Title
Assessment of assimilating SMOS soil moisture information into a distributed hydrologic model

ISBN
9789292213053

Authors
Thorstensen, A
Nguyen, P
Hsu, K
et al.

Publication Date
2016-08-01

ASSESSMENT OF ASSIMILATING SMOS SOIL MOISTURE INFORMATION INTO A DISTRIBUTED HYDROLOGIC MODEL

Andrea Thorstensen[(1)], Phu Nguyen[(1,2)], Kuolin Hsu[(1)], Soroosh Sorooshian[(1)]

[(1)] *University of California Irvine, The Henry Samueli School of Engineering, Civil and Environmental Engineering zotcode 2175, University of California, Irvine, Irvine, CA 92697-2175 United States of America, Email:athorste@uci.edu; ndphu@uci.edu; kuolinh@uci.edu; soroosh@uci.edu*
[(2)] *Nong Lam University- Ho Chi Min City, Thu Duc District, Ho Chi Min City, Vietnam, Email: ndphu@hcmuaf.edu.vn*

ABSTRACT

The role that soil moisture plays in terms of modulating hydrologic processes including infiltration and runoff generation makes it an essential component to capture for hydrologic modeling. This work aims to leverage information gained from SMOS to improve surface soil moisture simulations in the Russian River Basin (California, U.S.A). The basin's complex terrain offers a rigorous testing ground for SMOS soil moisture products. Data from seven in situ observation sites are used to assess model performance after assimilating SMOS-based soil saturation ratios. For a comparison of "best case" scenarios, the in situ observations themselves are assimilated. Results show that SMOS assimilated simulations shows modest improvement at most in situ locations. Despite the observed decrease in model performance at some locations, overall performance of simulations assimilated with SMOS-based saturation ratios remains high. Findings suggest that even in a complex environment, useful information may be extracted from SMOS estimates for hydrologic modeling.

1. INTRODUCTION

Several studies have already made efforts to examine how the incorporation of soil moisture observations through data assimilation into hydrologic, land surface, and hillslope models improves estimates and predictions of the soil moisture state. Using a distributed soil-vegetation-atmosphere transfer (SVAT) model and EnKF, [1] found that all data assimilation runs in their study provide an improvement over non assimilation runs, even when observation frequency was reduced from daily to once every 5 days. This result is particularly encouraging for satellite based soil moisture applications, since these observations may only be available every 1-3 days for a given location. Reference [2] also employs the EnKF but uses a synthetic experiment with a land surface model. These results yield reasonable soil moisture estimates even with relatively few ensemble members, suggesting a perhaps computationally efficient method. Using the Noah LSM, [3] express improvement, especially in the top soil layer estimates by incorporating AMSR-E surface soil moisture retrievals in a semi-arid region. Reference [4] shows significant reduction of surface soil moisture bias with some reduction of RMSE for over half the watershed in the hillslope tRIBS-VEGGIE model, which they use for assimilation of synthetic 3 km Soil Moisture Active Passive (SMAP) radar data.

2. MODEL

The distributed hydrologic model used in the study is the Hydrology Laboratory Research Distributed Hydrologic Model (HL-RDHM) developed by the U.S. National Weather Office of Hydrologic Development. HL-RDHM was developed and implemented for use over the continental United States on the Hydrologic Rainfall Analysis Project (HRAP) grid. The model can be run at a spatial resolution of 1 HRAP (~4 km), ½ HRAP, or ¼ HRAP. The model can be run at any desired time step as well [5]. For more detailed information about the general structure of HL-RDHM, readers are referred to [6], [5], and [7]. For this study, the model was run at 1 HRAP spatial resolution with an hourly time step.

Central to HL-RDHM and the work presented here is the Sacramento Soil Moisture Accounting (SAC-SMA) model (Fig. 1). HL-RDHM utilizes *a priori* SAC-SMA parameters derived from soil and land use data at each model pixel [8]. Recent developments include the estimation of a physically meaningful soil moisture profile and evapotranspiration from the soil column. Through this conversion, physics that loosely mimic those present in the Noah Land Surface Model allow for a heat transfer component to account for frozen ground processes to take place at each soil layer [9].

Proc. 'Living Planet Symposium 2016', Prague, Czech Republic, 9–13 May 2016 (ESA SP-740, August 2016)

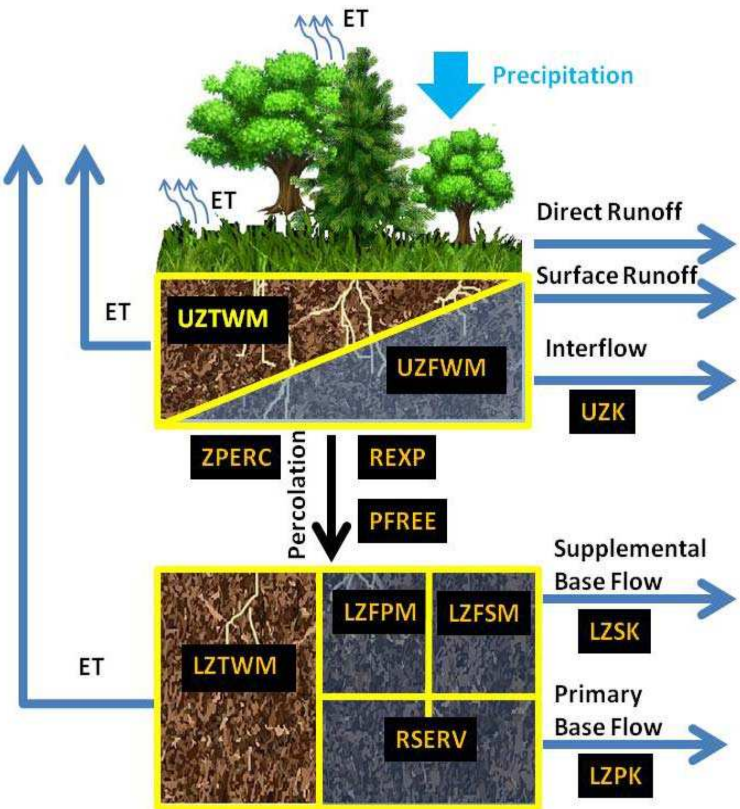

Figure 1. Formulation of basic SAC-SMA model and its parameters (black boxes) most relevant to soil moisture.

Advanced evapotranspiration estimation was introduced in the version dubbed the Sacramento Soil Moisture Accounting Heat Transfer component for Enhanced Evapotranspiration (SAC-HTET) [10]. Within SAC-HTET, accounting for photosynthetically active radiation, soil moisture and vapor pressure deficits, and air temperature takes place. Physical representations of

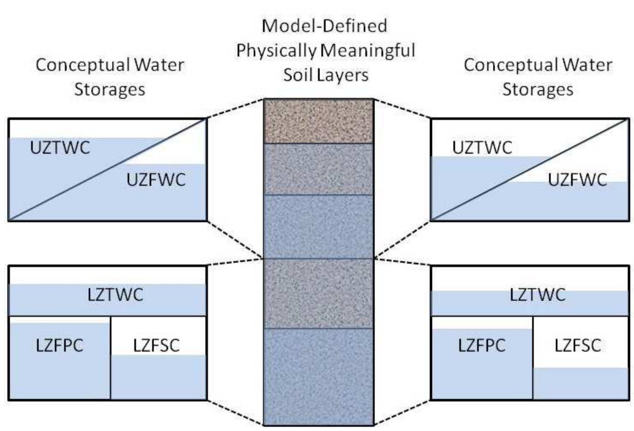

Figure 2. Conversion of SAC conceptual storages to model-prescribed physically meaningful soil layers

these additional variables are estimated through empirical relationships in order to keep the input data requirements low.

Following the adjustments to the soil moisture state at various physical layers, the estimates are converted back to SAC-SMA conceptual storages (Fig. 2) and changes due to free water exchange and runoff are made.

From these model-defined physically meaningful layers, interpolation to any desired depth within the model boundaries can be retrieved for comparisons to observations. However, user-defined depths have no bearing on the calculations within the model.

3. STUDY AREA

This study is conducted over the Russian River Basin in Northern CA and utilizes the United States National Oceanic and Atmospheric Administration ((NOAA) Hydrometeorology Testbed (HMT) program in situ measurements [11]. The network provides observations of soil moisture and temperature at several depths, but this study makes use of only the 10 cm depth (the shallowest layer for most of the sites). Using only the topmost observation layer is done to mimic what is captured by satellite-based retrievals.

Observation sites are dispersed throughout the approximately 3,800 km2 basin and include Willits (WLS) and Potter Valley (PTV) in the upper basin, Hopland (HLD) and Lake Sonoma (LSN) in the central basin, and Cazadero (CZC), Rio Nido (ROD), and Healdsburg (HBG) in the lower basin (Fig. 3).

Although the CZC site does not properly sit within the drainage area of the Russian River Basin, the observations from this site are still useful during for spreading innovation to pixels that are within the basin boundaries. Furthermore, as the model is run completely in “unconnected mode” with no routing scheme demanding lateral pixel interaction, inclusion of CZC in the procedure carries no adverse implications.

Figure 3. Russian River Basin and HMT sites.

4. METHOD

Soil saturation ratios are used as a substitute for soil moisture values to partially circumvent scaling discrepancies among SMOS pixel resolution, HL-RDHM resolution, and the point in situ measurements (although SMOS soil saturation ratios are further disaggregated using spatial patterns from HL-RDHM control runs). This is not an uncommon practice, and is the recommended this strategy for similar purposes using a lumped variant of SAC-HTET [12].

In the case of both in situ observations and satellite-based estimates, observations are not available at all locations within the basin at every observation time step (Fig. 4). Therefore, a recursive EnKF strategy is developed to update all pixels in the study basin in a manner that is respectful of the spatially heterogeneous nature of soil moisture dynamics. This is in contrast to previous studies that have (or assume) observations available at all locations (i.e. [4], [13], [14], [15]) or assign the same observation to all pixels but assign varying degrees of uncertainty according to spatial variations in soil moisture [16].

Considering a single observed model pixel scale, the state equation in the EnKF for this work is the running of HL-RDHM to project conceptual states to the next time step. The observations are observed soil moisture estimates either from in situ HMT soil moisture probes or SMOS observations disaggregated to the HRAP scale.

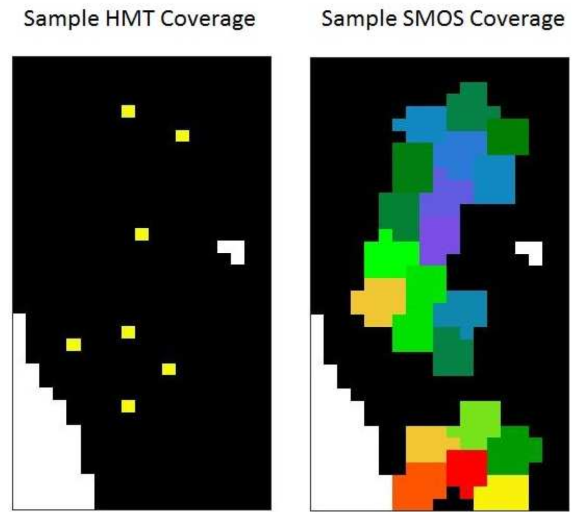

Figure 4. Sample coverage (colored areas) for a single time step of HMT stations (left) and SMOS coverage (right) in the Russian River Basin. Black area represents entire land domain that is being modeled but not observed, white areas represent water features that produce no simulation.

However, to reach the point of the conceptual storages at all simulated pixels being updated by the soil moisture data assimilation process, a second filtering step in which there is a shift in the observation equation is proposed. For this second step, the states are now the conceptual storages of the "unobserved" pixels and the observations are changed to be the ensemble of updated conceptual storages at model locations collocated with observations.

To evaluate the impact and practicality of soil moisture assimilation with the proposed double EnKF, the model is run for a 1-year spinup period (2012) and a 1-year data assimilation period (2013). Precipitation and temperature data for model forcing come from the California-Nevada River Forecast Center. A 25 member ensemble is generated by sampling a Gaussian distribution with a mean equal to that of the states at the end of the 1-year spinup period, and a variance of 0.25. This sizeable variance was chosen to represent a large initial uncertainty in the possible range of 0 to 1 and approaches a "worst case" scenario for prior understanding of the state. Ensemble members are sustained by perturbing the precipitation and temperature forcing data with noise sampled from a normal distribution. Similarly, Gaussian noise that reflects uncertainty associated with the soil moisture measurement is added to the HMT soil moisture observations and disaggregated SMOS estimates.

The evaluation is performed at the 7 HL-RDHM pixels collocated with the HMT observation sites in two phases. The first phase uses the same observation set for the assimilation and the assessment to test the impact of the first filtering step in which the observations come from the soil moisture probes. In this case, the states being estimated are the SAC-SMA upper zone conceptual storages at the pixels collocated with the volumetric soil moisture observations. For the second phase, only 6 of the 7 observation sites are used in the complete double EnKF process, with the 7th saved for a validation of the spreading of the innovation to "unobserved" pixels. For both parts, the RMSE, correlation, bias, and NSE are used as performance metrics.

The double EnKF is repeated using SMOS observations for assimilation rather than HMT station observations. There is no separate validation stage for these observations, as each SMOS pixel is not necessarily retrieved at the same locations for at each observation time. That is, at some assimilation time steps, a given pixel might be collocated with an observation and at others it may not and must rely on the second filtering step. Although SMOS observations are viable for the ~5 cm depth and are assimilated into HL-RDHM accordingly, they are compared to HMT observation sets at 10 cm depths.

5. RESULTS

Figs. 5, 6, and 7 feature results from both phases of the evaluation for the upper, central, and lower basin sites respectively. Phase 1 is represented on left panels and phase 2 is shown on the right. Control runs and HMT observations are also included in all plots. From Tab. 1, it can be seen that 5 of the 7 sites showed at least some improvement across all statistics for the experiment that included collocated observations (phase 1). The exceptions came from ROD and HBG, which show a slight degradation in bias. It is worth noting here that the control runs at these two sites were already

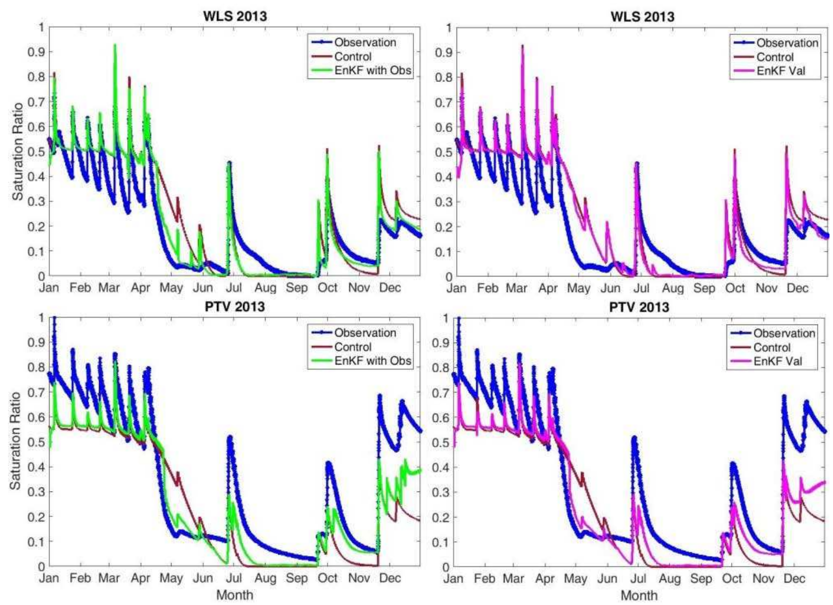

Figure 5. Double EnKF soil saturation ratio results at 10 cm for upper basin observation sites in the Russian River Basin. Left: Results with observations collocated at the site assimilated. Right: Validation of the second filter step with collocated observations removed from the assimilation.

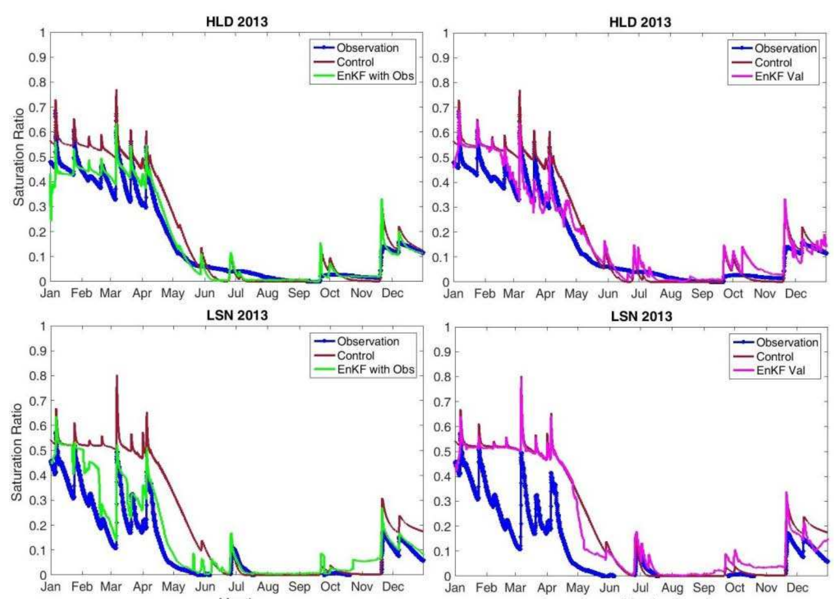

Figure 6. Same as Figure 5, but for central basin observation sites.

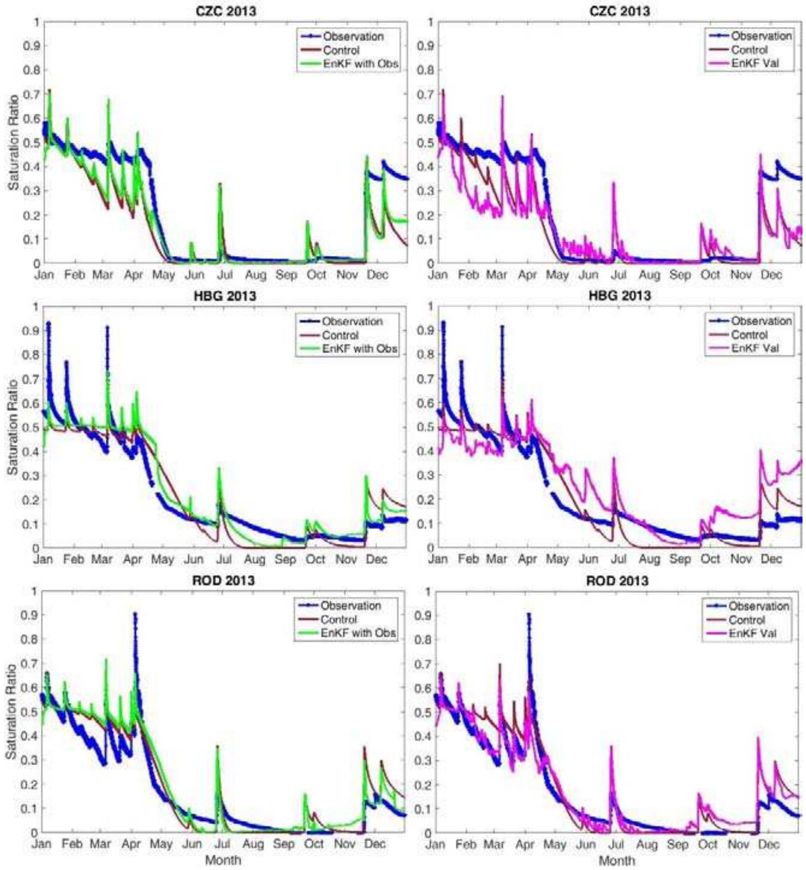

Figure 7. Same as Figure 5, but for lower basin observation sites.

performing high across all statistics, and even though the bias suffered mildly, the predictive capability in the form of NSE was unchanged at ROD and even had a modest 3% improvement at HBG. For the remaining sites, a 3-61% RMSE reduction, a 0-16% correlation increase, a 4-66% bias decrease, and a 1-186% NSE increase is seen. The site that clearly benefited the most from data assimilation of its collocated observation was LSN, which went from -0.84 NSE to 0.72 in addition to dramatic improvement in the other three metrics as well.

Where the experiments that use collocated observations for the assimilation and evaluation provide valuable insight as a sanity check/general proof of concept, their impact basin-wide is minimal unless the innovation spreading step can be demonstrated to be effective. This is especially true given the relatively sparse nature of the in situ observation network of this basin. Removing one station at a time to treat its corresponding HL-RDHM pixel as "unobserved" allows for validation of the second spreading step.

It is expected that a pixel collocated with an observation will exhibit improvement in soil moisture simulation, however, the minimalist goal for unobserved pixels after assimilation is that performance is not worse than the control run. The results for the validation are mixed in this sense with 4 sites (WLS, HLD, LSN, and ROD) outperforming the control run, 1 site (PTV) larger unchanged from the control run, and 2 sites (CZC and HBG) performing measurably worse than the control run. Even though the two sites downgraded, they still outperformed the LSN site validation, which even after the assimilation has an unacceptable NSE of -59 (due to a large bias), and even slightly edged out the PTV site in terms of NSE. For this reason, the validation experiments are still crowned "more successful than detrimental" overall.

Simulation results for SMOS assimilation experiments are presented in Figs. 8 through 10. The model control run and 10 cm depth HMT observations are provided in each plot. Noteworthy is the model tendency to overestimate soil moisture during the first large spring-time dry down period (April and May), especially at WLS, PTV, HLD, LSN, and HBG. Generally, the SMOS assimilation is able to push the ensemble mean toward the observation in this case. Improvement is also visible at sites with a control model run that overestimates in the first four months of the year (WLS, HLD, LSN, and ROD).

Several striking pessimistic features appear in the assimilation runs as well. At PTV for example, a large dip in soil saturation ratio appears at the end of February, despite the control run already underestimating at that site. Similarly, the SMOS assimilation simulation at HBG jumps in early November and remains higher than the control run even though the control run overestimates this until the end of the year. These features could be due to a number of factors including spurious correlations with observations not located at the pixel site, issues with

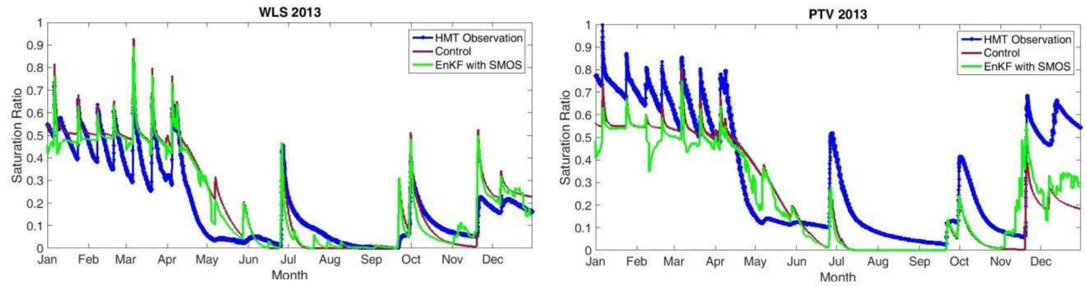

Figure 8. Double EnKF soil saturation ratio results at 5 cm for upper basin observation sites in the Russian River Basin using SMOS observations for assimilation.

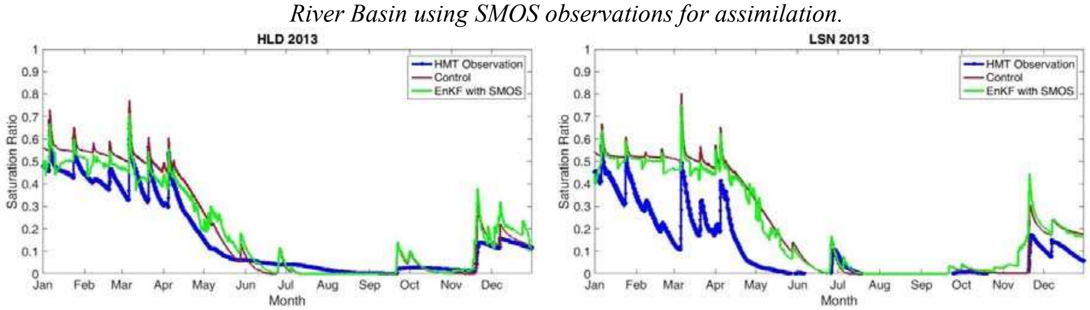

Figure 9. Same as Figure 8, but for central basin observation sites.

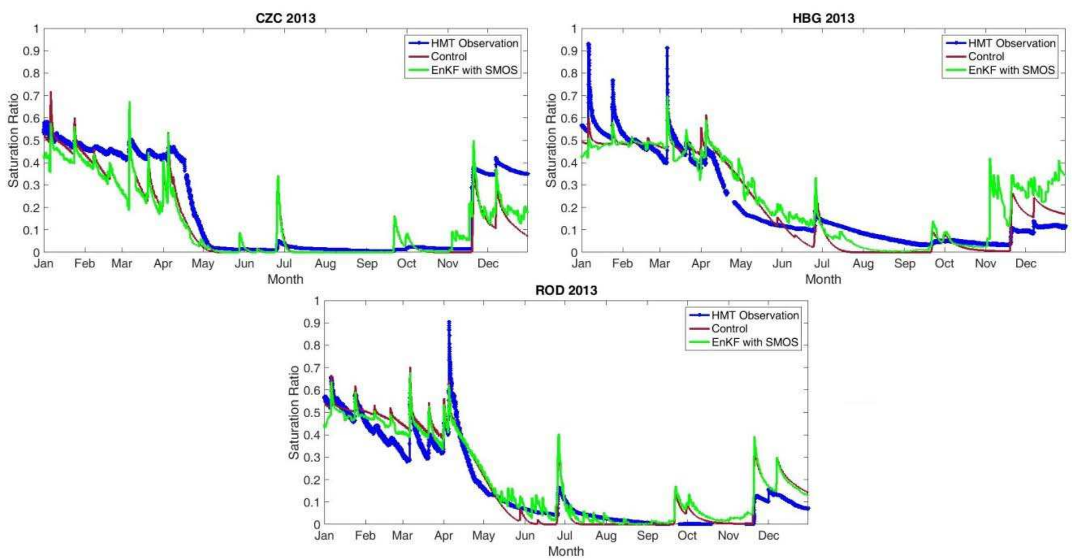

Figure 10. Same as Figure 8, but for lower basin observation sites.

SMOS observations at the pixel location, or even improper model parameter specification.

A statistical summary of control runs along with simulations assimilated with SMOS saturation ratios are presented in Tab 1. It is stressed that the catch with these results is that the HMT observations are used as a baseline and are at a 10 cm depth.

With the exception of HBG, all sites show an RMSE decrease (or remained unchanged at PTV). Correlation results were more mixed, with three sites showing a decrease in correlation and four showing an increase. With the exception of HBG, changes in either direction were less than 2.5%. Correlations for all simulations (with SMOS soil moisture assimilation or not) were high with the minimum value of 0.83 belonging to the control run of LSN. While bias results for HBG and ROD increased and PTV remained unchanged, improvement at the other stations ranged from 6% to 35%. NSE improved at all sites except for HBG with improvement ranging from 2% to 40%.

Table 1

Site ID	Model Run	RMSE [fract]	Corr	Bias	NSE
WLS	Control	0.10	0.91	0.21	0.70
	EnKF HMT	0.08	0.95	0.10	0.83
	HMT Validation	0.09	0.91	0.15	0.73
	EnKF SMOS	0.09	0.90	0.13	0.75
PTV	Control	0.17	0.86	-0.29	0.61
	EnKF HMT	0.12	0.95	-0.24	0.80
	HMT Validation	0.17	0.85	-0.27	0.59
	EnKF SMOS	0.17	0.88	-0.30	0.61
HLD	Control	0.07	0.98	0.26	0.82
	EnKF HMT	0.07	0.98	0.25	0.83
	HMT Validation	0.06	0.97	0.16	0.90
	EnKF SMOS	0.06	0.97	0.22	0.87
LSN	Control	0.20	0.83	1.00	-0.84
	EnKF HMT	0.08	0.97	0.34	0.72
	HMT Validation	0.18	0.84	0.90	-0.59
	EnKF SMOS	0.18	0.85	0.90	-0.49
CZC	Control	0.10	0.91	-0.29	0.75
	EnKF HMT	0.09	0.94	-0.24	0.82
	HMT Validation	0.13	0.87	-0.31	0.63
	EnKF SMOS	0.10	0.92	-0.29	0.75
HBG	Control	0.07	0.94	-0.02	0.88
	EnKF HMT	0.06	0.96	0.06	0.91
	HMT Validation	0.11	0.85	0.22	0.67
	EnKF SMOS	0.10	0.87	0.14	0.72
ROD	Control	0.07	0.95	0.12	0.87
	EnKF HMT	0.07	0.96	0.13	0.87
	HMT Validation	0.06	0.95	0.05	0.90
	EnKF SMOS	0.06	0.96	0.15	0.89

It should be noted that although LSN enjoyed the largest improvement in NSE through assimilation of SMOS soil moisture information, the simulation still produced a value of -0.49, which indicates no predictive ability.

6. DISCUSSION

A double EnKF technique was introduced as a means to update the conceptual storages at every pixel within the model domain without assuming observations are available at every location, and without having to rely on interpolation of soil moisture observations prior to assimilation. In the first step of the assimilation, conceptual model states at "observed" pixels are updated with observed near surface soil moisture observations. In the second step, the remaining pixels are update by treating the ensemble of the adjusted states from step one as the observations in the EnKF process. Tests with the HMT sites show consistent improvement for step one of the procedure, and the validation phase with station removal revealed mostly favorable results over the control run.

While it is expected that the more observations to contribute to the update of an unobserved state the better, given the formulation of this second EnKF step, there must be a correlation between the observation and unobserved state in order for it to be useful. Therefore, strategies related to maximizing the benefit of the most relevant observations to a particular unobserved state could be further investigated. This notion of data selection is discussed in [17] who utilize a cutoff radius to distinguish which analyzed points should be considered impacted by each observation in an atmospheric model. They also stress that the further a point becomes from an observation, the potential positive impact from updating can be expected to be small. The analogy to the hydrologic application in this sense is that the more dissimilar an observation location is (due to distance from the observation, physical properties leading to different drying rates, or differences in recent meteorological influences), the less of a positive impact that observation will have on an analyzed pixel. Although not fully examined here since performance in the in situ validation investigation was largely positive, localization techniques may be beneficial in preventing detrimental prediction skill results as seen at the HBG and CZC sites during the validation portion.

Following the development of the double EnKF procedure via HMT data was the testing of the procedure using SMOS satellite-based estimates. These tests were also evaluated against the HMT observations. Overall, the improvement over the control run was largely underwhelming. Nonetheless, with the exception of the HBG site, all of the sites experienced minor improvement in predictive capability as expressed by the NSE. Of the cases that did outperform the control run, bias showed largest degree of improvement. The slight decrease in correlation at some sites (and large decrease at the HBG site) may be attributed to the frequency of assimilation time steps, which was lower than the HMT study. Although improvement is marginal, this study suggests there is valuable information contained in the SMOS soil moisture retrievals for the Russian River Basin, despite the fairly complex terrain challenging the capabilities of the retrieval. However, SMOS assimilation in the context presented here does not render itself particularly useful.

7. ACKNOWLEDGEMENTS

This research was conducted with Government support under and awarded by DoD, National Defense Science and Engineering Graduate (NDSEG) Fellowship, 32 CFR 168a and NOAA National Weather Service.

8. REFERENCES

1. Merlin, O., Chehbouni, A., Boulet, G., and Kerr, Y. (2006). Assimilation of Disaggregated Microwave Soil Moisture into a Hydrologic Model Using Coarse-Scale Meteorological Data. *J. Hydrometeor.* **7**, 1308-1322. doi: http://dx.doi.org/10.1175/JHM552.1

2. Reichle, R.H., McLaughlin, D.B., Entekhabi, D. (2002). Hydrologic Data Assimilation with the Ensemble Kalman Filter. *Mon. Weather Rev.* **130**, 103-114. doi: http://dx.doi.org/ 10.1175/1520-0493(2002)130<0103:HDAWTE >2.0.CO;2

3. Hsu, K., Li, J., and Sorooshian, S. (2012). To Improve Model Soil Moisture Estimation in Arid/Semi-Arid Region Using In Situ and Remote Sensing Information. *Paddy and Water Environment.* **10**(3), 165-173. doi:10.1007/ s10333-011-0308-9

4. Flores, A.N., Bras, R.L., and Entekhabi, D. (2012). Hydrologic Data Assimilation with a Hillslope Scale-Resolving Model and L Band Radar Observations: Synthetic Experiments with the Ensemble Kalman Filter. *Water Resour. Res.*, **48**, doi:W08509.10.1029/2011 WR011500

5. National Weather Service (NWS), 2011: Hydrology Laboratory-Research Distributed Hydrologic Model (HL-RDHM) User Manual V. 3.2.0

6. Koren, V., Reed, S., Smith, M., Zhang, Z., and Seo, D.J. (2004). Hydrology Laboratory Research Modeling System (HL-RMS) of the US National Weather Service. *J. Hydrol.*, **291**, 297–318. doi:10.1016/j.jhydrol.2003.12.039

7. Smith, M., Koren, V., Zhang, Z., Zhang, Y., Reed, S., Cui, Z., Moreda, F., Cosgrove, B., Mizukami, N., Anderson, E., DMIP 2 Participants. (2012). Results of the DMIP 2 Oklahoma Experiments. *J. Hydrol.*,**418–419**,17-48. doi:10.1016/j.jhydrol.2011.08.056

8. Koren, V., Smith, M., and Duan, Q. (2003): Use of a priori Parameter Estimates in the Derivation of Spatially Consistent Parameter Sets of Rainfall–Runoff Models, Q. Duan, H. Gupta, S. Sorooshian, A. Rousseau, and R. Turcotte, Calibration of Watershed Models: Water Science and Application Series, **6**, *American Geophysical Union,*239-254.doi: 10.1002/9781118665671.ch18

9. Koren, V., Smith, M., Cui, Z., Cosgrove, B. (2007). Physically-Based Modifications to the Sacramento Soil Moisture Accounting Model: Modeling the Effects of Frozen Ground on the Rainfall-Runoff Process. *NOAA Tech. Report NWS 52*

10. Koren, V., Smith, M., Cui, Z., Cosgrove, B., Werner, K., and Zamora, R. (2010). Modification of Sacramento Soil Moisture Accounting Heat Transfer Component (SAC-HT) for Enhanced Evapotranspiration. *NOAA Tech. Report NWS 53*

11. Zamora, R., Ralph, F.M., Clar, E. and Schneider, T. (2011). The NOAA HydrometeorologyTestbed Soil Moisture Observing Networks:Design, Instrumentation, and Preliminary Results. *J. Atmos. Oceanic Technol.* **28**, 1129-1140. doi:10.1175/ 2010JTECHA1465.1

12. Koren, V., Moreda, F., and Smith, M. (2008). Use of Soil Moisture Observatins to Improve Parameter Consistency in Watershed Calibration*, Phys. Chem.EARTH PR A/B/C,* **33**(17-18), 1068-1080. doi:10.1016/j.pce.2008.01.003

13. Han, E., Merwade, V., and Heathman, G. C. (2012). Implementation of Surface Soil Moisture Data Assimilation with Watershed Scale Distributed Hydrological Model, *J. Hydrol.*, **416-417**, 98-117. doi:10.1016/j.jhydrol.2011.11.039

14. Sutanudjaja, E. H., van Beek, L. P. H., de Jong, S. M., van Geer, F. C., and Bierkens, M. F. P. (2014). Calibrating a Large-Extent High-Resolution Coupled Groundwater-Land Surface Model using Soil Moisture and Discharge Data, *Water Resour. Res.,* **50**, 687-705. doi: 10.1002/2013WR013807

15. Wanders, N., Bierkens, M., de Jong, S., de Roo, A. , and Kaarssenberg, D. (2014). The Benefits of Using Remotely Sensed Soil Moisture in Parameter Identification of Large-Scale Hydrological Models, *Water Resour. Res.*, **50**, 6874-6891. doi: 10.1002/2013WR014639

16. Lee, H., Seo, D., and Koren, V. (2011). Assimilation of Streamflow and in situ Soil Moisture Data into Operational Distributed Hydrologic Models: Effects of Uncertainties in the Data and Initial Model Soil Moisture States, *Adv. Water Resour.*, **34**,1597-1615.doi:10.1016/j.advwatres.2011.08.012

17. Hautekamer, P. L. and Mitchell, H. L. (1998). Data Assimilation Using an Ensemble Kalman Filter Technique. *Mon. Weather Rev.*, **126** (3), 796-811. doi: http://dx.doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2