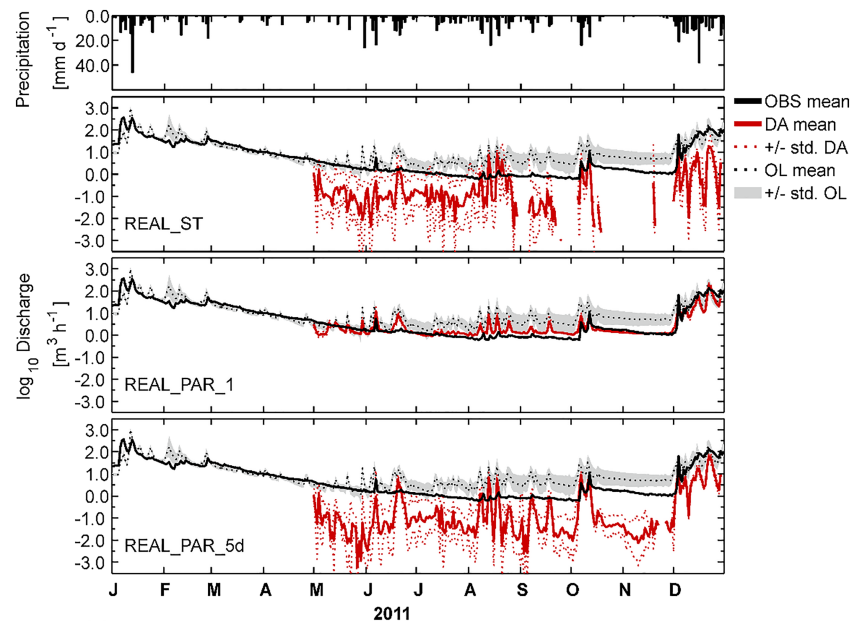


Figure 12. Simulated and observed daily mean logarithmic discharge for the assimilation period at the Rollesbroich catchment outlet. Open loop simulations are shown together with different updating scenarios for the real-world experiments. Precipitation is also indicated on a daily basis. DA = data assimilation.

Table 6
Discharge Model Performance for the Assimilation Period at the Catchment Outlet for Different Update Scenarios With Real-World Observations in 2011

	REAL_OL	REAL_ST	REAL_PAR_1	REAL_PAR_5d
NSE (–)	0.49	0.05	0.67	0.41
PBIAS (%)	64.98	–88.84	–23.55	–75.00
RMSE (m ³ /hr)	15.77	20.43	12.08	9.17

Note. The model performance is indicated by the mean Nash-Sutcliffe efficiency (NSE), percent bias, and root-mean-square error (RMSE).

more effective in updating the SWC (14% RMSE reduction averaged over all locations and layers for REAL_PAR_5d compared to open loop) than state update alone (6% RMSE reduction). However, a larger impact for parameter updating could have been expected. In contrast to their real-world counterparts, the synthetic scenarios exhibit a more significant improvement of SWC after DA showing a relatively strong RMSE reduction at the average over all SWC verification locations from 40% (SYN_ST) to 55% (SYN_PAR αn _1). This can be called satisfying, especially because the evaluations in these synthetic runs are made at verification locations not used in the assimilation.

The fact that for the synthetic case simulation results are better than for real-world case (with 40% lower layer-averaged RMSE of SWC for the open loop run) indicates that model structural errors play a role for the SWC estimation. One possible explanation, which was addressed with the simulation experiments, was the fact that all soil hydraulic parameters are uncertain, but only hydraulic conductivity is estimated. For a relatively simple 1-D subsurface flow model Li and Ren (2011) already demonstrated that the update of multiple Mualem-van Genuchten parameters (i.e., hydraulic conductivity, α , and n) is more effective than updating hydraulic conductivity alone. This issue has been analyzed by the SYN_PAR αn _1 and SYN_PAR αn _5_256 scenarios with perfect knowledge of Mualem-van Genuchten α and n . Although overall the SYN_PAR αn scenarios give smaller errors, the improvement with respect to their counterparts with uncertain α and n is small. This indicates that model simulations could be further improved by an estimation of these parameters but that their impact at least in this study is relatively small. However, under drier conditions the SWC could be more strongly governed by α and n , so that estimating the three parameters jointly could potentially improve SWC characterization more than in this study. Chaudhuri et al. (2018) showed that under certain conditions the joint estimation of multiple spatially distributed parameter fields can be successful, but it strongly depends on data availability and the joint presence of very dry and very wet conditions in the data set.

The performance differences between the real-world case and the synthetic case are also analyzed with a scenario for the synthetic case that departs from erroneous prior geostatistical parameters (saturated hydraulic conductivity and variogram range). It is found that modest errors already result in a declined performance of the DA, especially for the fluxes in this case. The too high saturated hydraulic conductivity for the upper soil layers creates drought stress and a significant underestimation of ET, which is not corrected by DA. This simulation experiment shows that erroneous values for the prior geostatistical parameters are a possible explanation for the relative poor performance in the real-world case.

Table 7
Observed and Simulated Yearly Water Balance Components: Tipping Bucket Precipitation (P), Evapotranspiration (ET), and Surface Runoff (Q) of the Rollesbroich Catchment for Different Real-World Simulation Scenarios

OBSERVATION										
Period		P [mm]	ET [mm]			Q [mm]			P-ET-Q [mm]	
2011		953	n/a			392			n/a	n/a
REAL-WORLD EXPERIMENTS										
		P [mm]	ET [mm]			Q [mm]			P-ET-Q [mm]	
Scenario	Period	Sum	Sum	Max	Min	Sum	Max	Min	Mean	%
REAL_OL	2011	953	488	513	437	546	1118	165	−81	−8.5
REAL_ST	2011	953	467	483	444	272	447	139	214	22.5
REAL_PAR_1	2011	953	471	486	444	342	558	180	140	14.7
REAL_PAR_5d	2011	953	467	483	445	262	423	137	224	23.5

Note. n/a = not available.

Such erroneous initial guesses of geostatistical parameters can, for example, also be related to the existence of preferential flow through macropores, which was already observed at the nearby Wüstebach site (Wiekenkamp et al., 2016) and other studies investigating in situ SWC observations (e.g., Martini et al., 2015; Poltoradnev et al., 2016). The underrepresentation of preferential flow in the model might explain the high bias at 50-cm depth for the real-world simulations having a large influence on the infiltration of a grassland with an extensive rooted top soil layer (Weiler & Naef, 2003). This issue was already indicated by Gebler et al. (2017) for a slightly different ParFlow-CLM setup for the Rollesbroich study site. Furthermore, the representation of the saprolite and the drainage system in the subsurface might be suboptimal in our simulation model and could be better represented with a dual porosity approach (e.g., Frey et al., 2012; Frey et al., 2016). Moreover, an unknown additional storage in the bedrock fractures potentially has influence on the baseflow of the Kieselbach (Hale et al., 2016). The representation of the drainage system in the model might be an additional cause for the model structural error, although no direct evidence in the SWC spatial error distribution was found.

Other limitations of the DA experiments are potentially associated with update frequency, ensemble size, and filter inbreeding. Whereas the larger ensemble size of 256 gives slightly better results than the ensemble size of 128, the overall effect of the increased ensemble size is small. Nevertheless, it is found that in the DA experiments, especially in case of daily updating and the real-world experiment, the ensemble variance decreases too fast and is too low. For the real-world case, in too many cases, the observations are not covered by the ensemble of model simulations. As a consequence, restricted weights are assigned to the measurements, which have too little influence on the state and parameter updates (e.g., Evensen, 2004; Hendricks Franssen & Kinzelbach, 2008; Houtekamer & Mitchell, 1998; Zhang et al., 2007). For the synthetic case, the observations are generally well covered by the ensemble of model simulations shown in Figure 5 (note that ± 1 standard deviation is plotted for the ensembles). The only exception is scenario SYN_PARKs_1 with biased hydraulic conductivity, indicating the negative impact of biases, which could also be the reason for the reduced performance of the real-world case. While for the synthetic case the ensemble spread for soil moisture is adequate in general, the parameter spread is sometimes too narrow. Figure 6 shows that this is the case for the shallow layers, for several of the scenarios with daily parameter updating. A possible explanation is that updating hydraulic conductivity compensates for uncertain α and n parameters, which are not updated. However, also for the scenario SYN_PARKs_1 the ensemble of estimated parameters does not include the reference values for Horizons 1 and 3. Therefore, also the synthetic case appears to be affected to some degree by filter inbreeding impacting the parameter estimation, in spite of applying dampening and tests with a relatively large ensemble size of 256. In a comparison study of seven EnKF variants, Keller et al. (2018) showed for two physical setups and 1,000 synthetic studies per physical setup that the introduction of the damping factor improves the parameter estimation for relatively small ensemble sizes, in comparison to the standard EnKF without damping factor. Although the damping factor and other simple inflation methods have been applied successfully (e.g., Aksoy et al., 2006; Erdal et al., 2014; Hendricks Franssen & Kinzelbach, 2008), these methods are subject of discussion (Houtekamer & Zhang, 2016). The reduced impact of the error statistics in the EnKF potentially obscures other sources of error and, therefore, can lead to inconsistencies in the derivation of the Kalman gain (Houtekamer & Zhang, 2016). The question is how the estimation of spatially distributed parameters can be further improved in the future avoiding filter inbreeding. In order to answer this question, we see three different strategies, which can also be applied in combination. First, the application of localization to reduce the influence of spurious correlations may be useful (Houtekamer & Mitchell, 2001; Anderson, 2007). A more sophisticated DA strategy including localization could potentially improve the results, as suggested by Rasmussen et al. (2015) who performed tests with integrated terrestrial models, but this option was not yet available and implemented in our TerrSysMP-PDAF framework. Second, the use of normal score ensemble Kalman filter (NS-EnKF; Zhou et al., 2012; Schöniger et al., 2012), which includes a step to make the marginal distributions of states and parameters Gaussian, may show improvements. The NS-EnKF update is performed then in terms of transformed variables. The transformation may be helpful in case of strongly skewed pressure head distributions in the vadose zone under dry conditions (Erdal et al., 2015; Zhang et al., 2018). Especially under extremely dry conditions it could be expected that NS-EnKF outperforms classical EnKF. However, our study area was not affected by these very dry conditions; thus, the advantage of using NS-EnKF, in combination with a relative small ensemble size, would probably have been limited. NS-EnKF is not implemented yet in combination

with TerrSysMP-PDAF because it is associated with additional challenges in terms of (complex) parallelization. Third, in very recent work, we found that combining the EnKF with a pilot points method (PP-EnKF) can preserve ensemble spread much better than (other variants of) EnKF. The PP-EnKF method does not need more compute time than classical EnKF. In this method, the update at measurement locations is done with numerically estimated covariances, and between measurement locations analytical covariance functions are applied for parameters. The method still needs to be implemented in the TerrSysMP-PDAF framework.

Discharge was in general not well reproduced in the experiments showing that the SWC characterization is not the main factor for improving discharge estimation. Nevertheless, for the best performing scenarios in the synthetic case the considerable NSE improved from -0.04 to 0.61 is related to assimilation of SWC-data and parameter estimation. For this small hillslope scale where precipitation uncertainty plays a minor role, still a higher impact of SWC-assimilation could have been expected. Particularly for the real-world case where the improvement of discharge estimation related to DA and parameter estimation was smaller. The representation of the Kieselbach in the model is critical (Gebler et al., 2017). Given the small spatial extent of the channel (0.3 – 1.0 m) the interaction between surface water and groundwater is highly sensitive to the vertical pressure gradient which controls the reinfiltration of groundwater into the channel. The assimilation of discharge data, combined with the updating of Manning's roughness coefficient as well as other Mualem-van Genuchten soil hydraulic properties, could further improve discharge estimation. Baatz et al. (2017) demonstrated the benefit of estimating Manning's roughness coefficients on the discharge simulation of a 2-D regional-scale synthetic ParFlow model.

The assimilation of SWC data had in most scenarios no significant effect on the reproduction of the annual actual ET in the synthetic simulations, which was already very good in the open loop. However, the scenario with a systematic too high saturated hydraulic conductivity for the upper soil layers created drought stress and underestimation of ET, showing that SWC assimilation for drier conditions with limited water supply has potentially more impact.

Overall, this study indicates that DA with integrated physically based terrestrial systems models can potentially strongly improve SWC and parameter characterization, but that (realistic) biases in prior (geo)statistical parameters can already degrade the performance considerably. Although DA under those conditions with bias still improves SWC characterization, the estimation of fluxes like discharge and ET might degrade. This points to the high importance of the assimilation of additional data, including ET measurements or proxies of ET, which is until now not very common in integrated terrestrial systems modeling, as these data are often only used for verification.

6. Conclusions

This study investigated the assimilation of high-resolution SWC data with the EnKF in the fully coupled land surface-subsurface model TerrSysMP. The assimilation experiments were performed for the small headwater grassland catchment Rollesbroich in the Eifel (Germany), or alternatively, for a synthetic reality which mimics this catchment. DA experiments were performed with an ensemble size of 128 (256) model realizations at 10×10 -m lateral resolution and a variable vertical resolution (0.025 – 0.575 m) which resulted in a problem size of 0.3 million unknowns. The individual realizations were set up with a fully heterogeneous subsurface with geostatistical realizations of Mualem-van Genuchten soil hydraulic properties. The following DA experiments were performed: (i) daily state updating, (ii) daily joint state parameter updating, and (iii) daily state updating combined with parameter updating each 5 days and a damping factor. In addition, for the synthetic test case also experiments were performed where (i) Mualem-van Genuchten parameters α and n were deterministic and only saturated hydraulic conductivity uncertain, and (ii) prior geostatistical parameters like the mean saturated hydraulic conductivity and the range of all soil hydraulic parameters were systematically biased. In all experiments SWC data from 61 sensor network locations and three depths (5, 20, and 50 cm) were assimilated.

Considerable improvement was found for the synthetic case, which showed on average 40–55% RMSE reduction for the verification locations and for the different DA scenarios. Also the discharge estimation was improved with a NSE of -0.04 for the open loop and ~ 0.61 for the DA scenarios, which involved parameter estimation. It was found that estimated saturated hydraulic conductivity was much closer to the

reference values (after joint state-parameter updating). However, in case of biased prior geostatistical parameters the characterization of SWC and saturated hydraulic conductivity improved less, and the estimation of discharge and ET was even degraded with DA.

For the real-world case the EnKF in combination with joint updating of model states and hydraulic conductivity was more efficient in updating the SWC than state updating alone. For joint state-parameter updating the average RMSE for the overall SWC sensor locations decreased by 14% for the real-world scenario with joint state-parameter updating and damping while state updating reduced the RMSE only by 6%. The improvement of the SWC characterization for the real-world case was limited. The uppermost layer of 5-cm depth exhibited a performance impoverishment, indicated by strong RMSE increases. The discharge for the verification period was only marginally improved, although some discharge improvement was found, if both states and parameters were updated with DA. In particular, the systematic bias up to 65% between real-world model simulations and measurements was not significantly improved.

In order to explain the very different results for the real-world case and the synthetic case, it is important to point to the very different RMSE for the open loop runs for the real-world case and the synthetic case. The RMSE for the open loop run, calculated over the 61 measurement locations and upper three layers, was 53% lower for the synthetic case than for the real-world case. This can be related to model structural errors, and these same model structural errors might have inhibited the positive impact of DA as updated parameters just might have compensated for model structural errors. In this context, it was investigated whether results could be better, if besides saturated hydraulic conductivity also Mualem-van Genuchten parameters α and n were estimated. This was tested for the synthetic case by assuming that α and n were perfectly known and only saturated hydraulic conductivity was uncertain. Results improved only very slightly compared to the scenarios where α and n were unknown and point to the fact that for this case the uncertain α and n were not a main additional limitation to improve SWC estimation by DA. However, from the synthetic experiment with relative modest biases in the prior geostatistical parameters, for example related to preferential flow, we conclude that such biases could explain the relative poor performance of the DA experiments for the real-world case.

In summary, the potential of assimilating SWC observations at the hillslope scale to improve the characterization of spatially distributed SWC and discharge has been demonstrated for the synthetic case, which mimicked the real-world case with all its complexities including 3-D fully distributed fields of saturated hydraulic conductivity and Mualem-van Genuchten parameters α and n . The fact that for the real-world case the improvements induced by DA were much smaller we attribute mainly to model structural errors like biases in prior geostatistical parameters and also other model structural errors like the representation of drainage might have played some role. These results might be site specific, and a similar approach could give better results at another hillslope site. However, it also points to the importance of precise estimates of prior geostatistical parameters and the representation of small-scale processes (e.g., macropore flow), which are difficult to capture, even with a physically based integrated terrestrial systems model. The study suggests that for predictions with integrated terrestrial system models it is essential to assimilate multiple data types, including (proxies of) ET.

Acknowledgments

This research had been supported by the TERENO research infrastructure. We thank the Transregio32 (Transregional Collaborative Research Centre 32-Patterns in Soil-Vegetation-Atmosphere Systems) for contributing data to this study and acknowledge all scientists and technicians accounting for the realization and maintenance of the research facilities. We particularly thank Ansgar Weuthen and Bernd Schilling for setup and maintenance of SoilNet sensors at the Rollesbroich study site. The authors further gratefully acknowledge the computing time granted by the JARA-HPC Vergabegremium and provided on the JARA-HPC Partition part of the supercomputer JUQUEEN at Forschungszentrum Jülich. Model setup data and simulation results are available online (<https://datapub.fz-juelich.de/slots/rollesbroich/>).

References

- Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell, P. E., & Rasmussen, J. (1986). An introduction to the European Hydrological System—Système Hydrologique Européen, “SHE”, 2: Structure of a physically-based, distributed modelling system. *Journal of Hydrology*, 87(1–2), 61–77. [https://doi.org/10.1016/0022-1694\(86\)90114-9](https://doi.org/10.1016/0022-1694(86)90114-9)
- Aksoy, A., Zhang, F., & Nielson-Gammon, J. W. (2006). Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model. *Monthly Weather Review*, 134(10), 2951–2970. <https://doi.org/10.1175/MWR3224.1>
- Anderson, J. L. (2007). Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D: Nonlinear Phenomena*, 230(1–2), 99–111. <https://doi.org/10.1016/j.physd.2006.02.011>
- Ashby, S. F., & Falgout, R. D. (1996). A parallel multigrid preconditioned conjugate gradient algorithm for groundwater flow simulations. *Nuclear science and engineering*, 124(1), 145–159. <https://doi.org/10.13182/NSE96-A24230>
- Baatz, D., Kurtz, W., Hendricks Franssen, H. J., Vereecken, H., & Kollet, S. J. (2017). Catchment tomography—An approach for spatial parameter estimation. *Advances in Water Resources*, 107, 147–159. <https://doi.org/10.1016/j.advwatres.2017.06.006>
- Bailey, R. T., & Baù, D. (2012). Estimating geostatistical parameters and spatially-variable hydraulic conductivity within a catchment system using an ensemble smoother. *Hydrology and Earth System Sciences*, 16(2), 287–304. <https://doi.org/10.5194/hess-16-287-2012>
- Bixio, A., Gambolati, G., Paniconi, C., Putti, M., Shestopalov, V., Bublias, V., et al. (2002). Modeling groundwater-surface water interactions including effects of morphogenetic depressions in the Chernobyl exclusion zone. *Environmental Geology*, 42(2–3), 162–177. <https://doi.org/10.1007/s00254-001-0486-7>

- Bogena, H., (2003). Die Grundwasserneubildung in Nordrhein-Westfalen. URL: http://wwwzb.fz-juelich.de/contentenrichment/inhaltsverzeichnis/2010/verlag/umwelt37_i.PDF. Schriften des Forschungszentrums Jülich. Reihe Umwelt, Forschungszentrum, Zentralbibliothek, Jülich, 148 pp.
- Bonan, G. B., Oleson, K. W., Vertenstein, M., Levis, S., Zeng, X. B., Dai, Y. J., et al. (2002). The land surface climatology of the community land model coupled to the NCAR community climate model. *Journal of Climate*, 15(22), 3123–3149. [https://doi.org/10.1175/1520-0442\(2002\)015<3123:TLSCOT>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<3123:TLSCOT>2.0.CO;2)
- Brown, R. N., Percivalle, C., Narkiewicz, S., & DeCuollo, S. (2010). Relative rooting depths of native grasses and amenity grasses with potential for use on roadsides in New England. *HortScience*, 45(3), 393–400.
- Burgers, G., van Jan Leeuwen, P., & Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, 126(6), 1719–1724. [https://doi.org/10.1175/1520-0493\(1998\)126<1719:ASITEK>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2)
- Burgess, T. M., & Webster, R. (1980). Optimal interpolation and isarithmic mapping of soil properties. *Journal of Soil Science*, 31(2), 333–341. <https://doi.org/10.1111/j.1365-2389.1980.tb02085.x>
- Camporese, M., Paniconi, C., Putti, M., & Orlandini, S. (2010). Surface-subsurface flow modeling with path-based runoff routing, boundary condition-based coupling, and assimilation of multisource observation data. *Water Resources Research*, 46, W02512. <https://doi.org/10.1029/2008WR007536>
- Camporese, M., Paniconi, C., Putti, M., & Salandin, P. (2009). Ensemble Kalman filter data assimilation for a process-based catchment scale model of surface and subsurface flow. *Water Resources Research*, 45, W070731. <https://doi.org/10.1029/2008WR007031>
- Carsel, R. F., & Parrish, R. S. (1988). Developing joint probability distributions of soil water retention characteristics. *Water Resources Research*, 24(5), 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>
- Chen, F., Crow, W. T., Starks, P. J., & Moriasi, D. N. (2011). Improving hydrologic predictions of a catchment model via assimilation of surface soil moisture. *Advances in Water Resources*, 34(4), 526–536. <https://doi.org/10.1016/j.advwatres.2011.01.011>
- Chen, Y., & Zhang, D. (2006). Data assimilation for transient flow in geologic formations via ensemble Kalman filter. *Advances in Water Resources*, 29(8), 1107–1122. <https://doi.org/10.1016/j.advwatres.2005.09.007>
- Cornelissen, T., Diekkrüger, B., & Bogena, H. R. (2014). Significance of scale and lower boundary condition in the 3D simulation of hydrological processes and soil moisture variability in a forested headwater catchment. *Journal of Hydrology*, 516, 140–153. <https://doi.org/10.1016/j.jhydrol.2014.01.060>
- Dickinson, R.E., Henderson-Sellers, A., & Kennedy, P.J., (1993). Biosphere-Atmosphere Transfer Scheme (BATS) version 1e as coupled to the NCAR community climate model. Version 1e as Coupled to the NCAR Community Climate Model. National Center for Atmospheric Research (NCAR) Technical Note United States.
- Duan, Q., Sorooshian, S., & Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, 28(4), 1015–1031. <https://doi.org/10.1029/91WR02985>
- Erdal, D., Neuweiler, I., & Wollschläger, U. (2014). Using a bias aware EnKF to account for unresolved structure in an unsaturated zone model. *Water Resources Research*, 50, 132–147. <https://doi.org/10.1002/2012WR013443>
- Erdal, D., Rahman, M. A., & Neuweiler, I. (2015). The importance of state transformations when using the ensemble Kalman filter for unsaturated flow modeling: Dealing with strong nonlinearities. *Advances in Water Resources*, 86, 354–365. <https://doi.org/10.1016/j.advwatres.2015.09.008>
- Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5), 10143–10162. <https://doi.org/10.1029/94JC00572>
- Evensen, G. (2004). Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynamics*, 54(6), 539–560. <https://doi.org/10.1007/s10236-004-0099-2>
- Faticchi, S., Katul, G. G., Ivanov, V. Y., Pappas, C., Paschalis, A., Consolo, A., et al. (2015). Abiotic and biotic controls of soil moisture spatiotemporal variability and the occurrence of hysteresis. *Water Resources Research*, 51, 3505–3524. <https://doi.org/10.1002/2014WR016102>
- Frey, S. K., Hwang, H. T., Park, Y. J., Hussain, S. I., Gottschall, N., Edwards, M., & Lapen, D. R. (2016). Dual permeability modeling of tile drain management influences on hydrologic and nutrient transport characteristics in macroporous soil. *Journal of Hydrology*, 535, 392–406. <https://doi.org/10.1016/j.jhydrol.2016.01.073>
- Frey, S. K., Rudolph, D. L., Lapen, D. R., & Ball Coelho, B. R. (2012). Viscosity dependent dual permeability modeling of liquid manure movement in layered, macroporous, tile drained soil. *Water Resources Research*, 48, W00L11. <https://doi.org/10.1029/2011WR010809>
- Gebler, S., Hendricks Franssen, H. J., Kollet, S. J., Qu, W., & Vereecken, H. (2017). High resolution modelling of soil moisture patterns with TerrSysMP: A comparison with sensor network data. *Journal of Hydrology*, 547, 309–331. <https://doi.org/10.1016/j.jhydrol.2017.01.048>
- Gebler, S., Hendricks Franssen, H. J., Pütz, T., Post, H., Schmidt, M., & Vereecken, H. (2015). Actual evapotranspiration and precipitation measured by lysimeters: A comparison with eddy covariance and tipping bucket. *Hydrology and Earth System Sciences*, 19(5), 2145–2161. <https://doi.org/10.5194/hess-19-2145-2015>
- Gómez-Hernández, J. J., & Journel, A. (1993). Joint sequential simulation of multi-Gaussian fields. In A. Soares (Ed.), *Geostatistics Tróia'92. Quantitative Geology and Geostatistics*, (Vol. 85, chap. 8, pp. 85–94). Netherlands: Springer.
- Graham, D. N., & Butts, M. B. (2005). Flexible, integrated watershed modelling with MIKE SHE. *Watershed models*, 245–272.
- Grayson, R. B., Western, A. W., Chiew, F. H. S., & Blöschl, G. (1997). Preferred states in spatial soil moisture patterns: Local and nonlocal controls. *Water Resources Research*, 33(12), 2897–2908. <https://doi.org/10.1029/97WR02174>
- Hale, V. C., McDonnell, J. J., Stewart, M. K., Solomon, D. K., Doolittle, J., Ice, G. G., & Pack, R. T. (2016). Effect of bedrock permeability on stream base flow mean transit time scaling relationships: 2. Process study of storage and release. *Water Resources Research*, 52, 1375–1397. <https://doi.org/10.1002/2014WR016124>
- Hendricks Franssen, H. J., & Kinzelbach, W. (2008). Real-time groundwater flow modeling with the ensemble Kalman filter: Joint estimation of states and parameters and the filter inbreeding problem. *Water Resources Research*, 44, W006505. <https://doi.org/10.1029/2007WR006505>
- Herbst, M., & Diekkrüger, B. (2003). Modelling the spatial variability of soil moisture in a micro-scale catchment and comparison with field data using geostatistics. *Physics and Chemistry of the Earth, Parts A/B/C*, 28(6–7), 239–245. [https://doi.org/10.1016/S1474-7065\(03\)00033-0](https://doi.org/10.1016/S1474-7065(03)00033-0)
- Herbst, M., Diekkrüger, B., & Vanderborght, J. (2006). Numerical experiments on the sensitivity of runoff generation to the spatial variation of soil hydraulic properties. *Journal of Hydrology*, 326(1–4), 43–58. <https://doi.org/10.1016/j.jhydrol.2005.10.036>

- Houtekamer, P. L., & Mitchell, H. L. (1998). Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126(3), 796–811. [https://doi.org/10.1175/1520-0493\(1998\)126<0796:DAUAEK>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2)
- Houtekamer, P. L., & Mitchell, H. L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129(1), 123–137.
- Houtekamer, P. L., & Zhang, F. (2016). Review of the ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 144(12), 4489–4532. <https://doi.org/10.1175/MWR-D-15-0440.1>
- Ivanov, V. Y., Fatchi, S., Jenerette, G. D., Espeleta, J. F., Troch, P. A., & Huxman, T. E. (2010). Hysteresis of soil moisture spatial heterogeneity and the “homogenizing” effect of vegetation. *Water Resources Research*, 46, WR008611. <https://doi.org/10.1029/2009WR008611>
- Jones, J. E., & Woodward, C. S. (2001). Newton–Krylov-multigrid solvers for large-scale, highly heterogeneous, variably saturated flow problems. *Advances in Water Resources*, 24(7), 763–774. [https://doi.org/10.1016/S0309-1708\(00\)00075-0](https://doi.org/10.1016/S0309-1708(00)00075-0)
- Keller, J., Hendricks Franssen, H. J., & Marquart, G. (2018). Comparing seven variants of the ensemble Kalman filter: How many synthetic experiments are needed? *Water Resources Research*, 54, 6299–6318. <https://doi.org/10.1029/2018WR023374>
- Kollet, S. J., & Maxwell, R. M. (2006). Integrated surface-groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model. *Advances in Water Resources*, 29(7), 945–958. <https://doi.org/10.1016/j.advwatres.2005.08.006>
- Kollet, S. J., & Maxwell, R. M. (2008). Capturing the influence of groundwater dynamics on land surface processes using an integrated, distributed watershed model. *Water Resources Research*, 44, W02402. <https://doi.org/10.1029/2007WR006004>
- Kollet, S. J., Maxwell, R. M., Woodward, C. S., Smith, S., Vanderborght, J., Vereecken, H., & Simmer, C. (2010). Proof of concept of regional scale hydrologic simulations at hydrologic resolution utilizing massively parallel computer resources. *Water Resources Research*, 46, W04201. <https://doi.org/10.1029/2009WR008730>
- Korres, W., Koyama, C. N., Fiener, P., & Schneider, K. (2010). Analysis of surface soil moisture patterns in agricultural landscapes using empirical orthogonal functions. *Hydrology and Earth System Sciences*, 14(5), 751–764. <https://doi.org/10.5194/hess-14-751-2010>
- Koyama, C. N., Korres, W., Fiener, P., & Schneider, K. (2010). Variability of surface soil moisture observed from multitemporal C-band synthetic aperture radar and field data. *Vadose Zone Journal*, 9(4), 1014–1024. <https://doi.org/10.2136/vzj2009.0165>
- Kumar, M., (2009): Toward a hydrologic modeling system. Ph.D. thesis, The Pennsylvania State University, 251 pp.
- Kurtz, W., He, G., Kollet, S. J., Maxwell, R. M., Vereecken, H., & Hendricks Franssen, H. J. (2016). TerrSysMP-PDAF (version 1.0): A modular high-performance data assimilation framework for an integrated land surface–subsurface model. *Geoscientific Model Development*, 9(4), 1341–1360. <https://doi.org/10.5194/gmd-9-1341-2016>
- Li, C., & Ren, L. (2011). Estimation of unsaturated soil hydraulic parameters using the ensemble Kalman filter. *Vadose Zone Journal*, 10(4), 1205–1227. <https://doi.org/10.2136/vzj2010.0159>
- Martini, E., Wollschläger, U., Kögler, S., Behrens, T., Dietrich, P., Reinstorf, F., et al. (2015). Spatial and temporal dynamics of hillslope-scale soil moisture patterns: Characteristic states and transition mechanisms. *Vadose Zone Journal*, 14(4). <https://doi.org/10.2136/vzj2014.10.0150>
- Maxwell, R. M. (2013). A terrain-following grid transform and preconditioner for parallel, large-scale, integrated hydrologic modeling. *Advances in Water Resources*, 53, 109–117. <https://doi.org/10.1016/j.advwatres.2012.10.001>
- Maxwell, R. M., Lundquist, J. K., Mirocha, J. D., Smith, S. G., Woodward, C. S., & Tompson, A. F. B. (2011). Development of a coupled groundwater–Atmosphere model. *Monthly Weather Review*, 139(1), 96–116. <https://doi.org/10.1175/2010MWR3392.1>
- Maxwell, R. M., & Miller, N. L. (2005). Development of a coupled land surface and groundwater model. *Journal of Hydrometeorology*, 6(3), 233–247. <https://doi.org/10.1175/JHM422.1>
- Montzka, C., Grant, J. P., Moradkhani, H., Franssen, H.-J. H., Weihermüller, L., Drusch, M., & Vereecken, H. (2013). Estimation of radiative transfer parameters from L-band passive microwave brightness temperatures using advanced data assimilation. *Vadose Zone Journal*, 12(3). <https://doi.org/10.2136/vzj2012.0040>
- Montzka, C., Moradkhani, H., Weihermüller, L., Franssen, H.-J., Canty, M., & Vereecken, H. (2011). Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter. *Journal of Hydrology*, 399(3–4), 410–421. <https://doi.org/10.1016/j.jhydrol.2011.01.020>
- Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), 885–900. <https://doi.org/10.13031/2013.23153>
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models Part I—A discussion of principles. *Journal of Hydrology*, 10(3), 282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Nerger, L., & Hiller, W. (2013). Software for ensemble-based data assimilation systems—Implementation strategies and scalability. *Computers & Geosciences*, 55, 110–118. <https://doi.org/10.1016/j.cageo.2012.03.026>
- Oleson, K., Bonan, G. B., Levis, S., Thornton, P., Vertenstein, M., Yang, Z., et al., (2004). Technical description of the Community Land Model (CLM). NCAR Technical Note NCAR/TN-461+STR.
- Oleson, K. W., Niu, G. Y., Yang, Z. L., Lawrence, D. M., Thornton, P. E., Lawrence, P. J., et al. (2008). Improvements to the Community Land Model and their impact on the hydrological cycle. *Journal of Geophysical Research*, 113, G000563. <https://doi.org/10.1029/2007JG000563>
- Pasetto, D., Niu, G.-Y., Pangle, L., Paniconi, C., Putti, M., & Troch, P. A. (2015). Impact of sensor failure on the observability of flow dynamics at the Biosphere 2 LEO hillslopes. *Advances in Water Resources*, 86, 327–339. <https://doi.org/10.1016/j.advwatres.2015.04.014>
- Pauwels, V. R. N., & De Lannoy, G. J. M. (2009). Ensemble-based assimilation of discharge into rainfall-runoff models: A comparison of approaches to mapping observational information to state space. *Water Resources Research*, 45, W08428. <https://doi.org/10.1029/2008WR007590>
- Poltoradnev, M., Ingwersen, J., & Streck, T. (2016). Spatial and temporal variability of soil water content in two regions of southwest Germany during a three-year observation period. *Vadose Zone Journal*, 15(6). <https://doi.org/10.2136/vzj2015.11.0143>
- Qu, W., Bogaen, H. R., Huisman, J. A., Martinez, G., Pachepsky, Y. A., & Vereecken, H. (2014). Effects of soil hydraulic properties on the spatial variability of soil water content: Evidence from sensor network data and inverse modeling. *Vadose Zone Journal*, 13(12). <https://doi.org/10.2136/vzj2014.07.0099>
- Qu, W., Bogaen, H. R., Huisman, J. A., Schmidt, M., Kunkel, R., Weuthen, A., et al. (2016). The integrated water balance and soil data set of the Rollesbroich hydrological observatory. *Earth System Science Data*, 8(2), 517–529. <https://doi.org/10.5194/essd-8-517-2016>
- Qu, W., Bogaen, H. R., Huisman, J. A., & Vereecken, H. (2013). Calibration of a novel low-cost soil water content sensor based on a ring oscillator. *Vadose Zone Journal*, 12(2). <https://doi.org/10.2136/vzj2012.0139>
- Qu, Y., & Duffy, C. J. (2007). A semidiscrete finite volume formulation for multiprocess watershed simulation. *Water Resources Research*, 43, W08419. <https://doi.org/10.1029/2006WR005752>

- Rasmussen, J., Madsen, H., Jensen, K. H., & Refsgaard, J. C. (2015). Data assimilation in integrated hydrological modeling using ensemble Kalman filtering: Evaluating the effect of ensemble size and localization on filter performance. *Hydrology and Earth System Sciences*, 19(7), 2999–3013. <https://doi.org/10.5194/hess-19-2999-2015>
- Reichle, R. H., McLaughlin, D. B., & Entekhabi, D. (2002). Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, 130(1), 103–114. [https://doi.org/10.1175/1520-0493\(2002\)130<0103:HDAWTE>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2)
- Richards, L. A. (1931). Capillary conduction of liquids through porous mediums. *Physics*, 1(1), 318–333.
- Robinson, D. A., Campbell, C. S., Hopmans, J. W., Hornbuckle, B. K., Jones, S. B., Knight, R., et al. (2008). Soil moisture measurement for ecological and hydrological watershed-scale observatories. *Vadose Zone Journal*, 7(1), 358–389. <https://doi.org/10.2136/vzj2007.0143>
- Schaap, M. G., Leij, F. J., & van Genuchten, M. T. (2001). rosetta: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology*, 251(3–4), 163–176.
- Schöniger, A., Nowak, W., & Hendricks Franssen, H. J. (2012). Parameter estimation by ensemble Kalman filters with transformed data: Approach and application to hydraulic tomography. *Water Resources Research*, 48, W04502. <https://doi.org/10.1029/2011WR010462>
- Sciuto, G., & Diekkrüger, B. (2010). Influence of soil heterogeneity and spatial discretization on catchment water balance modeling. *Vadose Zone Journal*, 9(4), 955–969. <https://doi.org/10.2136/vzj2009.0166>
- Shi, Y., Davis, K. J., Zhang, F., Duffy, C. J., & Yu, X. (2014). Parameter estimation of a physically based land surface hydrologic model using the ensemble Kalman filter: A synthetic experiment. *Water Resources Research*, 50, 706–724. <https://doi.org/10.1002/2013WR014070>
- Shi, Y., Davis, K. J., Zhang, F., Duffy, C. J., & Yu, X. (2015). Parameter estimation of a physically-based land surface hydrologic model using an ensemble Kalman filter: A multivariate real-data experiment. *Advances in Water Resources*, 83, 421–427. <https://doi.org/10.1016/j.advwatres.2015.06.009>
- Shrestha, P., Sulis, M., Masbou, M., Kollet, S., & Simmer, C. (2014). A scale-consistent Terrestrial Systems Modeling Platform based on COSMO, CLM, and ParFlow. *Monthly Weather Review*, 142(9), 3466–3483. <https://doi.org/10.1175/MWR-D-14-00029.1>
- Simmer, C., Thiele-Eich, I., Masbou, M., Amelung, W., Bogen, H., Crewell, S., et al. (2015). Monitoring and modeling the terrestrial system from pores to catchments: The transregional collaborative research center on patterns in the soil-vegetation-atmosphere system. *Bulletin of the American Meteorological Society*, 96(10), 1765–1787. <https://doi.org/10.1175/BAMS-D-13-00134.1>
- Šimůnek, J., van Genuchten, M. T., & Šejna, M. (2008). Development and applications of the HYDRUS and STANMOD software packages and related codes. *Vadose Zone Journal*, 7(2), 587–600. <https://doi.org/10.2136/vzj2007.0077>
- Tian, W., Li, X., Cheng, G. D., Wang, X. S., & Hu, B. X. (2012). Coupling a groundwater model with a land surface model to improve water and energy cycle simulation. *Hydrology and Earth System Sciences*, 16(12), 4707–4723. <https://doi.org/10.5194/hess-16-4707-2012>
- Tran, A. P., Vanclooster, M., Zupanski, M., & Lambot, S. (2014). Joint estimation of soil moisture profile and hydraulic parameters by ground-penetrating radar data assimilation with maximum likelihood ensemble filter. *Water Resources Research*, 50, 3131–3146. <https://doi.org/10.1002/2013WR014583>
- Valcke, S. (2013). The OASIS3 coupler: A European climate modelling community software. *Geoscientific Model Development*, 6(2), 373–388.
- van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal*, 44(5), 892–898. <https://doi.org/10.2136/sssaj1980.03615995004400050002x>
- Vereecken, H., Schnepf, A., Hopmans, J. W., Javaux, M., Or, D., Roose, T., et al. (2016). Modeling soil processes: Review, key challenges, and new perspectives. *Vadose Zone Journal*, 15(5). <https://doi.org/10.2136/vzj2015.09.0131>
- Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., & Verstraten, J. M. (2005). Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. *Water Resources Research*, 41, W01017. <https://doi.org/10.1029/2004WR003059>
- Whelan, B., McBratney, A., & Minasny, B. (2002). Vesper 1.5—spatial prediction software for precision agriculture, Precision Agriculture, Proc. 6th Int. Conf. on Precision Agriculture, ASA/CSSA/SSSA, Madison, WI, United States.
- Wiekenkamp, I., Huisman, J. A., Bogen, H. R., Lin, H. S., & Vereecken, H. (2016). Spatial and temporal occurrence of preferential flow in a forested headwater catchment. *Journal of Hydrology*, 534, 139–149. <https://doi.org/10.1016/j.jhydrol.2015.12.050>
- Weiler, M., & Naef, F. (2003). An experimental tracer study of the role of macropores in infiltration in grassland soils. *Hydrological Processes*, 17(2), 477–493.
- Wu, 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(3–4), 313–325. <https://doi.org/10.1016/j.jhydrol.2011.01.011>
- Wu, C., & Margulis, S. A. (2013). Real-time soil moisture and salinity profile estimation using assimilation of embedded sensor datastreams. *Vadose Zone Journal*, 12(1). <https://doi.org/10.2136/vzj2011.0176>
- Zacharias, S., Bogen, H., Samaniego, L., Mauder, M., Fuss, R., Puetz, T., et al. (2011). A network of terrestrial environmental observatories in Germany. *Vadose Zone Journal*, 10(3), 955–973. <http://doi.org/10.2136/vzj2010.0139>
- Zeng, X. B., Zhao, M., & Dickinson, R. E. (1998). Intercomparison of bulk aerodynamic algorithms for the computation of sea surface fluxes using TOGA COARE and TAO data. *Journal of Climate*, 11(10), 2628–2644. [https://doi.org/10.1175/1520-0442\(1998\)011<2628:IOBAAF>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<2628:IOBAAF>2.0.CO;2)
- Zeng, X., Barlage, M., Dickinson, R. E., Dai, Y., Wang, G., & Oleson, K. (2005). Treatment of Undercanopy Turbulence in Land Models. *Journal of Climate*, 18(23), 5086–5094.
- Zhang, D., Lu, Z., & Chen, Y. (2007). Dynamic reservoir data assimilation with an efficient, dimension-reduced Kalman filter. *Society of Petroleum Engineers*, 12(1), 108–117. <https://doi.org/10.2118/95277-PA>
- Zhang, D., Madsen, H., Ridler, M. E., Refsgaard, J. C., & Jensen, K. H. (2015). Impact of uncertainty description on assimilating hydraulic head in the MIKE SHE distributed hydrological model. *Advances in Water Resources*, 86, 400–413. <https://doi.org/10.1016/j.advwatres.2015.07.018>
- Zhang, H., Kurtz, W., Kollet, S., Vereecken, H., & Hendricks Franssen, H.-J. (2018). Comparison of different assimilation methodologies of groundwater levels to improve predictions of root zone soil moisture with an integrated terrestrial system model. *Advances in Water Resources*, 111, 224–238. <https://doi.org/10.1016/j.advwatres.2017.11.003>
- Zhou, H., Li, L., Franssen, H. J. H., & Gómez-Hernández, J. J. (2012). Pattern recognition in a bimodal aquifer using the normal-score ensemble Kalman filter. *Mathematical Geosciences*, 44(2), 169–185.