Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-41
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 4 March 2019

Hydrology and Earth System Sciences
Open Access
Discussions

Dual state/rainfall correction via soil moisture assimilation for improved streamflow simulation: Evaluation of a large-scale implementation with SMAP satellite data

Yixin Mao[1], Wade T. Crow[2] and Bart Nijssen[1]

1: Department of Civil and Environmental Engineering, University of Washington, Seattle, WA

2: Hydrology and Remote Sensing Laboratory, Agricultural Research Service, USDA, Beltsville, MD

Corresponding author: Bart Nijssen (nijssen@uw.edu)

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-41
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 4 March 2019

Abstract

Soil moisture (SM) measurements contain information about both pre-storm hydrologic states and within-storm rainfall estimates, both are essential for accurate streamflow simulation. In this study, an existing dual state/rainfall correction system is extended and implemented in a large basin with a semi-distributed land surface model. The latest Soil Moisture Active Passive (SMAP) satellite surface SM retrievals are assimilated to simultaneously correct antecedent SM states in the model and rainfall estimates from the latest Global Precipitation Measurement (GPM) mission. While the GPM rainfall is corrected slightly to moderately, especially for larger events, the correction is smaller than that reported in past studies because of the improved baseline quality of the new GPM satellite product. The streamflow is corrected slightly to moderately via dual correction across 8 Arkansas-Red sub-basins. The correction is larger at sub-basins with poorer GPM rainfall and poorer open-loop streamflow simulations. Overall, although the dual data assimilation scheme is able to nudge streamflow simulations in the correct direction, it corrects only a relatively small portion of the total streamflow error. Systematic modeling error accounts for a larger portion of the overall streamflow error, which is uncorrectable by standard data assimilation techniques. These findings suggest that we may be reaching a point of diminishing returns for applying data assimilation approaches to correct random errors in streamflow simulations. More substantial streamflow correction would rely on future research efforts aimed at reducing the systematic error and developing higher-quality satellite rainfall products.

1. Introduction

Accurate streamflow simulation is important for water resources management applications such as flood control and drought monitoring. Reliable streamflow simulation requires accurate soil moisture (SM) conditions that control the partitioning of infiltration and surface runoff during rainfall events as well as longer-memory subsurface flow [Freeze and Harlan, 1969; Western et al., 2002; Aubert et al., 2003]. Good streamflow simulations also require realistic rainfall time series estimates.

SM measurements, if available, contain information about both antecedent hydrologic states and preceding rainfall events. With the advance of in-situ and satellite-measured SM products, researchers have started to explore the potential of using SM measurements to improve both aspects. For example, a number of studies have attempted to assimilate SM measurements to improve antecedent SM states in hydrologic models via Kalman-filter-based techniques [e.g., Francois et al., 2003; Brocca et al., 2010, 2012; Wanders et al., 2014; Alvarez-Garreton et al., 2014; Lievens et al., 2015, 2016; Massari et al., 2015; Mao et al., 2019]. Other studies have explored approaches to using SM measurements to back-calculate rainfall or to correct existing rainfall products [e.g., Crow et al., 2011; Chen et al., 2012; Brocca et al., 2013; Brocca et al., 2014; Brocca et al., 2016; Koster et al., 2016].

In the recent decade, so-called dual state/rainfall correction systems have been implemented that combine both the state update and rainfall correction schemes to optimally improve streamflow simulations [e.g., Crow and Ryu, 2009; Chen et al., 2014; Alvarez-Garreton et al., 2016]. Specifically, SM measurements (typically from satellite observation) are used to simultaneously update model states and correct a rainfall product (also typically satellite-observed). The updated antecedent states and corrected rainfall are then combined as inputs into a hydrologic model to produce an improved streamflow simulation (see Fig. 1 for illustration of the dual correction system). Past studies have suggested that such systems generally outperform either state-update-only or rainfall-correction-only schemes [Crow and Ryu, 2009; Chen et al., 2014; Alvarez-Garreton et al., 2016], with the rainfall correction contributing more during high-flow events and the state update during low flow periods [also see Massari et al., 2018].

While these past studies had encouraging findings, they applied the dual correction system only to catchment-scale, lumped hydrologic models. In this study, a semi-distributed land surface model, the Variable Infiltration Capacity (VIC) model, is implemented instead. The VIC model, compared to the previous lumped models, includes a more detailed representation of both energy and water balance processes [Liang et al., 1994; Hamman et al., 2018]. The macroscale grid-based VIC also better matches the spatial resolution of satellite SM measurements and provides a means for correcting large-scale streamflow analysis. In addition, earlier dual correction studies used previous-generation satellite products such as the Advanced Scatterometer (ASCAT) satellite SM data, the Soil Moisture Ocean Salinity (SMOS) satellite SM data and the Tropical Rainfall Measuring Mission (TRMM) precipitation data. Here, we use data products from the more recent Global Precipitation Measurement (GPM) mission [Hou et al., 2014] and the NASA Soil Moisture Active Passive (SMAP) mission [Entekhabi et al., 2010]. Both the SMAP and GPM products provide near-real-time measurements over much of the global land surface, making them especially useful for regions with scarce in-situ rainfall and SM observations.

The main objective of this study is to assess the effectiveness of such a dual correction system to improve streamflow simulations using the latest satellite SM and precipitation products. To address this main objective, we introduced a number of methodological advances. Specifically, we 1) extended the system to provide a probabilistic streamflow estimate via ensemble simulations (past studies focused solely on deterministic improvement), 2) updated the rainfall correction scheme to take advantage of the higher accuracy and higher temporal resolution of the satellite data, and 3) investigated the potential cross-correlation of errors in the dual system and validated the theoretical correctness of the system design. These methodological contributions will be presented throughout the paper.

The remainder of this paper is organized as follows. Section 2 describes the dual correction system and our novel methodological contributions, as well as the study domain, hydrologic model, and datasets used. Results are presented in Sect. 3. Section 4 discusses a few remaining issues and takeaways from the study, and Sect. 5 summarizes our conclusions.

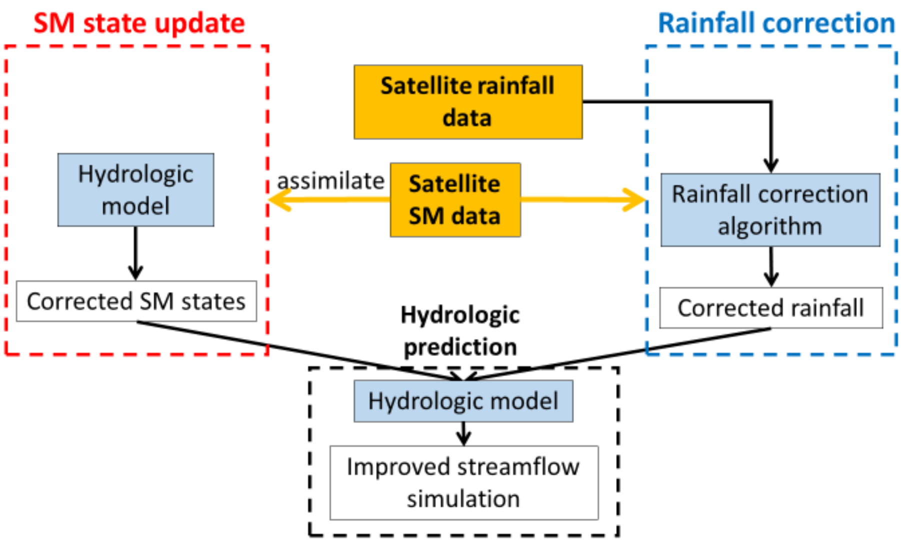

Figure 1. The dual state/rainfall correction framework applied in this study. Satellite-based soil moisture (SM) data is integrated into a hydrological simulation system via two correction schemes: 1) a standard data assimilation system to correct modeled SM states (shown in the red box on the left), and 2) a rainfall correction algorithm to correct rainfall forcing data (shown in the blue box on the right). Finally, these two contributions are combined to improve streamflow simulations (shown in the black box at the bottom).

2. Methods

2.1. Study domain

The dual state/rainfall correction system is applied in the Arkansas-Red River basin (approximately 605,000 km^2) located in the south-central United States (Fig. 2). This basin consists of the Arkansas River and the Red River, both converging eastward into the Mississippi River. This domain has a strong climatic gradient and is wetter in the east and drier in the west (Fig. 2). The basin experiences little snow cover in winter except for the mountainous areas along its far western edge. Vegetation cover tends to be denser in the east (deciduous forest) than in the west (wooded grassland, shrubs, crops and grassland).

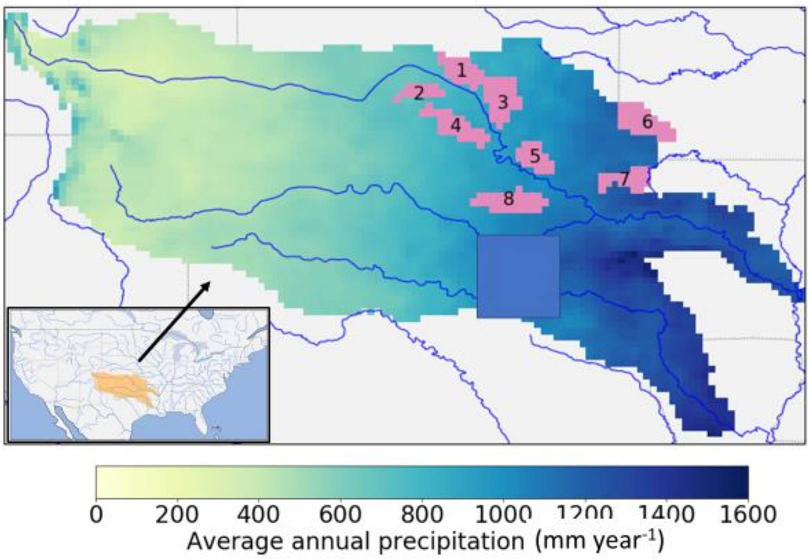

Figure 2. The Arkansas-Red River basin with climatology-averaged annual precipitation (calculated from NLDAS-2 precipitation data over 1979-2017). The pink shaded areas show the upstream sub-basins of the 8 USGS streamflow sites evaluated in this study, with basin numbers labeled on the plot (see Table 1 for basin numbers and corresponding sites).

2.2. Data

2.2.1. SMAP satellite SM data

The SMAP mission provides SM estimates for the top 5 centimeters of the soil column, with an average revisit time of 2-3 days, a resolution of 36 km and a 50-hour data latency. Both ascending (PM) and descending (AM) retrievals from the SMAP L3 Passive product [O'Neill et al., 2016] (data Version 4) from Mar 31, 2015 to December 31, 2017 were used in this study. A few SMAP pixels with obvious quality flaws (i.e., near-constant retrieval values) were manually masked out. The internal quality flags provided by the SMAP mission were not applied in this study to preserve the measurements in the east half of the domain, where the data quality of the entire region is flagged as unrecommended due to relatively heavy vegetation cover. The native

36-km SMAP retrievals were used throughout the study without spatial remapping or temporal aggregation.

2.2.2 GPM satellite precipitation data

The Integrated Multi-satellitE Retrievals for GPM (IMERG) Level 3 Version 05 Early Run precipitation data was used in this study [Huffman et al., 2018]. IMERG merges multiple satellite observations and provides a near-global precipitation product with a spatial resolution of 0.1º [Huffman et al., 2015]. The "Early Run" version of this product was used in this study since its short latency (4 hours) makes it suitable for near-real-time assimilation applications. We aggregated the original 30-minute precipitation product to our 3-hourly modeling timestep and remapped it onto our 1/8º model resolution.

2.2.3. Other meteorological forcing data

Other than precipitation, the VIC model requires air temperature, shortwave and longwave radiation, air pressure, vapor pressure and wind speed as forcing inputs. These variables were obtained from the 1/8º gridded North American Land Data Assimilation System Phase 2 (NLDAS-2) meteorological forcing data product [Xia et al., 2009]. We aggregated the original hourly NLDAS-2 meteorological variables to the 3-hourly modeling timestep.

2.2.4. Validation data

Daily streamflow data at 8 USGS streamflow sites in the study domain [USGS, 2018] was used to evaluate the streamflow time series from the dual correction system (Fig. 2 and Table 1). These 8 sites were selected for their lack of human regulation and their dense rain gauge coverage (see Crow et al. [2017] for details). We separately evaluated the rainfall correction scheme, in which the gauge-informed NLDAS-2 precipitation data was treated as the benchmark.

2.3. Hydrologic modeling

We used Version 5 of the VIC model [Liang et al., 1994; Hamman et al., 2018]. VIC is a large-scale, semi-distributed model that simulates various land surface processes. In this study, the VIC model was implemented in the Arkansas-Red River basin with the same setup as in Mao et al. [2019]. Specifically, the model was set up at 1/8° spatial resolution with each grid cell further divided into multiple vegetation tiles via statistical distributions. Each grid cell was simulated by VIC separately using a soil column discretized into 3 vertical layers (with domain-average thicknesses of 0.10 m, 0.40 m and 0.93 m, respectively). Runoff can be generated by fast-response surface runoff and by slow-response runoff from the bottom soil layer. All vegetation cover and soil property parameters in the model were taken from Maurer et al. [2002], which were calibrated against streamflow observations at the most downstream outlet of the combined Arkansas and Red River basins. The simulation period was from March 2015 to December 2017 when both the SMAP and GPM products are available. The VIC model was spun-up by running the period 1979-2015 twice.

The local runoff simulated by VIC at each grid cell was routed through the stream channels using the RVIC routing model [Hamman et al., 2017]. RVIC is an adapted version of the routing model developed by Lohmann et al. [1996, 1998].

2.4. The dual correction system

In this section, we describe our methodological updates to the rainfall correction scheme, followed by a description of the state update scheme. Next, we describe how the two schemes are combined to produce the final ensemble streamflow analysis.

2.4.1. The SMART rainfall correction scheme updates and adaption

The Soil Moisture Analysis Rainfall Tool (SMART) rainfall correction algorithm [Crow et al., 2009; 2011; Chen et al., 2012] is based on sequential assimilation of SM measurements into a simple Antecedent Precipitation Index (API) model:

$$API_t = \gamma API_{t-1} + P_t \qquad (1)$$

where t is a timestep index; P is the original IMERG precipitation observation; and γ is a loss coefficient. We implemented a 3-hourly version of SMART (instead of the daily version in past studies) to receive the 3-hourly IMERG rainfall input and both the ascending (PM) and descending (AM) SMAP retrievals at the correct time of day. We also extended the ensemble Kalman filter (EnKF) version of SMART introduced by Crow et al. [2011] to an ensemble

Kalman smoother (EnKS), in which the API state is not only updated at timesteps when SMAP is available, but also updated during measurement gaps (see Supplemental Material Sect. S1 for mathematical details of the SMART EnKS). We set γ to 0.98 [3 hours^{-1}] such that the un-corrected API time series approximately captures the dynamics of SMAP retrievals (i.e., with high correlation). SMAP was rescaled to the API regime through cumulative distribution function (CDF) matching over the 2.5-year simulation period prior to assimilation.

The SMART algorithm then uses the API increment, δ_t, to estimate the rainfall correction amount via a simple linear relation. We implemented an ensemble rainfall correction rather than the single deterministic rainfall correction used in past SMART applications:

$$P^{(j)}_{corr,t} = P^{(j)}_{pert,t} + \lambda \delta^{(j)}_t \tag{2}$$

where the superscript *(j)* denotes the *j*th ensemble member (ensemble size M = 32); $P_{corr,t}$ is the corrected precipitation for time t; $P_{pert,t}$ is the perturbed IMERG precipitation; λ is a scaling factor that linearly relates API increment to rainfall correction, which was set to a domain-constant of 0.1 [-] (see Supplemental Material Sect. S2 for discussion on the choice of λ). We applied rainfall correction only at timesteps when the original IMERG rainfall observation is non-zero, taking advantage of the enhanced rain/no rain detection accuracy of IMERG [Gebregiorgis et al., 2018]. This tactic mitigates the degradation of the rainfall estimates during low-rainfall timesteps introduced by SMART (see also Sect. 3.1). Finally, following Crow et al. [2009; 2011], negative $P_{corr,t}$ values were set to zero, and the final corrected precipitation time series was multiplicatively rescaled to be unbiased over the entire simulation period against the original IMERG estimates.

In this study, the SMART algorithm was run at each of the 36-km SMAP pixels individually. The original 0.1^{o} IMERG product was remapped to the coarser 36-km resolution prior to SMART, and the corrected 36-km rainfall was then downscaled to the VIC 1/8^{o} modeling resolution. In our implementation of an EnKS-based SMART system, the original IMERG precipitation was multiplicatively perturbed by log-normally distributed noise with mean and standard deviation equal to one. SMAP measurement error ranges from 0.03 to 0.045 m^3/m^3 across domain, which was estimated from the SMAP ground validation studies [e.g., Colliander et al., 2017; Chan et al., 2017] and its spatial distribution was set to be proportional to leaf area index (LAI) (denser vegetation cover corresponds to larger SMAP error). The API state was directly perturbed by zero-mean Gaussian noise to represent API model error. The perturbation variance was set to 0.3 mm^2 over the entire domain such that the normalized filter innovation has variance of approximately one (which is a necessary condition for proper error assumptions in a Kalman filter; see Mehra [1971] and Crow and Bolten [2007]). See Supplemental Material Sect. S1 for mathematical details of these error assumptions.

2.4.2. State updating via EnKF

As illustrated in Fig. 1 (the red box on the left), the SMAP SM retrievals were also assimilated into the VIC model to update model states using the EnKF method. The EnKF implementation in this study generally follows Mao et al. [2019]. Specifically, a 1D filter was implemented for each 36-km SMAP pixel separately and at each pixel SMAP was assimilated to update the SM states of multiple underlying finer 1/8° VIC grid cells. Only the upper two layers of SM states in VIC were updated during EnKF (following Lievens et al. [2015; 2016] and Mao et al. [2019]), although the bottom layer SM does respond to the update of the upper two layers through drainage. An ensemble of 32 model run replicates was used to represent the probabilistic estimate of corrected SM states.

The SMAP retrievals were rescaled to match the 2.5-year mean and standard deviation of the VIC-simulated surface-layer SM time series prior to assimilation. The error statistics of IMERG precipitation and unscaled SMAP retrievals were assumed to be the same as used in SMART (Sect. 2.4.1). The VIC SM states of all three layers were directly perturbed during EnKF by zero-mean Gaussian noise with standard deviation of 0.5 mm over the entire study domain (following Mao et al. [2019]), which represents VIC modeling errors. Although VIC modeling errors are likely to contain spatial auto-correlation, consideration of this did not result in significantly better filter performance in our case and therefore not implemented here. This finding is consistent with Gruber et al. [2015] which described the limited benefit of a 2-D filter when assimilating distributed SM retrievals into a land surface model. We will further discuss this in Sect. 4.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-41
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 4 March 2019

2.4.3. Combining the state update and the rainfall correction schemes

The ensemble of updated model states and the corrected rainfall forcing were combined to produce final streamflow results (black box in the bottom of Fig. 1). We first randomly paired ensemble members of corrected rainfall and updated VIC states and selected 32 such pairs to balance competing considerations of computational cost and statistical stability. For each pair, the VIC model was re-run with the updated states inserted sequentially over time and forced by the corrected rainfall. Other meteorological forcings were kept unchanged. The runoff output from VIC for each pair was then routed to the gauge locations, resulting in an ensemble of basin-outlet streamflow time series for evaluation. To further separate the relative contribution of the state update and the rainfall correction schemes to overall streamflow improvement, two additional streamflow simulations were performed. The first was the "state-updated streamflow" case, where VIC was re-run with the updated states and forced by the original IMERG precipitation. The resulting streamflow reflects only the impact of state updating on streamflow simulations. The second was the "rainfall-corrected streamflow" case, where VIC was forced by the SMART-corrected rainfall ensemble but without inserting the updated states. The resulting streamflow reflects only the effect of SMART rainfall correction.

Although the state and rainfall correction schemes were performed separately with no feedback to each other to mitigate correlated error [Crow et al., 2009], error correlation still potentially exists in the dual system since the two schemes are informed by the same SM measurement data. Such cross-correlated error could potentially be amplified when combining the two schemes and degrading streamflow estimates. Massari et al. [2018] intentionally avoided combining the state and rainfall correction schemes due to this concern. To investigate this, we performed a set of synthetic experiments where we compared the following two scenarios: 1) a single set of synthetically generated SM measurements were assimilated into the state and rainfall correction schemes, mimicking the real dual correction system; 2) two SM measurements with mutually independent errors were assimilated separately into the two schemes, thereby avoiding error cross-correlation in the system. Results show that the two scenarios achieve very similar streamflow correction performance. This suggests that it is safe to assimilate a single SM measurement product into both schemes without significantly degrading the final streamflow performance (see Sect. S3 in Supplemental Material).

2.5. Evaluation strategies and metrics

We evaluated the rainfall correction results in addition to the dual-corrected streamflow results in terms of both deterministic and probabilistic metrics.

The 1/8^{o} gauge-informed NLDAS-2 precipitation data was remapped to the 36-km SMART resolution grid as the benchmark for evaluating rainfall. Deterministically, the ensemble-mean SMART-corrected rainfall was compared to the original IMERG precipitation (remapped to 36 km), and its improvement was evaluated in terms of: 1) correlation coefficient (*r*) of time series; 2) percent error reduction (PER) in terms of the root-mean-squared error (RMSE); 3) Categorical skill metrics, including false alarm ratio (FAR), probability of detection (POD) and threat score (TS) [Wilks, 2011; Crow et al., 2011; Chen et al., 2012; Brocca et al., 2016]. Probabilistically, the normalized ensemble skill (NENSK) was calculated, which measures the ensemble-mean error normalized by ensemble spread:

$$NENSK = \frac{ENSK}{ENSP} \tag{3}$$

where the ensemble skill (ENSK) is the temporal mean of ensemble-mean squared error, and the ensemble spread (ENSP) is the temporal mean of ensemble variance [De Lannoy et al., 2006; Brocca et al., 2012; Alvarez-Garreton et al., 2014; Mao et al., 2019]. Ideally, if an ensemble time series correctly represent the uncertainty of analysis, NENSK should be 1 [Talagrand et al., 1997; Wilks, 2011]. NENSK > 1 indicates an under-dispersed ensemble while NENSK < 1 indicates an over-dispersed ensemble. For all metrics, precipitation datasets were aggregated to multiple temporal accumulation periods (the native 3-hour period without aggregation; 1-day; 3-day) for evaluation.

The dual-corrected streamflow was evaluated at the 8 USGS sites shown in Fig. 2. Deterministically, the ensemble-median corrected streamflow was compared to the baseline streamflow, or the so-called “open-loop” streamflow, which is simply the single VIC simulation forced by IMERG precipitation without any correction, in terms of 1) PER; and 2) the Kling-Gupta efficiency (KGE) [Gupta et al. 2009] which combines the performance of correlation, variance and bias. Ensemble-median instead of ensemble-mean streamflow was used for more stable evaluation results in the case of a skewed streamflow ensemble caused by model nonlinearity. Probabilistically, NENSK was calculated for streamflow ensembles.

3. Results

3.1. SMART rainfall correction

3.1.1. The impact of SMART methodological choices

Figure 3 shows the rainfall improvement in terms of r based on EnKS (the left column) compared to EnKF (the right column). For EnKF, both δ and P in Eq. (2) were aggregated to 3-day windows prior to correction to ensure SM data availability in every correction window. EnKF results in less r improvement than EnKS overall, which confirms the benefit of applying SMART using a smoothing approach.

The impact of our choice of only correcting rainfall at non-zero IMERG timesteps is demonstrated by the domain-median categorical metrics (Fig. 4). If every timestep is corrected (Fig. 4 Column 1), FAR is largely degraded (by 0.1 – 0.4) at low rainfall thresholds especially with shorter accumulation periods (3-hour and 1-day; see Fig. 4a). This is likely due to the issue of SMART misinterpreting SM retrieval noise as small rainfall corrections [Chen et al., 2014]. POD is improved at these low thresholds (Fig. 4b), but not enough to compensate for the large FAR degradation. Therefore, TS, which accounts for both false alarms and missed events, is also degraded at low thresholds (by as large as 0.2 at 3-hourly). In contrast, when we only correct rainfall at non-zero IMERG timesteps (Fig. 4 Column 2), the FAR degradation is much less (note the different y-axes in the two columns in Fig. 4). While it does sacrifice POD at low thresholds (Fig. 4e), the overall TS for 1-day and 3-day aggregation is improved over most of the event thresholds, especially at higher ones. As mentioned in Sect. 2.4.1, the success of this SMART choice is likely due to the improved rain/no rain detection quality of the baseline IMERG precipitation product, which was found to have superior miss-rain, false-rain and hit rate relative to TRMM TMPA-RT over the Continental U.S. [Gebregiorgis et al., 2018]. It is thus more beneficial to retain the IMERG's rain/no rain detection than to use SMART to correct it.

3.1.2. Rainfall correction evaluation

After rainfall correction at 1-day and 3-day accumulation periods, PER exhibits a domain-median error reduction of ~8% (Fig. 5 Column 1). The PER improvement is consistent with the improvement of the categorical metrics at high-event thresholds (Fig. 4 Column 2), since PER is more sensitive to high rainfall values. Three-hourly PER shows little improvement (Fig. 5a), suggesting that the deterministic correction is more effective at an accumulation period that more closely matches the SMAP retrieval interval. The same finding can also be drawn from the correlation and categorical results (Fig. 3 Column 2 and Fig. 4 Column 2).

Overall, SMART improves the IMERG rainfall product, but the improvement is generally smaller than found in previous SMART studies, especially in terms of correlation r (domain-median improvement of 0.01 to 0.02). The relatively smaller improvement is likely due to the improved accuracy of the baseline IMERG precipitation product. Table 2 summarizes the past SMART studies in literature, including the baseline and benchmark rainfall products, the SM product assimilated, baseline correlation r and its improvement, and baseline RMSE and its reduction (PER). Over the past decade, the quality of the baseline satellite-derived rainfall product has improved considerably, from TRMM 3B40-RT used in Crow et al. [2009] and Crow et al. [2011] with r = ~0.5, to TRMM 3B42-RT used in Brocca et al. [2016] with r = ~0.6 – 0.7, to IMERG used in our study with r over 0.8. Gebregiorgis et al. [2018] also used a direct comparison study to show the improved accuracy of IMERG relative to TRMM over the Continental U.S. in terms of correlation, RMSE, bias and categorical metrics. The marginal value of SMART is known to decrease as a function of increased baseline rainfall accuracy [Crow et al., 2011]. Although SMAP presumably provides more reliable SM measurements than the older satellite SM products used in previous SMART applications, its benefit does not appear sufficient to substantially correct the current generation of satellite-derived rainfall products. The high correlation may also be approaching that of the NLDAS-2 rainfall benchmark (which itself does not have perfect accuracy), thus undermining our ability to detect improvements in SMART rainfall estimates.

Finally, the probabilistic metric NENSK (Fig. 5 Column 2) is less than one for most of the domain at a 3-hour timestep, indicating an over-dispersed ensemble on average. However, when evaluating at 1-day and 3-day accumulation periods, NENSK is closer to one, indicating a

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-41
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 4 March 2019

better representation of the uncertainty of rainfall estimates. As we aggregate over longer accumulation windows (e.g., 3-day), NENSK becomes slightly greater than 1 (i.e., under-dispersed ensemble), since the SMART algorithm only assumes random rainfall error but not systematic bias, and therefore slightly underestimates the uncertainty range over longer-term periods.

In summary, SMART is able to use the SMAP retrievals to correct IMERG rainfall at relatively larger events, with slight to moderate deterministic improvement. SMART correction is less successful for small rainfall events and can even lead to slight degradation. The correction is more effective and ensemble representation is better when rainfall estimates are temporally aggregated to periods consistent with SMAP retrieval intervals (i.e., 1-day to 3-day accumulation periods), while the raw 3-hourly correction is less successful.

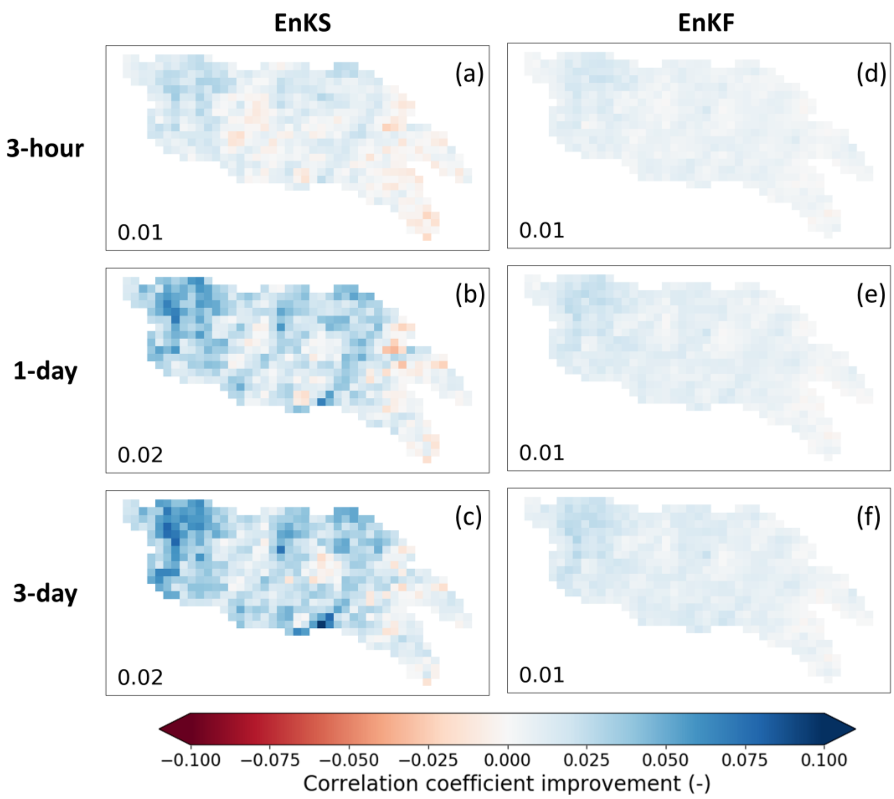

Figure 3. Maps of correlation coefficient improvement after SMART rainfall correction. The left column shows the SMART EnKS experiments (*a*, *b*, *c*), and the right column shows the EnKF experiments (*d*, *e*, *f*). Each row shows results based on different temporal accumulation period: 3-hourly, 1-day and 3-day aggregation, respectively. The number on the lower left corner of each subplot shows the domain-median correlation improvement.

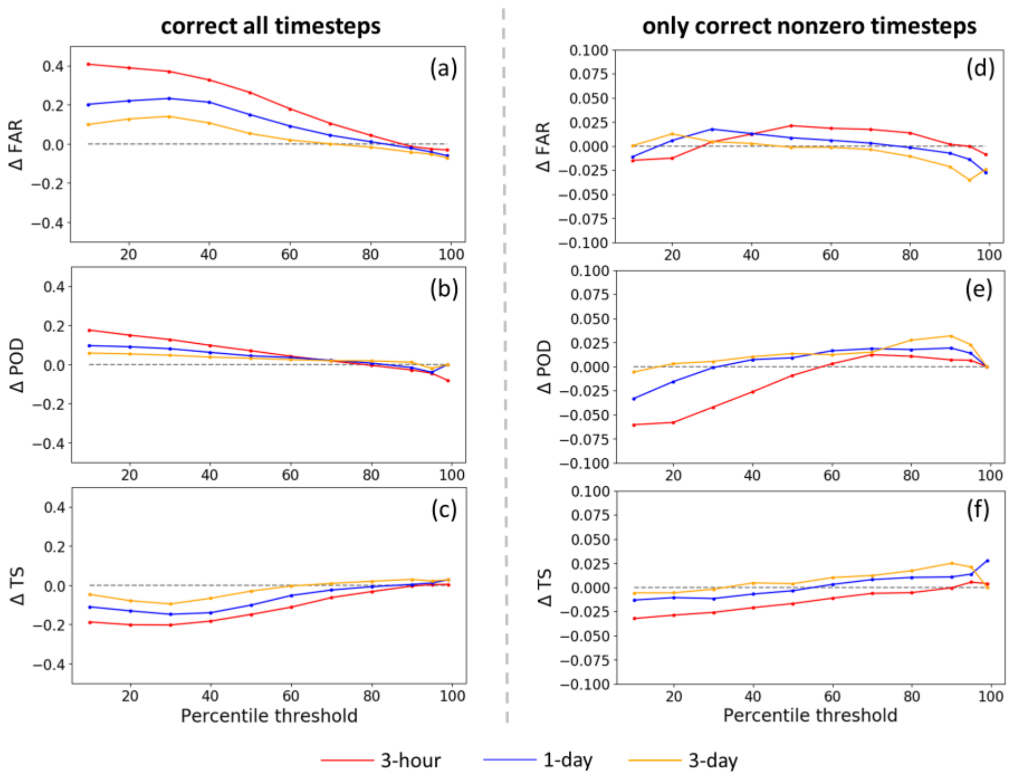

Figure 4. Change in categorical metrics (FAR, POD and TS) before and after SMART correction for 3-hourly, 1-day and 3-day accumulation periods. Metrics at different event thresholds are shown on the *x* axis. The left column (*a*, *b*, *c*) is for SMART with rainfall corrected at all timesteps; the right column (*d*, *e*, *f*) is for SMART with rainfall corrected only at non-zero timesteps. Note that the y-axis range is different for the two columns.

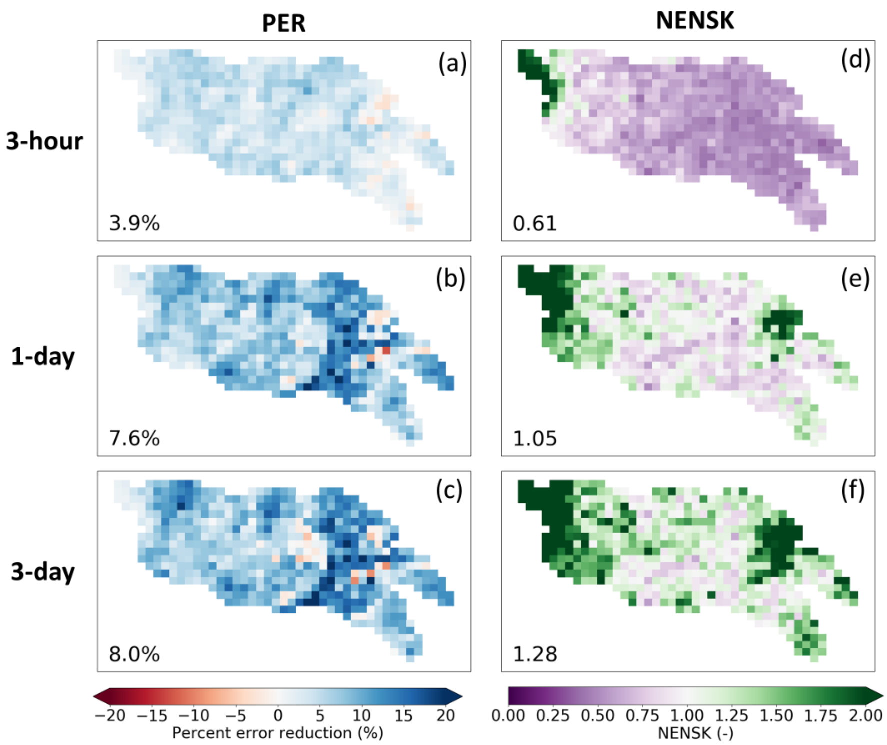

Figure 5. Maps of SMART rainfall correction results (with $\lambda = 0.1$, EnKS, and rainfall corrected only at non-zero timesteps). Each column shows the following metrics, respectively: percent RMSE reduction (PER) (*a*, *b*, *c*), and ensemble NENSK (*d*, *e*, *f*). Each row shows results based on different temporal accumulation period: 3-hourly, 1-day and 3-days, respectively. The number on the lower left corner of each subplot shows the domain-median statistic.

3.2. Streamflow from the dual correction system

3.2.1. Evaluation of streamflow improvement

The final daily streamflow performance from the dual correction system is listed in Table 3 (the "*dual*" columns) for each sub-basin. Overall, streamflow estimates are improved but with large variability across sub-basins. Specifically, PER ranges from approximately 6% to 34% and KGE improvement ranges from slightly negative to +0.95 across all sub-basins. If using the open-loop KGE (listed in Table 3) as a measure of baseline streamflow performance without any correction, we observe that at sub-basins with better open-loop streamflow simulations (i.e., Ninnescah, Walnut and Chikaskia, all with positive baseline KGE), the relative improvement after the dual correction is generally smaller.

Table 3 also summarizes the streamflow improvement from each of the correction schemes alone (the "*state update only*" and "*rainfall correction only*" columns). For sub-basins with relatively better open-loop model performance (the three with positive KGE as well as the Little Arkansas with slightly negative baseline KGE), the contribution of state updating in general surpasses that of rainfall correction. Conversely, at sub-basins with relatively poorer open-loop model performance (i.e., Bird, Spring, Illinois and Deep), streamflow improvement is primarily attributable to the SMART rainfall correction scheme.

3.2.2. Impact of rainfall forcing error

To further understand the relationship between open-loop simulation performance, rainfall forcing error and correction performance, we forced the VIC model by the NLDAS-2 benchmark rainfall (without state update). The subsequent streamflow improvement level is the maximum achievable by rainfall correction alone (Table 3 "*NLDAS2-forced*" columns). While almost all sub-basins show an obvious streamflow improvement simply by switching to the NLDAS-2 rainfall forcing, the improvement is larger at sub-basins with poorer open-loop streamflow. For example, at the four sub-basins with worse open-loop streamflow, PER is over 65% and the negative open-loop KGE improves to near zero or positive. This suggests that the poor open-loop streamflow simulations at these sub-basins are largely caused by the poor IMERG rainfall forcing. While the state update is still beneficial at these sub-basins, the SMART rainfall correction scheme is particularly important.

In contrast, the sub-basins with better open-loop streamflow demonstrate a reduced capability of streamflow improvement when switching to the NLDAS-2 rainfall forcing. The sub-basin with best open-loop streamflow, Chikaskia, even experiences smaller streamflow improvement when forced by the NLDAS-2 rainfall than when forced by SMART-corrected rainfall (Table 3). One possible reason is that the NLDAS-2 benchmark rainfall at this sub-basin is not obviously superior than the IMERG baseline. Therefore, switching to the NLDAS-2 rainfall forcing does not benefit streamflow much, but SMART is still able to extract information from SMAP and slightly correct IMERG rainfall and subsequent streamflow.

3.2.3. Impact of model parameterization

The dual correction scheme presented in this study is designed to only correct the random error existing in the simulation system, but not systematic error or overall bias. Figure 6 shows example time series of the open-loop, USGS-observed and dual-corrected streamflow at three sub-basins with various levels of open-loop performance. It is readily apparent from the time series that, although the dual system often nudges the simulated streamflow in the correct direction (especially during high-flow periods) and results in overall improved evaluation statistics, obvious systematic error (in the model process representation as well as rainfall forcing) exists. This systematic error, although difficult to quantify, cannot be corrected by the data assimilation approach discussed here. The NENSK statistic partly reflects such systematic error. NENSK is significantly above one at most sub-basins, indicating an under-dispersed ensemble on average. In other words, at most sub-basins the ensemble spread created by the dual system only represents the random uncertainty around the open-loop streamflow, but not the systematic error which accounts for much of the total error.

The level of systematic error is tied closely to the quality of the hydrologic model parameters, often estimated through calibration. The VIC parameters used in this study were taken from Maurer et al. [2002] and derived based on streamflow at the outlets of large basins. To further examine the effect of systematic error on data assimilation, we instead calibrated the model parameters for the 8 sub-basins separately using streamflow acquired from the USGS (Table 1). Specifically, VIC parameters that control infiltration, soil conductivity and baseflow generation as well as the recession rate of the grid-cell-scale unit hydrograph in RVIC were calibrated using the MOCOM multi-objective autocalibration method [Yapo et al., 1998]. Basin-constant parameters were calibrated toward USGS streamflow time series during 2015 to 2017 (forced by the baseline IMERG precipitation) to optimize daily KGE and monthly bias. Only a subset of the 8 sub-basins were able to achieve better-than-open-loop streamflow results via this traditional calibration method, mainly due to the large IMERG forcing error at some sub-basins that makes the calibration scheme incapable of finding an improved parameterization. Figure 7 shows three example sub-basins with relatively good calibration outcome as demonstration. Comparing Fig. 6 and Fig. 7, all three sub-basins exhibit a similar or smaller magnitude of streamflow correction after parameter calibration. While a good calibration itself can

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-41
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 4 March 2019

significantly improve baseline performance, a poor calibration does not degrade (and sometimes even benefit) the relative added value of the dual correction.

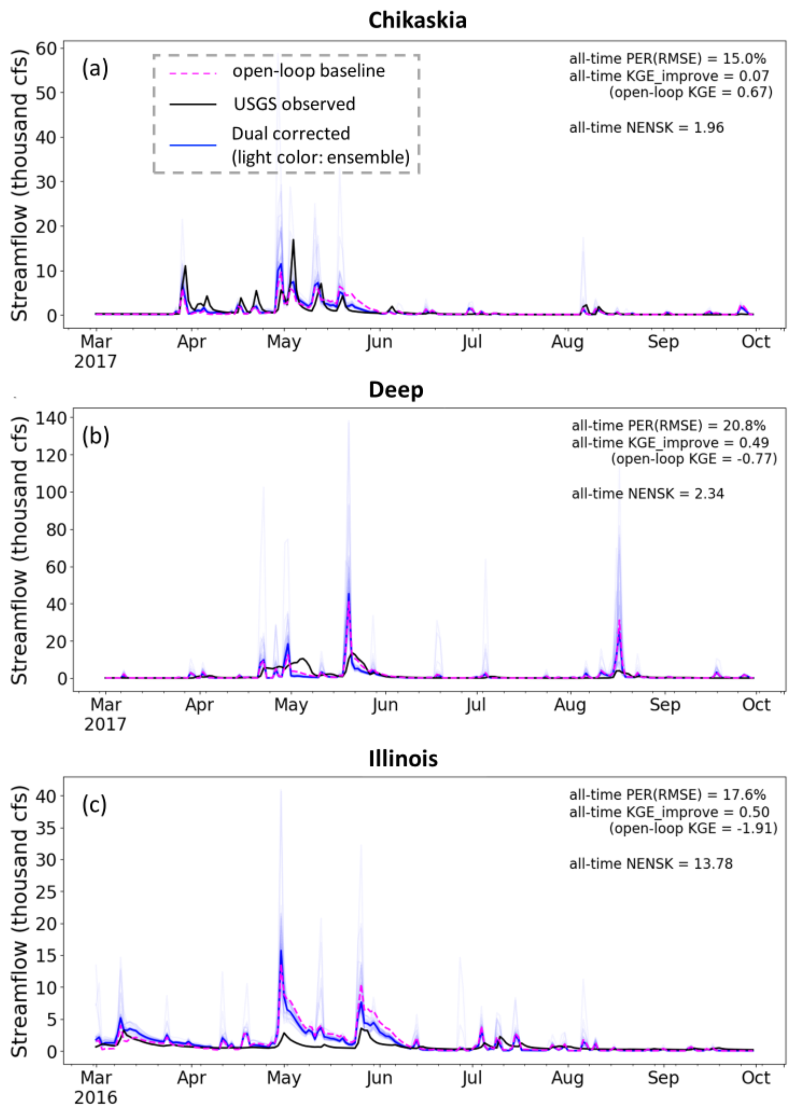

Figure 6. Example time series of streamflow results from the dual correction system. *Black line*: USGS observed streamflow; *magenta line*: baseline VIC simulation; *light blue lines*: ensemble updated streamflow results; *solid blue line*: ensemble-mean updated streamflow. Only part of the

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-41
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 4 March 2019

simulation period is shown for clear display. Statistics shown on each panel are based on the entire simulation period (approximately 2.5 years).

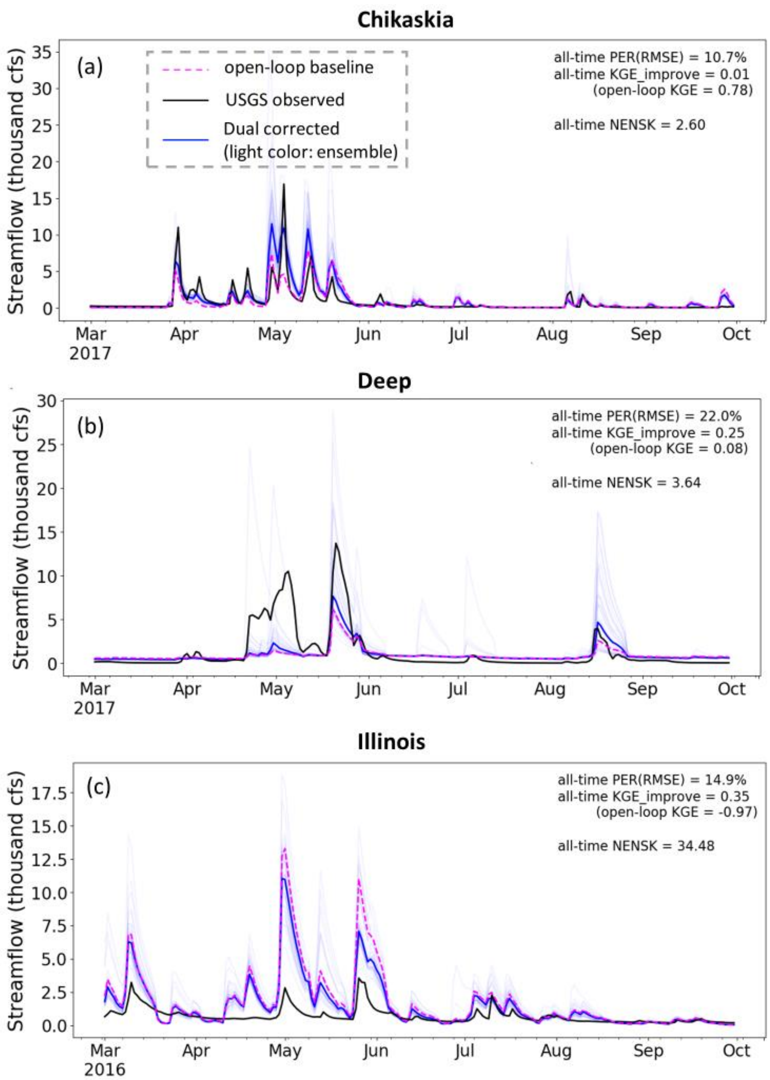

Figure 7. Time series of simulated open-loop, corrected and observed streamflow at three example sub-basin outlets with calibrated model parameters. All lines and notations are the same as in Fig. 6.

4. Discussion

Although we applied the dual correction system to the entire Arkansas-Red basin, we selected 8 smaller basins for our streamflow evaluation due to the limited availability of unregulated streamflow observations at basin outlets. Additional research is needed to fully investigate the impact of error spatial correlation on downstream streamflow performance before extending our findings to large-scale river systems. Specifically, while a 1-D filter with spatially white model representation error may be appropriate for small-basin correction, ignoring the spatial correlation structure of errors could potentially have a more profound impact on the correction performance at large river outlets where streamflow originates from runoff from a large number of grid cells. A number of studies have investigated the effects of spatial error patterns in hydrologic data assimilation. For example, Reichle and Koster [2003] investigated the impact of spatial error correlation in the model SM states on its assimilation performance; Gruber et al. [2015] examined the impact of a 2-D filter with spatially auto-correlated error versus a 1-D filter on SM updating quality; Pan et al. [2009] and Pan and Wood [2009; 2010] evaluated the surface SM assimilation performance with VIC by comparing a 1-D filter, a 2-D filter and a multiscale autoregressive filtering approach, as well as considering spatial and temporal structure of precipitation error. However, these studies focused exclusively on the performance of SM simulations. Direct assessment of the impact of spatial error patterns on the routed streamflow results is needed, especially from a probabilistic perspective since the ignorance of spatial error patterns may potentially cause error cancelation at large outlets and therefore incorrect ensemble representation of uncertainty.

Nevertheless, this study leads to a number of valuable insights. We have shown that the dual correction approach is able to correctly nudge streamflow simulation, especially during relatively high flow events in areas with poor IMERG data. However, the magnitude of this correction is generally small for two reasons. First, the latest generation of satellite rainfall products (e.g., IMERG) has significantly improved precision compared to its predecessors. The already high-quality rainfall estimates are more difficult for SM retrievals to contribute substantial rainfall correction skill. Second, the dual correction approach is designed to correct only the zero-mean random error component in the total streamflow error but not systematic error or bias. However, systematic error sources, typically associated with inaccurate model structure and/or parameterization and large rainfall bias, can account for a significant fraction of overall streamflow error. The existence of systematic error is particularly problematic from a probabilistic perspective, since the ensemble streamflow produced by the dual system only represents random error, and therefore largely underestimates simulation uncertainty.

Given the above considerations, we may be approaching a point of diminishing returns for applying data assimilation techniques that are aimed solely at reducing random error sources in streamflow simulations. This insight provides few recommendations for future research:

1) More sophisticated data assimilation techniques aimed solely at random error sources are unlikely to substantially reduce streamflow error further, since random errors sometimes account for only a relatively small portion of the total error;

2) Instead, approaches that reduce systematic errors in streamflow simulation are needed. To date this is still a challenging task in large-scale hydrologic modeling, since calibration is difficult to perform with limited streamflow data and a large number of distributed parameters. With the availability of the near-global and distributed satellite products such as SMAP and IMERG, more creative methods need to be developed to extract useful information from the large volume of remote sensing observations. For example, characteristics of SM dynamics and its response to rainfall can be directly extracted from the datasets themselves, which can potentially inform hydrologic model representation. These areas of research are less studied but have the potential to improve hydrologic modeling beyond correcting random errors;

3) It is worthwhile to continue to develop future generation of higher-quality, near-real-time rainfall products, since rainfall plays a dominant role in streamflow simulations and its error is not easily and substantially reduced by the current correction methods that use SM measurement information.

5. Conclusion

In this paper, we applied a dual state/rainfall correction data assimilation system in the Arkansas-Red River basin. Built upon the dual system developed in past studies, we have made several methodological advances. First, we implemented the dual correction system with a more complexed, semi-distributed land surface model, the VIC model, and applied it in a regional-scale basin. Second, the latest satellite products, the SMAP SM product and the IMERG rainfall product, were incorporated into the system. Third, the existing dual correction algorithm was extended to maximize the use of information contained in the more accurate and temporally finer satellite data products, and also to produce an ensemble streamflow product. Fourth, we confirmed via a formal synthetic experiment that error cross-correlation that potentially exists in the dual correction system does not cause noticeable degradation of streamflow improvement, and the dual correction scheme applied here is optimal.

Our results show that, overall, IMERG rainfall and streamflow are improved to some extent but not substantially via dual correction. For rainfall, the improvement is primarily from the correction of larger events via SMART, while smaller events are slightly degraded. Rainfall correction is more effective at daily to multi-daily time scales than at a 3-hourly scale. The ensemble produced by the correction scheme represents the rainfall uncertainty relatively well at daily to multi-daily scale. For streamflow, the dual correction reduces the random errors in simulated streamflow across the 8 test sub-basins, ranging from near zero improvement to moderate error reduction. Sub-basins with relatively poorer open-loop streamflow simulations, due mainly to poor IMERG rainfall forcing quality, exhibit relatively larger correction, and the correction is mainly contributed by the SMART rainfall correction scheme. Sub-basins with relatively better IMERG and open-loop streamflow show less relative correction, and the correction is attributable more to state updating. The streamflow ensemble produced by the dual correction system largely underestimates error uncertainty, because the system accounts only for the random error components and not systematic error (resulting, e.g., from incorrect model structure or parameterization). Finally, we demonstrated that model parameterization errors that commonly exist in large-scale distributed models in general does not degrade (and sometimes actually benefits) the relative added value of the dual correction scheme.

These findings suggest that we are approaching a point of diminishing returns for SM data assimilation techniques aimed solely at the reduction of random errors in simulated streamflow. More sophisticated SM data assimilation techniques may lead to additional marginal improvement, but more substantial streamflow reduction likely require future research efforts on reducing systematic modeling errors via, e.g., innovative ways of achieving better model representation as well as obtaining higher-quality satellite rainfall products.

Code availability

The VIC model used in the study can be found at https://github.com/UW-Hydro/VIC. Specifically, we used VIC version 5.0.1 (doi:10.5281/zenodo.267178) with a modification to the calculation of drainage between soil layers (https://github.com/UW-Hydro/VIC/releases/tag/Mao_etal_stateDA_May2018). The DA code used in this study is available at https://github.com/UW-Hydro/dual_DA_SMAP.

Author contribution

All co-authors designed the experiments. Yixin Mao developed the system code and carried out the experiments. Wade T. Crow and Bart Nijssen supervised the study. Yixin Mao prepared the manuscript with contributions from all co-authors.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

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 No. EAR-1338606. We would also like to thank Andrew Wood from NCAR for help on calibration.

References

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-41
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 4 March 2019

Hydrology and Earth System Sciences
Open Access
Discussions

Alvarez-Garreton, C., D. Ryu, A. W. Western, 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, doi:10.1016/j.jhydrol.2014.07.041.

Alvarez-Garreton, C., D. Ryu, A. W. Western, W. T. Crow, 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(7), 5357-5375, doi:10.1002/2015WR018429.

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(1-4), 145-161, doi:10.1016/S0022-1694(03)00229-4.

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, doi:10.5194/hess-14-1881-2010.

Brocca, L., 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(7), 2542-2555,doi:10.1109/TGRS.2011.2177468.

Brocca, L., T. Moramarco, F. Melone, and W. Wagner (2013), A new method for rainfall estimation through soil moisture observations, *Geophys. Res. Lett.*, 40, 853–858, doi:10.1002/grl.50173.

Brocca, L., L. Ciabatta, C. Massari, T. Moramarco, S. Hahn, S. Hasenauer, R. Kidd, W. Dorigo, W. Wagner, and V. Levizzani (2014), Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data, *J. Geophys. Res. Atmos.*, 119, 5128–5141, doi:10.1002/2014JD021489.

Brocca, L., T. Pellarin, W. T. Crow, L. Ciabatta, C. Massari, D. Ryu, C.-H. Su, C. Rüdiger, and Y. Kerr (2016), Rainfall estimation by inverting SMOS soil moisture estimates: A comparison of different methods over Australia, *J. Geophys. Res. Atmos.*, 121, 12,062–12,079, doi:10.1002/2016JD025382.

Chan, S. et al. (2017), Development and validation of the SMAP enhanced passive soil moisture product, Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International, doi:10.1109/IGARSS.2017.8127512.

Chen F., W. T. Crow, and T. R. H. Holmes (2012), Improving long-term, retrospective precipitation datasets using satellite-based surface soil moisture retrievals and the Soil Moisture Analysis Rainfall Tool, *J. Appl. Remote Sens.*, 6(1), 063604, doi:10.1117/1.JRS.6.063604.

Chen, F., W. T. Crow, and D. Ryu (2014), Dual forcing and state correction via soil moisture assimilation for improved rainfall–runoff modeling, *J. Hydrometeorol.*, *15*(5), 1832–1848, doi:10.1175/JHM-D-14-0002.1.

Colliander, A. et al. (2017), Validation of SMAP surface soil moisture products with core validation sites, *Remote Sens. Environ.*, 191, 215-231, doi: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, doi:10.1029/2007GL029450.

Crow, W. T., and D. Ryu (2009), A new data assimilation approach for improving hydrologic prediction using remotely-sensed soil moisture retrievals, *Hydrol. Earth Syst. Sci.*, 12(1-16), doi:10.5194/hess-13-1-2009.

Crow W. T., G. J. Huffman, R. Bindlish, and T. J. Jackson (2009), Improving satellite-based rainfall accumulation estimates using spaceborne surface soil moisture retrievals, *J. Hydrometeorol.*, 10, 199-212, doi:10.1175/2008JHM986.1.

Crow, W. T., 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, doi:10.1029/ 2011WR010576.

Crow, W. T., 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, doi:10.1002/2017GL073642.

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, doi:10.1029/2005JD006367.

Entekhabi et al. (2010), The Soil Moisture Active and Passive (SMAP) Mission, *Proceedings of the IEEE*, 98(5), 704-716, doi:10.1109/JPROC.2010.2043918.

Francois, C., Quesney, A., 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(2), 473-487, doi: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(3), 237-258, doi:10.1016/0022-1694(69)90020-1.

Gebregiorgis, A. S., P.-E. Kirstetter, Y. E. Hong, J. J. Gourley, G. J. Huffman, W. A. Petersen, X. Xue, and M. R. Schwaller (2018), To what extent is the day 1 GPM IMERG satellite precipitation estimate improved as compared to TRMM TMPA-RT?, *J. Geophys. Res. Atmos.*, 123, 1694–1707, doi: 10.1002/2017JD027606.

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, doi: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, doi: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.*, 122(3), 1683-1701, doi:10.1002/2016JC012323.

Hamman, J. J., B. Nijssen, 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, doi:10.5194/gmd-11-3481-2018.

Hou, A. Y., R. K. Kakar, S. Neeck, A. A. Azarbarzin, C. D. Kummerow, M. Kojima, R. Oki, K. Nakamura, and T. Iguchi (2014), The Global Precipitation Measurement mission, *Bull. Amer. Meteor. Soc.*, 95(5), 701-722, doi:10.1175/BAMS-D-13-00164.1.

Huffman, G. J., D. T. Bolvin, and E. J. Nelkin (2015), Integrated Multi-Satellite Retrievals for GPM (IMERG) Technical Documentation. Tech. Doc., NASA GSFC. [Available online at https://docserver.gesdisc.eosdis.nasa.gov/public/project/GPM/IMERG_doc.05.pdf.]

Huffman, G. J., E. F. Stocker, D. T. Bolvin, and E. J. Nelkin (2018), last updated 2018: IMERG L3 Early Run Data Sets. NASA/GSFC, Greenbelt, MD, USA, Accessed 2018-08-29, https://gpm1.gesdisc.eosdis.nasa.gov/opendap/hyrax/GPM_L3/GPM_3IMERGHHL.05/.

Koster, R. D., L. Brocca, W. T. Crow, M. S. Burgin, and G. J. M. De Lannoy (2016), Precipitation estimation using L-band and C-band soil moisture retrievals, *Water Resour. Res.*, 52, 7213–7225, doi:10.1002/2016WR019024.

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(D7), 14415-14428, doi:10.1029/94JD00483.

Lievens, H., et al. (2015), SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia, *Remote Sens. Environ.*, *168*, 146-162, doi:10.1016/j.rse.2015.06.025.

Lievens, H., G. J. M. De Lannoy, A. Al Bitar, M. Drusch, G. Dumedah, H.-J. Hendricks Franssen, Y. H. Kerr, S. K. Tomer, B. Martens, O. Merlin, M. Pan, J. K. Roundy, H. Vereecken, and J. P. Walker (2016), Assimilation of SMOS soil moisture and brightness temperature products into a land surface model, *Remote Sens. Environ.*, 180, 292-304, doi: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 parametrization schemes, Tellus, 48(A), 708-721, doi:10.1034/j.1600-0870.1996.t01-3-00009.x.

Lohmann, D., 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(1), 131-141, doi:10.1080/02626669809492107.

Mao Y., W. T. Crow, and B. Nijssen (2019), A framework for diagnosing factors degrading the streamflow performance of a soil moisture data assimilation system, *J. Hydrometeorol.*, 20(1), 79-97, doi:10.1175/JHM-D-18-0115.1.

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, 11403-11433, doi:10.3390/rs70911403.

Massari, C., S. Camici, L. Ciabatta, and L. Brocca (2018), Exploiting satellite-based surface soil moisture for flood forecasting in the Mediterranean area: State update versus rainfall correction, *Remote Sens.*, 10, 292, doi:10.3390/rs10020292.

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

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-41
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 4 March 2019

United States, *J. Clim.*, 15(22), 3237-3251, doi: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. Control.*, 16(1), 12-21, doi:10.1109/TAC.1971.1099621.

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, Boulder, Colorado USA, NASA National Snow and Ice Data Center Distributed Active Archive Center, Accessed 2018-01-18, doi:10.5067/OBBHQ5W22HME.

Pan, M., E. F. Wood, D. B. McLaughlin, and D. Entekhabi (2009), A multiscale ensemble filtering system for hydrologic data assimilation. Part I: Implementation and synthetic experiment, *J. Hydrometeorol.*, 10, 794-806, doi:0.1175/2009JHM1088.1.

Pan, M., and E. F. Wood (2009), A multiscale ensemble filtering system for hydrologic data assimilation. Part II: Application to land surface modeling with satellite rainfall forcing, *J. Hydrometeorol.*, 10, 1493-1506, doi:10.1175/2009JHM1155.1.

Pan, M., and E. F. Wood (2010), Impact of accuracy, spatial availability, and revisit time of satellite-derived surface soil moisture in a multiscale ensemble data assimilation system, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 3 (1), 49-56, doi:10.1109/JSTARS.2010.2040585.

Reichle, R. H., and R. D. Koster (2003), Assessing the impact of horizontal error correlations in background fields on soil moisture estimation, *J. Hydrometeorol.*, 4 (6), 1229-1242, doi:10.1175/1525-7541(2003)004<1229:ATIOHE>2.0.CO;2.

Talagrand, O., R. Vautard, and B. Strauss (1997), Evaluation of probabilistic prediction systems, technical report, Eur. Cent. for Medium-Range Weather Forecast., Reading, UK.

United States Geological Survey (USGS) (2018), USGS Surface-water daily data for the nation. [Available at https://waterdata.usgs.gov/nwis/dv/?referred_module=sw.]

Wanders, N., D. Karssenberg, 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*(6), 2343-2357, doi: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*(1), 149-180, doi:10.1146/annurev.earth.30.091201.140434.

Wilks, D. S. (2011), Statistical methods in the atmospheric sciences (3rd edition), Elsevier/Academic Press, Amsterdam; Boston.

Xia, Y. et al., NCEP/EMC (2009), NLDAS Primary Forcing Data L4 Hourly 0.125 x 0.125 degree V002, Edited by David Mocko, NASA/GSFC/HSL, Greenbelt, Maryland, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed 2018-02-27, doi:10.5067/6J5LHHOHZHN4.

Yapo, P. O., H. V. Gupta, and S. Sorooshian (1998), Multi-objective global optimization for hydrologic models, *J. Hydrol*. 2014, 83-97, doi:10.1016/S0022-1694(97)00107-8.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-41
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 4 March 2019

Table 1. List of USGS streamflow sites used for verification.

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

Table 2. Review of SMART rainfall correction results in literature along with the results in this study.

Literature	Baseline rainfall product	Benchmark rainfall product	SM product	Domain	Accumulation period	Baseline correlation r	r improvement	Baseline RMSE (mm)	PER
Crow et al. [2009]	TRMM 3B40RT	CPC rain gauge analysis	AMSR-E	Southern Great Plain	3-day	~ 0.5	~ + 0.2	13.0	~ 30%
				CONUS	3-day	~ 0.55	~ + 0.05	11.8	~ 15%
Crow et al. [2011]	TRMM 3B40RT	CPC rain gauge analysis	AMSR-E	CONUS	3-day	~ 0.55	~ + 0.1	13.1	~ 20%
Chen et al. [2012]	Princeton Global Forcing Dataset	CPC rain gauge analysis	SMMR, SMM/I, ERS	Global	10-day	~ 0.35	~ + 0.15	-	-
Brocca et al. [2016]	TRMM 3B42RT	AWAP rain gauge product	SMOS	Australia	1-day	0.62	+0.01	5.6	7%
					5-day	0.71	+0.05	14.0	14%
This study	IMERG Early Run	NLDAS-2	SMAP L3 Passive	Arkansas-Red	1-day	0.80	+0.02	6.1	8%
					3-day	0.82	+0.02	11.0	8%

Table 3. Daily streamflow results from the dual correction system for the 8 USGS sub-basins shown in Fig. 1. In addition to the deterministic KGE improvement, PER and probabilistic NENSK results from the dual system ("*dual*" columns), the table also lists the open-loop streamflow KGE ("*open-loop KGE*" column), KGE improvement and PER as a result of state update or rainfall correction scheme alone ("*state update only*" and "*rainfall correction only*" columns, respectively), and KGE improvement and PER when forced by the NLDAS-2 benchmark precipitation without state update ("*NLDAS-2 forced*" column).

	Open-loop KGE	KGE improvement				PER				NENSK
		Dual	State update only	Rainfall correction only	NLDAS2-forced	Dual	State update only	Rainfall correction only	NLDAS2-forced	Dual
L Arkansas	-0.12	+0.17	+0.23	-0.01	+0.57	7.3%	10.8%	1.2%	40.0%	1.98
Ninnescah	0.25	+0.15	+0.06	+0.16	+0.20	14.0%	5.5%	13.7%	30.4%	0.35
Walnut	0.54	-0.02	-0.03	+0.03	-0.23	5.8%	5.7%	2.8%	23.3%	2.70
Chikaskia	0.67	+0.07	+0.05	+0.02	-0.45	15.0%	11.1%	6.6%	2.2%	1.96
Bird	-1.49	+0.95	+0.58	0.63	+0.95	33.5%	17.0%	25.8%	68.9%	2.01
Spring	-3.64	+0.83	+0.65	+0.33	+3.93	13.2%	8.7%	7.0%	83.4%	13.11
Illinois	-1.91	+0.50	+0.36	+0.26	+2.72	17.6%	7.4%	12.9%	81.8%	13.78
Deep	-0.77	+0.49	+0.39	+0.37	+1.55	20.8%	13.1%	21.2%	68.3%	2.34