A Framework for Diagnosing Factors Degrading the Streamflow Performance of a Soil Moisture Data Assimilation System

YIXIN MAO

Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington

WADE T. CROW

Hydrology and Remote Sensing Laboratory, Agricultural Research Service, USDA, Beltsville, Maryland

BART NIJSSEN

Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington

(Manuscript received 8 June 2018, in final form 8 October 2018)

ABSTRACT

Data assimilation (DA) techniques have been widely applied to assimilate satellite-based soil moisture (SM) measurements into hydrologic models to improve streamflow simulations. However, past studies have reached mixed conclusions regarding the degree of runoff improvement achieved via SM state updating. In this study, a synthetic diagnostic framework is designed to 1) decompose the random error components in a hydrologic simulation, 2) quantify the error terms that originate from SM states, and 3) assess the effectiveness of SM DA to correct these random errors. The general framework is illustrated through a case study in which surface Soil Moisture Active Passive (SMAP) data are assimilated into a large-scale land surface model in the Arkansas–Red River basin. The case study includes systematic error in the simulated streamflow that imposes a first-order limit on DA performance. In addition, about 60% of the random runoff error originates directly from rainfall and cannot be corrected by SM DA. In particular, fast-response runoff dominates in much of the basin but is relatively unresponsive to state updating. Slow-response runoff is strongly controlled by the bottom-layer SM and therefore only modestly improved via the assimilation of surface measurements. Combined, the total runoff improvement in the synthetic analysis is small (<10% over the basin). Improvements in the real SMAP-assimilated case are further limited due to systematic error and other factors such as inaccurate error assumptions and SMAP rescaling. Findings from the diagnostic framework suggest that SM DA alone is insufficient to substantially improve streamflow estimates in large basins.

1. Introduction

Soil moisture (SM) is a key component of the hydrologic cycle. For hydrologic prediction purposes, SM plays an important role in runoff generation both by partitioning infiltration and surface runoff during rainfall events and by controlling the rate of subsurface flow during interstorm periods (e.g., Freeze and Harlan 1969; Western et al. 2002; Aubert et al. 2003). Therefore,

improved estimation of SM states in hydrologic models is potentially beneficial for streamflow prediction.

A number of studies have explored the potential to assimilate SM measurements into hydrologic models for runoff simulation improvement, either via synthetic twin experiments (e.g., Reichle et al. 2008; Crow and Ryu 2009; Chen et al. 2011; Brocca et al. 2012) or the assimilation of real SM measurements, especially remotely sensed products (e.g., Pauwels et al. 2002; Francois et al. 2003; Parajka et al. 2006; Brocca et al. 2010; Wanders et al. 2014; Alvarez-Garreton et al. 2014; Kumar et al. 2014; Lievens et al. 2015; Massari et al. 2015). However, these studies disagree on the benefits of assimilating SM data for streamflow improvement. Some studies have found improved streamflow simulation

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JHM-D-18-0115.s1>.

Corresponding author: Bart Nijssen, nijssen@uw.edu

with SM assimilation (e.g., Pauwels et al. 2002; Francois et al. 2003; Brocca et al. 2010, 2012; Wanders et al. 2014; Massari et al. 2015), while others have shown limited improvement or degraded streamflow after SM assimilation, or improved streamflow only under certain conditions. For example, Parajka et al. (2006) found that runoff was not improved, and sometimes even degraded, after satellite SM assimilation; Reichle et al. (2008) found (at best) marginally improved runoff in a synthetic study; Alvarez-Garreton et al. (2014) found varying levels of improvement when looking at specific flood events; Kumar et al. (2014) found marginal improvements via SM assimilation; Lievens et al. (2015) found improved peak flows but degraded flow at a daily scale; Chen et al. (2011) found moderate streamflow improvement in a synthetic experiment but generally no improvement in a real-data case; and Brocca et al. (2012) found improved streamflow when assimilating root-zone SM products, but limited success when only assimilating surface SM.

Multiple factors control the magnitude of runoff improvement, including hydrologic model structure, basin characteristics, data assimilation (DA) techniques, error assumptions, and the type of SM data product assimilated [see Massari et al. (2015) and Brocca et al. (2010) for overviews of these factors]. Some past studies have diagnosed performance limitations for their specific DA applications. For example, Chen et al. (2011) argued that the weak vertical coupling in the Soil and Water Assessment Tool (SWAT) limited the success of updating deeper-layer SM (and subsequently subsurface flow) in a synthetic twin experiment, and systematic model bias uncorrectable by DA further limited the streamflow improvement in a real-data case; Brocca et al. (2012) found much larger streamflow improvement via assimilating a root-zone SM product than via assimilating surface SM. Other studies such as Reichle et al. (2008), Alvarez-Garreton et al. (2014), and Massari et al. (2015) discussed the impact of error assumptions, rescaling techniques, study basins, seasonality, and hydrologic conditions on streamflow results in their specific cases. All together, these studies pointed out key features that could impact runoff improvement.

Despite this progress, the diagnostic methods used by these past studies are typically only relevant to their specific application, resulting in conclusions that are mostly qualitative. Motivated by an SM DA experiment which yielded no discernible improvement in runoff estimates (see below, especially Fig. 1), we aim to design a formal diagnostic framework that draws upon synthetic experiments to decompose error terms in streamflow simulation and to provide quantitative understanding of limiting factors in an SM DA system.

We formally identify a number of error source terms that occur in the runoff simulation (see Fig. 2):

- 1) error in meteorological forcing (primarily precipitation), which is partly reflected in SM states and subsequently in runoff, and partly contributes directly to runoff without going through antecedent states;
- 2) SM state error caused by an incorrect model representation of SM response to forcing or an incorrect representation of the temporal evolution of SM; and
- 3) additional runoff error caused by an incorrect model representation of runoff generation processes or streamflow routing.

Built upon this error decomposition, we design a synthetic DA diagnostic framework to address two challenges. First, we distinguish the systematic and random error components (each error term in Fig. 2 consists of both components), since Kalman-filter-based DA techniques can only correct random error. Second, we further investigate the random error and distinguish the random error associated with SM state variables from other sources of random error. We present a framework that serves as a general diagnostic approach for SM DA problems, which is illustrated by a case study in the Arkansas–Red River basin with a semidistributed hydrologic model.

Most studies in hydrologic DA focus solely on deterministic analysis results. However, a probabilistic analysis can also be applied to gain additional insight into DA performance. For example, De Lannoy et al. (2006) used several ensemble verification metrics commonly used in meteorology to assess the ensemble representation of SM. Several other DA studies used ensemble verification metrics to verify or calibrate error assumptions (e.g., Pauwels and De Lannoy 2009; Brocca et al. 2012; Alvarez-Garreton et al. 2014), but typically reported only deterministic results as the final DA performance metric. Throughout this study, we use ensemble metrics not only as a verification and model selection tool, but also to evaluate DA performance from a probabilistic perspective.

The remainder of this paper is organized as follows. Section 2 describes our synthetic diagnostic framework as well as the study domain, hydrologic model, and data used for the case study. Section 3 presents results from our real-data and synthetic experiments and illustrates the usage of the latter to help diagnose the former. Section 4 contains our conclusions.

2. Methods and data

Sections 2a–c describe the study domain, hydrologic modeling approach, and data used in the case

FIG. 1. Streamflow time series at two example USGS gauges. The black line represents USGS observed streamflow, the red line represents open-loop simulated streamflow, and the blue lines are DA-updated ensemble streamflow. PER, PSR, and NENSK metrics based on the entire simulation period (March 2015–December 2017) are labeled in each subplot (see [section 2f](#) and section S2 in the supplemental material for a full description of evaluation metrics).

study, respectively, for both the real-data and synthetic experiments. DA techniques and our framework design are then described in [sections 2d](#) and [2e](#), respectively. Finally, evaluation metrics are presented in [section 2f](#).

a. Case study domain

The Arkansas–Red River basin is located in the south-central United States and covers an area of approximately $605\,000\text{ km}^2$ ([Fig. 3](#)). The basin consists of two major subbasins, the Arkansas River and the Red River, and represents a major tributary of the Mississippi River. Within the basin, there is a clear east–west climatic gradient (wetter in the east and drier in the west). Most of the basin experiences little snow cover in winter except for mountainous areas on its far western edge. Vegetation cover varies from deciduous forest in the east to wooded grassland, shrubs, crops, and grassland in the west.

b. Case study hydrologic modeling and meteorological forcing data

1) HYDROLOGIC MODEL DESCRIPTION

Hydrologic model simulations are based on version 5 of the Variable Infiltration Capacity (VIC) model ([Liang et al. 1994](#); [Hamman et al. 2018](#)). The VIC model is a large-scale, semidistributed model that simulates various land surface processes including both water balance and energy balance components. The VIC model is typically implemented on a gridcell basis, with each grid cell further divided into tiles via statistical distributions to represent subgrid heterogeneity in elevation and vegetation cover. Particularly relevant to SM DA for runoff improvement, the soil column in each grid cell tile is discretized into L vertical layers, with water within each layer draining into the lower layer via a nonlinear gravity-driven drainage curve ([Brooks and Corey 1964](#)). The sum of water contained in the top $(L - 1)$ soil

FIG. 2. Illustration of error decomposition in a streamflow simulation system (see section 1 for a detailed explanation of each term). All error terms in the diagram consist of a systematic error component and a random error component. The red highlighted terms are the parts correctable by SM DA. The blue numbers/text describe random error decomposition results from the Arkansas–Red case study.

layers determines the partitioning of rainfall or snowmelt into infiltration and overland runoff (i.e., fast-response runoff generated over saturated soil) by the variable infiltration capacity curve (Zhao et al. 1980; Wood et al. 1992). Slow-response runoff generated as drainage from the deep soil is determined by the moisture level in the bottom layer via a nonlinear recession curve (Todini 1996; Nijssen et al. 2001).

In this study, the Arkansas–Red River basin was modeled at a spatial resolution of $1/8^\circ$ (3999 grid cells in total) with each grid cell further divided into multiple vegetation tiles. We did not consider heterogeneity in subgrid elevation. All model parameters regarding vegetation cover and soil properties were taken from Maurer et al. (2002) and calibrated for large-basin streamflow. The soil column in each grid cell was divided into three vertical layers, with domain-average thicknesses of 0.10, 0.40, and 0.93 m, respectively. The VIC model was run at a 3-hourly time step in water balance mode.

2) METEOROLOGICAL FORCING DATA

The real-data experiment was conducted for March 2015–December 2017, a period when NASA Soil Moisture Active Passive (SMAP; Entekhabi et al. 2010) SM retrieval data were available. The North American Land Data Assimilation System phase 2 (NLDAS-2) meteorological data (Xia et al. 2009) were used to force the simulations. This dataset was chosen because of its temporal coverage, temporal resolution, and data quality. Hourly NLDAS-2 meteorological variables required by VIC were aggregated to a 3-hourly time step. NLDAS-2 forcing data from 1979 to 2015 were used to spin up the VIC model.

Since the SMAP retrievals are only available after mid-2015, we selected a longer 10-yr period (1980–89) for the synthetic experiments in order to more robustly

sample performance evaluation metrics. We used daily precipitation and maximum and minimum temperatures from one of the realizations of a $1/8^\circ$ ensemble meteorology product (Newman et al. 2015). This forcing dataset was used because it considered observation uncertainty, making it suitable for DA applications. Daily wind speed for this period was taken from a $1/8^\circ$ gridded meteorological data product by Maurer et al. (2002). Other meteorological forcing data required by VIC, which includes shortwave and longwave radiation, air pressure, and humidity, were calculated using the MTCLIM algorithms and the Tennessee Valley Authority algorithm, and then disaggregated to a 3-hourly time step following Bohn et al. (2013). Prior to the synthetic simulation period, VIC was spun up from 1949 to 1979 using the Maurer et al. (2002) forcing data.

3) STREAMFLOW ROUTING

For the real-data experiments, we routed gridcell runoff through the stream channel using the RVIC

FIG. 3. The Arkansas–Red River basin. The pink shaded areas show the upstream subbasins of the eight USGS streamflow sites used in this study with basin numbers labeled on the plot (see Table 1 for the site name and USGS site number).

TABLE 1. List of subbasins evaluated in the study with their corresponding USGS gauged streamflow site information.

Basin No.	USGS site No.	USGS station name
1	07144200	Little Arkansas River at Valley Center, KS
2	07144780	Ninnescah River AB Cheney Re, KS
3	07147800	Walnut River at Winfield, KS
4	07152000	Chikaskia River near Blackwell, OK
5	07177500	Bird Creek Near Sperry, OK
6	07186000	Spring River near Wace, MO
7	07196500	Illinois River near Tahlequah, OK
8	07243500	Deep Fork near Beggs, OK

routing model (Hamman et al. 2017), which is a source-to-sink model based on the routing model typically used as a postprocessor for VIC model runoff simulations (Lohmann et al. 1996, 1998). In the synthetic experiments, locally generated runoff was evaluated without routing through stream channels.

c. SMAP retrievals and USGS streamflow data

The SMAP mission provides SM estimates for the top 5 cm of the soil column with a revisit time of 2–3 days, a resolution of 36 km, and a 50-h data latency. Here, SMAP L3 passive retrievals (O'Neill et al. 2016) were assimilated during the nonwinter seasons (March–October). A few SMAP pixels with near-constant retrieval values, indicating an obvious quality flaw, were manually masked out. Daily streamflow data acquired at eight USGS streamflow sites (USGS 2018) were used for evaluation in the real-data experiment (see Fig. 3 and Table 1 for site locations). These eight sites were selected due to their lack of human streamflow regulation and good rain gauge coverage [see Crow et al. (2017) for details].

d. Ensemble Kalman filter

The standard ensemble Kalman filter (EnKF) method is described in the supplementary material (section S1). In the following, we describe implementation details specific to our application.

1) STATE VECTOR, OBSERVATION OPERATOR, AND ENSEMBLE SIZE

Although SMAP only provides surface measurements, deeper layer SM updates are essential for successful runoff improvement because of their control on runoff generation processes in VIC (discussed further in section 3b). However, in the Lievens et al. (2015, 2016)

SM DA studies using VIC, the authors did not include the bottom-layer state in the Kalman filter to avoid unstable behavior. We will further examine this decision via the synthetic experiments in section 3c(1). However, unless otherwise noted, all DA results are based on the Lievens et al. (2015, 2016) approach, whereby only the first and second soil layers are updated. Furthermore, temperature states were not updated in this study.

Following Gruber et al. (2015), each pixel of observations was assimilated separately (i.e., 1D filtering) without consideration of spatially correlated errors. Gruber et al. (2015) found only limited performance enhancement when transitioning from a 1D to a 2D filter for the assimilation of remotely sensed SM. For the synthetic experiments, synthetic surface SM measurements were generated at the same spatial resolution as the VIC model setup and assimilated into each 1/8° VIC grid cell separately. In such a 1D filter, the measurement vector $\tilde{\mathbf{y}}$ [Eq. (S2) in the supplemental material] reduces to a scalar:

$$\tilde{y} = \text{SM}_1^{\text{obs}}, \quad (1)$$

where SM_1^{obs} is the synthetic observation of surface SM at 1/8° resolution. The state vector [$\mathbf{x}$ in Eqs. (S1) and (S2)] is simply a collection of SM states for the top two layers in all vegetation tiles within a grid cell. The observation operator [H in Eq. (S2)] calculates the tile-averaged first-layer SM state as a map from the state space to the observation space.

For the real-data experiment, SMAP retrievals were acquired at a coarser spatial resolution (36 km) than the VIC grid (1/8°), and the grid edges do not align. To resolve this spatial mismatch, every VIC grid cell is assigned to the nearest SMAP pixel. Then for each SMAP pixel, SMAP was assimilated to update all corresponding VIC grid cells via EnKF by augmenting the states from all those grid cells in the state vector. In this case, the measurement vector $\tilde{\mathbf{y}}$ is still a scalar as in Eq. (1), and the observation operator H calculates the areal-averaged first-layer SM state from the multiple VIC grid cells. Unless otherwise stated, a 32-member Monte Carlo ensemble was used to generate the EnKF.

Although potential bias in SM states as well as in runoff can, in theory, develop in the EnKF result due to the interaction of zero-mean noise with model nonlinearities, we did not implement bias correction schemes during EnKF (as in, e.g., Ryu et al. 2009) since including such a scheme did not improve filter performance in our case (results not shown). This decision does not significantly degrade our EnKF result because, as applied here, the random precipitation perturbations are relatively small and thus produce a correspondingly small EnKF propagation bias.

2) ERROR ASSUMPTIONS

Measurement errors were estimated to be in the range of $0.03\text{--}0.045\text{ m}^3\text{ m}^{-3}$ for the SMAP L3 retrievals, which is roughly consistent with those reported by SMAP validation studies (e.g., Colliander et al. 2017; Chan et al. 2017). For each SMAP pixel (or the synthetic $1/8^\circ$ measurements in the synthetic case), we linearly mapped its leaf area index (LAI) to the above error range, with larger error values corresponding to denser vegetation cover. A positive relationship between vegetation density and satellite-observed moisture error has been well documented (e.g., Crow et al. 2010). Generally, eastern portions of the Arkansas–Red River basin have relatively larger measurement errors than western areas due to their denser vegetation cover.

The EnKF method requires the generation of an ensemble to capture the error in model propagation. We perturbed precipitation forcing at a 3-hourly time step with multiplicative random error. The multiplier was drawn from a lognormal distribution with mean one and a dimensionless standard deviation of 0.3, and high temporal autocorrelation (with the underlying normal random variable generated by a first-order autoregressive process with a coefficient of 0.9, also dimensionless). The lognormal precipitation error form has been used frequently (e.g., Ciach and Krajewski 1999; Nijssen and Lettenmaier 2004; Chen et al. 2014). The standard deviation used here is smaller than used in some other studies [e.g., Chen et al. (2014) used a standard deviation of 1] to represent the good-quality, gauge-informed NLDAS-2 forcing dataset. Direct state perturbation was applied to all SM layers and kept constant over space without spatial or temporal autocorrelation (normally distributed with zero mean and a standard deviation of 0.5 mm for all three layers). This state perturbation was tuned in the real-data case to ensure that the overall normalized variance of the filter innovation is approximately one, which, assuming the correct specification of measurement error, is a necessary condition for proper error levels in a filter (see Mehra 1971; Crow and Bolten 2007). The state perturbation was assumed to be highly correlated between vegetation tiles and vertical layers (correlation coefficient = 0.9). These same model error statistics were also used for the synthetic experiments.

3) SMAP RESCALING

The SMAP SM measurements were rescaled prior to their assimilation in the real-data experiment. Specifically, we rescaled SMAP L3 to match the seasonal (31-day window) mean and long-term (the 3-yr simulation period) standard deviation of the VIC-simulated

surface-layer SM time series. Such moment-matching rescaling is one of the standard techniques in DA (e.g., Chen et al. 2011; Brocca et al. 2012; Alvarez-Garretton et al. 2013). The rescaling was applied to each SMAP pixel separately and to the ascending and descending retrievals separately. The same standard deviation ratio used for SM rescaling was also applied to rescale SMAP retrieval errors into the VIC model regime.

e. Experimental design

In this section we describe the real-data experiment followed by a set of synthetic experiments conducted to diagnose notable results from the real-data experiment. As motivated in the introduction, the synthetic experiments quantify the random error components in a simulation system only since these are the only components potentially correctable by the EnKF.

1) REAL-DATA EXPERIMENT

A so-called “open loop” run serves as the baseline for modeling performance without any DA. Here the open-loop run was defined as a single, unperturbed VIC simulation. Deterministic modeling improvement using DA will be evaluated by comparing to the open-loop run.

To evaluate DA performance probabilistically, we generated an ensemble of perturbed (but not updated) VIC simulations to serve as the baseline (hereafter called the “open-loop ensemble”). The error statistics of the perturbations added to the forcings and states in the open-loop ensemble were the same as those applied in the DA procedure (see section 2d). The size of the open-loop ensemble was also kept the same as the DA ensemble (i.e., 32 members unless otherwise noted).

The DA setup described in section 2d was applied to assimilate SMAP retrievals into VIC. The final DA setup for the real-data experiment was informed by findings from the synthetic twin experiment (see section 3c). The runoff values after DA were routed to the gauge locations and compared with streamflow observations. Streamflow improvement after DA was evaluated both deterministically and probabilistically.

2) SYNTHETIC EXPERIMENTS

(i) Open-loop, synthetic truth, and synthetic measurement generation

In the synthetic experiments, one single perturbed VIC realization was chosen as the synthetic “truth” (following Crow and Reichle 2008; Kumar et al. 2009). Synthetic measurements were then generated by degrading the true first-layer SM by adding random

observation errors (error statistics consistent with those described in [section 2d](#)). Synthetic measurements were generated at a 24-h temporal interval that roughly matches the average length of SMAP data gaps. The synthetic open-loop runs were produced as in the real-data case described above.

(ii) Synthetic perfect-state/forcing experiments

We conducted two sets of model experiments to attribute the total random runoff error to individual error sources.

- 1) In the “perfect state” analysis, we substituted the true antecedent states (including all-layer SM states, snow states, and all other initial conditions) into the VIC model at a 24-h interval and evaluated the error reduction of subsequent runoff simulations (relative to the open-loop runoff). Any such error reduction is attributable to the improved specification of antecedent states, where the uncorrected portion of runoff error can be attributed directly to rainfall error. The former is potentially correctable by SM updating (highlighted in red [Fig. 2](#)), while the latter is not (the gray parts of [Fig. 2](#)). To further decompose the error source from each individual SM layer, we conducted three additional perfect-single-state runs, where we substituted the true i th-layer SM (SM_i) (where $i = 1, 2$, and 3) states into the VIC model while keeping all the other antecedent states the same as in the open-loop baseline. We performed this experiment separately for each of the three soil layers.
- 2) In the “perfect forcing” analysis, we used the true precipitation forcing as input to the VIC model without inserting the true model states and evaluated the resulting error reduction in SM states. This error reduction in SM quantifies the part of the forcing error reflected in SM states and therefore potentially correctable by SM updating ([Fig. 2](#)). SM error reduction was also calculated for each layer separately for additional interpretation.

(iii) Synthetic identical twin experiment

In addition, a standard synthetic identical twin experiment was conducted to quantify the overall effectiveness of SM DA to correct random error components linked directly to antecedent SM states. Specifically, we assimilated synthetic surface SM retrievals into VIC and evaluated the subsequent improvement in SM states and runoff fluxes relative to the open-loop baseline. The effect of some EnKF methodological choices will be investigated in more detail in [section 3c\(1\)](#).

f. Evaluation metrics

We evaluated the improvement of SM states and local runoff fluxes in the synthetic experiment, and the improvement of routed streamflow in the real-data experiment. Runoff was aggregated to daily values after which we applied a logarithmic transformation:

$$\log\text{Runoff} = \log(\text{runoff} + 1). \quad (2)$$

The logarithmic transformation reduces the dominant influence of a few large flood events on the evaluation metric and the “+ 1” in the transformation has the effect that zero runoff values remain zero after transformation. SM state was not log transformed.

We used percent root-mean-square error reduction (PER) for deterministic evaluation. In addition, the Kling–Gupta efficiency (KGE; [Gupta et al. 2009](#)) was calculated for streamflow to quantify the baseline performance. KGE describes overall performance by integrating the evaluation of correlation, variance, and bias, and ranges from negative infinity to 1 with values closer to 1 indicating better performance.

In addition, two probabilistic metrics were used: percent continuous rank probability score reduction (PSR) and normalized ensemble skill (NENSK). PSR measures the reduction of continuous rank probability score (CRPS) which, in turn, measures the deviation of the cumulative distribution function (CDF) of an ensemble from that of a reference (i.e., observation in the real-data case or truth in the synthetic case; [Hersbach 2000](#)). NENSK measures the ensemble-mean error normalized by ensemble spread:

$$\text{NENSK} = \frac{\text{ENSK}}{\text{ENSP}}, \quad (3)$$

where the ensemble skill (ENSK) is the temporal mean of the ensemble-mean squared error, and the ensemble spread (ENSP) is the temporal mean of the ensemble variance ([De Lannoy et al. 2006](#); [Brocca et al. 2012](#); [Alvarez-Garreton et al. 2014](#)). In an ideal situation where an ensemble is a correct representation of analysis uncertainty, NENSK should be 1 ([Talagrand et al. 1997](#); [Wilks 2011](#)). $\text{NENSK} > 1$ indicates an underdispersed ensemble while $\text{NENSK} < 1$ indicates an overdispersed ensemble. Note that NENSK is calculated based on the updated ensemble alone and is not compared to the open-loop baseline. See the supplemental material (section S2) for mathematical details concerning the above metrics.

3. Arkansas–Red case study results and discussion

DA results for the real SMAP-assimilated case study will be presented in this section, followed by the

TABLE 2. Comparison of runoff metrics in the real-data and the synthetic experiments. NENSK is only shown for the real-data case. Baseflow index (calculated as the fraction of slow-response runoff in total runoff from the synthetic open-loop simulation) is also shown for each subbasin.

		Open-loop KGE		PER				PSR				NENSK
Basin No.	Baseflow index	Synthetic	Real data	Synthetic			Real data	Synthetic			Real data	Real data
				Total runoff	Fast response	Slow response		Total runoff	Fast response	Slow response		
1	0.28	0.72	0.44	50.7%	3.2%	71.7%	3.9%	8.4%	2.9%	24.6%	−0.4%	28.82
2	0.32	0.86	0.45	8.6%	3.7%	39.5%	−2.0%	8.3%	4.6%	18.8%	0.1%	46.52
3	0.45	0.70	0.30	15.9%	2.5%	20.1%	7.0%	13.0%	2.0%	20.2%	−0.6%	73.91
4	0.31	0.88	0.22	10.2%	4.0%	22.9%	2.7%	9.0%	3.2%	19.8%	−0.2%	114.23
5	0.55	0.83	0.68	11.5%	0.9%	17.9%	1.7%	12.5%	2.0%	20.0%	−0.2%	17.55
6	0.64	0.80	0.29	17.3%	1.3%	19.4%	5.8%	11.6%	1.7%	15.0%	2.2%	33.79
7	0.63	0.83	0.80	16.4%	0.1%	19.2%	3.0%	7.1%	1.0%	9.2%	0.1%	30.20
8	0.55	0.84	0.78	9.3%	0.0%	13.5%	0.9%	8.7%	1.8%	13.9%	−0.4%	27.60

diagnostic results from the synthetic framework. Note that we present the real-data results before the synthetic results (even if the former uses methodological choices informed by the latter) since we aim to use the synthetic framework as a tool to diagnose the streamflow improvement we see in the real-data case. Although the diagnostic conclusions we draw are case specific, we use this case study as a demonstration of the usage of our general diagnostic framework.

a. Open-loop streamflow simulation and SMAP assimilation performance

To begin, we present the NLDAS-2 forced, open-loop baseline streamflow simulation evaluated against USGS observations in the eight subbasins. This open-loop streamflow performance is compared with the corresponding open-loop runoff performance from the synthetic case. Because synthetic runoff is evaluated at each grid cell, subbasin median statistics are reported. Since the synthetic and the real-data experiments were conducted for different time periods, the comparison is primarily qualitative. The daily open-loop KGE results are shown in Table 2. For all sites, the real-data open-loop case performs comparably to or worse than the corresponding synthetic case. This discrepancy reflects the existence of systematic error in the real-data streamflow simulation, since the synthetic case is designed to only capture the random error components (see sections 1 and 2e). The systematic error is also visible from the open-loop time series at two example sites in Fig. 1 at Little Arkansas River at Valley Center, Kansas, and Deep Fork near Beggs, Oklahoma. For example, the red lines in both show simulated flow recessing too rapidly.

The real-data DA results assimilated by SMAP are presented in Table 2 (the “real data” columns). Note that this experiment was conducted using a DA setup

that was informed by our synthetic twin experiment, with 32-member EnKF and no SM3 Kalman update [see section 3c(1) below for a complete discussion of these methodological choices]. Overall the streamflow correction is small, with PER ranging from −2.0% to 7.0% and PSR close to zero across subbasins. The example time series in Fig. 1 illustrate that DA often nudges the simulated flow in the correct direction, but is unable to substantially correct the inaccurate flow behavior. Since the streamflow ensemble produced by DA only accounts for random error, the existence of the systematic error partly results in an ensemble with too little dispersion and consequently a large NENSK (>20 at all sites).

The systematic error we observe here is partly due to model parameterization error. Although the baseline VIC simulation could be improved by parameter calibration for each subbasin, we chose not to perform this additional calibration for the following reasons: 1) Maurer et al. (2002) calibrated the current VIC parameter set toward large-basin streamflow and this parameter set has been used widely in large-scale VIC applications, and 2) although it is possible to calibrate individual unregulated small basins based on streamflow observations, it is challenging to generalize such a calibration scheme to a large domain due to flow regulation, low-quality meteorological forcing data, and lack of streamflow observations. Therefore, we chose not to calibrate and to maintain the level of parameterization error we see here, so that the evaluation and investigation in this section is more relevant to large-scale DA applications where similar errors are likely to occur.

b. Decomposition of random runoff error via synthetic perfect-state/forcing analysis

As a first step to diagnose the above limited streamflow improvement via SM DA, in this section we

FIG. 4. Maps of PER of runoff for synthetic perfect-state analysis. (first row) The PER when we substitute true antecedent states of all VIC states. The PER when we only substitute the true (second row) SM1 state, (third row) SM2 state, and (fourth row) SM3 state only. (left) PER of total runoff fluxes, (center) fast-response runoff, and (right) slow-response runoff. The number on each panel is the domain-median PER value. Blue shading indicates improved runoff accuracy, and red shading indicates degraded runoff accuracy.

decompose the random error components in the runoff simulations, identify their sources, and discuss the theoretical limitation of the error reduction using a set of perfect-state/forcing synthetic experiments.

1) PERFECT-STATE ANALYSIS

Results of the perfect-state experiments are shown in Fig. 4. PER was calculated based on the runoff fluxes immediately after a state substitution and therefore reflects the runoff response to state substitution. Bluer shading in the figure indicates greater runoff improvement in response to the substitution of perfect antecedent state(s), thus indicating that more of the runoff error originates from the error in antecedent state(s).

The first row in Fig. 4 shows the runoff PER by substituting the true antecedent SM states for all layers

(and all other VIC model states). In addition to total runoff, fast-response and slow-response runoff are analyzed separately. Total runoff displays a domain-median PER of 38.4% relative to the open-loop baseline. The runoff error reduction primarily relies on slow-response runoff, whose error is almost eliminated (99.5% domain-median PER). Conversely, fast-response runoff is only modestly improved (13.4% domain-median PER) even if all antecedent VIC states are known perfectly.

This result suggests that only 38.4% of the total runoff error originates from antecedent states, while the remaining 61.6% is derived directly from precipitation forcing error (see “runoff error” box in the lower right side of Fig. 2; the case-study error decomposition results are marked in blue on the diagram). Therefore, this

error decomposition imposes a first-order upper limit for runoff improvement via SM updating. Further reduction of the runoff error that originates directly from the forcing requires additional techniques such as rainfall correction (e.g., Crow et al. 2011). This upper limit is especially restrictive for fast-response runoff since 86.6% of its error is attributed directly to within-storm rainfall forcing.

We further evaluate the drastically different error decomposition of fast- versus slow-response runoff by calculating the runoff PER from single-layer true state substitutions (rows 2–4 in Fig. 4). The (small) PER of fast-response runoff mainly comes from SM2. The lack of control of SM1 on fast-response runoff error reduction is due to the VIC model structure as well as its parameterization: VIC combines the moisture amount in the top two layers to determine fast-response runoff. Since the second layer is typically much thicker than the first layer, it is dominant in controlling the overland runoff/infiltration partitioning process. In contrast, slow-response runoff error originates almost entirely from the SM3 state alone, since VIC simulates slow-response runoff as a direct function of the SM3 state. The difference between fast- and slow-response runoff can also explain the spatial gradient in total runoff PER. The wetter and more vegetated areas in the eastern part of the basin are more dominated by subsurface flow, and thus runoff is more strongly linked to model states. In contrast, drier areas in the western part of the domain rely more on surface-excess, fast-response runoff, and thus its runoff error primarily originates directly from forcing.

2) PERFECT-FORCING ANALYSIS

Figure 5 shows PER in SM when VIC is forced by the synthetic true precipitation. Domain-median PER of total SM (sum of all three layers) is reduced by 39.7%, which represents the random SM error that originates from uncertain forcing. The remaining 60.3% of SM error results from errors in the model representation of SM-related processes itself (see the lower-left box in Fig. 2). In contrast to the forcing error directly contributing to runoff as quantified in section 3b(1), the SM error that arises from forcing error is potentially correctable by SM updating. We also note that surface-layer PER in response to perfect forcing substitution is smaller than PER in deeper layers (Fig. 5). This is likely due to the longer memory of deeper-layer SM, whose error, over time, reflects the accumulated error in precipitation forcing.

To summarize the perfect-state/forcing analysis, we find that, in our case study, approximately 60% of the random error in runoff directly originates from

FIG. 5. Maps of PER of SM for the synthetic perfect-forcing analysis. Each panel shows PER of the total SM (sum of all three layers), SM1, SM2, and SM3, respectively. The number on each panel is the domain-median PER value. Blue shading indicates improved SM accuracy and red shading indicates degraded SM accuracy.

precipitation forcing and is thus uncorrectable by SM updating. In particular, fast-response runoff error is largely attributable to direct forcing error and therefore cannot be substantially corrected by SM DA, while slow-response runoff is strongly state controlled and its error reduction relies largely on SM3 correction. The part of the precipitation error that is reflected in SM states is also quantified and is potentially correctable via SM DA. These case-specific conclusions serve to illustrate the usage of the general framework to decompose error sources and quantify theoretical limitations in the application of SM DA to enhance streamflow simulation.

FIG. 6. Example SM3 time series at one grid cell for the small-domain synthetic DA experiments (37.9375°N , 97.9375°W , which is located in the southwest part of the small domain). Black lines represent synthetic truth, red lines are the open-loop analysis, light blue lines are the updated ensemble, and dark blue lines are the updated ensemble-mean. The 10-yr PER, PSR, and NENSK statistics of the updated ensemble are labeled in each subplot. Shown are ensemble sizes N = (left) 16, (center) 32, and (right) 64; (top) the default DA setup and (bottom) the DA run with no Kalman SM3 update.

c. Effectiveness of random error reduction via identical twin experiments and diagnosis of real-data performance

The perfect-state/forcing analysis demonstrates that, in our case study, constraining the VIC SM3 state is critical for reducing random error in runoff via SM DA. In this section, we will first examine factors determining SM3 DA performance in a smaller subdomain via a synthetic identical twin experiment. Then, we will present the synthetic DA performance for the entire domain.

1) SMALL-DOMAIN INVESTIGATION ON SM3 DA PERFORMANCE

We arbitrarily chose a smaller domain to further investigate DA performance (a box around the Little Arkansas subbasin; see Fig. 3, subbasin 1). As mentioned in section 2d, we examine the effect of including SM3 in the EnKF state vector, which was excluded by Lievens et al. (2015, 2016). According to Kalman filter theory, we should be able to include any state variable in the state vector, even if it is only weakly correlated or uncorrelated with the available measurements. The EnKF algorithm should be able to correctly determine the level of correlation and use that information in the subsequent update. However, in practice SM3 ensembles tend to be underdispersed when we update SM3 as part of our EnKF (first row in Fig. 6, which shows runoff at a single grid cell with NENSK of the updated ensemble greater than 1). Possible reasons for this include 1) a finite ensemble size, which can result in an overestimate of the error correlation between the surface measurement and SM3, leading to an overconfident

update, and 2) nonlinearity of the model, which can result in a nonoptimal update after propagation. These problems are analogous to three-dimensional atmospheric DA where state vectors are commonly “localized” to exclude states geographically far from observations (e.g., Anderson 2007). In our context, excluding SM3 from EnKF essentially applies “vertical localization” since we assume the SMAP surface measurements contain little information about SM3.

To examine this issue, we conducted a set of synthetic DA analyses in the small subbasin altering 1) ensemble size and 2) the decision of whether or not to apply “vertical localization,” that is, excluding SM3 states from the Kalman state vector. For all these experiments, the assimilated synthetic measurements and the truth were kept unchanged as described in section 2e(2). In the synthetic truth, a weak vertical error correlation exists across layers (because of the way we generated the open-loop baseline and the synthetic truth), but the goal is to examine how much benefit there is in updating SM3.

Figures 7–9 show maps of the DA performance for three different DA setups for the small domain. The first row (default) includes the SM3 state update in EnKF, and the second row excludes the SM3 Kalman update. Figure 6 shows example time series at one grid cell for the same experiments. In the following we discuss the effect of each DA variation in detail:

1) The effect of ensemble size:

Comparing the columns in Figs. 6–9, a larger ensemble size clearly benefits the SM3 update both deterministically and probabilistically. Specifically, when using the default DA setup, substantial improvement in

FIG. 7. Maps of PER of SM3 for the small-domain synthetic DA experiments. The numbers on top of each subplot are the domain-median PER value. The subplot panel layout is the same as in Fig. 6. Blue shading refers to improved SM3 accuracy and red shading refers to degraded SM3 accuracy.

both PER and PSR is seen when the ensemble size increases from $N = 16$ to $N = 64$. NENSK also steadily drops closer to one with larger ensemble sizes (row 1 in Fig. 9). Interestingly, the benefit of a larger ensemble size is much smaller when we do not update SM3 in DA (row 2 in Figs. 6–9). We will further discuss the reason for this below.

2) The effect of vertical localization (i.e., excluding SM3 update in EnKF):

Excluding SM3 from the Kalman update generally results in better SM3 estimates (comparing row 2 with row 1 in Figs. 6–9). This confirms our hypothesis that SM3, which is only weakly correlated with the surface measurement, is prone to overconfident Kalman updates, because the finite ensemble tends to overestimate the vertical correlation. The benefit of vertical localization is less significant for larger ensemble sizes, which again confirms that it is mainly the finite ensemble size that causes the overconfident SM3 update. Note that even without direct Kalman updating, SM3 is still moderately improved by approximately 20% in terms of

PER and PSR due to its positive response to updating SM1 and SM2.

In terms of PER and PSR, vertical localization still enhances SM3 performance when N grows to 32 but shows no extra benefit beyond that. However, for NENSK, SM3 localization continues to benefit when we increase N from 32 to 64 (Fig. 9). In other words, the ensemble representation of SM3, quantified by NENSK, still suffers slightly from overconfident updates with $N = 64$. This conclusion is based on our ensemble evaluation and cannot be determined solely from a deterministic evaluation.

2) WHOLE-DOMAIN (ARKANSAS-RED) RESULTS FROM THE SYNTHETIC TWIN EXPERIMENT

Based on the previous small-domain studies, our final DA setup in the case study used 32 ensemble members and no SM3 updating in order to balance DA performance and computational cost. We expect that these choices will provide an optimal EnKF analysis. Figures 10 and 11 show PER and PSR results, respectively, of the whole-domain synthetic twin experiment.

FIG. 8. As in Fig. 7, but for PSR of SM3.

Both SM1 and SM2 states are effectively improved, with $>30\%$ domain-median PER and approximately 30% PSR. The fact that the update of SM2 is only slightly less effective than that of SM1 suggests that SM1 and SM2 are highly coupled in the VIC model, and that the information contained in the surface measurement can be used effectively to update SM2. SM3 states are moderately improved as well, with domain-median PER of 24.8% and PSR of 13.7%. PSR is in general smaller than PER, especially for SM3, possibly related to the challenge of a good ensemble representation in the bottom layer as discussed in section 3c(1). The deterministic SM3 improvement shows a clear spatial pattern, with greater improvement in some subregions in the east and much less improvement in the west (Fig. 10). Kumar et al. (2009) found that the effectiveness of deeper-layer SM updates via surface assimilation depends on the degree of vertical coupling between soil layers, which is, in turn, determined by the specific model used, model parameters, and climate. Therefore, the results in our case study suggest that the vertical coupling is weaker in the west of the basin and stronger in some subregions in the east.

Even in the synthetic twin experiment containing no systematic error and assuming perfect statistical knowledge of errors, fast-response runoff is essentially unimproved across the entire domain. This is unsurprising, given the finding from the perfect-state analysis in section 3b that most of the error in fast-response runoff does not originate from error in the antecedent states. Slow-response runoff, on the other hand, does show improvement, with a domain-median PER value of 20.0% and a value PSR of 13.0%. Combined, total runoff is on average slightly improved (domain-median PER of 7.8% and PSR of 7.1%) and shows the same spatial gradient noted before in the perfect-state experiments (i.e., slight improvement in the east and almost no improvement in the west).

3) DIAGNOSING THE REAL-DATA RESULTS USING THE SYNTHETIC EXPERIMENTS

We further compare the results from the synthetic SM DA with those from the SMAP-assimilated real-data experiment presented earlier in section 3a at the eight USGS sites (Table 1, the “synthetic, total runoff” and “real data” columns). In general, larger synthetic PER corresponds to larger real-data PER across subbasins

FIG. 9. As in Fig. 7, but for NENSK of SM3. Lighter color (either purple or green) indicates a NENSK closer to one and, therefore, better ensemble representation.

(this correspondence is less clear in PSR partly because real-data PSR is near-zero for most subbasins). However, the real-data runoff improvement is much smaller than the synthetic improvement at all subbasins because of significant systematic error existing in the real-data simulation case, which is not correctable by SM DA. SMAP data quality, SMAP rescaling, spatial grid-matching techniques, and incorrect error assumptions may further affect the real-data case. While these factors are interesting research topics themselves, they have been investigated by previous studies (e.g., Massari et al. 2015; Alvarez-Garreton et al. 2013) and are not the focus of this study.

Finally, we examine the synthetic fast- and slow-response runoff separately for each subbasin (Table 2). For the two example sites shown in Fig. 1, they have obviously different characteristics: in the Little Arkansas subbasin there is stronger vertical coupling and therefore better synthetic performance, while in the Deep River subbasin there is weaker vertical coupling and therefore poorer synthetic performance. The stronger vertical coupling at Little Arkansas is reflected in the real-data time series (Fig. 1) which shows a

baseflow recession more sensitive to the DA update than in the Deep River subbasin. This results in both a larger deterministic streamflow update and a larger ensemble spread that better covers the observation for the Little Arkansas. The flashy streamflow peaks are almost unaffected by DA at both sites because of the weak coupling between fast-response runoff and top-layer SM.

Assisted by the understanding gained from the synthetic framework, we are able to extend the small-subbasin results to parts of our study domain where unregulated streamflow is unavailable for model evaluation. We do not expect successful streamflow improvement from SMAP assimilation in the western part of the Arkansas–Red basin where fast-response runoff is dominant.

4. Conclusions

In this paper, we present a diagnostic framework that builds on a set of synthetic experiments to decompose the error terms in streamflow simulation and to provide quantified understanding of limiting factors in an SM DA system. Specifically, the synthetic framework

FIG. 10. Maps of PER of DA-updated SM states and daily runoff fluxes from the synthetic twin experiment. Domain-median PER values are labeled on each subplot.

captures the random error components in the simulation system, distinguishes the runoff error terms that originate from errors in SM states, and assesses the effectiveness of SM DA to correct them.

The usage of the general framework is illustrated through a VIC case study in the Arkansas–Red River basin. A significant level of systematic error was found in the real-data experiment, which imposes a first-order limit on the extent to which SMAP assimilation can improve streamflow. We further found through the synthetic framework that approximately 60% of the random runoff error in the basin comes directly from precipitation forcing rather than antecedent states and is

therefore not correctable by SM DA. In particular, fast-response runoff has weak coupling with antecedent states and is insensitive to state updating. Slow-response runoff, which is a smaller component in total runoff in much of the basin, relies strongly on the bottom-layer SM for its correction. This critical layer is excluded from the EnKF state vector to avoid suboptimal filter results, and its improvement relies on its response to SM1 and SM2 updating and is only moderately improved. Therefore, total runoff improvement is quite low—even in a synthetic twin experiment assuming purely random errors and a perfect statistical parameterization of these errors. Real SMAP assimilation

FIG. 11. As in Fig. 10, but for PSR.

exhibits even less successful streamflow improvement due to the additional presence of systematic error and other factors such as SMAP rescaling and inaccurate error assumptions.

The results from our case study highlight that state correction alone is not sufficient to improve streamflow substantially. Although this conclusion is drawn from the specific case study, it likely holds for a generic streamflow simulation system, especially in large-scale basins, where systematic error and rainfall error are inevitable. Therefore, in addition to more sophisticated SM DA techniques, substantial streamflow improvement would only be possible if future research focuses on reducing these error components that are not

correctable by SM DA. Systematic error reduction for large-scale basins is a challenging task but has its potential via, for example, innovative model parameterization techniques. For example, Mizukami et al. (2017) calibrated “transfer functions” instead of model parameters themselves, which can then be applied to distributed model parameterization. The increasing availability of large-scale, satellite-observed distributed hydrologic datasets such as SMAP and Global Precipitation Mission (GPM) data (Hou et al. 2014) is especially promising for informing model representation. Crow et al. (2018), for example, diagnosed systematic coupling errors between antecedent states and runoff generation processes in land surface models using the

SMAP data. As for rainfall error, a dual correction system has been developed in the past decade whereby SM measurements are used to simultaneously update antecedent SM states and correct rainfall estimates (e.g., Chen et al. 2014; Alvarez-Garretton et al. 2016). The findings from this study support the potential of such correction schemes.

Acknowledgments. This work was supported in part by NASA Terrestrial Hydrology Program Award NNX16AC50G to the University of Washington and NASA Terrestrial Hydrology Program Award 13-THP13-0022 to the United States Department of Agriculture, Agricultural Research Service. Yixin Mao also received a Pathfinder Fellowship by CUAHSI with support from the National Science Foundation (NSF) Cooperative Agreement EAR-1338606. The VIC model used in the study is available at <https://github.com/UW-Hydro/VIC>. Specifically, we used VIC version 5.0.1 (<https://doi.org/10.5281/zenodo.267178>) with a modification to the calculation of drainage between soil layers (https://github.com/UW-Hydro/VIC/releases/tag/Mao_et_al_stateDA_May2018). The DA code used in this study is available at https://github.com/UW-Hydro/dual_DA_SMAP.

REFERENCES

- Alvarez-Garretton, C., D. Ryu, A. W. Western, W. Crow, and D. Robertson, 2013: Impact of observation error structure on satellite soil moisture assimilation into a rainfall-runoff model. *20th Int. Congress on Modelling and Simulation*, Adelaide, Australia, Modelling and Simulation Society of Australia and New Zealand, 3071–3077, <https://www.mssanz.org.au/modsim2013/L19/alvarezgarretton.pdf>.
- , —, —, W. T. Crow, and D. E. Robertson, 2014: The impacts of assimilating satellite soil moisture into a rainfall-runoff model in a semi-arid catchment. *J. Hydrol.*, **519**, 2763–2774, <https://doi.org/10.1016/j.jhydrol.2014.07.041>.
- , —, —, —, C.-H. Su, and D. R. Robertson, 2016: Dual assimilation of satellite soil moisture to improve streamflow prediction in data-scarce catchments. *Water Resour. Res.*, **52**, 5357–5375, <https://doi.org/10.1002/2015WR018429>.
- Anderson, J. L., 2007: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D*, **230**, 99–111, <https://doi.org/10.1016/j.physd.2006.02.011>.
- Aubert, D., C. Loumagne, and L. Oudin, 2003: Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall-runoff model. *J. Hydrol.*, **280**, 145–161, [https://doi.org/10.1016/S0022-1694\(03\)00229-4](https://doi.org/10.1016/S0022-1694(03)00229-4).
- Bohn, T. J., B. Livneh, J. W. Oyler, S. W. Running, B. Nijssen, and D. P. Lettenmaier, 2013: Global evaluation of MTCLIM and related algorithms for forcing of ecological and hydrological models. *Agric. For. Meteorol.*, **176**, 38–49, <https://doi.org/10.1016/j.agrformet.2013.03.003>.
- Brocca, L., F. Melone, T. Moramarco, W. Wagner, V. Naeimi, Z. Bartalis, and S. Hasenauer, 2010: Improving runoff prediction through the assimilation of the ASCAT soil moisture product. *Hydrol. Earth Syst. Sci.*, **14**, 1881–1893, <https://doi.org/10.5194/hess-14-1881-2010>.
- , T. Moramarco, F. Melone, W. Wagner, S. Hasenauer, and S. Hahn, 2012: Assimilation of surface-and root-zone ASCAT soil moisture products into rainfall-runoff modeling. *IEEE Trans. Geosci. Remote Sens.*, **50**, 2542–2555, <https://doi.org/10.1109/TGRS.2011.2177468>.
- Brooks, R. H., and A. T. Corey, 1964: Hydraulic properties of porous media. Hydrology Paper 3, Colorado State University, 27 pp., https://mountainscholar.org/bitstream/handle/10217/61288/HydrologyPapers_n3.pdf?sequence=1&isAllowed=y.
- Chan, S., and Coauthors, 2017: Development and validation of the SMAP enhanced passive soil moisture product. *2017 IEEE Geoscience and Remote Sensing Symp.*, Ft. Worth, TX, IEEE, 2539–2542, <https://doi.org/10.1109/IGARSS.2017.8127512>.
- Chen, F., W. T. Crow, P. J. Starks, and D. N. Moriasi, 2011: Improving hydrologic predictions of a catchment model via assimilation of surface soil moisture. *Adv. Water Resour.*, **34**, 526–536, <https://doi.org/10.1016/j.advwatres.2011.01.011>.
- , —, and D. Ryu, 2014: Dual forcing and state correction via soil moisture assimilation for improved rainfall-runoff modeling. *J. Hydrometeorol.*, **15**, 1832–1848, <https://doi.org/10.1175/JHM-D-14-0002.1>.
- Ciach, G. J., and W. F. Krajewski, 1999: Radar-rain gauge comparisons under observational uncertainties. *J. Appl. Meteor.*, **38**, 1519–1525, [https://doi.org/10.1175/1520-0450\(1999\)038<1519:RRGCUO>2.0.CO;2](https://doi.org/10.1175/1520-0450(1999)038<1519:RRGCUO>2.0.CO;2).
- Colliander, A., and Coauthors, 2017: Validation of SMAP surface soil moisture products with core validation sites. *Remote Sens. Environ.*, **191**, 215–231, <https://doi.org/10.1016/j.rse.2017.01.021>.
- Crow, W. T., and J. D. Bolten, 2007: Estimating precipitation errors using spaceborne surface soil moisture retrievals. *Geophys. Res. Lett.*, **34**, L08403, <https://doi.org/10.1029/2007GL029450>.
- , and R. H. Reichle, 2008: Comparison of adaptive filtering techniques for land surface data assimilation. *Water Resour. Res.*, **44**, W08423, <https://doi.org/10.1029/2008WR006883>.
- , and D. Ryu, 2009: A new data assimilation approach for improving hydrologic prediction using remotely-sensed soil moisture retrievals. *Hydrol. Earth Syst. Sci.*, **13**, 1–16, <https://doi.org/10.5194/hess-13-1-2009>.
- , D. G. Miralles, and M. H. Cosh, 2010: A quasi-global evaluation system for satellite-based surface soil moisture retrievals. *IEEE Trans. Geosci. Remote Sens.*, **48**, 2516–2527, <https://doi.org/10.1109/TGRS.2010.2040481>.
- , M. J. van den Berg, G. J. Huffman, and T. Pellarin, 2011: Correcting rainfall using satellite-based surface soil moisture retrievals: The Soil Moisture Analysis Rainfall Tool (SMART). *Water Resour. Res.*, **47**, W08521, <https://doi.org/10.1029/2011WR010576>.
- , F. Chen, R. H. Reichle, and Q. Liu, 2017: L band microwave remote sensing and land data assimilation improve the representation of prestorm soil moisture conditions for hydrologic forecasting. *Geophys. Res. Lett.*, **44**, 5495–5503, <https://doi.org/10.1002/2017GL073642>.
- , —, R. J. Reichle, Y. Xia, and Q. Liu, 2018: Exploiting soil moisture, precipitation, and streamflow observations to evaluate soil moisture/runoff coupling in land surface models. *Geophys. Res. Lett.*, **45**, 4869–4878, <https://doi.org/10.1029/2018GL077193>.
- De Lannoy, G. J. M., P. R. Houser, V. R. N. Pauwels, and N. E. C. Verhoest, 2006: Assessment of model uncertainty for soil moisture through ensemble verification. *J. Geophys. Res.*, **111**, D10101, <https://doi.org/10.1029/2005JD006367>.

- Entekhabi, D., and Coauthors, 2010: The Soil Moisture Active and Passive (SMAP) Mission. *Proc. IEEE*, **98**, 704–716, <https://doi.org/10.1109/JPROC.2010.2043918>.
- Francois, C., A. Quesney, and C. Ottle, 2003: Sequential assimilation of ERS-1 SAR data into a coupled land surface-hydrological model using an extended Kalman filter. *J. Hydrometeorol.*, **4**, 473–487, [https://doi.org/10.1175/1525-7541\(2003\)4<473:SAOESD>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)4<473:SAOESD>2.0.CO;2).
- Freeze, R. A., and R. L. Harlan, 1969: Blueprint for a physically-based, digitally-simulated hydrologic response model. *J. Hydrol.*, **9**, 237–258, [https://doi.org/10.1016/0022-1694\(69\)90020-1](https://doi.org/10.1016/0022-1694(69)90020-1).
- Gruber, A., W. T. Crow, W. Dorigo, and W. Wagner, 2015: The potential of 2D Kalman filtering for soil moisture data assimilation. *Remote Sens. Environ.*, **171**, 137–148, <https://doi.org/10.1016/j.rse.2015.10.019>.
- Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez, 2009: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.*, **377**, 80–91, <https://doi.org/10.1016/j.jhydrol.2009.08.003>.
- Hamman, J., B. Nijssen, A. Roberts, A. Craig, W. Maslowski, and R. Osinski, 2017: The coastal streamflow flux in the Regional Arctic System Model. *J. Geophys. Res. Oceans*, **122**, 1683–1701, <https://doi.org/10.1002/2016JC012323>.
- , —, T. J. Bohn, D. R. Gergel, and Y. Mao, 2018: The Variable Infiltration Capacity model version 5 (VIC-5): Infrastructure improvements for new applications and reproducibility. *Geosci. Model Dev.*, **11**, 3481–3496, <https://doi.org/10.5194/gmd-11-3481-2018>.
- Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**, 559–570, [https://doi.org/10.1175/1520-0434\(2000\)015<0559:DOTCRP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2).
- Hou, A. Y., and Coauthors, 2014: The Global Precipitation Measurement mission. *Bull. Amer. Meteor. Soc.*, **95**, 701–722, <https://doi.org/10.1175/BAMS-D-13-00164.1>.
- Kumar, S. V., R. H. Reichle, R. D. Koster, W. T. Crow, and C. D. Peters-Lidard, 2009: Role of subsurface physics in the assimilation of surface soil moisture observations. *J. Hydrometeorol.*, **10**, 1534–1547, <https://doi.org/10.1175/2009JHM1134.1>.
- , and Coauthors, 2014: Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation. *J. Hydrometeorol.*, **15**, 2446–2469, <https://doi.org/10.1175/JHM-D-13-0132.1>.
- Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, 1994: A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res.*, **99**, 14 415–14 428, <https://doi.org/10.1029/94JD00483>.
- Lievens, H., and Coauthors, 2015: SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia. *Remote Sens. Environ.*, **168**, 146–162, <https://doi.org/10.1016/j.rse.2015.06.025>.
- , and Coauthors, 2016: Assimilation of SMOS soil moisture and brightness temperature products into a land surface model. *Remote Sens. Environ.*, **180**, 292–304, <https://doi.org/10.1016/j.rse.2015.10.033>.
- Lohmann, D., R. Nolte-Holube, and E. Raschke, 1996: A large-scale horizontal routing model to be coupled to land surface parameterization schemes. *Tellus*, **48A**, 708–721, <https://doi.org/10.3402/tellusa.v48i5.12200>.
- , E. Raschke, B. Nijssen, and D. P. Lettenmaier, 1998: Regional scale hydrology: I. Formulation of the VIC-2L model coupled to a routing model. *Hydrol. Sci. J.*, **43**, 131–141, <https://doi.org/10.1080/02626669809492107>.
- Massari, C., L. Brocca, A. Tarpanelli, and T. Moramarco, 2015: Data assimilation of satellite soil moisture into rainfall-runoff modelling: A complex recipe? *Remote Sens.*, **7**, 11 403–11 433, <https://doi.org/10.3390/rs70911403>.
- Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nijssen, 2002: A long-term hydrologically-based data set of land surface fluxes and states for the conterminous United States. *J. Climate*, **15**, 3237–3251, [https://doi.org/10.1175/1520-0442\(2002\)015<3237:ALTHBD>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2).
- Mehra, R. K., 1971: On-line identification of linear dynamic systems with applications to Kalman filtering. *IEEE Trans. Autom. Contr.*, **16**, 12–21, <https://doi.org/10.1109/TAC.1971.1099621>.
- Mizukami, N., M. P. Clark, A. J. Newman, A. W. Wood, E. D. Gutmann, B. Nijssen, O. Rakovec, and L. Samaniego, 2017: Towards seamless large-domain parameter estimation for hydrologic models. *Water Resour. Res.*, **53**, 8020–8040, <https://doi.org/10.1002/2017WR020401>.
- Newman, A. J., M. P. Clark, and J. Craig, 2015: Gridded ensemble precipitation and temperature estimates for the contiguous United States. *J. Hydrometeorol.*, **16**, 2481–2500, <https://doi.org/10.1175/JHM-D-15-0026.1>.
- Nijssen, B., and D. P. Lettenmaier, 2004: Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the Global Precipitation Measurement satellites. *J. Geophys. Res.*, **109**, D02103, <https://doi.org/10.1029/2003JD003497>.
- , G. M. O'Donnell, D. P. Lettenmaier, D. Lohmann, and E. F. Wood, 2001: Predicting the discharge of global rivers. *J. Climate*, **14**, 3307–3323, [https://doi.org/10.1175/1520-0442\(2001\)014<3307:PTDOGR>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<3307:PTDOGR>2.0.CO;2).
- O'Neill, P. E., S. Chan, E. G. Njoku, T. Jackson, and R. Bindlish, 2016: SMAP L3 radiometer global daily 36 km EASE-Grid soil moisture, version 4. National Snow and Ice Data Center Distributed Active Archive Center, accessed 18 January 2018, <https://doi.org/10.5067/OBBHQ5W22HME>.
- Parajka, J., V. Naemi, G. Blöschl, W. Wagner, R. Merz, and L. Scipal, 2006: Assimilating scatterometer soil moisture data into conceptual hydrologic models at the regional scale. *Hydrol. Earth Syst. Sci.*, **10**, 353–368, <https://doi.org/10.5194/hess-10-353-2006>.
- Pauwels, V. R. N., and G. J. M. De Lannoy, 2009: Ensemble-based assimilation of discharge into rainfall-runoff models: A comparison of approaches to mapping observational information to state space. *Water Resour. Res.*, **45**, W08428, <https://doi.org/10.1029/2008WR007590>.
- , R. Hoeben, N. Verhoest, F. De Troch, and P. Troch, 2002: Improvement of TOPLATS-based discharge predictions through assimilation of ERS-based remotely sensed soil moisture values. *Hydrol. Processes*, **16**, 995–1013, <https://doi.org/10.1002/hyp.315>.
- Reichle, R. H., W. T. Crow, and C. L. Keppenne, 2008: An adaptive ensemble Kalman filter for soil moisture data assimilation. *Water Resour. Res.*, **44**, W03423, <https://doi.org/10.1029/2007WR006357>.
- Ryu, D., W. T. Crow, X. Zhan, and T. J. Jackson, 2009: Correcting unintended perturbation biases in hydrologic data assimilation. *J. Hydrometeorol.*, **10**, 734–750, <https://doi.org/10.1175/2008JHM1038.1>.
- Talagrand, O., R. Vautard, and B. Strauss, 1997: Evaluation of probabilistic prediction systems. ECMWF Tech. Doc.,

- 25 pp., <https://www.ecmwf.int/en/elibrary/12555-evaluation-probabilistic-prediction-systems>.
- Todini, E., 1996: The ARNO rainfall-runoff model. *J. Hydrol.*, **175**, 339–382, [https://doi.org/10.1016/S0022-1694\(96\)80016-3](https://doi.org/10.1016/S0022-1694(96)80016-3).
- USGS, 2018: USGS surface-water daily data for the nation. U.S. Geological Survey, https://waterdata.usgs.gov/nwis/dv/?referred_module=sw.
- Wanders, N., D. Karssenbergh, A. De Roo, S. M. De Jong, and M. F. P. Bierkens, 2014: The suitability of remotely sensed soil moisture for improving operational flood forecasting. *Hydrol. Earth Syst. Sci.*, **18**, 2343–2357, <https://doi.org/10.5194/hess-18-2343-2014>.
- Western, A. W., R. B. Grayson, and G. Blöschl, 2002: Scaling of soil moisture: A hydrologic perspective. *Annu. Rev. Earth Planet. Sci.*, **30**, 149–180, <https://doi.org/10.1146/annurev.earth.30.091201.140434>.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 704 pp.
- Wood, E. F., D. P. Lettenmaier, and V. G. Zartarian, 1992: A land-surface hydrology parameterization with subgrid variability for general circulation models. *J. Geophys. Res.*, **97**, 2717–2728, <https://doi.org/10.1029/91JD01786>.
- Xia, Y., and Coauthors, 2009: NLDAS primary forcing data L4 hourly 0.125×0.125 degree V002. GES DISC, accessed 27 February 2018, <https://doi.org/10.5067/6J5LHHOHZHN4>.
- Zhao, R.-J., Y.-L. Zhang, L.-R. Fang, X.-R. Liu, and Q.-S. Zhang, 1980: The Xinanjiang model. *IAHS Publ.*, **129**, 351–356, http://hydrologie.org/redbooks/a129/iahs_129_0351.pdf.