Toward Improving Drought Monitoring using the Remotely Sensed Soil Moisture Assimilation: A Parallel Particle Filtering Framework

Hongxiang Yan, Mahkameh Zarekarizi, and Hamid Moradkhani*
Center for Complex Hydrosystems research
Department of Civil, Construction and Environmental Engineering, The University of Alabama

* Corresponding Author

Abstract

Drought is the costliest hazard among all natural disasters. Despite the significant improvements in drought modeling over the last decade, accurate provisions of drought conditions in a timely manner is still a major research challenge. In order to improve the current drought monitoring skills, this study presents a land data assimilation system by merging the remotely sensed surface soil moisture with the model simulations with the use of a recently developed particle Markov chain Monte Carlo (PMCMC) method. To cope with the computational complexity, a modular parallel particle filtering framework (PPFF) is developed which allows a large ensemble size in PMCMC applications. The implementation of the proposed system is demonstrated with the 2012 summer flash drought case study over the Contiguous United States (CONUS). Results from both synthetic and real case studies suggest that the land data assimilation system improves the soil moisture predictions and the drought monitoring skills. Compared with the U.S. Drought Monitoring (USDM), the land data assimilation can better capture the drought onset on May 2012 and the drought severity in June and July 2012. This study recommends that the proposed land data assimilation system based on a high-performance computing (HPC) infrastructure can better facilitate the drought preparation and response actions.

Keywords: Drought monitoring; Soil moisture; Particle Markov chain Monte Carlo; Remote sensing; Parallel computing

1 Introduction

Drought is a complex natural hazard that affects hydrological, environmental, ecological, and social systems in many ways. Currently, no universal definition of drought exists (Lloyd-Hughes, 2014). Several drought definitions can be found in Wilhite (2000), Keyantash and Dracup (2002), Mishra and Singh (2010), Sheffield and Wood (2011), and Van Loon (2015). Generally, drought can be described as a deficiency in precipitation, soil moisture, or surface/ground water over an extended period, which can have significant negative impacts on agricultural, ecological, and socio-economic systems. A drought event can be short, lasting for just a few months, or it can persist for multiple years.

Among all natural disasters, drought is the most costly hazard (Sheffield et al., 2014). For example, the North American drought in 1988 resulted in nearly $62 billion loss, which was more than the cost of the 1993 Mississippi River flood and Hurricane Andrew combined (Ross and Lott, 2003). The 2012 summertime flash drought event across the Central U.S. caused a major curtailment in crop yields, and resulted in about $12 billion economic loss (Hoerling et al., 2014). One possible reason for such huge losses from a drought event is the lack of prompt preparation and effective response actions due to insufficient knowledge of the drought development behavior. Different from other natural disasters, drought has a slow onset and develops over large areas, which makes it difficult to detect until severe damage has already occurred (Wood et al., 2015). Therefore, a drought monitoring system that can detect drought conditions in a timely manner is essential for drought preparedness and risk reduction (Ahmadalipour et al., 2017; Andreadis et al., 2005; Hao et al., 2014; Maurer et al., 2002; Sheffield et al., 2014).

The current operational drought monitoring systems generally use the simulated soil moisture from hydrologic models to monitor drought conditions. For instance, the Climate Prediction Center (CPC) soil moisture data sets operationally used in the U.S. Drought Monitoring (USDM) (Svoboda et al., 2002) is based on a one-layer leaky bucket model. Although model simulations can provide consistent long-term soil moisture data sets at a continental scale, these soil moisture estimates are potentially biased due to the errors in model parameters, forcing data, and deficiencies in the model structure (Chaney et al., 2015; DeChant and Moradkhani, 2014; Moradkhani and Sorooshian, 2008; Samaniego et al., 2013; Yan and Moradkhani, 2016). As a result, the biased soil moisture may lead to sub-optimal drought monitoring skills.

A plausible approach to improve simulated soil moisture is to exploit the remotely sensed observations to update the soil moisture states in the model. This method of integrating observations and model simulations is referred to as data assimilation (DA) (Moradkhani, 2008). The assimilation of satellite soil moisture into a hydrologic model has received increasing attention and there have been numerous studies that have investigated the effects of assimilation of remotely sensed data on soil moisture predictions (Brocca et al., 2012; De Lannoy et al., 2007; Draper et al., 2012; Montzka et al., 2011; Reichle et al., 2008; Yan et al., 2015). While these studies have suggested improvements on soil moisture predictions, few of the studies actually quantified the improvements on the end-use application such as drought monitoring skill (Kumar et al., 2014a). Different from the studies focused on the soil moisture predictions, which can be performed at point or watershed scale, drought develops at regional to continental scales and therefore, it generally requires hydrologic modeling at large-scale (DeChant and Moradkhani, 2015; Hoerling et al., 2014). Compared to the model forward run, the large-scale DA is far more computationally expensive. Because of this computational complexity, the majority of previous large-scale satellite soil moisture DA studies were based on the ensemble Kalman filter (EnKF) with the use of small ensemble size (12–20) (Kumar et al., 2014b, 2009; Pan and Wood, 2010; Yin et al., 2015).

Although the successful applications of the EnKF have been reported in the above studies, the EnKF technique has some inherent features resulting in sub-optimal performances in hydrologic applications (Abbaszadeh et al., 2018; DeChant and Moradkhani, 2012; Dong et al., 2015; Leisenring and Moradkhani, 2011; Lorentzen and Naevdal, 2011; Yan et al., 2017). First, the EnKF uses only the first- and second-order statistical moments and assumes the model and observation error distribution to be Gaussian, which is violated in the nonlinear and non-Gaussian hydrologic system. Second, the updating step within the EnKF is based on a linear equation. As is the case in most hydrologic models, the observation model (or observation operator) is nonlinear, therefore, such an updating rule may not be correct since the posterior ensemble is not a sample from the posterior probability density function resulted from the Bayes' law (Lorentzen and Naevdal, 2011). Third, the EnKF technique violates mass conservation (not preserving the water balance) because water is removed from or added to the model by the updating formulation, which may lead to non-physical model state values. For drought applications, the closure of the water balance is especially important because a moderate change of soil water due to imbalanced water budget can result in a category change of drought severity. The use of small ensemble size (12–20) in these studies further deteriorates the EnKF performances in drought monitoring because such small ensemble size is insufficient to represent the posterior distributions.

To overcome the above EnKF problems, data assimilation by means of particle filter (PF) has been recommended as an alternative approach in hydrologic applications (Dong et al., 2015; Montzka et al., 2011; Moradkhani et al., 2012; Noh et al., 2011; Plaza et al., 2012; Yan et al., 2017). In comparison to the EnKF, the PF can preserve the water balance and relaxes the Gaussian assumption of error distributions, which allows the PF to potentially characterize skewed or multimodal posterior distributions. This is accomplished by resampling the model state or state-parameter ensemble, as opposed to the linear updating rule of the EnKF. In other words, the PF can lead to a more complete representation of the posterior distribution for a nonlinear and non-Gaussian hydrologic system. In the literature, a few studies have compared the effectiveness and robustness of the EnKF and PF in hydrologic predictions, and they suggested that the PF is a more effective and robust data assimilation technique. For instances, Leisenring and Moradkhani (2011) examined the performances of EnKF and PF on snow water equivalent (SWE) predictions with the assimilation of SNOTEL observations. They found out that the PF reduced the root-mean-square-error (RMSE) of SWE predictions from the EnKF by about 33%. DeChant and Moradkhani (2011) assessed both the EnKF and PF techniques on SWE predictions with the assimilation of satellite brightness temperature data. Their results suggested that PF provided more accurate predictions than EnKF. Van Delft et al. (2009) compared the use of EnKF and PF on streamflow predictions and suggested PF approach led to a lower RMSE value. Pasetto et al. (2012) found out that PF is a more robust method in soil moisture assimilation because Gaussian approximation in the EnKF led to a state estimation that is inconsistent with the physics of the model. DeChant and Moradkhani (2012) further provided a comprehensive robust assessment between the EnKF and PF on streamflow predictions. They demonstrated that the PF is a more robust technique because the streamflow predictions from the

EnKF were consistently overconfident, and occasional filter divergence was also identified in the EnKF.

Despite the advantages of PF, few satellite soil moisture DA studies used PF approach and the majority of these PF studies were limited to point to watershed scale (Montzka et al., 2011; Plaza et al., 2012; Yan et al., 2015; Yan and Moradkhani, 2016), and fewer have attempted to implement PF for large-scale drought analysis (e.g., Yan et al., 2017). The main obstacle for the PF to be applied in large-scale drought application is the limited computational power of modern computers, which means that we cannot have enough model ensembles to simulate the posterior distributions and avoid weight degeneration (filter collapse due to few particles having significant weight). Compared to the EnKF method, the successful application of PF requires a larger ensemble size. The current EnKF based large-scale drought monitoring system (with the use of small ensemble size) is already compute-intensive; while the PF needs to increase that demand by 1–2 orders of magnitude. One possible solution allowing the use of PF in large-scale drought applications is to benefit from the parallel computing technique in a high-performance computing (HPC) cluster infrastructure, which requires the parallel implementations of DA algorithm. Currently, open source parallel DA library is available in the community such as the Parallel Data Assimilation Framework (PDAF) (Nerger and Hiller, 2013). The PDAF is developed at the Computing Center of the Alfred Wegener Institute based on Fortran code compilable in the Unix/Linux environment and provides fully implemented and optimized data assimilation based on the Kalman filtering algorithms (Ridler et al., 2014). Although the PDAF can provide parallel simulation for large-scale DA applications on a HPC cluster, the PF technique is not included in the PDAF up to date.

Given the above discussions, the main motivation of this study is to advocate the use of PF approach in large-scale drought applications by using the parallel computing technique in the HPC cluster. Specifically, a parallel PF modular is needed to be compatible with modern HPC infrastructure. Therefore, the goals of this study are to:

1. develop a modular parallel particle filtering framework (PPFF) which allows a large ensemble size in large-scale (continental to global scale) drought applications;
2. examine the effectiveness of PPFF by comparing its soil moisture predictions to the results from the EnKF based data assimilation system; and
3. implement the large-scale assimilation of remotely sensed soil moisture into a distributed hydrologic model and provide a quantitative assessment of its impact toward drought monitoring skills with the use of PPFF."

The remaining of the paper is organized as follows: section 2 describes the framework of the proposed PF based drought monitoring system, which includes the dynamical hydrologic modeling, the DA algorithm, and the parallelization of the DA. Section 3 assesses the drought monitoring system over the Contiguous United States (CONUS) for both synthetic and real case studies. Finally, section 4 concludes the paper.

2 Drought Monitoring System

In this study, the proposed drought monitoring system is composed of two parts: dynamical hydrologic molding and land data assimilation system. First, the hydrologic model is calibrated for the study region of interest, then the remotely sensed soil moisture observations are assimilated into the calibrated hydrologic model through the land data assimilation system. Figure 1 illustrates the framework of the proposed drought monitoring system.

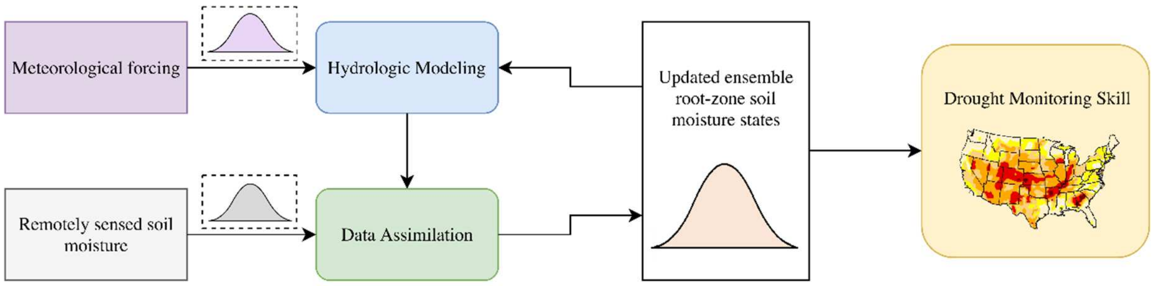

Figure 1. The flowchart of the proposed drought monitoring system. The three probability distributions associated with the meteorological forcing, remotely sensed soil moisture, and soil moisture states represent the corresponding uncertainties.

2.1 Dynamical Hydrologic Modeling

The Variable Infiltration Capacity (VIC) model is a physically based and semi-distributed macroscale hydrologic model. The VIC model was originally developed by Liang et al. (1994) and later improved by Lohmann et al. (1998) and Liang and Xie (2001). The VIC model includes both water-balance and energy-balance parameterizations and two types of runoff-yielding mechanisms. In this model, the land surface is simulated as a gird of large, flat, and uniform cells. The VIC model balances both surface energy and water over each grid cell. The VIC model represents sub-grid variability in soils, topography, and vegetation, and this allows representation of the non-linear dependence of the partitioning of precipitation into infiltration and direct runoff as determined by soil moisture in the upper layer and its spatial heterogeneity. The VIC model partitions the vadose zone into three soil layers. The first soil layer has a fixed depth of 10 cm and responds quickly to changes in surface conditions and precipitation. The second and third soil layer depths are spatially varied and the same as in the Land Data Assimilation System (LDAS) retrospective simulations (Maurer et al., 2002). Moisture movement between the three soil layers is governed by gravity drainage, with diffusion from the second to the first layer allowed in unsaturated conditions. Water drained from the second layer to the third layer is entirely controlled by gravity. Base flow is a non-linear function of the soil moisture content of the third layer (Andreadis et al., 2005; Andreadis and Lettenmaier, 2006; Shukla et al., 2011). The minimum meteorological forcing data for VIC model are the time series of daily or sub-daily precipitation, maximum and minimum air temperature, and wind speed. In this study, the root-zone soil moisture is estimated as the total column soil moisture (the sum of the three soil layers) as suggested by Shukla et al. (2011).

2.2 Data Assimilation Algorithm

Following Moradkhani (2008), the state-space models that describe the generic earth system are as follows:

$$x_t = f(x_{t-1}, u_t, \theta) + q_t \quad (1)$$

$$y_t = h(x_t) + r_t \quad (2)$$

where $x_t \in \mathbb{R}^n$ is a vector of the uncertain state variables at current time step, $y_t \in \mathbb{R}^m$ is a vector of observation data, u_t is the uncertain forcing data, $\theta \in \mathbb{R}^d$ is the model parameters, $h(\cdot)$ is a non-linear function that relates the states x_t to the observations y_t, q_t represents the model error, and r_t indicates the observation error. The errors q_t and r_t are assumed to be white noise with mean zero and covariance Q_t and R_t, respectively.

2.2.1 Ensemble Kalman Filter

The EnKF is an ensemble version of the Kalman filter, which does not require for a liner model and the estimation of priori model covariance (Evensen, 1994). In the EnKF, an ensemble of state vectors is propagated forward in time. Each time an observation becomes available, the state vector of each ensemble member is updated as follows:

$$x_t^{i+} = x_t^{i-} + K\left(y_t - Hx_t^{i-} + r_t\right) \tag{3}$$

where x_t^{i+} is the posterior state vector for the i^{th} ensemble member, x_t^{i-} is the priori state vector, H is the linearized observation operator ($H = \partial h/\partial x$) to translate from model space to observation space, and K is the Kalman gain factor which is calculated as:

$$K = P^{-}H^{T}(HP^{-}H^{T} + R_t)^{-1} = C_{xy}\left(C_{yy} + R_t\right)^{-1} \tag{4}$$

where P^{-} is the model state error covariance, $P^{-}H^{T} = C_{xy}$ is the covariance of the state ensembles with the predicted observations, and $HP^{-}H^{T} = C_{yy}$ is the variance of the predicted observations.

2.2.2 Particle Filter Markov Chain Monte Carlo

Under the assumption of independence in time series, the posterior distribution of the state variables x_t given a realization of the observations $y_{1:t}$ is as follows:

$$\begin{aligned} p(x_t|y_{1:t}) = p(x_t|y_{1:t-1}, y_t\,) &= \frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{p(y_t|y_{1:t-1})} \\ &= \frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{\int p(y_t|x_t)p(x_t|y_{1:t-1})dx_t} \end{aligned} \tag{5}$$

$$\begin{aligned} p(x_t|y_{1:t-1}) = \int p(x_t, x_{t-1}|y_{1:t-1})dx_{t-1} = \\ \int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1} \end{aligned} \tag{6}$$

where $p(y_t|x_t)$ is the likelihood, $p(x_t|y_{1:t-1})$ is the prior distribution, and $p(y_t|y_{1:t-1})$ is the normalization factor. The marginal likelihood function $p(y_{1:t})$ can be computed as:

$$p(y_{1:t}) = p(y_1)\prod_{k=2}^{t} p(y_k|y_{1:k-1}) \tag{7}$$

where the normalization factor $p(y_t|y_{1:t-1})$ is:

$$p(y_t|y_{1:t-1}) = \int p(y_t, x_t|y_{1:t-1})dx_t = \int p(y_t|x_t)\, p(x_t|y_{1:t-1})dx_t \tag{8}$$

Equation (5) shows mathematically that a posterior conditional probability distribution of model predicted states x_t, given all previous observations $y_{1:t-1}$ and the current observation y_t can be computed sequentially in time. In practice, equation (5) does not have an analytic solution except for few special cases. Instead, the posterior distribution $p(x_t|y_{1:t})$ is usually approximated using a set of Monte Carlo (MC) random samples as:

$$p(x_t|y_{1:t}) \approx \sum_{i=1}^{N} w_t^{i+} \delta\left(x_t - x_t^i\right) \tag{9}$$

where w_t^{i+} is the posterior weight of the i^{th} particle, δ is the Dirac delta function, and N is the ensemble size. The normalized weights are calculated as follows:

$$w_t^{i+} = w_t^{i-} \frac{p\left(y_t|x_t^i\right)}{\sum_{i=1}^{N} w_t^{i-} p\left(y_t|x_t^i\right)} \tag{10}$$

where w_t^{i-} is the prior particle weights, and $p\left(y_t|x_t^i\right)$ can be computed from the likelihood $L\left(y_t|x_t^i\right)$. Generally, a Gaussian distribution is used to estimate $L\left(y_t|x_t^i\right)$:

$$L\left(y_t|x_t^i\right) = \frac{1}{\sqrt{(2\pi)^m |R_t|}} exp\left[-\frac{1}{2}\left(y_t - h\left(x_t^i\right)\right)^T {R_t}^{-1} \left(y_t - h\left(x_t^i\right)\right)\right] \tag{11}$$

A resampling operation is necessary to minimize the weight degeneration, where all but few of the importance weights are close to zero. Moradkhani et al. (2005) suggests resampling the particles with a probability greater than the uniform probability. After resampling, all the particle weights are set equal to $1/N$.

Regardless of the resampling approach, the successful application of PF still requires a large ensemble size to avoid the potential weight degeneration. For large-scale drought applications, however, a relatively small ensemble size (50 to maybe 500) can be affordable given the capabilities of the user's HPC resource and study domain. To further reduce the weight degeneration problem in large-scale applications, the particle filter with sampling importance resampling (PF-SIR) algorithm can be combined with Markov chain Monte Carlo (MCMC) (Moradkhani et al., 2012). The particle Markov chain Monte Carlo (PMCMC) was first proposed in statistical literature by Andrieu et al. (2010). Moradkhani et al. (2012) re-designed the PMCMC to address both states and parameters while integrating the variable variance multiplier approach for a more objective perturbation of observation during the assimilation process (Leisenring and Moradkhani, 2012). In comparison to PF-SIR, the PMCMC achieved higher accuracy in hydrologic predictions and required a smaller ensemble size (Moradkhani et al., 2012). As such, PMCMC is more robust than PF-SIR as it is less prone to the weight degeneracy and sample impoverishment problems (loss of diversity in ensembles) (Abbaszadeh et al., 2018; Andrieu et al., 2010; Moradkhani et al., 2012), which makes the filter more efficient for computationally-intensive large-scale applications.

In this paper we just consider the use of PMCMC for state updating (Andrieu et al., 2010). The PMCMC is an extension of the PF-SIR and uses the PF-SIR to design efficient proposal distributions for MCMC algorithm. The PMCMC consists of the following three steps: 1) initialization ($j = 0$): run PF-SIR targeting $p(x_t|y_{1:t})$, sample $X_t(0) \sim p(x_t|y_{1:t})$ and let $p(y_{1:t})(0)$ denote the corresponding marginal likelihood estimate, 2) iteration ($j \geq 1$): sample $X_t^* \sim p(x_t|y_{1:t})$ again and let $p(y_{1:t})^*$ denote the corresponding marginal likelihood estimate, and 3) calculation of the acceptance ratio as:

$$min\left\{1, \quad \frac{p(y_{1:t})^*}{p(y_{1:t})(j-1)}\right\} \tag{12}$$

and set $X_t(j) = X_t^*$ and $p(y_{1:t})(j) = p(y_{1:t})^*$; otherwise set $X_t(j) = X_t(j-1)$ and $p(y_{1:t})(j) = p(y_{1:t})(j-1)$.

2.3 Parallelization of Data Assimilation

Data assimilation with dynamical distributed hydrologic models (such as the VIC model) are computationally expensive and often need to be run on high-performance computing (HPC) architectures. Especially for simulations which are performed at a large-scale with a high spatial resolution, code parallelization is an essential requirement to reduce computational times. Fortunately, the natural parallelism in the 1D ensemble DA algorithm can be used to implement parallel programming, since each ensemble member can be simulated independently from the others.

For code parallelization, depending on the HPC architecture, either a shared memory or a distributed memory paradigm can be used. The Open Multi-Processing (OpenMP) application programming interface (API) is an implementation of multithreading in a shared memory environment, which means that the memory is accessible by all threads simultaneously. As an alternative, the Message Passing Interface (MPI) is a *de facto* standardized and portable message-passing system running on a distributed memory system, such as computer clusters. Each processor in the distributed memory system has its own private memory and only the data belonging to a processor itself can be directly accessible. For data belonging to other processors, an explicit communication of data between the processors can be performed via the ethernet networks.

Depending on the HPC architecture, two parallelization strategies for the ensemble DA can be devised: the *model decomposition* and the *domain decomposition* (Nerger and Hiller, 2013). The *model decomposition* schema is to distribute the ensemble members of DA over all available processors, indicating that the study domain grid cell is simulated in sequentially while the ensemble members within each grid cell are simulated in parallel, namely ensemble member simulation is set as a precedent in parallel computing. For instance, if we have a 300 processors cluster and want to run DA on a study domain with 300 grid cells where each grid cell is associated with 300 ensemble members; the *model decomposition* schema first distributes the 300 ensemble simulations associated with the first grid cell over the 300 processors. It starts to simulate the second grid cell only if the simulations of the first grid cell are finished. However, the *model decomposition* schema would lead to a significant amount of data communication between different processors in each state updating step of DA. The frequent message-passing between all the processors in each time step may produce a communication overhead in MPI. Therefore, the model decomposition schema is more suitable for small-scale DA applications using OpenMP in a computer system with shared-memory (Ridler et al., 2014).

As an alternative, the *domain decomposition* implementation gains more popularity in the community due to the avoidance of the huge amount of message-passing. The *domain decomposition* distributes the large-scale study domain over all available processors, indicating that the grid cell is simulated in parallel while the ensemble member within each grid cell is simulated sequentially, that is grid cell simulation is set as a precedent in parallel computing. Use the same example as the above paragraph, the *domain decomposition* schema first distributes the 300 study domain grid cells over the 300 processors and all the grid cells are simulated simultaneously. Since each grid cell is associated with only one processor, the 300 ensemble simulations are simulated sequentially in each processor. As a result, each processor only updates the grid cell states which belong to its own and other grid cell states can be updated concurrently. The huge amount of message-passing at each time step is therefore relaxed in *domain decomposition* implementation. Currently, the *domain decomposition* is a standard parallelization strategy in the large-scale dynamical models using MPI.

Figure 2 illustrates the *domain decomposition* implementation in the proposed parallel particle filtering framework (PPFF). The PPFF is processed in the following five steps:

1. Initialization of MPI;
2. Initialization of the PMCMC;
3. Model initialization for VIC;
4. For time = start date to end date
   a. parallel simulation of fluxes and states in each gird cell to the next time step;
   b. for each grid cell, sequential simulation of fluxes and states for each ensemble member;
   c. for each grid cell, filtering the fluxes and states by PMCMC;
   d. for each grid cell, updating the initial conditions for the VIC run in next time step;
5. Finishing the VIC simulation and PPFF.

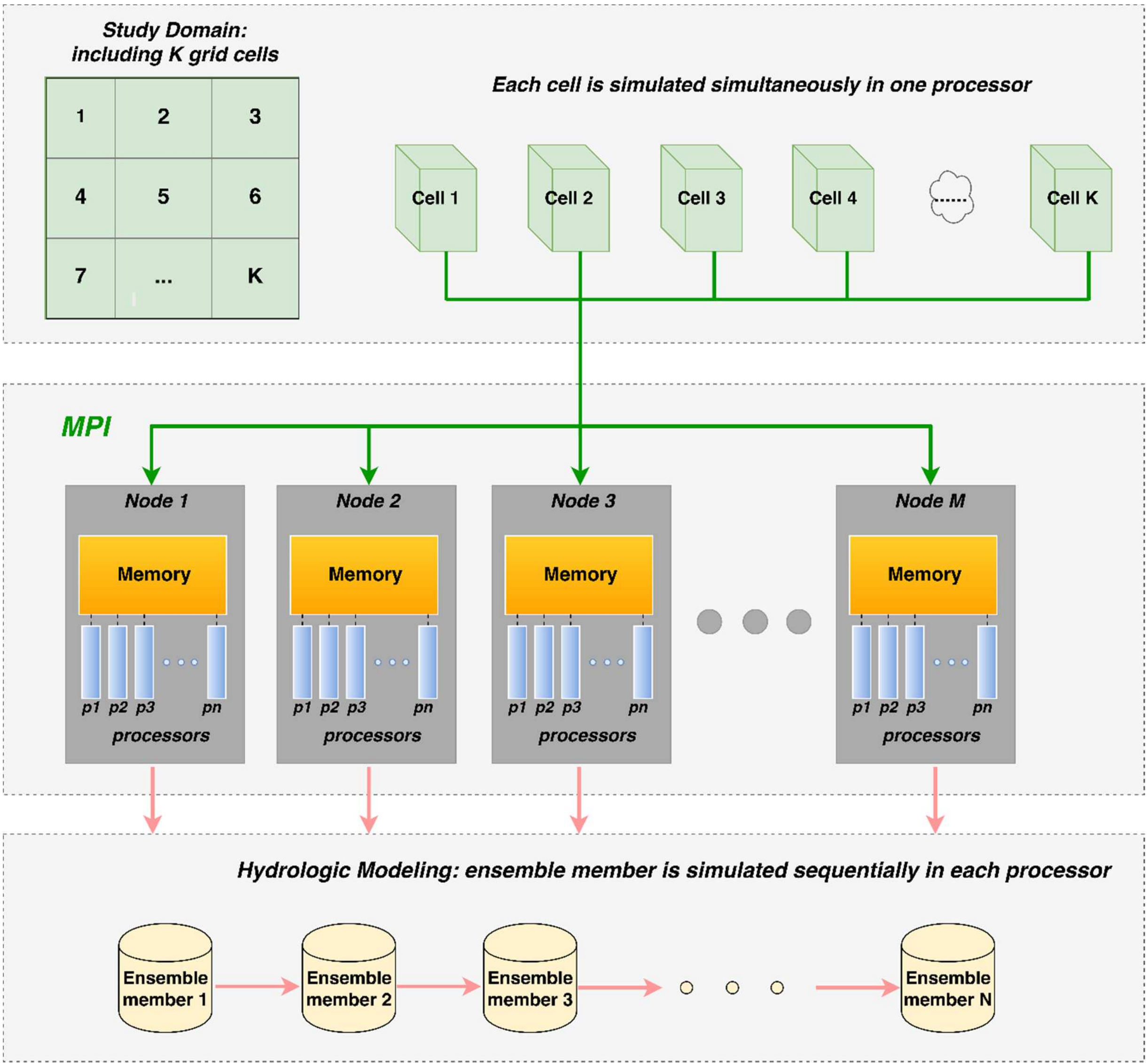

Figure 2. The flow diagram of the parallel particle filtering framework (PPFF) using Message Passing Interface (MPI). The *domain decomposition* parallel strategy is used in the PPFF: each grid cell is simulated in parallel while each ensemble member is simulated sequentially.

3 Case Study

3.1 Study Area and Drought Event

The present study aims to implement the drought monitoring framework over the Contiguous United States (CONUS). The 2012 summertime (May–August) flash drought over the Central U.S. is selected as the case drought event in this study. The 2012 summer drought event developed rapidly in May and reached peak intensity in August. The two main causes for this drought event were the precipitation deficits and high temperatures (Hoerling et al., 2013). Concurrence of substantial precipitation deficits and high temperatures resulted in decreased soil moisture and propagation of drought from meteorological drought to agricultural drought. According to the USDM, during May–August, over three-quarters of CONUS experienced at least abnormally dry drought conditions and the Central U.S. experienced severe to exceptional drought conditions.

This drought event garnered the attention of stakeholders, water managers, and policy makers due to two reasons. First, it is the most severe seasonal drought event in 117 years with significant economic loss (about $12 billion) and impacts on food security and commodity prices (PaiMazumder and Done, 2016). Using Arkansas state as an example, until 13 July 2012, the U.S. Department of Agriculture (USDA) had issued a drought disaster declaration for 69 out of 75 Arkansas counties. Second, the current operational drought monitoring and forecasting systems failed or underestimated the 2012 summer drought event. The National Oceanic and Atmospheric Administration (NOAA) CPC's Seasonal Drought Outlook (SDO) issued on 17 May 2012 failed to forecast this event (Hoerling et al., 2014), and the USDM also did not capture this event until late June 2012 (Mo and Lettenmaier, 2015).

In addition to the severe 2012 summer drought case, we also examine the performance of the drought monitoring system on a non-drought year of 2010 to check whether the land data assimilation system corrects false positive drought. According to the USDM, the percentage area of CONUS experiencing moderate to exceptional drought were only between 8% and 10% through the 2010 summer. This summer minimum is the smallest percent area in moderate to exceptional drought for the CONUS during the 10-year history of the USDM since 2000.

3.2 *Meteorological Forcing Data*

The precipitation, maximum and minimum temperature, and wind speed data (January 1, 1979 to present) were acquired from the Phase 2 of the North American Land Data Assimilation System (NLDAS-2) (Xia et al., 2012). The NLDAS-2 was developed upon the first phase of NLDAS (NLDAS-1) project, which was initiated to generate reliable initial land surface states to coupled atmosphere-land models and improve weather predictions. The majority of NLDAS atmospheric forcing data is derived from the North American Regional Reanalysis (NARR) which features a 32-km spatial resolution and a 3-hour temporal resolution. Besides the precipitation, temperature, and wind speed, the NARR forcing data also includes the specific humidity, surface pressure, incoming solar radiation, and incoming longwave radiation. The NLDAS software is used to interpolate the coarse-resolution NARR data to the finer-scale 1/8° NLDAS grid and to the one-hour NLDAS temporal resolution (Xia et al., 2012). In this study, the hourly NLDAS-2 primary forcing data were aggregated into daily time step for DA applications.

3.3 *Remotely Sensed Soil Moisture*

The blended microwave soil moisture climate change initiative (CCI) products v02.2 released on February 2016 are used in this study (Dorigo et al., 2017; Y. Y. Liu et al., 2011). The CCI soil moisture are merged from four passive and two active microwave products, including the Scanning Multichannel Microwave Radiometer (SMMR), Special Sensor Microwave Imager (SSM/I), Tropical Rainfall Measure Mission (TRMM) Microwave Imager (TMI), Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E), Advanced Microwave

Instrument (AMI), and Advanced Scatterometer (ASCAT). It is noted that more updated CCI products exist at present, which are expected to provide better performances.

According to Y. Y. Liu et al. (2012), the CCI merged soil moisture products are obtained by rescaling the active and passive retrievals into a common Noah simulated soil moisture climatology in the Global Land Data Assimilation System version 1 dataset (GLDAS-1-Noah), with the use of quantile mapping or cumulative distribution function (CDF) matching approach (Reichle and Koster, 2004). This method consists of three steps: 1) merging the four passive products into one dataset from 1978 to 2014; 2) merging the two active products into one dataset from 1991 to 2014; 3) blending both merged products into one final dataset from 1978 to 2014. In the final blended soil moisture, passive and active merged products are used for sparsely and moderately vegetated regions, separately. For transition areas where passive and active products show similar performances, the average of both products are taken. In this way, the blended soil moisture can take advantage of active and passive microwave sensors, as active products have higher accuracy over moderately vegetated regions while passive produces show better performance over sparsely vegetated regions (Albergel et al., 2009; Wagner et al., 2013).

The combined CCI soil moisture products are provided at a daily time step with a spatial resolution of 0.25°. Quality flags of both soil moisture have been used to mask pixels affected by snow cover, temperatures below 0 °C, dense vegetation, and pixels where the soil moisture retrieval failed (Dorigo et al., 2015). It is noted that the CCI soil moisture products are selected in this study due to data availability, since this study focuses on the drought event in 2012. In addition, the blended CCI product can have more number of observations (higher temporal resolution) than any single sensor product leading to a better performance of DA application (Yan et al., 2017). The latest Soil Moisture Active Passive (SMAP) L3 radiometer L-band soil moisture products (Das et al., 2011; Entekhabi et al., 2010) issued from 31 March 2015 will be used for future operational drought monitoring.

3.4 Drought Monitoring Validation

The proposed PPFF LDAS is mainly for agricultural drought monitoring applications. In this study, the simulated drought event is characterized with the root-zone soil moisture percentile. We focus on soil moisture variable because agriculture heavily depends on soil moisture reserves during crop growths. According to the USDM, five drought intensity categories are defined based on the soil moisture magnitude:

D0—abnormally dry, if the soil moisture percentile ≤ 30%;

D1—moderate drought, if the soil moisture percentile ≤ 20%;

D2—severe drought, if the soil moisture percentile ≤ 10%;

D3—extreme drought, if the soil moisture percentile ≤ 5%; and

D4—exceptional drought, if the soil moisture percentile ≤ 2%.

One of the main challenges for drought studies is the lack of an objective benchmark for validations, because no "true" drought data exists in real case studies. In this study, the USDM, the USDA's disaster declaration, and the drought economic loss are all used as references to assess the simulated drought monitoring skills. According to the USDM, during May–August 2012, over three-quarters of the CONUS experienced at least D0 drought conditions and the Central U.S. experienced D2–D4 drought conditions. The summer drought intensity in Central U.S can be classified as D3–D4 as it resulted in major curtailment of crop yields, and caused nearly $12 billion economic damage (Hoerling et al., 2014). Further, the USDA's drought disaster declarations are used to quantify the D3–D4 drought extent. USDA declared drought disaster in seven Central U.S. states in July and August 2012, including Nebraska, Iowa, Kansas, Missouri, Oklahoma, Arkansas, and Illinois. Table 1 presents the percentages of counties in each state under USDA's drought disaster declarations between July and August 2012. In Table 1, it is noted that in July 2012, over 70% of counties declared drought disaster for four out of the seven Central U.S. states (Arkansas, Kansas, Missouri, and Oklahoma). In August 2012, more than 80% of the counties declared drought disaster for the remaining three states (Iowa, Illinois, and Nebraska).

Table 1. The percentage of counties under drought disaster declarations by U.S. Department of Agriculture in July and August 2012 for the seven states in Central U.S.

States	Percentage of Counties under Drought Declarations issued by USDA		
	July	August	Total
Arkansas	92.0	8.0	100
Iowa	0	84.8	84.8
Illinois	19.6	80.4	100
Kansas	86.7	13.3	100
Missouri	100	0	100
Nebraska	15.1	84.9	100
Oklahoma	72.7	27.3	100

3.5 Synthetic Study

To objectively assess the potential benefit of assimilating satellite surface soil moisture into a hydrologic model, a synthetic study is first conducted through observing system simulation experiment (OSSE) (Moradkhani, 2008). The synthetic study includes the following four steps: 1) a "truth" run of hydrologic model with the pre-calibrated model parameters; 2) simulated satellite surface soil moisture observations, which are generated from the truth run by incorporating the observation errors; 3) an open loop (OL) run with the perturbed forcing data without DA; and 4) the DA step that assimilates the simulated surface soil moisture observations from step 2 to the model. Then the OL and DA results are compared against the truth simulation to evaluate the impact of satellite surface soil moisture assimilation. Figure 3 presents the flowchart of the synthetic study in this study using VIC model.

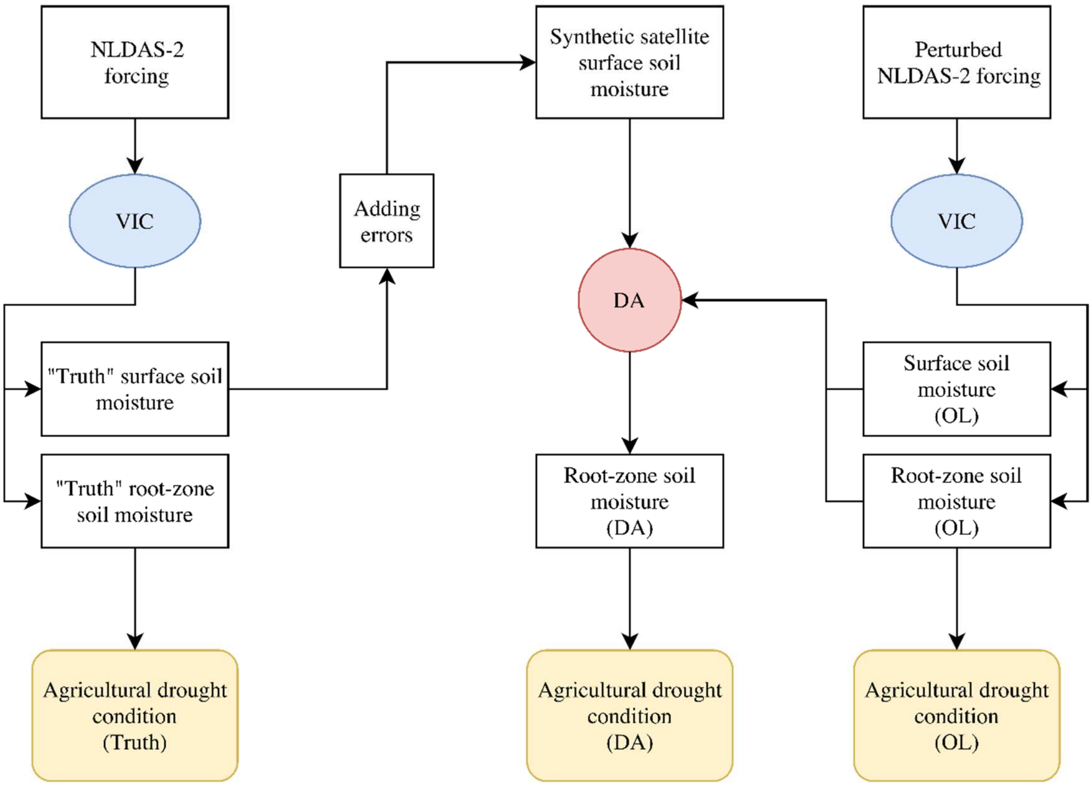

Figure 3. The flowchart of the synthetic study using VIC model to assess the potential benefit of assimilation of satellite surface soil moisture on drought monitoring.

3.5.1 Synthetic Truth Simulations

In this study, the VIC model is performed to simulate the soil moisture and reconstruct the 2012 drought conditions over the CONUS domain, at a 1/8° spatial resolution. The model parameter files, including the elevations, soil properties, and vegetation cover, are acquired from the VIC retrospective land surface dataset (Maurer et al., 2002). For synthetic truth simulation, the VIC model is run in a daily time step using the NLDAS-2 forcing data from 1 January 1979 to 31 December 2015 to produce long-term girded surface and root-zone soil moisture. In this study, the VIC model is run in water balance mode, which means that the surface temperature is set equal to the surface air temperature rather than iteratively solving the surface energy budget.

3.5.2 Error Models in Data Assimilation

To account for uncertainties in forcing data due to sensor errors and spatial heterogeneity, both precipitation and wind speed errors are assumed to be heteroscedastic and lognormal with a variance of 25% of the variable's magnitude. Both maximum and minimum temperature are assumed to be homoscedastic and normal with a standard deviation of 3 °C. For both synthetic and the real case studies, the same perturbation errors are used, and the initial state noise is free. The model structure is considered perfect for synthetic study; while for real case study, the model structural error is assumed to be normally distributed with a standard deviation equal to 10% of the prediction value. All errors in this study are assumed to be uncorrelated. Error assumptions are applied with the same magnitude in both the EnKF and PMCMC applications. The form and magnitude of these errors are based on previous DA studies, where the values were determined through a manual tuning to achieve reliable predictions (Abbaszadeh et al., 2018; DeChant and Moradkhani, 2014; Yan et al., 2015; Yan and Moradkhani, 2016). We acknowledge that these values may not necessarily be optimum.

Following Kumar et al. (2014b), the white noise (standard deviation) for the CCI satellite soil moisture is assumed to be 0.04 m^3/m^3 over all grid cells over the CONUS. To account for spatial correlations in CCI soil moisture errors, the constant CCI soil moisture error standard deviation (0.04 m^3/m^3) is scaled by the ratio of the soil moisture time series standard deviation of the VIC model to that of the real CCI soil moisture data (separately for each grid cell), resulting in CCI soil moisture error standard deviations that are spatially distributed (Kumar et al., 2014b; Q. Liu et al., 2011; Reichle et al., 2007). The synthetic CCI soil moisture are then generated from synthetic truth simulations by incorporating the scaled CCI soil moisture errors. We acknowledge that the constant soil moisture error (0.04 m^3/m^3) assumption before scaling is somewhat arbitrary. More advanced approaches, such as the triple collocation (TC) method, may be used to provide better estimations of satellite soil moisture error structure (Gruber et al., 2016). However, successful applications of TC require large numbers of coincident soil moisture from three independent time series (satellite soil moisture, model simulations, and *in-situ* observations), homogeneity of their linear relationships, and error structures. In general, these assumptions are difficult to realize in practice because of the infrequent spatiotemporal sampling of satellite and *in-situ* sensors. (Su et al., 2014).

Contrary to the small ensemble size (12–20) used in the majority of previous satellite soil moisture DA studies (Kumar et al., 2014b, 2009; Pan and Wood, 2010; Yin et al., 2015), a large ensemble size of 100 is used in this study, in order to fully quantify the soil moisture posteriors. The PPFF module is written in Python3 script and all the DA simulations are run on the Linux Hydra Cluster (24 nodes, 384 processors) located at the Office of Information Technology (OIT), Portland State University (PSU).

3.5.3 EnKF versus PMCMC

Before assessing the drought monitoring skill, we first compare the performances of the EnKF and PMCMC for root-zone soil moisture predictions. Both experiments are performed by assimilation of the synthetic satellite surface soil moisture for the period of 1 January 2012 to 31 December 2012. Since the main trust of this study is to investigate whether the PMCMC system can improve drought monitoring skill for summer 2012, the EnKF versus PMCMC comparison is based only on one basin as a proof-of-concept. In order to assess the PMCMC performance over the EnKF, the normalized information contribution (NIC) metric (Kumar et al., 2014b) is used in this study. The NIC for root-mean-square-error (RMSE) is defined as follows:

$$NIC = \frac{RMSE_{EnKF} - RMSE_{PMCMC}}{RMSE_{EnKF}} \qquad (13)$$

where $RMSE_{EnKF}$ indicates the RMSE values between EnKF and synthetic truth, and $RMSE_{PMCMC}$ indicates the RMSE values between PMCMC and synthetic truth. If NIC>0, the PMCMC shows improvement with respect to the EnKF predictions; if NIC<0, the PMCMC shows degraded performance as compared with the EnKF predictions; and if NIC=1, the PMCMC results in perfect predictions.

Figure 4a displays the NIC values in the root-zone soil moisture and their spatial distributions across the Columbia River Basin. The majority of the grid cells show positive NIC values indicating a better performance of the PMCMC over EnKF. The daily domain-averaged root-zone soil moisture RMSE (m3/m3) for the EnKF is 0.0039, and it decreases to 0.0019 with PMCMC (about 51% improvement). To better illustrate this difference, Figure 4b shows the time series of synthetic truth, EnKF, and PMCMC root-zone soil moisture predictions for one grid cell with NIC=0.61 (marked as yellow star in Figure 4a). It is observed that though EnKF and PMCMC show similar performances in winter and spring seasons, the EnKF underestimates the soil moisture during summer while the PMCMC predictions are closer to the synthetic truths. As mentioned in the introduction section, this can be explained as the inherent drawbacks of the EnKF technique. It is mainly due to inadequacy of linear updating rule for a nonlinear hydrologic model, leading to over adjustments in the updates (Harlim and Majda, 2010; DeChant and Moradkhani, 2012). These results come as no surprise and are consistent with previous findings which had demonstrated the superiority of the PF DA techniques over the EnKF in hydrologic applications (DeChant and Moradkhani, 2012, 2011; Leisenring and Moradkhani, 2011).

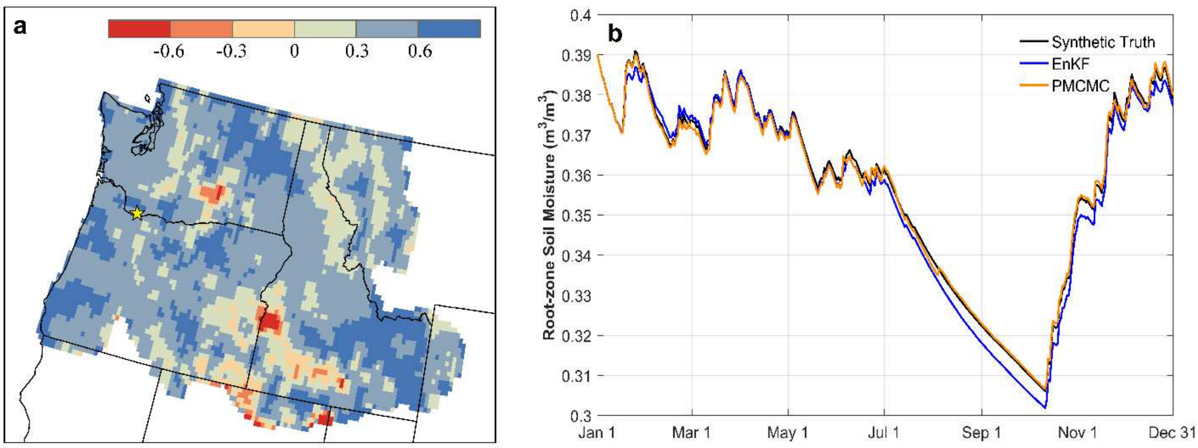

Figure 4. (a) The normalized information contribution (NIC) value between the EnKF and PMCMC root-zone soil moisture predictions (Eq. 13). Positive values indicate that the PMCMC improves soil moisture prediction as compared to the EnKF; negative values indicate the degradation over the EnKF. (b) Time series of the marked (yellow star) grid cell root-zone soil moisture (m3/m3) for the synthetic truth, EnKF, and PMCMC simulation for the period of 1 January 2012 to 31 December 2012. Both NIC values and time series are generated using the posterior means.

3.5.4 PMCMC Drought Monitoring

Here we examine the potential benefit of assimilating satellite surface soil moisture on drought monitoring skill according to the synthetic study flowchart shown in Figure 3. The OL simulation is run with the perturbed NLDAS-2 forcing data for the period of 1 January 2012 to

31 December 2012. The ensemble size and perturbation errors in the OL simulation are the same as in the DA. Figure 5 presents the NIC values (OL versus DA) in the surface and root-zone soil moisture and their spatial distributions across the CONUS for the 2012. The majority of the grid cells show the positive NIC values indicating the added-value of the DA. Generally, the improvements in the surface soil moisture field are consistent with the improvements in the root-zone soil moisture field, with more prominent improvements in the surface soil moisture field. The improvements in surface field are higher than root-zone field, which is mainly caused by the weak and highly non-linear cross covariance between the two layers (Kumar et al., 2014a). For surface soil moisture, the daily domain-averaged RMSE (m3/m3) for the OL is 0.0042, and it decreases to 0.0027 with DA (about 36% decrease). Similarly, the daily domain-averaged root-zone soil moisture RMSE (m3/m3) value decreases from 0.0034 in the OL to 0.0024 after DA (about 29% decrease). It is to be noted that several grid cells in Figure 5, show negative NIC values. From a theoretical perspective, the DA is expected to show improvements over the OL skills for all grid cells in a synthetic study framework knowing that the observation errors are synthetically introduced with predefined statistical properties. Here, the negative NIC values can be explained due to suboptimal choice of the uncertainty magnitude in the synthetic CCI soil moisture observations. Specifically, in the particle weight estimation using the Gaussian likelihood function, if the scaled soil moisture observation error is too small (e.g., 0.01 m3/m3), the PF may result in degenerated particles that is a few particles with larger weights dominate the rest of the particles and eventually collapse to one particle which poorly approximates the soil moisture posteriors and eventually resulting in negative NIC values.

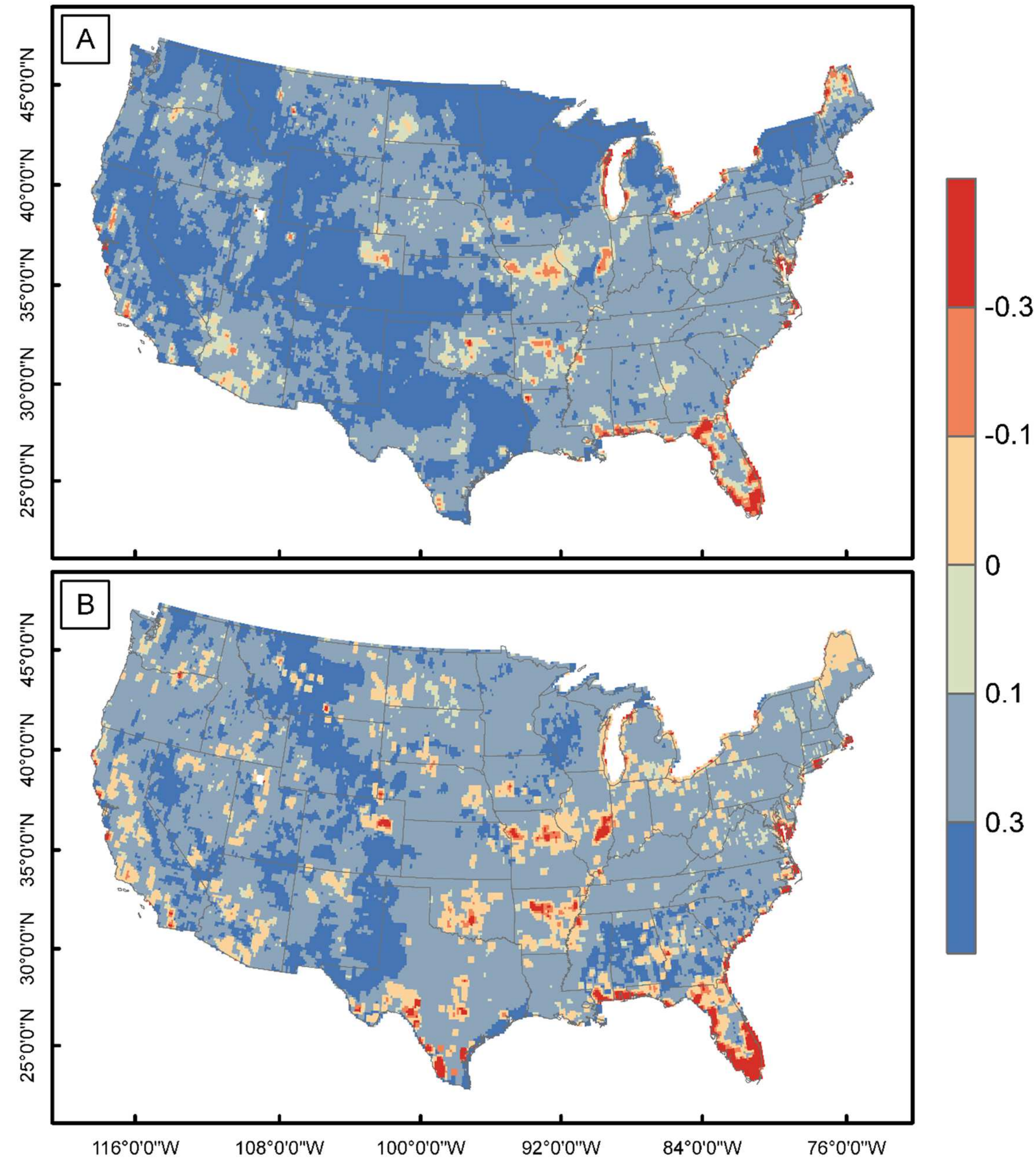

Figure 5. The normalized information contribution (NIC) value between the Open-Loop (OL) and Data Assimilation (DA). The NIC values are generated based on posterior means. The positive value indicates that the DA improves soil moisture prediction against OL; negative value indicates the degradation over the OL. (a): Surface soil moisture field. (b): Root-zone soil moisture filed.

In the implementation of the DA, the identifications of model and observation errors are very unintuitive and the inappropriate selections of observation errors may lead to over/under-confident soil moisture predictions (Y. Liu et al., 2012; Massari et al., 2015). For small-scale DA applications, such as one soil column DA study (Montzka et al., 2011; Yan et al., 2015), it is straightforward to manually tune the observation error to ensure that the synthetic framework performs well. For large-scale DA applications, however, it is not feasible to conclude on a single ideal manner an optimal error implementation, and some degraded cells are unavoidable. Similar results from a synthetic study using the EnKF-based DA showed the inferior performance of DA with respect to OL) were also reported by Kumar et al. (2014a) using the EnKF method.

Leaving the first year as model spin-up period, the soil moisture climatology is the soil moisture simulations from 1 January 1980 to 31 December 2015. Then, the soil moisture percentiles generated from the OL and DA monthly integrations are compared against the corresponding synthetic truth. For instance, to estimate the monthly soil moisture percentile in October 2010, it is first to take all October soil moisture values over 1980–2015 to construct the climatological distribution for each grid cell. Then, for any specific grid cell soil moisture value in October 2010, the corresponding soil moisture percentile can be estimated from the climatological distribution. Drought monitoring skills between the OL and DA are examined using the drought extent bias (%), which is estimated as the absolute difference between the percentage of detected drought area from OL/DA over CONUS and the synthetic true drought area over the CONUS.

Figures 6 and 7 present the spatial distribution of drought intensities and the drought extent bias (%), respectively, for five drought categories (D0–D4) over CONUS. The severe drought events from May–August 2012 across Central U.S. (Nebraska, Kansas, Oklahoma, Iowa, Missouri, Arkansas, and Illinois) can be clearly seen in the synthetic truth. In all comparisons of the four-month drought monitoring, the DA estimates show systematic improvements over the OL estimates. For the May 2012 case, the OL underestimates the intensity of drought across the Nevada, Utah, Colorado, and New Mexico whereas DA improves these representations. The drought extent bias (%) for D0–D4 between the OL and synthetic truth is 5.11%, and it decreases to 1.28% (about 75% decrease) with DA. For the June and July 2012 cases, the OL underestimates the intensity of drought across the Central U.S. (i.e., the Kansas, Nebraska, and Colorado), and DA helps to reduce these large biases. Similarly, drought extent biases (%) decrease from 5.25% and 4.69% in the OL to 1.02% (about 81% decrease) and 0.53% (about 89% decrease) in the DA, respectively. Although several grid cells in Central U.S. (i.e., Kansas) show relatively high soil moisture values in August 2012, the DA still helps reduce these biases. The drought extent bias (%) decrease from 3.02% to 0.34% (about 89% decrease). Overall, these results are consistent with the trends in Figure 5, which shows the improvements obtained by DA.

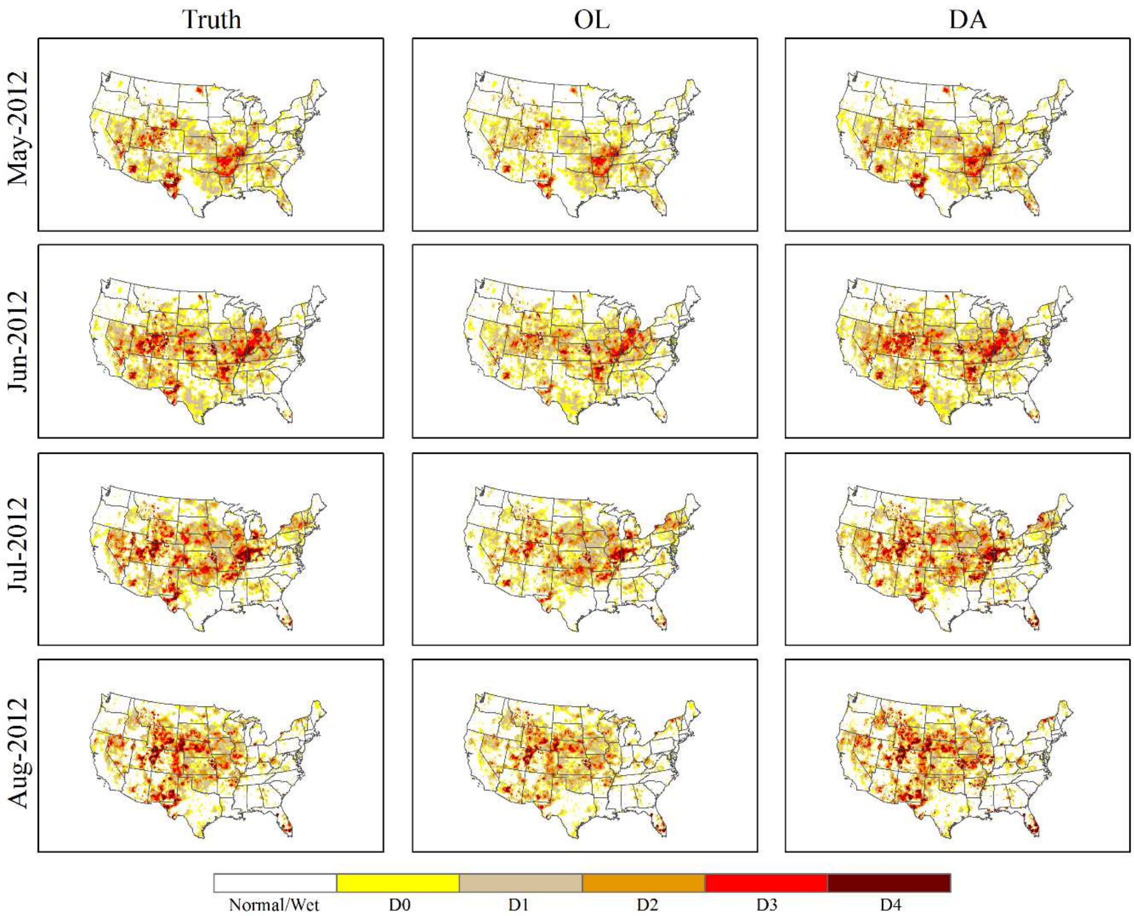

Figure 6. Comparison of the drought monitoring skill between Open-Loop (OL) and Data Assimilation (DA) for May–August 2012.

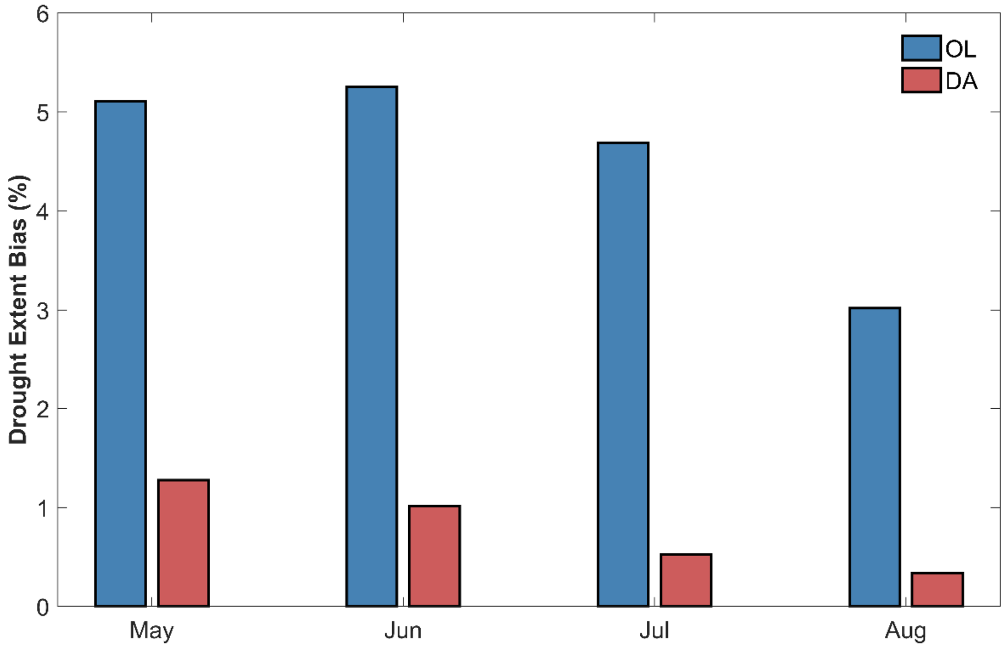

Figure 7. The absolute bias of drought extent (%) against the synthetic truth between Open-Loop (OL) and Data Assimilation (DA) over the CONUS for the period of May–August 2012.

3.6 Real Case Study

When assimilating the real satellite soil moisture data, the systematic biases between the satellite-based and model-based soil moisture cannot be avoided (Reichle and Koster, 2004; Su et al., 2014; Yan et al., 2015; Yilmaz and Crow, 2013). Proper treatment of these systematic biases is important, as the DA algorithm is designed to work with errors that are strictly random. The most common approach, the CDF matching (Reichle and Koster, 2004), is implemented in this study to rescale the satellite observations to the model's climatology. The CDF matching approach can correct all the moments of the distribution regardless of its shape. Leaving out the first year as model spin-up period, the CCI soil moisture products are rescaled from 1980–2014. For each model gird cell over the CONUS, the model CDF and the satellite observation CDF are generated for the entire period of 1 January 1980 to 31 December 2014. The CDF matching approach is then used to rescale the satellite observations to the model's climatology, separately for each grid cell.

For real data study, the perturbation errors are the same as the synthetic study except for the model error. As mentioned in section 3.5.2, the model structure error is assumed to be normally distributed with a standard deviation equal to 10% of the prediction value. The DA is performed by assimilation of the rescaled real satellite surface soil moisture for six-month duration from 1 February to 31 August for both 2010 non-drought and 2012 drought years.

3.6.1 2012 Summer Drought Monitoring

Figure 8 presents the OL and DA monthly drought monitoring results across CONUS in May/June/July/August 2012. Note that the drought monitoring from the DA in these comparisons are generated using the posterior means. For the purpose of comparison of drought intensity, the USDM weekly monitoring result in the middle of each month is also presented in Figure 8. It is noted that although the USDM maps provide a useful monitoring of current drought conditions, they should not be considered as the "truth" like in the synthetic study since they are largely based on subjective information about drought at multiple time scales and for a wide range of impacts (Otkin et al., 2013). In addition, the USDM did not capture the 2012 Central U.S. drought until late June 2012 (Mo and Lettenmaier, 2015; Otkin et al., 2013). Currently, an objective and quantitative measurement of true drought condition is still unconcluded in the community, due to inherent complexity of drought phenomenon. However, the comparisons with USDM are still helpful to assess the capacity of our proposed system to detect drought events at some level (Anderson et al., 2011; Kumar et al., 2016, 2014b; Otkin et al., 2013).

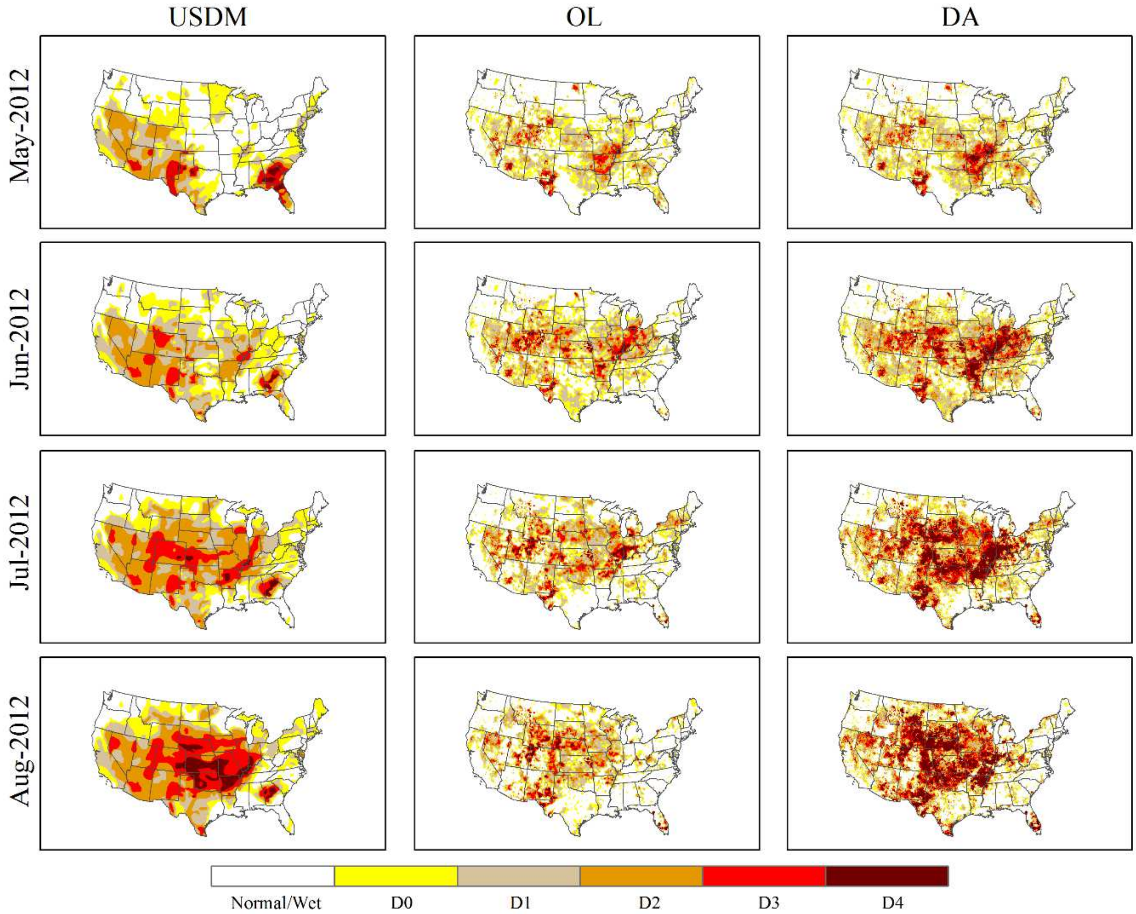

Figure 8. Comparison of the 2012 summer drought intensity over the CONUS from Open-Loop (OL), Data Assimilation (DA), and U.S. Drought Monitoring (USDM).

Because the USDM did not capture the 2012 summertime drought (May–August) until late June (Mo and Lettenmaier, 2015), it is important to investigate whether the proposed DA drought monitoring system can better monitor this drought event, especially in May. It is noted that the OL results in the real case study are the VIC forward model run with the predefined parameters and the unperturbed NLDAS-2 forcing data. These results were treated as the "truth" in the synthetic study. Figure 8 illustrates the added value of assimilating remotely sensed soil moisture for improving drought monitoring skill. For the May 2012 case (the onset of the 2012 summer drought), the USDM completely missed the drought onset in Central U.S. (such as Arkansas, Missouri, Oklahoma, and Kansas). Although the OL shows some improvements over the USDM, the soil moisture DA provides a better estimate of drought severity, especially for D0 and D1 categories (over Nebraska). For the August 2012 case (the 2012 drought event reached peak intensity in August), this time the USDM successfully captures this severe drought events in Central U.S. whereas the OL underestimates the drought intensity and DA provides similar results as the USDM. For the July 2012 case, since all the counties in Missouri and about 73% counties in Oklahoma had issued drought disaster declaration (due to severe damage to the crops) (Table 1), it is reasonable to conclude that the USDM underestimate the intensity in Missouri and Oklahoma. The USDM only provides D1–D2 drought in these two states, while DA suggests D3–D4 drought, which is more consistent with the USDA drought declaration. Similar pattern can be found for June 2012 case, since the crops were damaged in this month, the USDM and OL underestimate the drought intensity and DA provides a better estimate of drought severity. Overall, in all four months, DA estimation predicts more intense drought over Central U.S. and captures the spatial pattern of the intense drought very well relative to the OL and USDM.

Figure 9 summarizes the detected drought areas over CONUS based on OL, DA, and USDM. The USDM results are presented in Figure 9 only as a “reference” not “truth”. For the May 2012 case, the USDM missed the drought events in parts of Central U.S., however, the DA helps to correct these biases. The detected drought areas over CONUS for OL and DA are 43.82% and 47.91%, respectively. For the August 2012 case, the OL underestimates the drought intensity whereas DA improves these representations. The detected drought areas increase from 48.17% in the OL to 62.66% after the DA, respectively. Similarly, the DA adds the drought monitoring skills for the June and July 2012 cases, from 55.27% and 54.45% by OL to 62.49% and 65.73% by DA. As a result, compared with OL, the DA provides a more accurate estimate of drought areas, more consistent with the USDM map of this drought event.

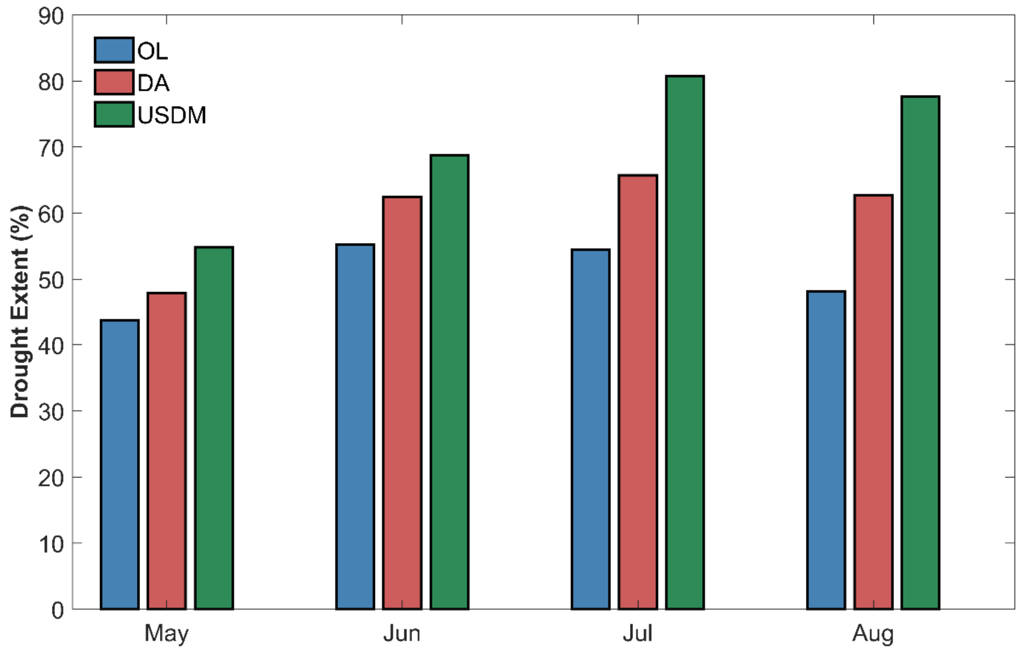

Figure 9. Comparison of the drought extent (%) over the CONUS from Open-Loop (OL), Data Assimilation (DA), and U.S. Drought Monitoring (USDM) for May/June/July/August 2012.

After the demonstration of the added-value of DA over OL in Figure 9, Figure 10 examines the drought monitoring skill between DA and USDM in details over Central U.S. Figure 10 shows the percentage area under drought over Central U.S. from the DA and USDM for each drought category. The monthly USDM drought extent is estimated as the average of the four weekly USDM products. It is noted that the drought extents presented in Figure 10 are comprised 7-State region of Nebraska, Iowa, Kansas, Missouri, Oklahoma, Arkansas, and Illinois. In the May 2012 case, the USDM missed the drought onset and detected less than half

(44.39%) of the Central U.S. under drought (D0–D4). While DA improves the detection of drought onset and warns that 78.52% of the Central U.S. is abnormally dry (D0–D4) and more than half (59.50%) of the Central U.S. is under moderate drought or worse (D1–D4). Although the USDM started monitoring the severe (D2) and extreme (D3) droughts since 26 June 2012, the USDM still underestimated the intensity of drought in June and July 2012. Most of the states in Central U.S. issued the drought disaster declaration in July, which indicated that the crop plants had been damaged in June (Table 1). For the June 2012 case, only 3.33% of the Central U.S. is under extreme (D3) and exceptional (D4) drought according to the USDM, while DA detects 36% of the Central U.S. under extreme (D3) and exceptional (D4) droughts. Same result can also be found for July 2012 case where DA provides a better estimate of drought severity showing 58.77% of the Central U.S. under D3–D4 drought, compared to 38.88% with the USDM. In the August 2012 case, nearly all the counties in Central U.S. issued drought disaster declaration (Table 1) and USDM provided a close exceptional drought (D4) monitoring skill as DA.

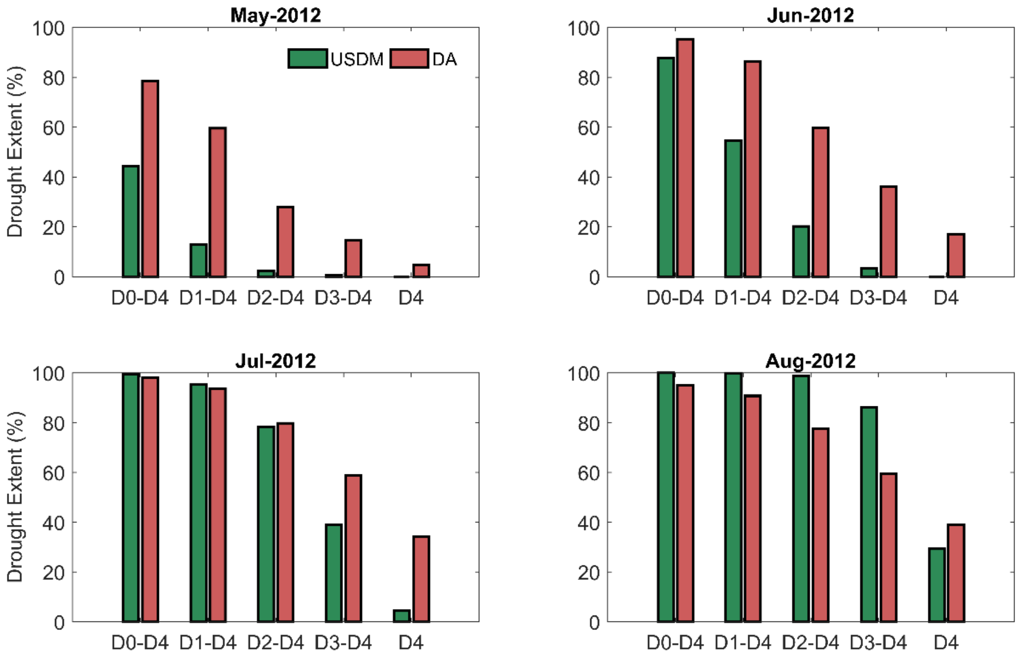

Figure 10. Comparison of U.S. Drought Monitoring (USDM) and Data Assimilation (DA) drought monitoring extent for five different drought categories (D0–D4) over the Central U.S. The drought extent is comprised of the 7-State region of Nebraska, Iowa, Kansas, Missouri, Oklahoma, Arkansas, and Illinois.

In summary, compared with the OL, the DA systematically improves the drought monitoring skill for 2012 drought event from May to August. Compared with the USDM, the DA could better capture the drought onset in May, and the drought intensity in June and July, and provide similar results in August. These results demonstrate the added-value of DA to facilitate the state drought preparation and effective response actions.

3.6.2 2010 Summer Drought Monitoring

Figure 11 presents the OL and DA monthly drought monitoring results derived based on the root-zone soil moisture posterior means across the CONUS in May/June/July/August 2010. Similar to the 2012 drought results, the USDM weekly monitoring result in the middle of each month is presented in the figure as a reference. The OL/DA monitoring results and USDM show a general agreement with each other except for the southwest US. For the June 2010 case, the USDM shows D0–D1 drought intensity in Arizona and New Mexico, while both OL and DA results suggest D3–D4 intensities. As the USDM data is a blended product of multiple indicators and data, we do not expect the modeled drought intensities derived based on a single variable (soil moisture) to match the USDM results exactly. Though the DA does not agree with the USDM drought intensity, in the southwestern US, it does not over-predict the drought of summer 2010. For the May 2010 case, the percentages of D0–D4 drought extent over the CONUS are 25% and 24% from DA and OL, respectively. This result suggests that for a non-drought year event like 2010, the DA performs similar to OL and do not increase false drought detections.

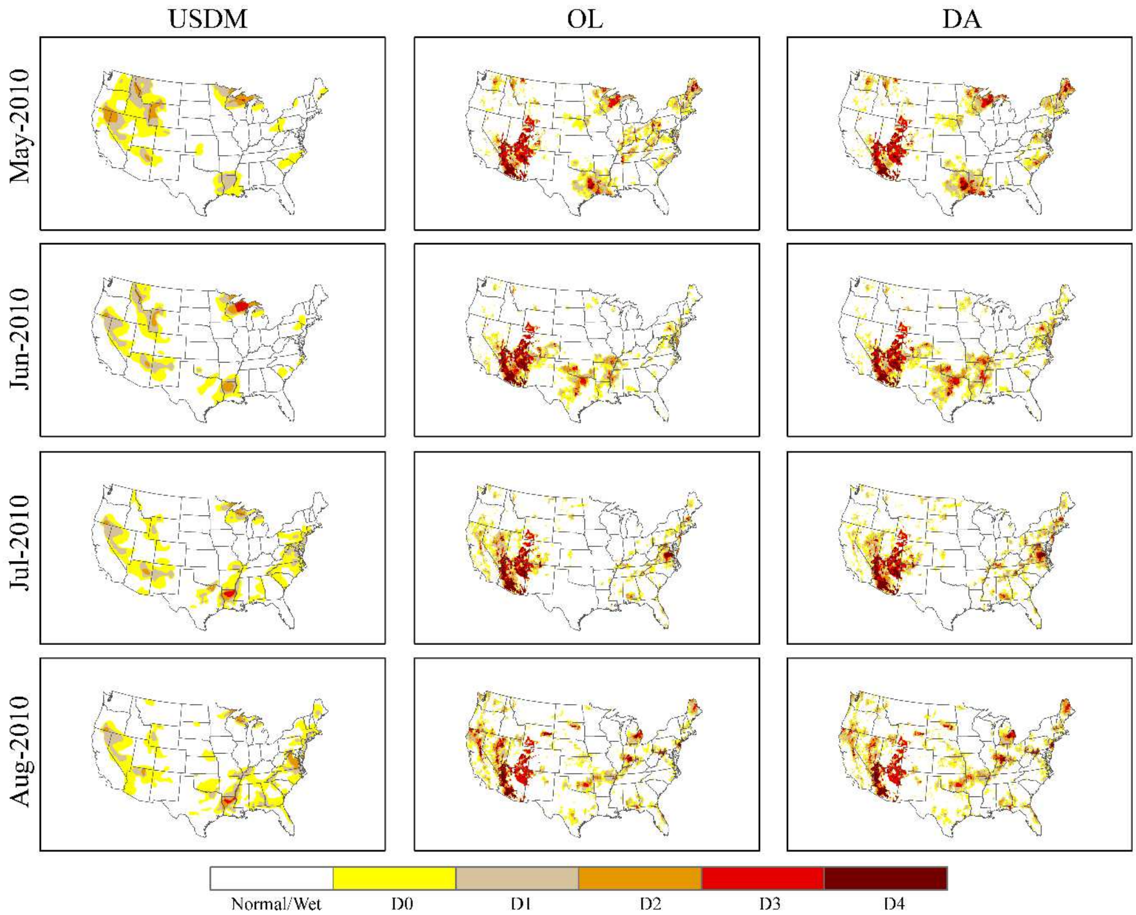

Figure 11. Comparison of the 2010 summer drought intensity over the CONUS from Open-Loop (OL), Data Assimilation (DA), and U.S. Drought Monitoring (USDM).

4 Conclusion

In this study, a land data assimilation system is proposed through the assimilation of remotely sensed soil moisture into a distributed dynamical hydrologic model to improve the drought monitoring skill. The recent developed PMCMC technique is implemented in this study to quantify the soil moisture posteriors. In order to cope with the large computational demand required by PMCMC, a modular parallel particle filtering framework is developed in this study which allows the use of large ensemble size in the PMCMC applications. The proposed drought monitoring system is assessed based on the 2012 summer flash drought and 2010 non-drought events over the Contiguous United States. The impacts of assimilating remotely sensed surface soil moisture on improving soil moisture predictions and drought monitoring are examined. Results from both synthetic and real case studies suggest that the proposed drought monitoring system improves the drought monitoring skill and can facilitate the state drought preparation and declaration. Compared with the state-of-the-art U.S. Drought Monitoring, the proposed system can better capture the drought onset in May, and the drought intensity in June and July 2012. For the 2010 non-drought year case study, the proposed system does not lead to over-detection of false positive drought.

A limitation of this study is the short data assimilation periods (two years) due to the limited HPC computational resources. As a result, only two case studies (a drought year and a non-drought year) are presented in this paper to assess the efficiency of this monitoring system. We acknowledge that further examination of the proposed approach with more case studies is desired. A practical path forward that we are currently implementing is to make this system operational monthly scale and compare the results with the U.S. Drought Monitor and other resources from related agencies.

Acknowledgement

Partial financial support for this project was provided by the National Oceanic and Atmospheric Administration (NOAA) Modeling, Analysis, Predictions, and Projections (MAPP) (Grant No. NA140AR4310234).

Appendix A. Coupling Data Assimilation with Hydrologic Model

Generally, there are two strategies to couple the hydrologic model and DA algorithm: *online coupling* and *offline coupling* (Kurtz et al., 2016; Nerger and Hiller, 2013). For *online coupling*, the DA algorithm is a subroutine of the dynamical model routines and they are all compiled into a single program. Data is exchanged via the main memory. The main advantage of the *online coupling* is that it is computationally more efficient since the exchange of data using files can be avoided. In addition, the model only needs to execute the start-up phase once and there is no additional start-up cost. However, the *online coupling* schema requires to modify the model source code to make it compatible with the DA subroutine.

For *offline coupling* schema, there are two separate programs used for model execution and DA, respectively. Data is then exchanged via the input/output (I/O) files of the model. Obviously, the *offline coupling* is more *ad hoc* since there is no need to modify the model source code. In addition, the *offline coupling* is the only option when the source code of the model is not available or not open source. One limitation of the *offline coupling* is to generate a large amount of files on the local drive and produce a lot of I/O overhead. One possible solution is to use the RAM disk to catch these I/O files.

In general, an ideal parallel DA framework should provide a generic DA environment that can be coupled with any hydrologic model. Therefore, the *offline coupling* schema is used in this study to provide a generic parallel DA framework for the VIC model. Figure A1 describes the detailed PMCMC-VIC *offline coupling* interface in this study. In Figure A1, a one-day VIC model simulation and PMCMC filtering is presented in details.

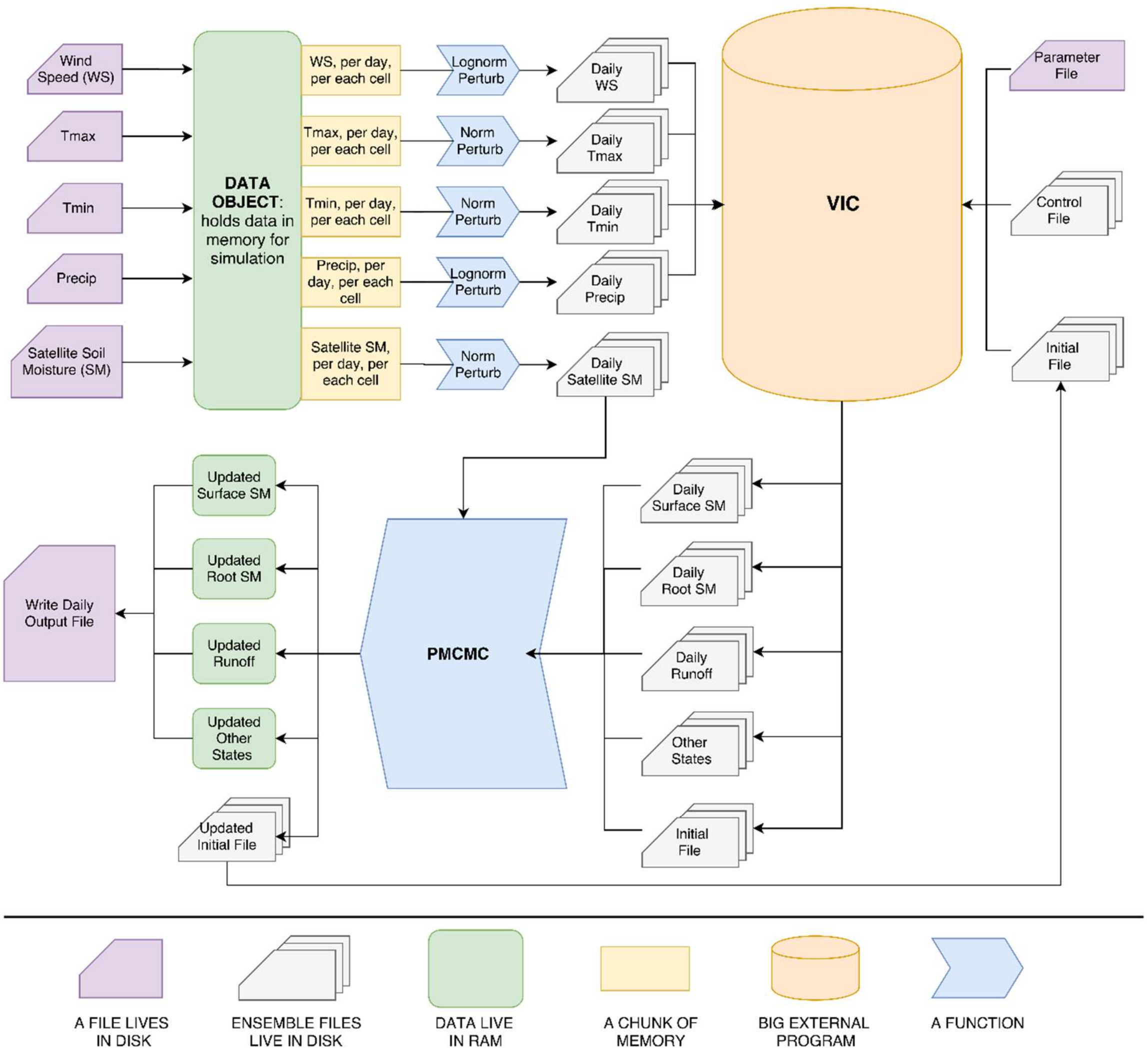

Figure A1. The flowchart of the *offline coupling* interface of the VIC model and the PMCMC data assimilation algorithm. The ensemble files indicate the ensemble members of the PMCMC algorithm. A one-day VIC model simulation and PMCMC updating are shown in this flowchart.

From Figure A1, the model grid, parameters, and initial conditions are first prepared in the model initialization phase. The forcing data (precipitation, maximum and minimum temperature, and wind speed) and remotely sensed observations for all the time steps are held in the main memory as one data object waiting for simulation. Subsequently, the PMCMC module initializes the ensemble members and the DA integration is conducted step by step. The remotely sensed observations and ensemble forcing data are perturbed with the *Normal Perturb* and *Lognormal Perturb* function routines, respectively. After the VIC model simulation for each ensemble member, the PMCMC module is called again to filter the specified model output (i.e., soil moisture) and also the initial condition files. At this point, the DA integration is finished for one-time step and the users can choose whether or not to output the ensemble members of required variables into the local drive, depending on the available main memory.

Reference

Abbaszadeh, P., Moradkhani, H., Yan, H., 2018. Enhancing hydrologic data assimilation by evolutionary Particle Filter and Markov Chain Monte Carlo. Adv. Water Resour. 111, 192–204. doi:10.1016/j.advwatres.2017.11.011

AghaKouchak, A., 2014. A baseline probabilistic drought forecasting framework using standardized soil moisture index: application to the 2012 United States drought. Hydrol. Earth Syst. Sci. 18, 2485–2492. doi:10.5194/hess-18-2485-2014

Ahmadalipour, A., Moradkhani, H., Yan, H., Zarekarizi, M., 2017. Remote Sensing of Drought: Vegetation, Soil Moisture, and Data Assimilation, in: Lakshmi, V. (Ed.), Remote Sensing of Hydrological Extremes. Springer International Publishing Switzerland. DOI: 10.1007/978-3-319-43744-6_7, pp. 121–149. doi:10.1007/978-3-319-43744-6_7

Albergel, C., Rüdiger, C., Carrer, D., Calvet, J.-C., Fritz, N., Naeimi, V., Bartalis, Z., Hasenauer, S., 2009. An evaluation of ASCAT surface soil moisture products with in-situ observations in Southwestern France. Hydrol. Earth Syst. Sci. 13, 115–124. doi:10.5194/hess-13-115-2009

Anderson, M.C., Hain, C., Wardlow, B., Pimstein, A., Mecikalski, J.R., Kustas, W.P., 2011.

Evaluation of Drought Indices Based on Thermal Remote Sensing of Evapotranspiration over the Continental United States. J. Clim. 24, 2025–2044. doi:10.1175/2010JCLI3812.1

Andreadis, K.M., Clark, E.A., Wood, A.W., Hamlet, A.F., Lettenmaier, D.P., 2005. Twentieth-century drought in the conterminous United States. J. Hydrometeorol. 6, 985–1001. doi:10.1175/JHM450.1

Andreadis, K.M., Lettenmaier, D.P., 2006. Trends in 20th century drought over the continental United States. Geophys. Res. Lett. 33. doi:10.1029/2006GL025711

Andrieu, C., Doucet, A., Holenstein, R., 2010. Particle Markov chain Monte Carlo methods. J. R. Stat. Soc. Ser. B (Statistical Methodol. 72, 269–342. doi:10.1111/j.1467-9868.2009.00736.x

Brocca, L., Moramarco, T., Melone, F., Wagner, W., Hasenauer, S., Hahn, S., 2012. Assimilation of surface- and root-zone ASCAT soil moisture products into rainfall-runoff modeling. IEEE Trans. Geosci. Remote Sens. 50, 2542–2555. doi:10.1109/TGRS.2011.2177468

Chaney, N.W., Herman, J.D., Reed, P.M., Wood, E.F., 2015. Flood and drought hydrologic monitoring: the role of model parameter uncertainty. Hydrol. Earth Syst. Sci. 19, 3239–3251. doi:10.5194/hess-19-3239-2015

Das, N.N., Entekhabi, D., Njoku, E.G., 2011. An algorithm for merging SMAP radiometer and radar data for high-resolution soil-moisture retrieval. Geosci. Remote Sensing, IEEE Trans. 49, 1504–1512.

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

DeChant, C.M., Moradkhani, H., 2015. Analyzing the sensitivity of drought recovery forecasts to land surface initial conditions. J. Hydrol. 526, 89–100. doi:10.1016/j.jhydrol.2014.10.021

DeChant, C.M., Moradkhani, H., 2014. Toward a reliable prediction of seasonal forecast uncertainty: Addressing model and initial condition uncertainty with ensemble data assimilation and sequential Bayesian combination. J. Hydrol. 519, 2967–2977. doi:10.1016/j.jhydrol.2014.05.045

DeChant, C.M., Moradkhani, H., 2012. Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. Water Resour. Res. 48, W04518. doi:10.1029/2011WR011011

DeChant, C.M., Moradkhani, H., 2011. Radiance data assimilation for operational snow and streamflow forecasting. Adv. Water Resour. 34, 351–364. doi:10.1016/j.advwatres.2010.12.009

Dong, J., Steele-Dunne, S.C., Judge, J., van de Giesen, N., 2015. A particle batch smoother for soil moisture estimation using soil temperature observations. Adv. Water Resour. 83, 111–122. doi:10.1016/j.advwatres.2015.05.017

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, P.D., Hirschi, M., Ikonen, J., de Jeu, R., Kidd, R., Lahoz, W., Liu, Y.Y., Miralles, D., Mistelbauer, T., Nicolai-Shaw, N., Parinussa, R., Pratola, C., Reimer, C., van der Schalie, R., Seneviratne, S.I., Smolander, T., Lecomte, P., 2017. ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. Remote Sens. Environ. 203, 185–215. doi:10.1016/j.rse.2017.07.001

Dorigo, W.A., Gruber, A., De Jeu, R.A.M., Wagner, W., Stacke, T., Loew, A., Albergel, C., Brocca, L., Chung, D., Parinussa, R.M., Kidd, R., 2015. Evaluation of the ESA CCI soil moisture product using ground-based observations. Remote Sens. Environ. 162, 380–395.

doi:10.1016/j.rse.2014.07.023
Draper, C.S., Reichle, R.H., De Lannoy, G.J.M., Liu, Q., 2012. Assimilation of passive and active microwave soil moisture retrievals. Geophys. Res. Lett. 39, L04401. doi:10.1029/2011GL050655
Entekhabi, D., Njoku, E.G., O'Neill, P.E., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Entin, J.K., Goodman, S.D., Jackson, T.J., Johnson, J., Kimball, J., Piepmeier, J.R., Koster, R.D., Martin, N., McDonald, K.C., Moghaddam, M., Moran, S., Reichle, R., Shi, J.C., Spencer, M.W., Thurman, S.W., Tsang, L., Van Zyl, J., 2010. The soil moisture active passive (SMAP) mission. Proc. IEEE 98, 704–716. doi:10.1109/JPROC.2010.2043918
Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. 99, 10143–10162. doi:10.1029/94JC00572
Gruber, A., Su, C.-H., Zwieback, S., Crow, W., Dorigo, W., Wagner, W., 2016. Recent advances in (soil moisture) triple collocation analysis. Int. J. Appl. Earth Obs. Geoinf. 45, 200–211. doi:10.1016/j.jag.2015.09.002
Hao, Z., AghaKouchak, A., Nakhjiri, N., Farahmand, A., 2014. Global integrated drought monitoring and prediction system. Sci. Data 1, 1–10. doi:10.1038/sdata.2014.1
Harlim, J., Majda, A.J., 2010. Catastrophic filter divergence in filtering nonlinear dissipative systems. Commun. Math. Sci. 8, 27–43.
Hoerling, M., Eischeid, J., Kumar, A., Leung, R., Mariotti, A., Mo, K., Schubert, S., Seager, R., 2014. Causes and predictability of the 2012 Great Plains drought. Bull. Am. Meteorol. Soc. 95, 269–282. doi:10.1175/BAMS-D-13-00055.1
Hoerling, M., Schubert, S., Mo, K., 2013. An Interpretation of the Origins of the 2012 Central Great Plains Drought, Assessment Report. NOAA Drought Task Force.
Keyantash, J., Dracup, J.A., 2002. The quantification of drought: An evaluation of drought indices. Bull. Am. Meteorol. Soc. 83, 1167–1180. doi:10.1175/1520-0477(2002)083<1191:TQODAE>2.3.CO;2
Kumar, S. V., Harrison, K.W., Peters-Lidard, C.D., Santanello, J. a., Kirschbaum, D., 2014a. Assessing the impact of L-band observations on drought and flood risk estimation: A decision-theoretic approach in an OSSE environment. J. Hydrometeorol. 15, 2140–2156. doi:10.1175/JHM-D-13-0204.1
Kumar, S. V., Peters-Lidard, C.D., Mocko, D., Reichle, R., Liu, Y., Arsenault, K.R., Xia, Y., Ek, M., Riggs, G., Livneh, B., Cosh, M., 2014b. Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation. J. Hydrometeorol. 15, 2446–2469. doi:10.1175/JHM-D-13-0132.1
Kumar, S. V., Reichle, R.H., Koster, R.D., Crow, W.T., Peters-Lidard, C.D., 2009. Role of subsurface physics in the assimilation of surface soil moisture observations. J. Hydrometeorol. 10, 1534–1547. doi:10.1175/2009JHM1134.1
Kumar, S. V., Zaitchik, B.F., Peters-Lidard, C.D., Rodell, M., Reichle, R., Li, B., Jasinski, M., Mocko, D., Getirana, A., De Lannoy, G., Cosh, M.H., Hain, C.R., Anderson, M., Arsenault, K.R., Xia, Y., Ek, M., 2016. Assimilation of gridded GRACE terrestrial water storage estimates in the North American Land Data Assimilation System. J. Hydrometeorol. 17, 1951–1972. doi:10.1175/JHM-D-15-0157.1
Kurtz, W., He, G., Kollet, S.J., Maxwell, R.M., Vereecken, H., Franssen, H.J.H., 2016. TerrSysMP-PDAF (version 1.0): A modular high-performance data assimilation framework for an integrated land surface-subsurface model. Geosci. Model Dev. 9, 1341–1360.

doi:10.5194/gmd-9-1341-2016

Leisenring, M., Moradkhani, H., 2012. Analyzing the uncertainty of suspended sediment load prediction using sequential data assimilation. J. Hydrol. 468–469, 268–282. doi:10.1016/j.jhydrol.2012.08.049

Leisenring, M., Moradkhani, H., 2011. Snow water equivalent prediction using Bayesian data assimilation methods. Stoch. Environ. Res. Risk Assess. 25, 253–270. doi:10.1007/s00477-010-0445-5

Liang, X., Lettenmaier, D.P., Wood, E.F., Burges, S.J., 1994. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. J. Geophys. Res. 99, 14415–14428. doi:10.1029/94JD00483

Liang, X., Xie, Z., 2001. A new surface runoff parameterization with subgrid-scale soil heterogeneity for land surface models. Adv. Water Resour. 24, 1173–1193. doi:10.1016/S0309-1708(01)00032-X

Liu, Q., Reichle, R.H., Bindlish, R., Cosh, M.H., Crow, W.T., de Jeu, R., De Lannoy, G.J.M., Huffman, G.J., Jackson, T.J., 2011. The contributions of precipitation and soil moisture observations to the skill of soil moisture estimates in a land data assimilation system. J. Hydrometeorol. 12, 750–765. doi:10.1175/JHM-D-10-05000.1

Liu, Y., Weerts, a. H., Clark, M., Hendricks Franssen, H.J., Kumar, S., Moradkhani, H., Seo, D.J., Schwanenberg, D., Smith, P., Van Dijk, a. I.J.M., Van Velzen, N., He, M., Lee, H., Noh, S.J., Rakovec, O., Restrepo, P., 2012. Advancing data assimilation in operational hydrologic forecasting: Progresses, challenges, and emerging opportunities. Hydrol. Earth Syst. Sci. 16, 3863–3887. doi:10.5194/hess-16-3863-2012

Liu, Y.Y., Dorigo, W.A., Parinussa, R.M., De Jeu, R.A.M., Wagner, W., McCabe, M.F., Evans, J.P., Van Dijk, A.I.J.M., 2012. Trend-preserving blending of passive and active microwave soil moisture retrievals. Remote Sens. Environ. 123, 280–297. doi:10.1016/j.rse.2012.03.014

Liu, Y.Y., Parinussa, R.M., Dorigo, W. a., De Jeu, R. a M., Wagner, W., M. Van Dijk, a. I.J., McCabe, M.F., Evans, J.P., 2011. Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals. Hydrol. Earth Syst. Sci. 15, 425–436. doi:10.5194/hess-15-425-2011

Lloyd-Hughes, B., 2014. The impracticality of a universal drought definition. Theor. Appl. Climatol. 117, 607–611. doi:10.1007/s00704-013-1025-7

Lohmann, D., Raschke, E., Nijssen, B., Lettenmaier, D.P., 1998. Regional scale hydrology: I. Formulation of the VIC-2L model coupled to a routing model. Hydrol. Sci. J. 43, 131–141. doi:10.1080/02626669809492107

Lorentzen, R.J., Naevdal, G., 2011. An Iterative Ensemble Kalman Filter. IEEE Trans. Automat. Contr. 56, 1990–1995. doi:10.1109/TAC.2011.2154430

Luo, L., Wood, E.F., 2007. Monitoring and predicting the 2007 U.S. drought. Geophys. Res. Lett. 34, L22702. doi:10.1029/2007GL031673

Madadgar, S., Moradkhani, H., Garen, D., 2014. Towards improved post-processing of hydrologic forecast ensembles. Hydrol. Process. 28, 104–122. doi:10.1002/hyp.9562

Massari, C., Brocca, L., Tarpanelli, A., Moramarco, T., 2015. Data assimilation of satellite soil moisture into rainfall-runoff modelling: A complex recipe? Remote Sens. 7, 11403–11433. doi:10.3390/rs70911403

Maurer, E.P., Wood, A.W., Adam, J.C., Lettenmaier, D.P., Nijssen, B., 2002. A long-term hydrologically based dataset of land surface fluxes and states for the Conterminous United

States*. J. Clim. 15, 3237–3251. doi:10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2

Mishra, A.K., Singh, V.P., 2010. A review of drought concepts. J. Hydrol. 391, 202–216. doi:10.1016/j.jhydrol.2010.07.012

Mo, K.C., Lettenmaier, D.P., 2015. Heat wave flash droughts in decline. Geophys. Res. Lett. 42, 2823–2829. doi:10.1002/2015GL064018

Montzka, C., Moradkhani, H., Weihermüller, L., Franssen, H.-J.H., Canty, M., Vereecken, H., 2011. Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter. J. Hydrol. 399, 410–421. doi:10.1016/j.jhydrol.2011.01.020

Moradkhani, H., 2008. Hydrologic remote sensing and land surface data assimilation. Sensors 8, 2986–3004. doi:10.3390/s8052986

Moradkhani, H., Dechant, C.M., Sorooshian, S., 2012. Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method. Water Resour. Res. 48, W12520. doi:10.1029/2012WR012144

Moradkhani, H., Hsu, K.-L., Gupta, H., Sorooshian, S., 2005. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. Water Resour. Res. 41, W05012. doi:10.1029/2004WR003604

Moradkhani, H., Sorooshian, S., 2008. General review of rainfall-runoff modeling: model calibration, data assimilation, and uncertainty analysis, in: Hydrological Modelling and the Water Cycle. Springer Berlin Heidelberg, pp. 1–24.

Nerger, L., Hiller, W., 2013. Software for ensemble-based data assimilation systems-Implementation strategies and scalability. Comput. Geosci. 55, 110–118. doi:10.1016/j.cageo.2012.03.026

Noh, S.J., Tachikawa, Y., Shiiba, M., Kim, S., 2011. Applying sequential Monte Carlo methods into a distributed hydrologic model: lagged particle filtering approach with regularization. Hydrol. Earth Syst. Sci. 15, 3237–3251. doi:10.5194/hess-15-3237-2011

Otkin, J.A., Anderson, M.C., Hain, C., Mladenova, I.E., Basara, J.B., Svoboda, M., 2013. Examining Rapid Onset Drought Development Using the Thermal Infrared–Based Evaporative Stress Index. J. Hydrometeorol. 14, 1057–1074. doi:10.1175/JHM-D-12-0144.1

PaiMazumder, D., Done, J.M., 2016. Potential predictability sources of the 2012 US drought in observations and a regional model ensemble. J. Geophys. Res. Atmos. doi:10.1002/2016JD025322

Pan, M., Wood, E.F., 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. Top. Appl. Earth Obs. Remote Sens. 3, 49–56. doi:10.1109/JSTARS.2010.2040585

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

Plaza, D.A., De Keyser, R., De Lannoy, G.J.M., Giustarini, L., Matgen, P., Pauwels, V.R.N., 2012. The importance of parameter resampling for soil moisture data assimilation into hydrologic models using the particle filter. Hydrol. Earth Syst. Sci. 16, 375–390. doi:10.5194/hess-16-375-2012s

Reichle, R.H., Crow, W.T., Keppenne, C.L., 2008. An adaptive ensemble Kalman filter for soil moisture data assimilation. Water Resour. Res. 44, W03423. doi:10.1029/2007WR006357

Reichle, R.H., Koster, R.D., 2004. Bias reduction in short records of satellite soil moisture.

Geophys. Res. Lett. 31, L19501. doi:10.1029/2004GL020938

Reichle, R.H., Koster, R.D., Liu, P., Mahanama, S.P.P., Njoku, E.G., Owe, M., 2007. Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). J. Geophys. Res. 112, D09108. doi:10.1029/2006JD008033

Ridler, M.E., van Velzen, N., Hummel, S., Sandholt, I., Falk, A.K., Heemink, A., Madsen, H., 2014. Data assimilation framework: Linking an open data assimilation library (OpenDA) to a widely adopted model interface (OpenMI). Environ. Model. Softw. 57, 76–89. doi:10.1016/j.envsoft.2014.02.008

Ross, T., Lott, N., 2003. A climatology of 1980-2003 extreme weather and climate events. National Ocanic and Atmospheric Administration.

Samaniego, L., Kumar, R., Zink, M., 2013. Implications of parameter uncertainty on soil moisture drought analysis in Germany. J. Hydrometeorol. 14, 47–68. doi:10.1175/JHM-D-12-075.1

Sheffield, J., Wood, E.F., 2011. Drought: past problems and future scenarios. Routledge.

Sheffield, J., Wood, E.F., Chaney, N., Guan, K., Sadri, S., Yuan, X., Olang, L., Amani, A., Ali, A., Demuth, S., Ogallo, L., 2014. A drought monitoring and forecasting system for sub-sahara african water resources and food security. Bull. Am. Meteorol. Soc. 95, 861–882. doi:10.1175/BAMS-D-12-00124.1

Shukla, S., Steinemann, A.C., Lettenmaier, D.P., 2011. Drought monitoring for Washington State: indicators and applications. J. Hydrometeorol. 12, 66–83. doi:10.1175/2010JHM1307.1

Su, C.H., Ryu, D., Crow, W.T., Western, A.W., 2014. Beyond triple collocation: Applications to soil moisture monitoring. J. Geophys. Res. Atmos. 119, 6419–6439. doi:10.1002/2013JD021043

Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., Rippey, B., Tinker, R., Palecki, M., Stooksbury, D., Miskus, D., Stephens, S., 2002. The drought monitor. Bull. Am. Meteorol. Soc. 83, 1181–1190. doi:10.1175/1520-0477(2002)083<1181:TDM>2.3.CO;2

van Delft, G., El Serafy, G.Y., Heemink, A.W., 2009. The ensemble particle filter (EnPF) in rainfall-runoff models. Stoch. Environ. Res. Risk Assess. 23, 1203–1211. doi:10.1007/s00477-008-0301-z

Van Loon, A.F., 2015. Hydrological drought explained. Wiley Interdiscip. Rev. Water 2, 359–392. doi:10.1002/wat2.1085

Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., Figa-Saldaña, J., de Rosnay, P., Jann, A., Schneider, S., 2013. The ASCAT soil moisture product: A review of its specifications, validation results, and emerging applications. Meteorol. Zeitschrift 22, 5–33.

Wagner, W., Naeimi, V., Scipal, K., Jeu, R., Martínez-Fernández, J., 2007. Soil moisture from operational meteorological satellites. Hydrogeol. J. 15, 121–131. doi:10.1007/s10040-006-0104-6

Wilhite, D.A., 2000. Drought as a natural hazard: concepts and definitions, in: Wilhite, D.A. (Ed.), Drought: A Global Assessment. Routledge, London, pp. 3–18.

Wood, E.F., Schubert, S.D., Wood, A.W., Peters-Lidard, C.D., Mo, K.C., Mariotti, A., Pulwarty, R.S., 2015. Prospects for advancing drought understanding, monitoring, and prediction. J.

Hydrometeorol. 16, 1636–1657. doi:10.1175/JHM-D-14-0164.1

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D., Koren, V., Duan, Q., Mo, K., Fan, Y., Mocko, D., 2012. Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. J. Geophys. Res. Atmos. 117, D03109. doi:10.1029/2011JD016048

Yan, H., DeChant, C.M., Moradkhani, H., 2015. Improving Soil Moisture Profile Prediction With the Particle Filter-Markov Chain Monte Carlo Method. IEEE Trans. Geosci. Remote Sens. 53, 6134–6147. doi:10.1109/TGRS.2015.2432067

Yan, H., Moradkhani, H., 2016. Combined assimilation of streamflow and satellite soil moisture with the particle filter and geostatistical modeling. Adv. Water Resour. 94, 364–378. doi:10.1016/j.advwatres.2016.06.002

Yan, H., Moradkhani, H., Zarekarizi, M., 2017. A probabilistic drought forecasting framework: A combined dynamical and statistical approach. J. Hydrol. 548, 291–304. doi:10.1016/j.jhydrol.2017.03.004

Yilmaz, M.T., Crow, W.T., 2013. The optimality of potential rescaling approaches in land data assimilation. J. Hydrometeorol. 14, 650–660. doi:10.1175/JHM-D-12-052.1

Yin, J., Zhan, X., Zheng, Y., Liu, J., Fang, L., Hain, C.R., 2015. Enhancing Model Skill by Assimilating SMOPS Blended Soil Moisture Product into Noah Land Surface Model. J. Hydrometeorol. 16, 917–931. doi:10.1175/JHM-D-14-0070.1

5 List of Captions

Figure 1 The flowchart of the proposed drought monitoring system. The three probability distributions associated with the meteorological forcing, remotely sensed soil moisture, and soil moisture states represent the corresponding uncertainties

Figure 2. The flow diagram of the parallel particle filtering framework (PPFF) using Message Passing Interface (MPI). The domain decomposition parallel strategy is used in the PPFF: each grid cell is simulated in parallel while each ensemble member is simulated sequentially.

Figure 3. The flowchart of the synthetic study using VIC model to assess the potential benefit of assimilation of satellite surface soil moisture on drought monitoring.

Figure 4. (a) The normalized information contribution (NIC) value between the EnKF and PMCMC root-zone soil moisture predictions (Eq. 13). Positive values indicate that the PMCMC improves soil moisture prediction as compared to the EnKF; negative values indicate the degradation over the EnKF. (b) Time series of the marked (yellow star) grid cell root-zone soil moisture (m3/m3) for the synthetic truth, EnKF, and PMCMC simulation for the period of 1 January 2012 to 31 December 2012. Both NIC values and time series are generated using the posterior means

Figure 5. The normalized information contribution (NIC) value between the Open-Loop (OL) and Data Assimilation (DA). The NIC values are generated based on posterior means. The positive value indicates that the DA improves soil moisture prediction against OL; negative

value indicates the degradation over the OL. (a): Surface soil moisture field. (b): Root-zone soil moisture filed.

Figure 6. Comparison of the drought monitoring skill between Open-Loop (OL) and Data Assimilation (DA) for May–August 2012.

Figure 7. The absolute bias of drought extent (%) against the synthetic truth between Open-Loop (OL) and Data Assimilation (DA) over the CONUS for the period of May–August 2012.

Figure 8. Comparison of the 2012 summer drought intensity over the CONUS from Open-Loop (OL), Data Assimilation (DA), and U.S. Drought Monitoring (USDM).

Figure 9. Comparison of the drought extent (%) over the CONUS from Open-Loop (OL), Data Assimilation (DA), and U.S. Drought Monitoring (USDM) for May/June/July/August 2012.

Figure 10. Comparison of U.S. Drought Monitoring (USDM) and Data Assimilation (DA) drought monitoring extent for five different drought categories (D0–D4) over the Central U.S. The drought extent is comprised of the 7-State region of Nebraska, Iowa, Kansas, Missouri, Oklahoma, Arkansas, and Illinois.

Figure 11. Comparison of the 2010 summer drought intensity over the CONUS from Open-Loop (OL), Data Assimilation (DA), and U.S. Drought Monitoring (USDM).

Figure A1. The flowchart of the offline coupling interface of the VIC model and the PMCMC data assimilation algorithm. The ensemble files indicate the ensemble members of the PMCMC algorithm. A one-day VIC model simulation and PMCMC updating are shown in this flowchart.