EGU General Assembly 2020

Towards an Effective and Scalable Hybrid Data Assimilation for Hydrogeophysical Applications

Hamid Moradkhani*, Peyman Abbaszadeh and Keyhan Gavahi

Center for Complex Hydrosystems Research (CCHR)
Department of Civil, Construction and Environmental Engineering

hmoradkhani@ua.edu, http://www.moradkhani.net/

THE UNIVERSITY OF ALABAMA

EGU2020-18135, May 5, Sharing Geoscience Online
Session HS6.11 – Coupled Data Assimilation for Earth System Models

NSF

NOAA

Introduction

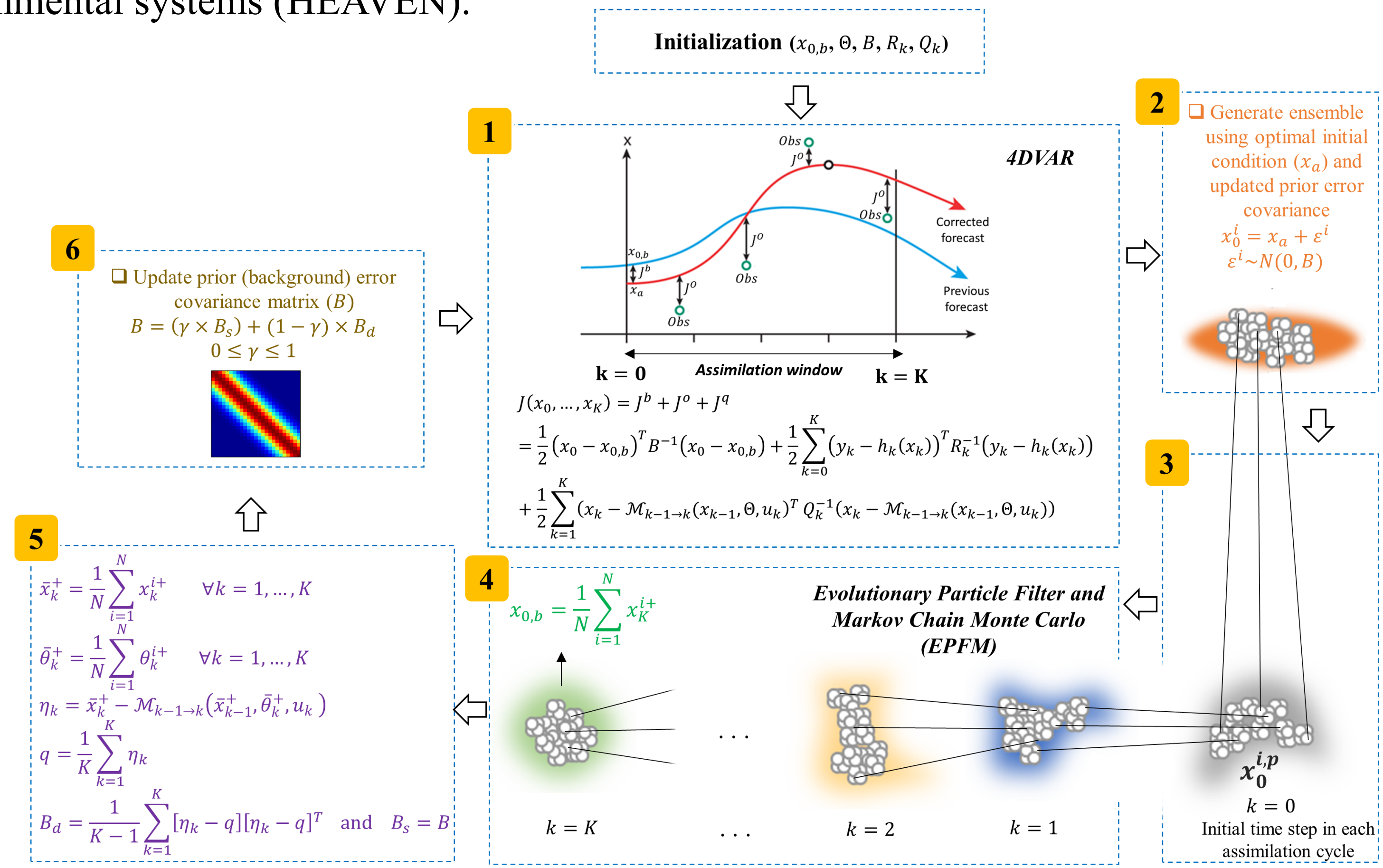
This presentation is a summary of the two recent development we recently published in Water Resources Research (Abbaszadeh et al., 2019) and also Advances in Water Resources (Abbaszadeh et al., 2018a). The applications of this method is presented for both flood forecasting and drought monitoring while utilizing the remotely sensed observations.

Here, we introduce a novel approach that couples a deterministic four-dimensional variational (4DVAR) assimilation method with an evolutionary ensemble filtering that together significantly improve the estimation of storages and fluxes, hence better forecasting skill. The Evolutionary Particle Filter with MCMC (EPFM) (Abbaszadeh et al., 2018a) uses the Genetic Algorithm (GA) to effectively sample the particles to better represent the posterior distribution of model prognostic variables and parameters. This is followed by coupling EPFM and 4DVAR which results in a superior DA approach, the so-called Hybrid Ensemble and Variational Data Assimilation framework for Environmental systems (HEAVEN) (Abbszadeh et al., 2019). The method explicitly accounts for model structural error during the assimilation process.

HEAVEN

(Abbaszadeh et al., 2019, WRR)

This figure illustrates the Hybrid Ensemble and Variational Data Assimilation framework for Environmental systems (HEAVEN).



HEAVEN

Observation error covariance matrix R_k at each time step can be specified as follows, where λ is the error percentage in observations and Obs_k is observation at time k. B is prior error covariance matrix.

$$R_k = [\max\{(\lambda \times Obs_k), 1\}]^2 \quad B = \mathrm{diag}\left(\left[\Omega \times x_{0,b}\right]^2\right) \quad Q_k = \Gamma \times \mathrm{diag}\left(\left[\pi \times x_{0,b}\right]^2\right)$$

where Ω is the error percentage in initial state variables and $x_{0,b}$ is the deterministic initial guess for state variables. Similarly, the model error covariance is Q_k, where π is the error percentage in model structure and Γ is the model error covariance inflation ($\Gamma \geq 1$) or deflation factor ($\Gamma \leq 1$).

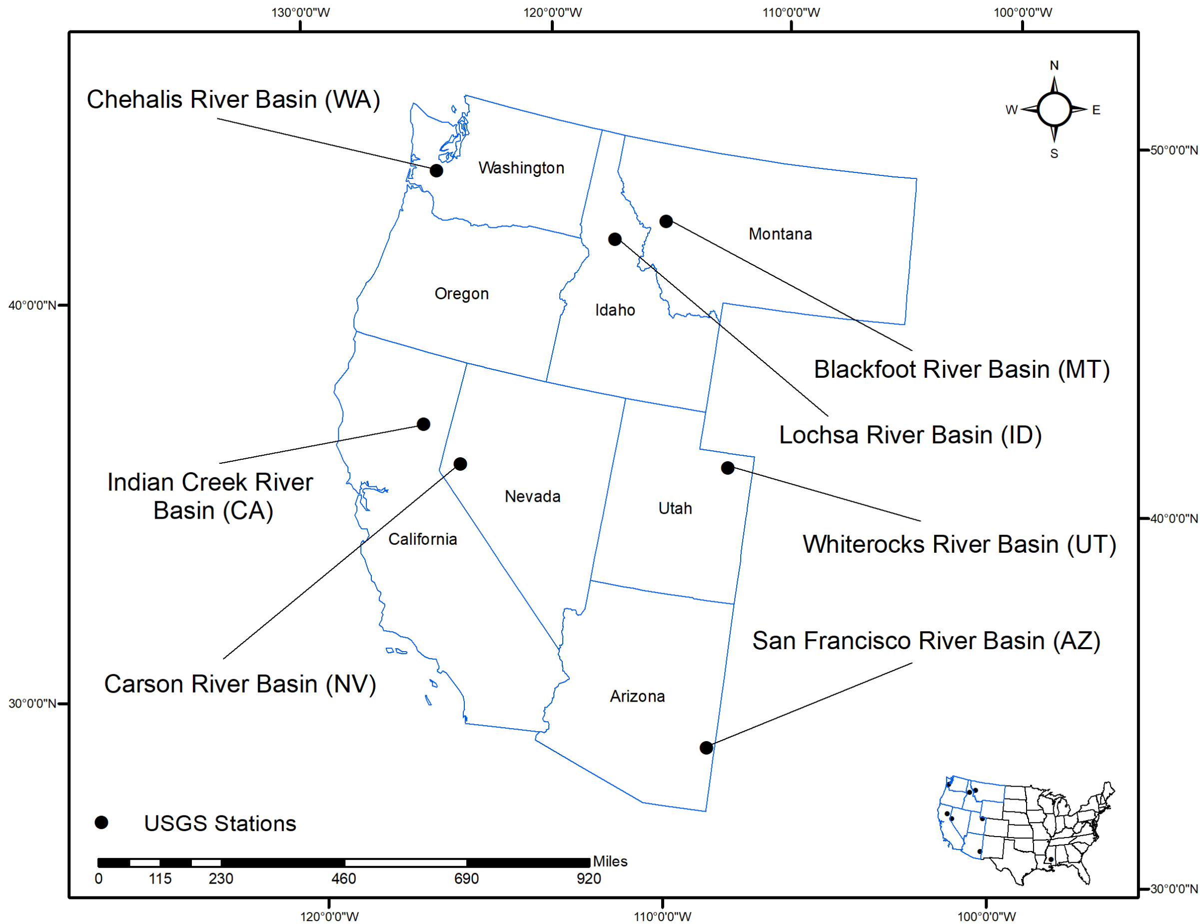
For real case study (model is imperfect), we use the Weak-constraint 4DVAR formula (1), while for synthetic study (model is perfect), the strong-constraint formulation (2) should be used.

$$\begin{aligned} J(x_0, \dots, x_K) &= J^b + J^o + J^q \\ &= \frac{1}{2}\left(x_0 - x_{0,b}\right)^T B^{-1}\left(x_0 - x_{0,b}\right) + \frac{1}{2}\sum_{k=0}^{K}\left(y_k - h_k(x_k)\right)^T R_k^{-1}\left(y_k - h_k(x_k)\right) \\ &+ \frac{1}{2}\sum_{k=1}^{K}(x_k - \mathcal{M}_{k-1\to k}(x_{k-1}, \Theta, u_k))^T Q^{-1}(x_k - \mathcal{M}_{k-1\to k}(x_{k-1}, \Theta, u_k)) \end{aligned} \qquad (1)$$

$$J(x_0) = J^b + J^o = \frac{1}{2}\left(x_0 - x_{0,b}\right)^T B^{-1}\left(x_0 - x_{0,b}\right) + \frac{1}{2}\sum_{k=0}^{K}\left(y_k - h_k(x_k)\right)^T R_k^{-1}\left(y_k - h_k(x_k)\right) \qquad (2)$$

HEAVEN

The basins located in different environmental settings (i.e., climate zones and hydrological properties) to fully examine the effectiveness and usefulness of the proposed method.



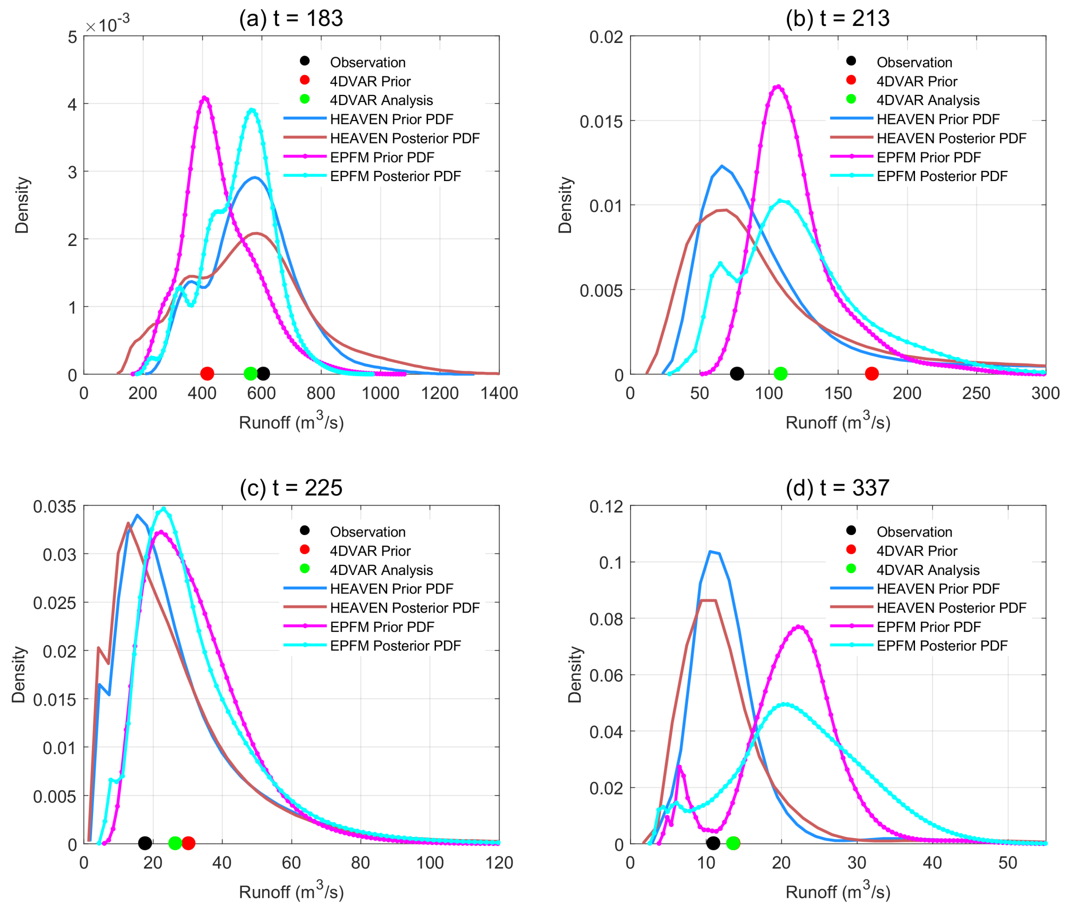
THE UNIVERSITY OF ALABAMA®

HEAVEN

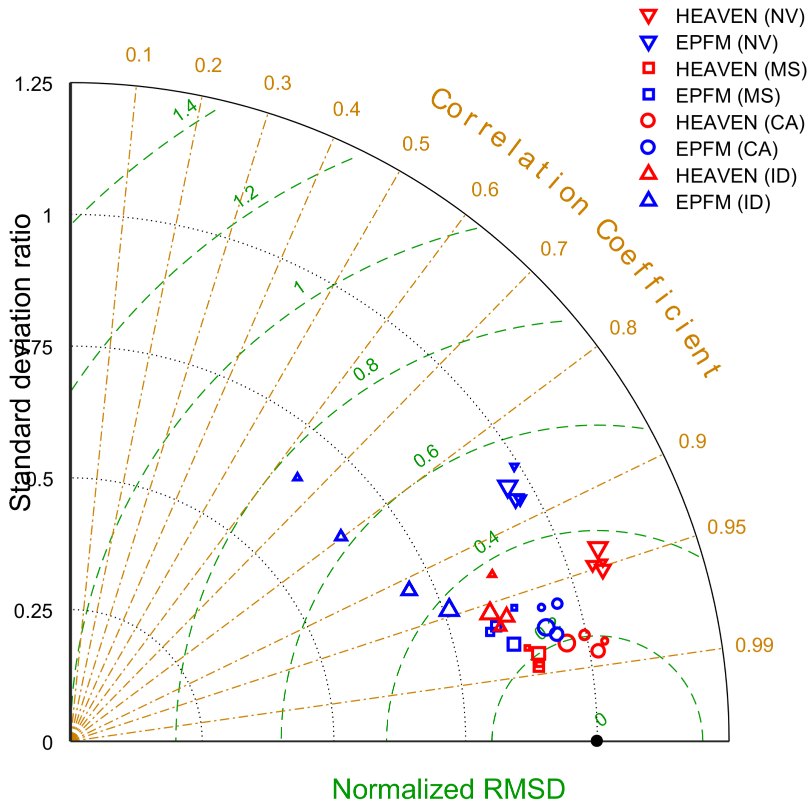
- The prior and posterior distributions at four daily time steps (t = 183, 213, 225 and 337 days).

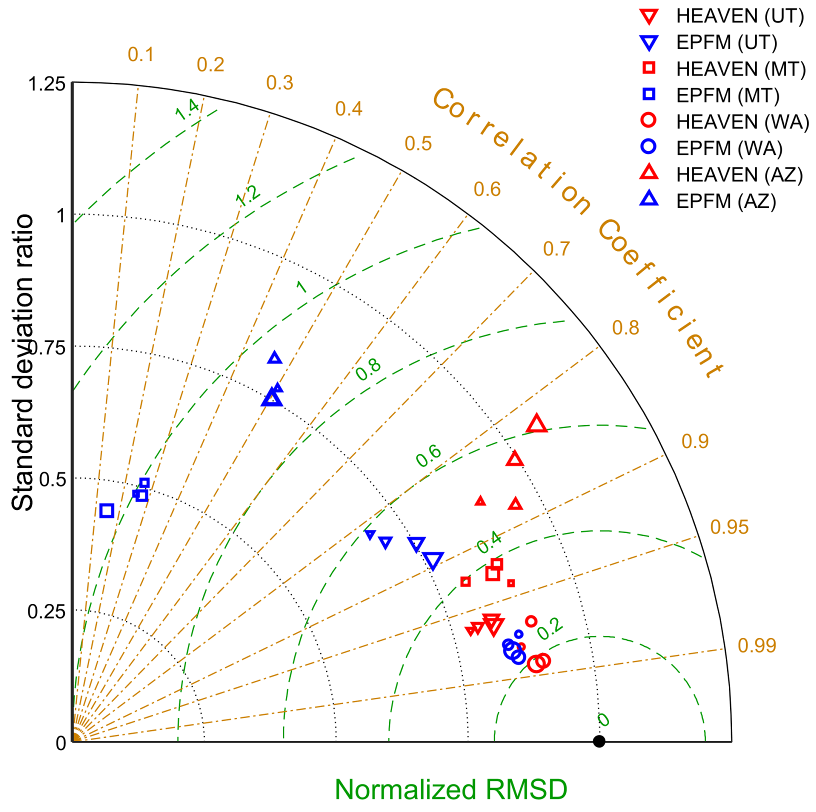
- These four days represent the initial time of four different assimilation cycles, which were chosen according to different streamflow regimes.

- In each assimilation cycle 4DVAR moves the prior (red point) to an optimal location (green point) at which the EFPM provides the best estimates of posterior distributions for both state variables and model parameters.



HEAVEN


HEAVEN (NV)
EPFM (NV)
HEAVEN (MS)
EPFM (MS)
HEAVEN (CA)
EPFM (CA)
HEAVEN (ID)
EPFM (ID)
Correlation Coefficient
Standard deviation ratio
Normalized RMSD


HEAVEN (UT)
EPFM (UT)
HEAVEN (MT)
EPFM (MT)
HEAVEN (WA)
EPFM (WA)
HEAVEN (AZ)
EPFM (AZ)
Correlation Coefficient
Standard deviation ratio
Normalized RMSD

Drought Monitoring

(Xu et al., 2020, RSE)

In this study, we use our Data Assimilation (DA) approach to assimilate Soil Moisture Active Passive (SMAP) soil moisture data into Variable Infiltration Capacity land surface model to provide more reliable topsoil layer moisture over the Continental United States.

Also, we used a multivariate probability distribution based on a Copula function to integrate the posterior soil moisture, precipitation and evapotranspiration information to develop a new integrated drought index, i.e., the SPESMI (Standardized Precipitation, Evapotranspiration and Soil Moisture Index).

In this study, we assimilate two remotely sensed data, namely SMOPS, and MODIS evapotranspiration (MODIS16 ET), at 1-km spatial resolution, into the VIC land surface model.

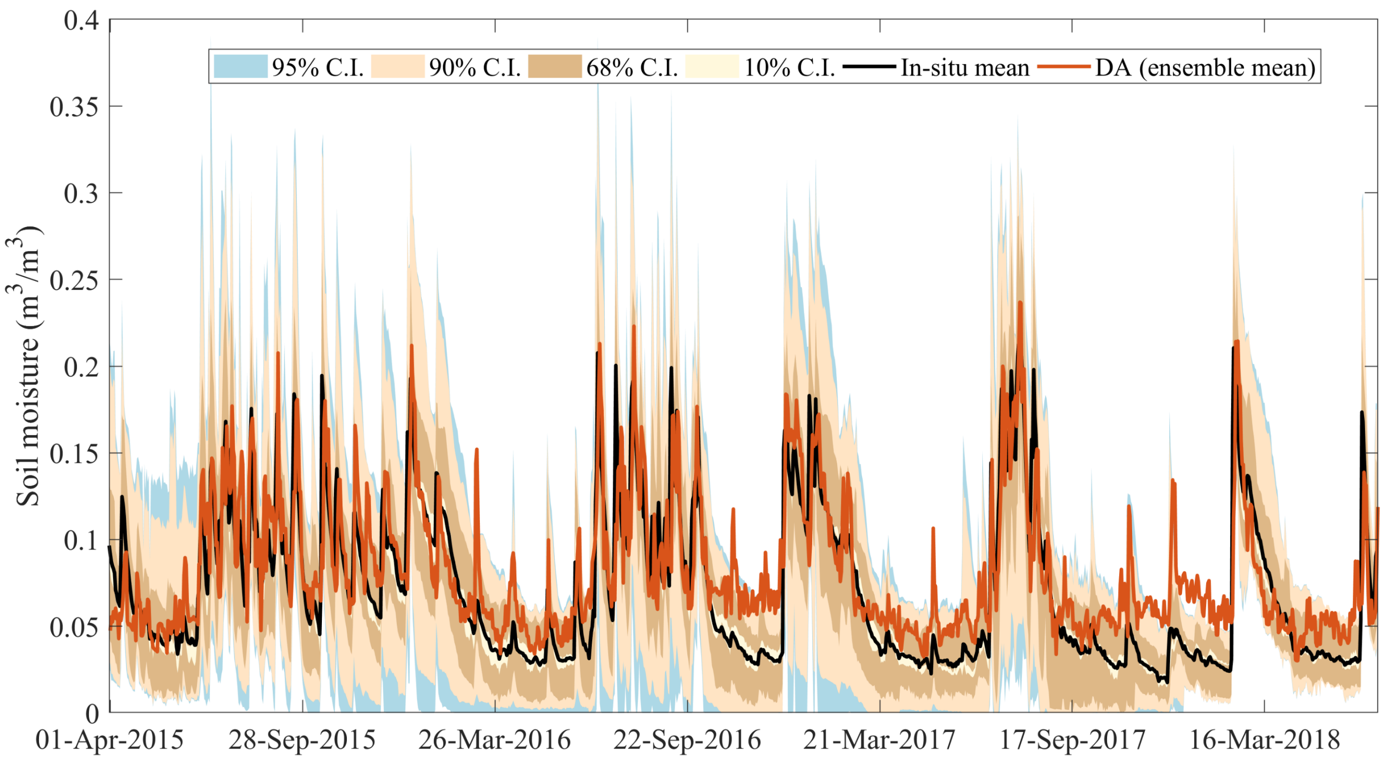
To validate the usefulness of the developed integrated drought index, we compared the drought events detected by this index with those reported by the United States Drought Monitor (USDM).

we noticed that our approach could identify some severe to extreme drought events that had been underestimated by the USDM.

Drought Monitoring

A comparison of the assimilated soil moisture and the in-situ soil moisture data at the Walnut Gulch Watershed in the state of Arizona, USA.

Most of the DA obtained soil moisture is well within the 68% confidence interval, i.e. the range of one standard deviation, indicating the DA results are strongly consistent with in-situ data.

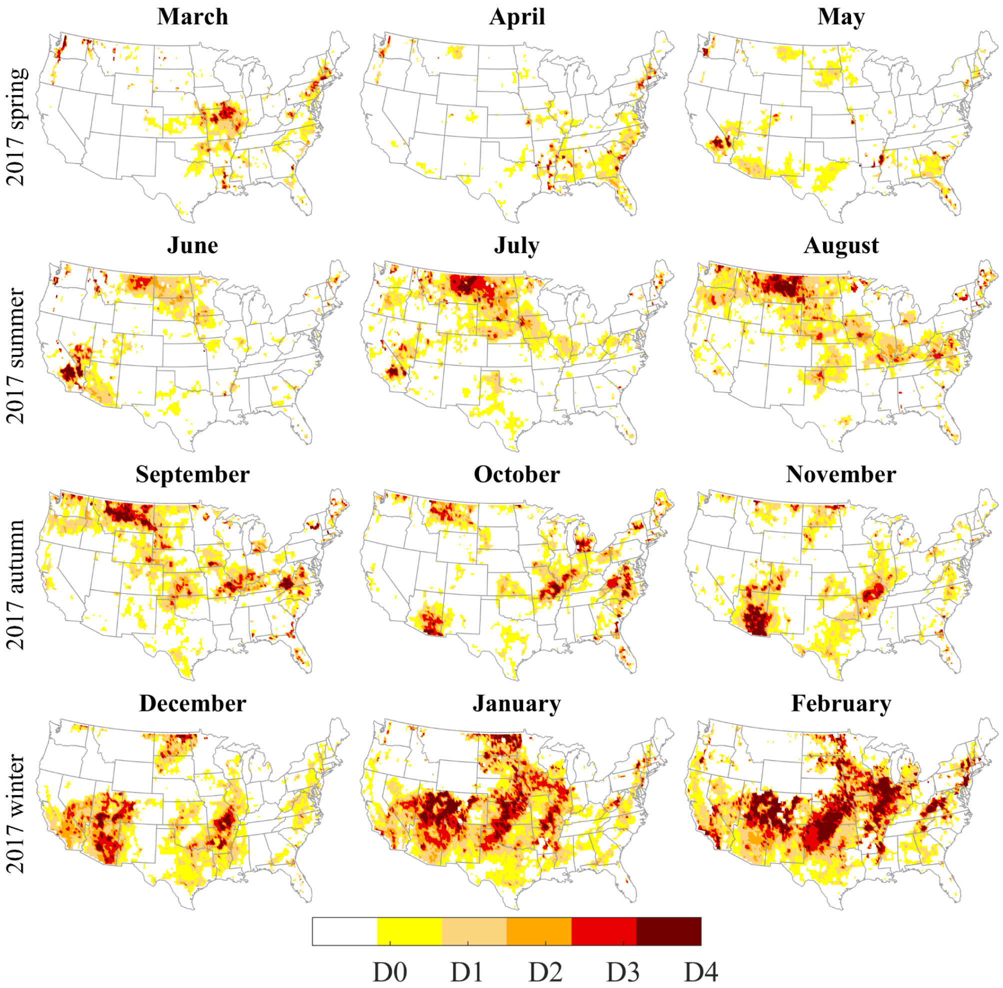


Drought Monitoring

- The Standardized Precipitation, Evapotranspiration and Soil Moisture Index (SPESMI), is developed based on the precipitation, MODIS PET and the posterior soil moisture.

- According to the SPESMI, mild drought spread out in the southern CONUS and the Midwest in the early spring, especially in Missouri and southern Illinois.

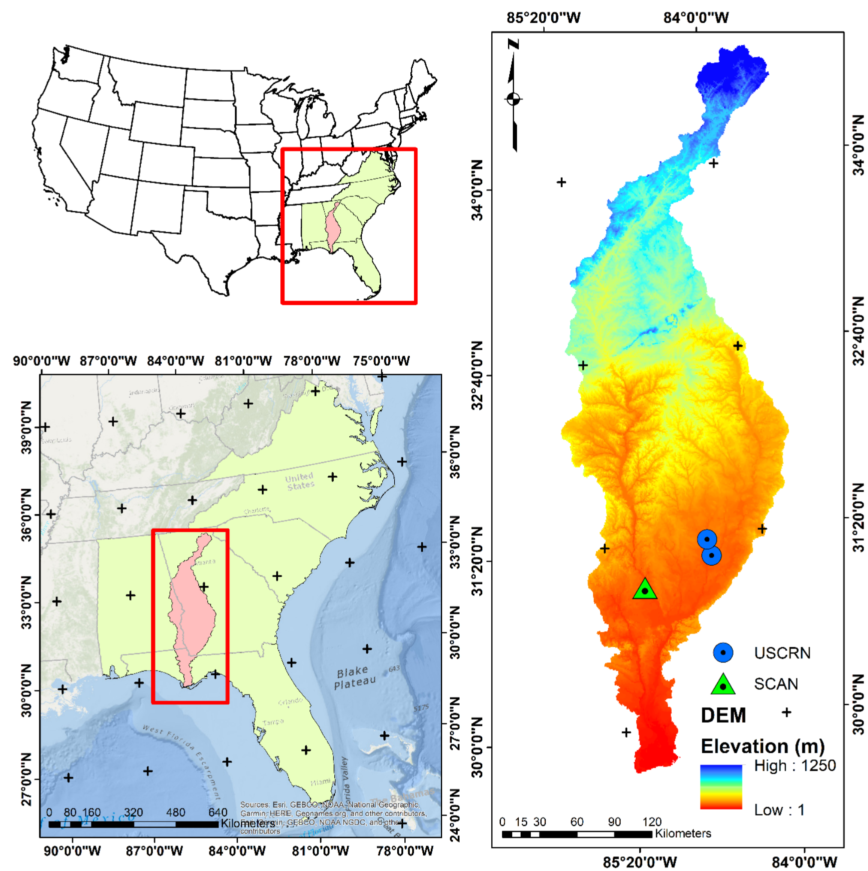
- During late summer, severe to extreme drought prevailed in northwestern CONUS, especially in Montana, consistent with the results from USDA topsoil moisture observations.



Drought Monitoring

(Gavahi et al., 2020, JHM)

- Soil moisture (SM) and evapotranspiration (ET) are key variables of the terrestrial water cycle with a strong relationship.
- This study examines remotely sensed soil moisture and evapotranspiration data assimilation (DA) with the aim of improving drought monitoring.
- Although numerous efforts have gone into assimilating satellite soil moisture observations into land surface models to improve their predictive skills, little attention has been given to the combined use of soil moisture and evapotranspiration to better characterize hydrologic fluxes.

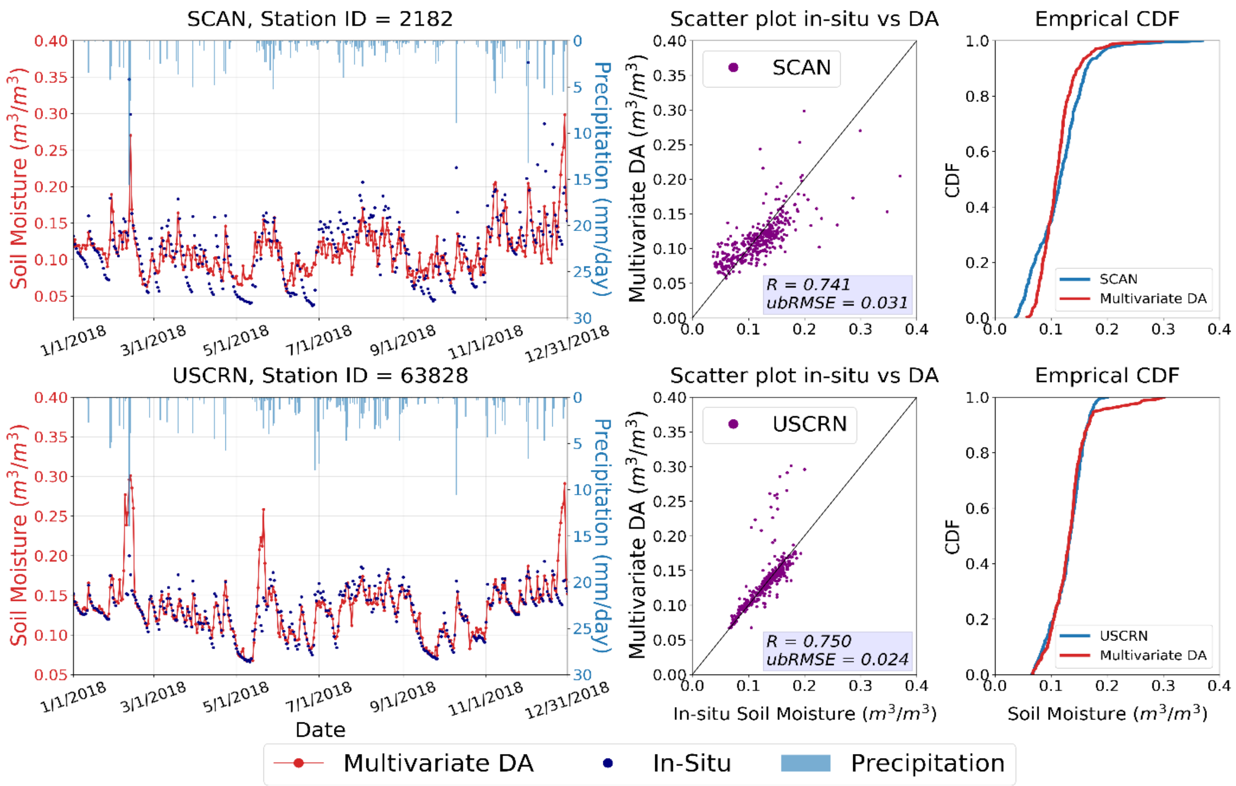


The location of the Apalachicola–Chattahoochee–Flint (ACF) basin in the Southeast United States alongside the Digital Elevation Map (DEM) of the basin.

Drought Monitoring

Comparison between multivariate DA results and two SCAN and USCRN stations within the ACF region.

Two of the USCRN and one of the SCAN stations are located inside the ACF basin, although only one of the USCRN stations provided data for the analysis period of the current study.

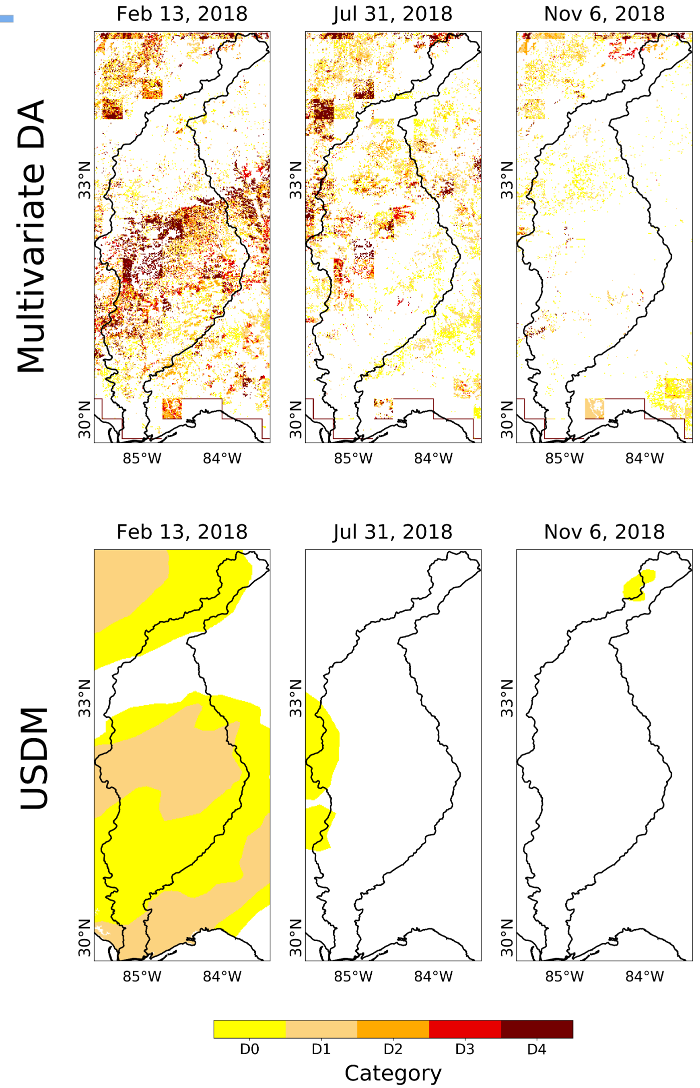


Drought Monitoring

This shows a comparison between USDM and the drought categories derived from SM percentiles of multivariate DA for three different weeks.

It is noteworthy to mention that the spatial resolution of multivariate DA is 1 km which provides us with a more detailed depiction of the drought extension over the ACF region.

The week starting on February 13 (Feb-13 to Feb-19) shows only D0 and D1 categories on the USDM maps whereas the same areas in these categories are classified as D2 to D4 in multivariate DA. The discrepancies arise mostly in summer.

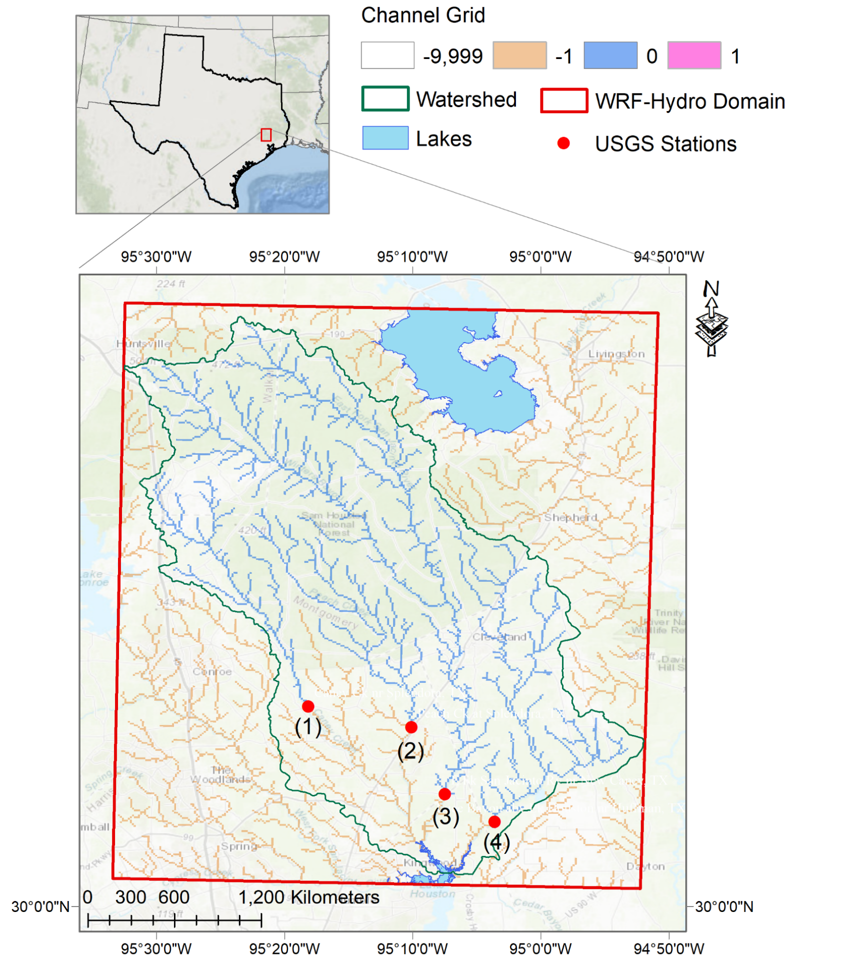


Flood Prediction: Hurricane Harvey

(Abbaszadeh et al., 2020, ADWR)

- This Figure illustrates location of the study area in the Southeast Texas along with watershed boundary, WRF- Hydro geogrid domain, lakes, stream networks, major rivers, and USGS streamflow gauges.

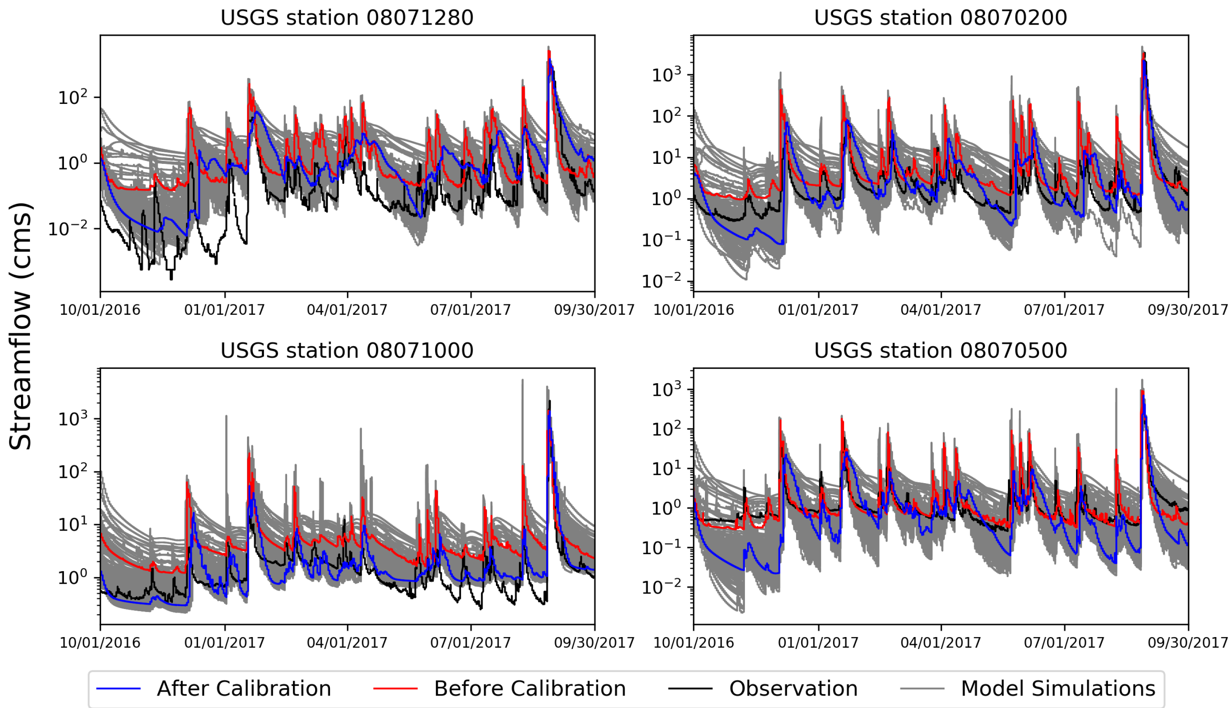
- In this study, we use an ensemble based Data Assimilation (DA) approach to explore the benefit of independently and jointly assimilating remotely sensed SMAP (Soil Moisture Active Passive) soil moisture (at different spatial resolutions) and USGS streamflow observations to improve the accuracy and reliability of WRF-Hydro model predictions while accounting for uncertainties.



(1) Caney Ck nr Splendora, TX
(2) Peach Ck at Splendora, TX
(3) E Fk San Jacinto Rv nr New Caney, TX
(4) Luce Bayou abv Lk Houston nr Huffman, TX

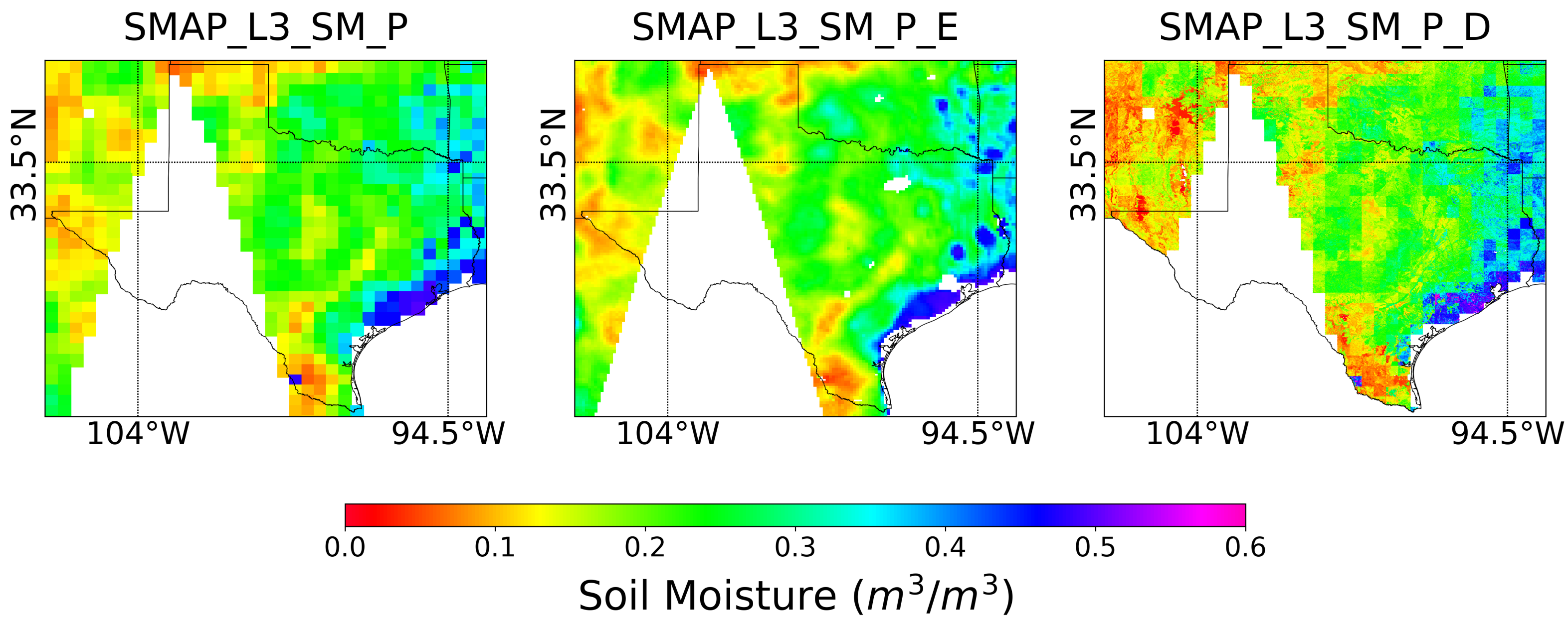
Flood Prediction: Hurricane Harvey

In this study, we used all the four USGS stations operated within the watershed to calibrate the WRF-Hydro model parameters. For the model calibration, the Noah-MP time step was set to one hour, which is the standard of the operational NWM. WRF-Hydro model calibration is performed by optimizing hourly streamflow using Dynamically Dimension Search (DDS) algorithm



Flood Prediction: Hurricane Harvey

This figure, used as an example for conceptualization, illustrates SMAP soil moisture data at three different spatial resolutions across the state of Texas on 26 August 2017, when the hurricane Harvey hit the southeast region of this state.

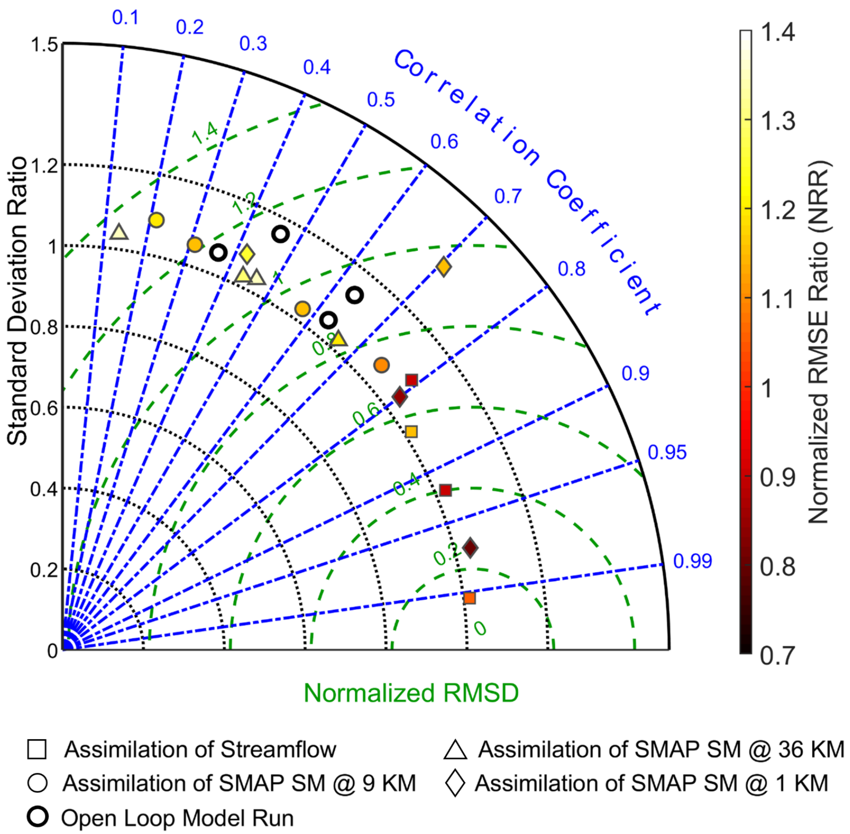


Right panel: SMAP_L3_SM_P (SMAP L3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture, Version 5) at 36 km spatial resolution. Middle panel: SMAP_L3_SM_P_E (SMAP Enhanced L3 Radiometer Global Daily 9 km EASE-Grid Soil Moisture, Version 2) at 9 km spatial resolution. SMAP_L3_SM_P_D (SMAP Radiometer Downscaled Product (Abbaszadeh et al., 2019a)) at 1 km spatial resolution.

Flood Prediction: Hurricane Harvey

Taylor diagram showing the deterministic and probabilistic performance measures of univariate assimilation configurations at four USGS gauges.

Comparing the univariate assimilation with OL simulations at four USGS stations over the entire period of study confirmed that although the assimilation of streamflow and soil moisture at 1 km spatial resolution could significantly outperform the OL predictions, assimilation of soil moisture at coarser spatial resolutions (i.e., 9 km or 36 km) had marginal performance compared to OL simulation.

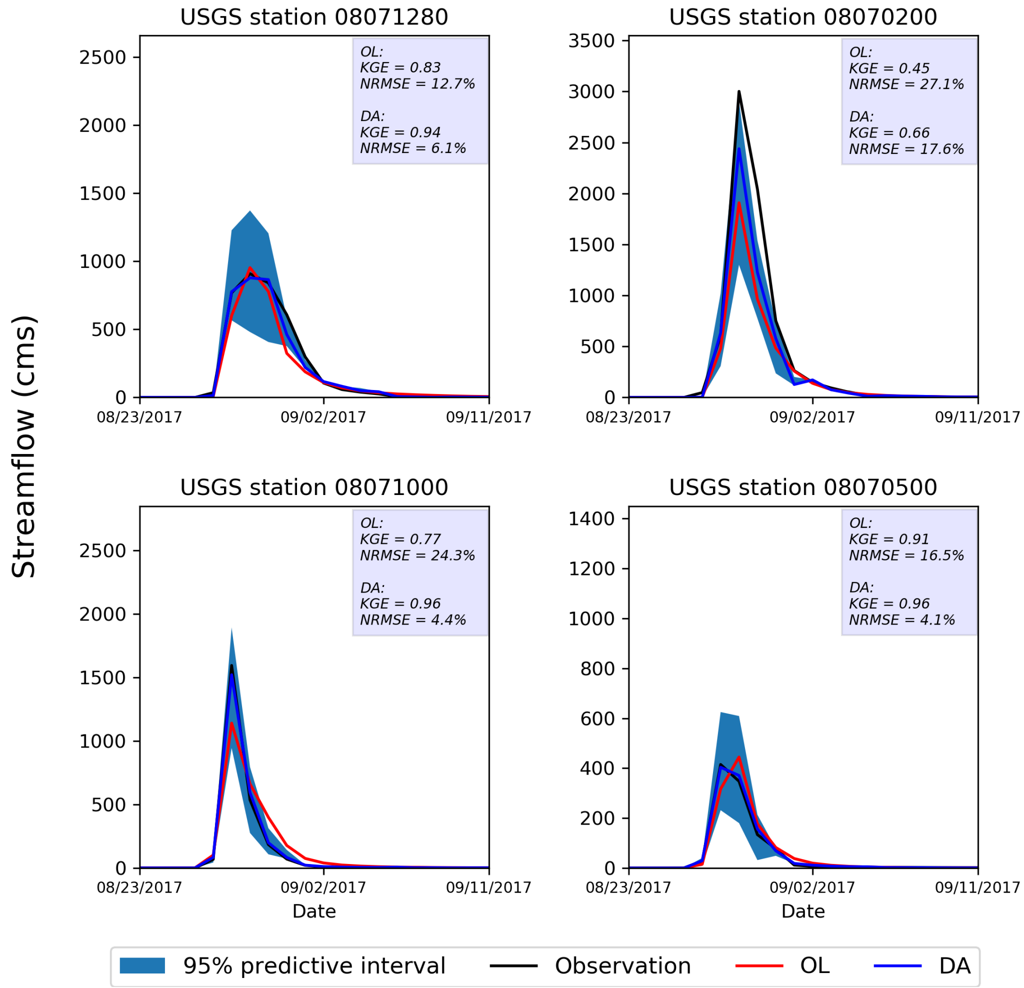


THE UNIVERSITY OF ALABAMA®

Flood Prediction: Hurricane Harvey

This figure illustrates the benefit of multivariate assimilation of satellite soil moisture and streamflow observation in improving the WRF-Hydro streamflow simulation during the period of hurricane Harvey.

The findings showed that the multivariate assimilation scenario results in better streamflow simulation regardless of the watershed's streamflow regime. It should be noted that during the hurricane Harvey assimilation of soil moisture at different spatial resolutions had similar impact on improving the streamflow simulation.



Results

The proposed approach:

- Characterizes model structural uncertainty by incorporating an explicit form of model error covariance matrix (Q) in the 4DVAR cost function.
- Propagates the prior error covariance matrix (B), which consists of a linear combination of static (Bs) and dynamic (Bd) error covariance matrices, from one cycle to another cycle over the entire assimilation period to fully account for a wide range of uncertainties in model predictions, and thus lead to more accurate and reliable posterior distributions.

Drought Monitoring:

- Compared to the commonly used data assimilation techniques, the proposed approach can improve the soil moisture estimation in terms of the correlation, ubRMSE and reliability in most areas of CONUS. The results indicated a strong temporal consistency of the drought areas detected by our approach and the USDM over the entire period of study (April 2015 to June 2018).
- Over ACF region, the results are consistent with the USDM maps during the winter and spring season considering the drought extents, but the severity of drought estimated by DA is slightly higher compared to USDM archives. During summer, however, the USDM maps show a complete drought-free condition whereas our findings show some areas with moderate to severe drought conditions.

Flood Prediction:

- Investigated the model performance during the hurricane Harvey and post-Harvey periods and realized that although DA, whether univariate or multivariate, and OL model runs have shown similar results in characterizing the onset of flooding, the DA outperforms in prediction of onset and demise of flooding (streamflow recession period).

Journal Papers

[1] Abbaszadeh, P., Moradkhani, H., Yan, H., (2018a). Enhancing hydrologic data assimilation by evolutionary Particle Filter and Markov Chain Monte Carlo. *Advances in Water Resources*. 111, 192–204.

[1] Abbaszadeh, P., Moradkhani, H., Zhan, X. (2018b). Downscaling SMAP radiometer soil moisture over the CONUS using an ensemble learning method. *Water Resources Research*. 55, 324-344.

[2] Abbaszadeh, P., Moradkhani, H., Daescu, D. N. (2019). The Quest for Model Uncertainty Quantification: A Hybrid Ensemble and Variational Data Assimilation Framework. *Water Resources Research*. 55. 2407-2431.

[4] Abbaszadeh, P., Gavahi, K., Moradkhani, H., (2020). Multivariate Remotely Sensed and In-situ Data Assimilation for Enhancing Community WRF-Hydro Model Forecasting. *Advances in Water Resources*. (Under Review)

[5] Lei, X., Abbaszadeh, P., Moradkhani, M.., Chen, N., Zhang, X., Chen, Z., (2020). Drought monitoring over CONUS using SMAP soil moisture, data assimilation and an integrated drought index. *Journal of Hydrometeorology*. (Under Review)

[6] Gavahi, K., Abbaszadeh, P., Moradkhani, M.., Chen, N., Zhang, X., Chen, Z., (2020). Multivariate Assimilation of Remotely Sensed Soil Moisture and Evapotranspiration for Drought Monitoring. *Remote Sensing of Environment.* (Under Review)

Journal Papers

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

ELSEVIER

Enhancing hydrologic data assimilation by evolutionary Particle Filter and Markov Chain Monte Carlo

Peyman Abbaszadeh, Hamid Moradkhani*, Hongxiang Yan

Water Resources and Remote Sensing Lab, Department of Civil and Environmental Engineering, Portland State University, 1930 SW 4th Ave., suite 200, Portland, OR 97201, United States

A R T I C L E I N F O

Keywords:
Particle Filters
Markov Chain Monte Carlo
Genetic algorithm
Hydrologic prediction

A B S T R A C T

Particle Filters (PFs) have received increasing attention by researchers from different disciplines including the hydro-geosciences, as an effective tool to improve model predictions in nonlinear and non-Gaussian dynamical systems. The implication of dual state and parameter estimation using the PFs in hydrology has evolved since 2005 from the PF-SIR (sampling importance resampling) to PF-MCMC (Markov Chain Monte Carlo), and now to the most effective and robust framework through evolutionary PF approach based on Genetic Algorithm (GA) and MCMC, the so-called EPFM. In this framework, the prior distribution undergoes an evolutionary process based on the designed mutation and crossover operators of GA. The merit of this approach is that the particles move to an appropriate position by using the GA optimization and then the number of effective particles is increased by means of MCMC, whereby the particle degeneracy is avoided and the particle diversity is improved. In this study, the usefulness and effectiveness of the proposed EPFM is investigated by applying the technique on a conceptual and highly nonlinear hydrologic model over four river basins located in different climate and geographical regions of the United States. Both synthetic and real case studies demonstrate that the EPFM improves both the state and parameter estimation more effectively and reliably as compared with the PF-MCMC.

1. Introduction

Accurate and reliable estimation of prognostic variables, such as streamflow and soil moisture, has always been one of the main challenges for hydrologists. Although hydrologic modeling can provide estimates of these quantities, the simulation results are potentially biased or erroneous given the following uncertainties:1) forcing data uncertainty due to the limitation of measurements and spatio-temporal representativeness of the data; 2) parameter uncertainty due to conceptualization of the model and non-uniqueness of parameters; 3) model structural uncertainty due to the imperfect representation of a real system; 4) initial and boundary condition uncertainty. Therefore, hydrologic predictions are better generated within a probabilistic framework, providing a mechanism to estimate the uncertainties involved in all layers of hydrologic predictions (Moradkhani et al., 2012). Most often, this is performed through Bayesian inference. Bayesian methods have been well acknowledged and used in numerous efforts to estimate the uncertainties in hydrologic model predictions (e.g., Kuczera and Parent, 1998; Marshal et al., 2004; Moradkhani et al., 2005; DeChant and Moradkhani, 2014; Yan et al., 2015; Pathiraja et al., 2016a; Pathiraja et al., 2016b).

Data Assimilation (DA) has been recognized as one of the effective methods to improve hydrologic prediction. Currently, the most widely used DA technique in the hydrologic community is the ensemble Kalman filter (EnKF) (Reichle et al., 2002; Crow and Wood, 2003; De Lannoy et al., 2007). Although the successful application of the EnKF and its variants has been reported in hydrologic literature, this technique has some inherent features resulting in sub-optimal performance. These include the Gaussian assumption of errors, linear updating rule within the EnKF and violation of water balance that limit its superiority (Moradkhani et al., 2005; Matgen et al., 2010; Noh et al., 2011; Plaza et al., 2012; DeChant and Moradkhani, 2012; Yan and Moradkhani, 2016). Given these concerns, data assimilation by means of Particle Filter (PF) as a viable alternative to the EnKF has garnered increasing attention in literature (e.g., Noh et al., 2011; Moradkhani et al., 2012; Montzka et al., 2013; Dong et al., 2015; Yan et al., 2017). The PF approach can relax the Gaussian assumption of error distributions by potentially characterizing multimodal or skewed distribution in state variables and parameters. Therefore, it can provide a thorough representation of the posterior distribution for a given nonlinear and non-Gaussian system. DeChant and Moradkhani (2012) presented a detailed performance assessment between the EnKF and PF, and found more

* Corresponding author.
E-mail address: hamidm@pdx.edu (H. Moradkhani).

https://doi.org/10.1016/j.advwatres.2017.11.011
Received 5 April 2017; Received in revised form 16 October 2017; Accepted 12 November 2017
Available online 13 November 2017

Water Resources Research

RESEARCH ARTICLE
10.1029/2018WR023629

Key Points:
- A joint sequential and variational data assimilation method was developed for superior and robust dual-state-parameter estimation
- The proposed HEAVEN approach accounts for all sources of uncertainties involved in model predictions
- The effectiveness and usefulness of HEAVEN was evaluated by both deterministic and probabilistic measures

Supporting Information:
- Supporting Information S1

Correspondence to:
H. Moradkhani,
hmoradkhani@ua.edu

Citation:
Abbaszadeh, P., Moradkhani, H., & Daescu, D. N. (2019). The Quest for Model Uncertainty Quantification: A Hybrid Ensemble and Variational Data Assimilation Framework. *Water Resources Research*, *55*, 2407–2431. https://doi.org/10.1029/2018WR023629

Received 5 JUL 2018
Accepted 19 FEB 2019
Accepted article online 25 FEB 2019
Published online 25 MAR 2019

The Quest for Model Uncertainty Quantification: A Hybrid Ensemble and Variational Data Assimilation Framework

Peyman Abbaszadeh[1], Hamid Moradkhani[1], and Dacian N. Daescu[2]

[1]Center for Complex Hydrosystems Research, Department of Civil, Construction and Environmental Engineering, University of Alabama, Tuscaloosa, AL, USA, [2]Fariborz Maseeh Department of Mathematics and Statistics, Portland State University, Portland, OR, USA

Abstract This article presents a novel approach to couple a deterministic four-dimensional variational (4DVAR) assimilation method with the particle filter (PF) ensemble data assimilation system, to produce a robust approach for dual-state-parameter estimation. In our proposed method, the Hybrid Ensemble and Variational Data Assimilation framework for Environmental systems (HEAVEN), we characterize the model structural uncertainty in addition to model parameter and input uncertainties. The sequential PF is formulated within the 4DVAR system to design a computationally efficient feedback mechanism throughout the assimilation period. In this framework, the 4DVAR optimization produces the maximum a posteriori estimate of state variables at the beginning of the assimilation window without the need to develop the adjoint of the forecast model. The 4DVAR solution is then perturbed by a newly defined prior error covariance matrix to generate an initial condition ensemble for the PF system to provide more accurate and reliable posterior distributions within the same assimilation window. The prior error covariance matrix is updated from one cycle to another over the main assimilation period to account for model structural uncertainty resulting in an improved estimation of posterior distribution. The premise of the presented approach is that it (1) accounts for all sources of uncertainties involved in hydrologic predictions, (2) uses a small ensemble size, and (3) precludes the particle degeneracy and sample impoverishment. The proposed method is applied on a nonlinear hydrologic model and the effectiveness, robustness, and reliability of the method is demonstrated for several river basins across the United States.

1. Introduction

Soil moisture and streamflow are among those key environmental variables that greatly affect flood forecasting, drought monitoring, and agricultural production that all collectively control the land and atmospheric system. Although, theoretically, these quantities can be estimated through hydrologic modeling, in practice they are often biased or erroneous due to the presence of uncertainties in all layers of hydrologic predictions. Data assimilation (DA) has been well received in the hydrologic community as one of the most effective methods in characterizing the aforementioned uncertainties while estimating parameters, prognostic, and diagnostic variables (Abbaszadeh et al., 2018; Clark et al., 2008; Moradkhani, Sorooshian, et al., 2005; Moradkhani et al., 2018; Pathiraja et al., 2016; Vrugt et al., 2006).

Generally, DA is defined as the application of Bayes' theorem to probabilistically condition the states of a dynamical model on observations (Moradkhani et al., 2018). A plethora of techniques is available to assimilate observations into a model for better initialization of the system and quantification of model parameter uncertainties. They all have some overlapping features making it difficult to define a clear-cut classification. Bayesian data assimilation seeks probabilistic estimates of state variables of interest in order to characterize their uncertainties. These probability distributions are sequentially adjusted according to the Bayes' theorem to better match the observations. In the hydrologic community, the best known and ubiquitous Bayesian approach is the ensemble Kalman filter (EnKF; Crow & Wood, 2003; De Lannoy et al., 2007; Moradkhani, Hsu, et al., 2005; Reichle et al., 2002). Despite the widespread use of the EnKF and its different variants in numerous hydrologic applications, it is subject to some inherent limitations that result in suboptimal performance of this technique. These include (1) the linear updating rule, (2) Gaussian assumption of errors in observations, and (3) violation of water balance (e.g., DeChant & Moradkhani, 2012; Matgen et al., 2010; Noh et al., 2011; Plaza et al., 2012). PF as an effective alternative to EnKF has emerged for applications in nonlinear and non-Gaussian systems (Abbaszadeh et al., 2018; DeChant & Moradkhani,