

The Value of SMAP for Long-Term Soil Moisture Estimation With the Help of Deep Learning

Kuai Fang^{ID}, Ming Pan^{ID}, and Chaopeng Shen^{ID}

Abstract—The Soil Moisture Active Passive (SMAP) mission measures important soil moisture data globally. SMAP's products might not always perform better than land surface models (LSM) when evaluated against *in situ* measurements. However, we hypothesize that SMAP presents added value for long-term soil moisture estimation in a data fusion setting as evaluated by *in situ* data. Here, with the help of a time series deep learning (DL) method, we created a seamlessly extended SMAP data set to test this hypothesis and, importantly, gauge whether such benefits extend to years beyond SMAP's limited lifespan. We first show that the DL model, called long short-term memory (LSTM), can extrapolate SMAP for several years and the results are similar to the training period. We obtained prolongation results with low-performance degradation where SMAP itself matches well with *in situ* data. Interannual trends of root-zone soil moisture are surprisingly well captured by LSTM. In some cases, LSTM's performance is limited by SMAP, whose main issue appears to be its shallow sensing depth. Despite this limitation, a simple average between LSTM and an LSM Noah frequently outperforms Noah alone. Moreover, Noah combined with LSTM is more skillful than when it is combined with another LSM. Over sparsely instrumented sites, the Noah-LSTM combination shows a stronger edge. Our results verified the value of LSTM-extended SMAP data. Moreover, DL is completely data driven and does not require structural assumptions. As such, it has its unique potential for long-term projections and may be applied synergistically with other model-data integration techniques.

Index Terms—Deep learning (DL), hindcasting, Soil Moisture Active Passive (SMAP), soil moisture.

I. INTRODUCTION

SOIL moisture (θ) critically controls various environmental and ecosystem processes, such as photosynthesis, evapotranspiration, runoff, soil respiration [1], flood [2], and land-atmosphere interactions [3]. For agricultural planning and other purposes, soil moisture is routinely measured by *in situ* monitoring networks. However, these networks have limited coverage in space and are typically sparsely instrumented. Soil moisture has a high spatial variability so that the scale of

these measurements poorly matches the needs of large-scale modeling and monitoring.

To offer large-scale monitoring of soil moisture, multiple soil moisture sensing satellite missions have been launched. Notable missions include the Advanced Microwave Scanning Radiometer for EOS (AMSR-E) [4], the Advanced Scatterometer [5], the Soil Moisture and Ocean Salinity (SMOS) [6], and the Soil Moisture Active Passive (SMAP) [7], among others. The data from these missions have been utilized for a variety of applications in hydrology, ecology, meteorology, and agriculture [8]. Among these missions, SMOS was reported to suffer significantly from unexpected radio frequency interference (RFI) [9], [10]. Launched in the year 2015, SMAP mitigates the RFI effects by utilizing the combination of spaceflight instrument hardware and ground-based science processing algorithms [11]. Initial evaluations suggest that SMAP's passive product has reached its design accuracy (~ 0.04) at most validation sites [12]–[14]. In some parts of the world, the quality of SMAP product might not exceed the modeled soil moisture from state-of-the-art LSMs forced with high-quality meteorological data [15], [16]. However, SMAP provides an independent source of observation derived from a different physical mechanism (radiative transfer). Independent information often helps reducing errors and uncertainties in a data fusion setting [17].

While SMAP has a global coverage, its uneven revisit schedule and short timespan significantly limit its utility. Longer records are needed to study drought trends or relate soil moisture to past environmental events, e.g., wildfires and landslides. Satellite data can also improve the current soil moisture simulations via data assimilation (DA) [18], [19], but DA has little impact on long-term soil moisture estimation. Previously, Liu *et al.* [20] merged multiple satellite products by matching earlier satellite data's cumulative distribution function (CDF) to AMSR-E's. They also replaced older product's seasonality based on AMSR-E's when overlaps exist. Later work further weighed different satellite data based on their random error variance [8], [21]. While this series of work provided valuable satellite soil moisture estimates back to 1979, the data quality is still affected by the retrieval skills of earlier satellites. The retrieval skills have been much improved in the more recent, dedicated soil moisture missions (SMOS and SMAP) [16]. As some earlier satellites have limited or no overlap with SMOS or SMAP, it is difficult to apply data fusion techniques such as CDF-matching with SMOS or SMAP.

In our earlier work, we developed a deep learning (DL)-based dynamical modeling system, which predicts the

Manuscript received July 24, 2018; revised August 15, 2018; accepted September 1, 2018. This work was supported in part by the Penn State Institute of CyberScience and in part by the National Science Foundation under Grant 1832294. The work of M. Pan was supported by the National Aeronautics and Space Administration (NASA) under Grant NNX14AH92G. (Corresponding author: Chaopeng Shen.)

K. Fang and C. Shen are with the Department of Civil and Environmental Engineering, Pennsylvania State University, University Park, PA 16802 USA (e-mail: cshen@engr.psu.edu).

M. Pan is with the Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ 08544 USA.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2018.2872131

SMAP surface soil moisture product with high fidelity [22]. DL refers to large-sized neural networks with the ability to automatically extract features using its intermediate layers [23]. DL is rapidly transforming various industries and scientific disciplines (see a review in [24]). The inputs to our model include atmospheric forcing data, static physiographic attributes, and (optionally) simulated fluxes and states from LSMs. The model was built on a form of deep recurrent neural network (RNN) called the LSTM architecture, which is typically employed for natural language processing. We showed that, when trained with a year of data, this system was capable of predicting SMAP level 3 passive data product in another year. LSTM can extend SMAP to spatiotemporally seamless coverage of continental US (CONUS) with high fidelity to SMAP. Despite its large parameter space, the system had higher accuracy and better robustness than simpler statistical methods. We called our method a prolongation of SMAP via DL. However, it must be noted that DL only captures the systematic soil moisture response pattern as observed by SMAP. It cannot extend the stochastic signal in satellite observations that could be used to correct forcing errors.

Our main hypothesis is that LSTM-prolonged SMAP data presents added value beyond LSMs and combining it with LSMs will help improve the long-term soil moisture predictions, as verified by *in situ* data. In addition, we seek to answer several important technical questions. First, we study whether the DL model can perform well in a multiyear scale projection setting. We check if it can avoid performance degradation over time like DA would. When the training data only contain two or three years of records, it is unclear whether multiyear trends can be learned and predicted in case they exist. Second, all the previous evaluation metrics of LSTM were calculated against SMAP data. The DL system has yet to show that it is reasonably close to the *in situ* data. If SMAP data have significant noise, the extent to which such errors would misinform the training process and amplify during long-term simulation has not been explored. Finally, SMAP also offers a root-zone soil moisture (RZSM, 0–100 cm depth) product via assimilating surface brightness temperature into model simulations [25]. Past evaluations show that the RZSM product has achieved the design accuracy of $0.04 \text{ cm}^3/\text{cm}^3$ [14]. We ask whether LSTM can also reproduce RZSM, which has a much longer system memory.

This paper uses LSTM to hindcast SMAP surface and root-zone the soil moisture products. In contrast to [22], here we:

- 1) prolonged the SMAP data to several years beyond the available data to test the model's ability to maintain good performance in long-term estimation;
- 2) additionally evaluated the RZSM hindcast product;
- 3) combined LSTM with Noah to test our main hypothesis above;
- 4) evaluated different products against *in situ* data instead of against SMAP;
- 5) removed Noah simulation from the list of inputs.

The *in situ* soil moisture data include both SMAP core validation sites [26], where dense measurements provide a means of upscaling, and a network of sparsely instrumented

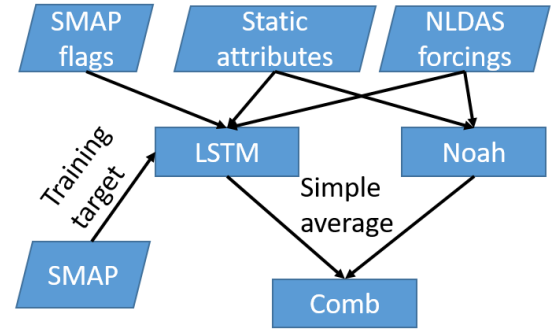


Fig. 1. Information flow diagram for the models compared in this paper. Both Noah and LSTM use forcing data from North American Land DA System (NLDAS). LSTM uses SMAP as the main training target along with land surface characteristic masks from the SMAP mission among the inputs. The static input attributes to LSTM are described in Section II-A.

sites. We examined whether some important characteristics of data, e.g., interannual trends, are reproduced in the hindcasts. Here, we chose to test the hindcast product because it can be evaluated using *in situ* data for its long-term behaviors. Given the future climate projections, a similar system can be similarly employed to make future soil moisture projections.

II. METHODS AND DATA SETS

The surface and root-zone LSTM models were both trained from April 2015 (SMAP launch) to April 2018. Inputs to the DL models include atmospheric forcings and constant geophysical attributes (Fig. 1). We then ran the trained models in forward mode using atmospheric forcings from 2010 to 2018. The hindcast products are validated over SMAP core validation sites (from 2012) and sparse networks (using available data since 2010) over CONUS.

A. Data Sets: Targets and Predictors

We used the SMAP level 3 radiometer product (L3SMP, version 4) as the target of surface soil moisture. This product is based on L-band passive observations of surface brightness temperature and was designed to represent top 5 cm soil moisture on a 36 km Equal-Area Scalable Earth Grid (EASE-Grid). As the RZSM cannot be directly sensed by satellite, SMAP level 4 soil moisture product (L4SM, version 3) integrates SMAP L-band brightness temperature observations into the catchment land model in Goddard Earth Observing System (GEOS-5) [27], [28] through DA. The model outputs column soil moisture of roughly 0–100 cm depth on a 9 km EASE-Grid, which is used as the target of our root-zone LSTM model.

LSTM receives three kinds of inputs, which have been selected from the list of inputs used in [22] via sensitivity analysis. First, atmospheric forcings, including precipitation, temperature, long-wave and short-wave radiation, specific humidity, and wind speed, were obtained from NLDAS phase II (NLDAS-2) [29]. Second, for static physiographic attributes, we obtained soil properties such as sand, silt, and clay percentages, bulk density, and soil water capacity from World

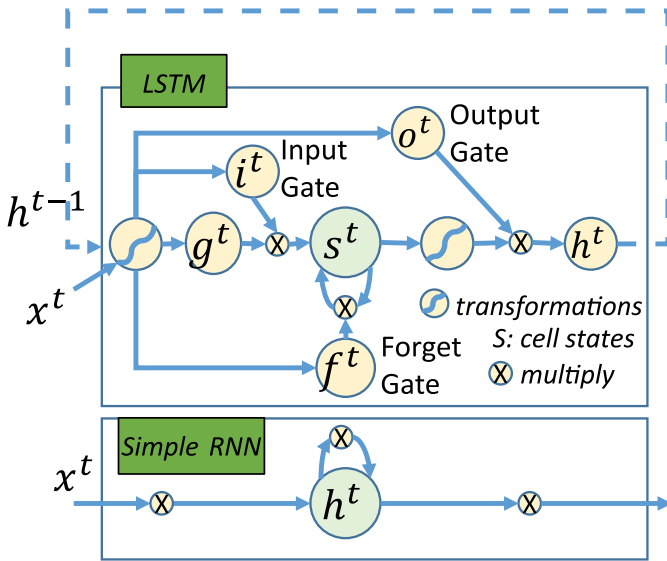


Fig. 2. Comparing an LSTM unit with simple RNN. The transformations from inputs to i , f , and o are sigmoidal functions. From inputs to g and from s to h , the transformation is $\tanh$. $\otimes$ means multiplication by weights. The figure is reproduced from [22] with reprint permission.

Soil Information (ISRIC-WISE) database [30]. The third type of inputs are SMAP masks that indicate mountainous terrain, ice, surface roughness, urban areas, water bodies, land cover classes, and vegetation density. Incorporating these masks as inputs is hypothesized to enable LSTM to implicitly assign less importance to high-uncertainty regions. LSTM is thought to automatically analyze, based on the data, which mask is useful and discards the unimportant masks. In [22], we included time-averaged SMAP data quality mask as an input. However, since our preliminary sensitivity test suggests this field was not influential, we did not include it in the model reported here. Its information has already been encompassed in the other inputs.

NLDAS-2 provides several LSM which simulate land surface states and fluxes. We selected Noah for comparison and model fusion purposes [31]. Since here one of our objectives is to measure the value of the independent information from LSTM, we did not include the Noah-simulated variables as LSTM inputs.

B. Time-Series Deep Learning Model

LSTM is an RNN that underlies the sequence learning applications, such as handwriting recognition [32], and speech recognition in digital assistant, such as Google Voice [33]. As a type of RNN, LSTM makes use of sequential information by updating states based on both inputs of the current time step (x^t) and network states of previous time steps. However, the distinguishing architectural difference from a “vanilla” RNN is that LSTM has “memory states” units and “gates” that are trained to decide when to forget some information, when and what to output, and what inputs are used (Fig. 2). These gates are also responsible for learning to engineer features that are important for soil moisture memory. Following the notation in [34], we can write an LSTM as LSTM: $x_0^{(t)}, h^{(t-1)}, s^{(t-1)} \rightarrow h^{(t)}, s^{(t)}$. The sequential update formulas are given as follows.

1) Input Transf

$$x = \text{ReLU}(W_{xx}x_0^{(t)} + b_{xx}). \quad (1)$$

2) Input Node

$$g^{(t)} = \tanh(W_{gx}\mathcal{D}(x^{(t)}) + W_{gh}\mathcal{D}(h^{(t-1)}) + b_g). \quad (2)$$

3) Input Gate

$$i^{(t)} = \sigma(W_{ix}\mathcal{D}(x^{(t)}) + W_{ih}\mathcal{D}(h^{(t-1)}) + b_i). \quad (3)$$

4) Forget Gate

$$f^{(t)} = \sigma(W_{fx}\mathcal{D}(x^{(t)}) + W_{fh}\mathcal{D}(h^{(t-1)}) + b_f). \quad (4)$$

5) Output Gate

$$o^{(t)} = \sigma(W_{ox}\mathcal{D}(x^{(t)}) + W_{oh}\mathcal{D}(h^{(t-1)}) + b_o). \quad (5)$$

6) Cell State

$$s^{(t)} = \mathcal{D}(g^{(t)}) \odot i^{(t)} + s^{(t-1)} \odot f^{(t)}. \quad (6)$$

7) Hidden Gate

$$h^{(t)} = \tanh(s^{(t)}) \odot o^{(t)}. \quad (7)$$

8) Output Layer

$$y^{(t)} = W_{hy}h^{(t)} + b_y \quad (8)$$

where $x_0^{(t)}$ is the vector of raw inputs for the time step t , ReLU is the rectified linear units, σ is the activation function, $\mathcal{D}$ is the dropout operator that is explained in the following, $\odot$ refers to the point-wise multiplication, W 's are connection weights in the network, b 's are constant parameters, y is the output and will be compared to soil moisture observations, h is the hidden cells, and s is memory cells, which is designed to hold long-term memory. To reduce overfitting, we employed dropout regularization [35], which stochastically sets some of the network connections to zero. Instead of introducing uncontrollable noise into the training process, this procedure induces robustness, as connections are randomly broken, weights cannot coadapt with others during training to fit the results [36]. Coadaptation, in which multiple weights are changed together to fit the target, is a primary culprit of overfitting. $\mathcal{D}$ applies dropout with constant dropout masks to recurrent connections [37], i.e., the connections that are set to 0 stay the same throughout each training instance. As compared to the default LSTM algorithm, we added the ReLU transformation and a dropout on part of the memory cells [the $\mathcal{D}(g^{(t)})$ term], following [38]. However, different from [38], we applied a constant mask for this dropout operation as well. While not a focus of this paper, we found that dropout of the memory cell reduced test errors.

The network outputs one scalar value ($y^{(t)}$) after each time step and computes a loss function by comparing it to the SMAP product $y^{*(t)}$, which is stated in the following:

$$L = \frac{1}{\rho} \sum_{t=1}^{\rho} \mathbf{1}_o(t) [y^{(t)} - y^{*(t)}]^2 \quad (9)$$

where $\mathbf{1}_o(t)$ is an indicator function (1 when time step t has SMAP observation and 0 otherwise) and ρ is the length of each

training instance. We added the indicator function because SMAP data have uneven revisit schedules. This network was implemented by adapting from the FastLSTM package based on Torch7 in the Lua programming language [39]. The network parameter updates were controlled by ADADELTA, which is an adaptive learning rate scheme [40].

For the surface moisture product, we trained a network with the SMAP data over CONUS at the 36 km resolution on SMAP EASE-Grid, forming a training set of 6321 pixels. The time series for each SMAP pixel was split into multiple training instances for the network. For the 9-km resolution root-zone product, we took one pixel out of every 4×4 pixels to obtain a training set of 7365 pixels. Such spatial downsampling was performed for the sake of computational efficiency. The training was conducted on NVidia 1080 Ti Graphical Processing Units.

C. Validation: SMAP Core Validation Sites, Sparse Network, and Evaluation Metrics

Due to high spatial heterogeneity inherent with soil moisture [41], the point-scale measurements cannot be directly compared to the SMAP data. SMAP partnered with *in situ* networks that have sufficiently dense soil moisture sensors to be reliably aggregated to a larger spatial scale. These sites are called core validation sites [26]. Nine core sites are distributed over CONUS and each of them contains multiple pixels matching the SMAP footprint. Consistent with past calibration and validation procedures, we created Thiessen polygons for measurement sensors on validation pixels, and *in situ* observations are the average of all the sensors using weights determined by the fraction of area covered by their Thiessen polygons. As sensors may stop working temporarily, especially under cold weather, we only retain the *in situ* records when at least half of the pixel is covered by the Thiessen polygons of the operational sensors. We also removed the records when the measured *in situ* soil temperatures were lower than 4 °C, as sensors' measurements are not reliable under cold weather.

We also employed the U.S. Climate Reference Network (CRN) [42], which are sparser, to validate the models. CRN sites were first deployed in 2000 and completed full deployment of 114 sites in 2011. We removed 10 coastal sites because they are not covered by SMAP footprint or the SMAP signals are missing for more than 75% days on the corresponding pixel. Out of 104 CRN sites employed in validation, 82 installed soil moisture probes at standard levels suggested by World Meteorological Organization (5, 10, 20, 50, and 100 cm), whereas 22 of them were instrumented only at the top two layers. Both types of sites were used for the evaluation of surface soil moisture, while only those with deeper sensors were used for evaluating RZSM. At each CRN site and every point of depth, three probes were planted 1.5 m away from the base at 0°, 120°, and 240° compass direction, and the soil moisture measurement employed was the average of those three sensors.

We evaluated three products: LSTM, Noah, and a simple average of LSTM and Noah, denoted by "Comb." Ideally, to fuse the two data sets, we can obtain both of their variances.

We can weigh their values using the inverse of these variances during averaging, i.e., using an Bayesian model averaging scheme [43] or instrumental variable regression [44]. However, at this stage, the uncertainty estimation methods for the DL model are not developed enough for such an application. The simple average here only represents an initial evaluation of the value of LSTM-extended SMAP data. Better schemes can certainly be evaluated in the future.

For each site, we calculated several metrics for two pairs of data: LSTM-predicted versus *in situ* data and SMAP versus *in situ* data. The metrics included the root-mean-square error (RMSE), bias (the difference between the means), the Pearson's correlation coefficient, R, and the unbiased RMSE (ubRMSE), which is RMSE calculated for two time series after means have been subtracted. The latter two metrics were unaffected by model bias. We calculated the metrics for the training period (April 1, 2015-April 1, 2017) and test period (site's operation start time-April 4, 2015) separately. When added, the ^{AL} superscript denotes the after-launch (AL) training period, while the ^{PL} superscript denotes the prelaunch (PL) test period. Also, ΔubRMSE denotes the difference in ubRMSE between training and hindcast periods, i.e., $\text{ubRMSE}^{\text{PL}}(\text{LSTM}, \text{in situ}) - \text{ubRMSE}^{\text{AL}}(\text{LSTM}, \text{in situ})$. It must be noted here again that the RZSM product results from assimilating SMAP brightness temperature into the Catchment land model with GEOS-5 FP forcing data [45]. Therefore, it is not as independent from model dynamics as the L3 product. Simultaneously, the innovations still take into account the actual SMAP observations.

We examined how interannual trends observed in *in situ* networks were captured by LSTM. We focused our trends analysis on the root-zone product, as it has longer memory than the surface soil moisture. The trend comparison was calculated for the maximum period of available data since 2005, which varies from site to site. We removed sites whose operational period is shorter than three years, or more than 10% of records are missing. To include more data for the trend analysis, we treated sensor readings when measured soil temperatures were lower than 0 °C as missing. For RZSM, we performed the nonparametric Mann-Kendall test [46] and computed the Sen's slopes using the longest available data. The starting date of the data occurs between 2000 and 2011. The Mann-Kendall test examines the null hypothesis that there is no trend in the data.

For RZSM, the SMAP root-zone product represents the accumulated water content in 0–100 cm, while *in situ* probes are located only at several depths. Following the guidelines mentioned in [14], we extrapolated the *in situ* measurements using weights that were proportional to sensor depths within the root-zone layer. Soil moisture was discretized using the moisture probes as layer centers, and the probe-measured moisture was used as the mean value for the corresponding layer. For example, if a station has sensors at depth $d_1, d_2, \dots, d_n$ cm, the weight of those sensors will be proportional to $(d_1 + d_2)/2, (d_2 + d_3)/2 - (d_1 + d_2)/2, \dots, 100 - (d_{n-1} + d_n)/2$. There are only five core sites that have soil moisture measurements at multiple depths. For the CRN network, we selected sites that had soil moisture sensors at

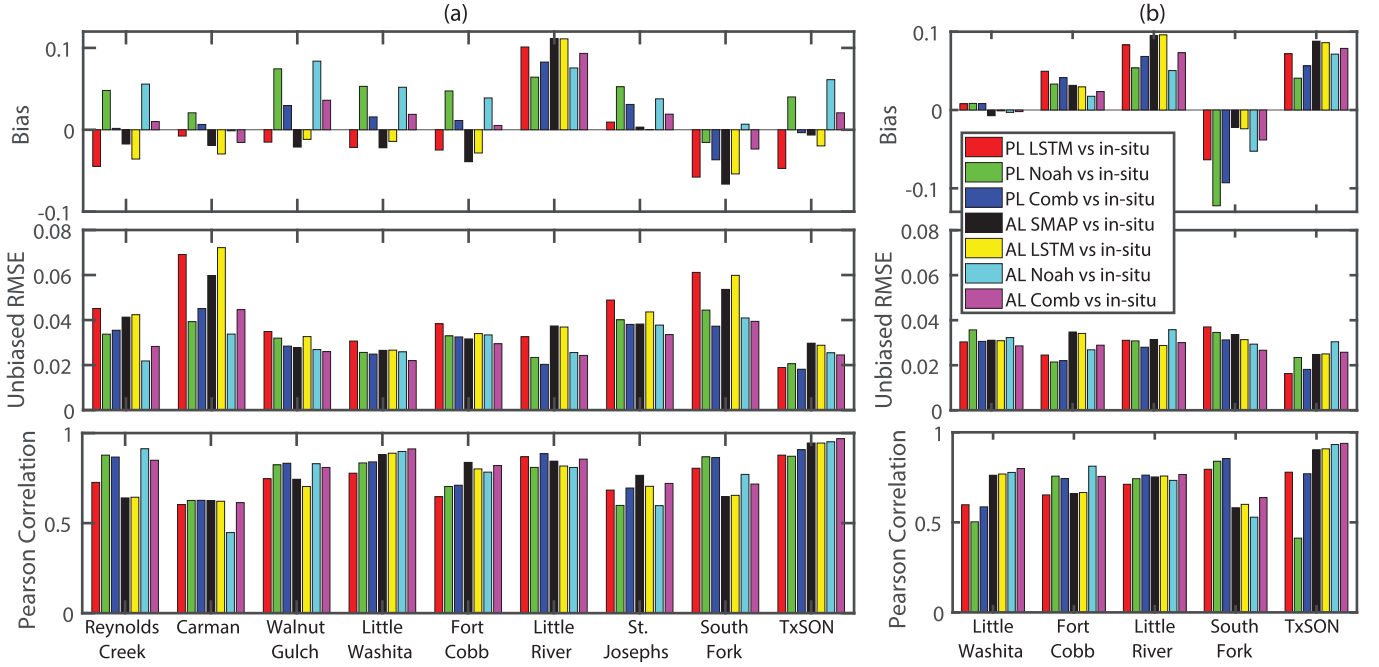


Fig. 3. Bias, ubRMSE, and Pearson correlation of (a) surface and (b) RZSM as evaluated by SMAP core-site *in situ* data for the AL training period and the PL validation periods. Comb means the simple average of LSTM and Noah. In general, ubRMSE between *in situ* data and SMAP (black) is slightly smaller than that between LSTM and *in situ* during the training period (yellow), which is in turn smaller than that between LSTM and *in situ* during the hindcast period (red). At sites where $R^{\text{AL}}(\text{SMAP}, \text{in situ})$ is higher than $R^{\text{AL}}(\text{Noah}, \text{in situ})$ (yellow higher than cyan), LSTM or Comb can have a higher R than Noah during hindcast, with Fort Cobb being an exception. At other sites, Noah is stronger than SMAP and LSTM, but Comb might still be better.

all design depths (5, 10, 20, 50, and 100 cm). The 72 sites remained after removing the ones with more than 50% of missing measurements or less than one year of records before the SMAP launch.

III. RESULTS AND DISCUSSION

A. Evaluation at the Core Sites

1) *Statistical Comparison Between Training and Hindcast Periods:* We first focus on the comparison of LSTM error statistics between the training period (yellow bars in Fig. 3) and the hindcast period (red bars) for the surface product. For most of the sites, we found only a small increase in ubRMSE from the training to the hindcast period [Fig. 3(b)]. Five sites showed $\text{ubRMSE}^{\text{AL}}(\text{SMAP}, \text{in situ})$ values that are ≤ 0.04 , and for all of these sites, $\text{ubRMSE}^{\text{PL}}(\text{LSTM}, \text{in situ})$ are also ≤ 0.04 . For more than half of the sites, ΔubRMSE is less than 0.006. At Little River and TxSON, the LSTM-to-*in situ* error actually reduces from training to test period, which is attributed to randomness and some missing data. Four sites have ΔubRMSE of more than 0.006: ubRMSE increases from 0.028–0.035 at Walnut Gulch, from 0.059–0.069 at Carman, from 0.038–0.048 at St. Josephs, and from 0.054–0.061 at South Fork.

As it is normal for the test error to be larger than the training error, a small and positive ΔubRMSE is to be expected. We note that when SMAP itself has a relatively large disagreement with the *in situ* data (> 0.04), as in Carman, St. Josephs, and South Fork, ΔubRMSE also tends to be larger. Presumably, when SMAP has a large difference from the *in situ* data, there might be some systematic deviations at those sites, e.g., the cold season behaviors do not match the *in situ* data as well,

which is discussed in the following. When such discrepancies exist, there are less detectable patterns to learn and more noise to interfere with the training. As a result, the effectiveness of learning weakens. The SMAP-*in situ* discrepancies mostly affect the temporal fluctuations, as witnessed by the degrading ubRMSE and anomaly correlation during the training, and it has little impact on the hindcast bias. In contrast, the bias between SMAP and LSTM appears to have little impact on ΔubRMSE . If we examine Little River, which has a large positive bias, the ubRMSE and Pearson anomaly correlation in the hindcast period are as good as the training period. Despite the mismatch between SMAP and *in situ* and a mild amplification of error, none of the increase in ubRMSE is greater than 0.011.

LSTM hindcast of RZSM also shows similar error statistics to SMAP level 4 root-zone product. ubRMSEs for RZSM are generally smaller than the surface product due to smaller soil moisture temporal variability in the root zone. Hindcast ubRMSEs at all the five sites are smaller than 0.04. For three sites (Little Washita, Little River, and South Fork), the $\text{ubRMSE}^{\text{PL}}(\text{LSTM}, \text{in situ})$ is almost identical to $\text{ubRMSE}^{\text{AL}}(\text{SMAP}, \text{in situ})$, but the correlation degraded at Little Washita and TxSON (Fig. 3).

2) *Statistical Comparison of Performances Between LSTM, Noah, and Comb:* We now switch our attention to the comparison between LSTM, Noah, and their averages. It should be recalled by the following.

- 1) LSTM learns from SMAP. Thus, in theory, SMAP is the “performance ceiling” of LSTM when evaluated against *in situ* data.
- 2) SMAP senses a moisture-dependent depth that is shallower than 5 cm.

For the surface soil moisture, we found that for four of the nine core sites, SMAP has a higher R than Noah during the training period (Carman, Fort Cobb, Little River, and St. Josephs) (compare black to cyan bars). Outside of these sites, LSTM has little chance of outperforming Noah. However, SMAP might still bring in new information to help improve the Comb. For the RZSM, SMAP has advantages for only two of the sites (Little River and South Fork).

With the above information in mind, we found that while LSTM has limited chance to surpass Noah for some sites, Comb shows noticeable advantages over Noah. During the training period, Comb has a higher R than Noah at six out of nine core sites for the surface soil moisture. For the remaining three sites, Comb and Noah are also quite close. In contrast, LSTM only outperforms Noah at two sites (Little River and St. Josephs), where SMAP shows noticeable advantage (Fig. 3). For the hindcast period, Comb is higher at five of nine sites. For the other four sites, Comb is similar to Noah. Comb's advantage is more clear with the RZSM. During the training period, Comb outperforms Noah at four of the five sites during the training period, even though LSTM only surpasses Noah at South Fork, which is the only site where SMAP's root-zone product has a higher R . However, during the hindcast period, Comb is more skillful at all of the five sites, while LSTM is higher than Noah at three of the five sites.

It perhaps is not surprising that the ensemble model average works better than each of the individual models. However, we tested the average of Noah and the variable infiltration capacity (VIC) model, which showed lower R than that of LSTM+Noah for six of the nine sites (Fig. 4). Therefore, simply combining the models together does not guarantee better skills. Since LSTM not only helps reduce bias but also improves the temporal dynamics as reflected in R , the results suggest SMAP and LSTM have brought in unique and valuable information. In time series analysis (Section III-A3), we examine how LSTM, despite being generally lower than R , offers this unique contribution.

The mixed comparison is in contrast to our previous analysis in [22], where LSTM was always found to outperform Noah when SMAP was used as the benchmark. Indeed, when we trained LSTM in one year and test it in another, the 1-year test R (SMAP, LSTM), which describes the match between LSTM and SMAP in prolongation period, is mostly 0.05–0.1 higher than R (SMAP, *in situ*), making it the highest among all the bars for most of the sites (Fig. 3). This comparison suggests while LSTM is sufficiently close to SMAP, there are some significant differences between SMAP and *in situ* data, which limited the LSTM's performance. These results show that the LSM is a powerful simulator for the temporal dynamics of soil moisture, which confirms the earlier findings [47]. Currently, it is difficult for a remote-sensing product to completely surpass Noah where high-quality meteorological forcings are available.

3) *Time Series Comparisons*: To gain further insights beyond the statistical comparisons, Fig. 5 shows the time series comparisons of surface soil moisture at five core sites. In general, LSTM mimics the well-performing characteristics of SMAP as well as some systematic discrepancies concerning

the *in situ* data. For example, at Carman, LSTM inherits the high-frequency, high-amplitude fluctuations from SMAP. These fluctuations appear unphysical because much smaller fluctuations can be observed with Noah and *in situ* data. Another example is, at Walnut Gulch, both the peaks and troughs are well captured by LSTM, but both the drydowns of SMAP/LSTM tend to be faster than the *in situ* data (Fig. 5). Similar faster drydowns can also be clearly observed with LSTM at Little Washita, Fort Cobb, and Little River. As noted previously, SMAP senses' moisture between the surface and a moisture-dependent depth is often 5 cm or less, whereas the probes are often centered at 5 cm and measure soil water between 3.5 and 6.5 cm [48]. Moisture near the surface dries faster. Thus the lower sensing depths of SMAP could explain lower R with SMAP and LSTM.

In contrast, the drydown in the Noah-simulated moisture is more in line with the *in situ* data. This is most obvious at Walnut Gulch and Little Washita, where Noah has higher R than LSTM. It seems Noah's simulation better matches the depths of the *in situ* probes. However, at Walnut Gulch, we noted that Noah tends to predict too much drydown toward the end of the dry spell, while LSTM and *in situ* both have a flatter bottom. It is possible that soil water holding parameters in Noah were not correctly specified here. SMAP data were better able to capture such trends. This pattern is also why, at this site, averaging Noah with LSTM helps Comb outperform Noah.

When there are fewer reliable data points for winter in the North, LSTM is not well trained like the other periods. At St. Josephs, we witnessed the periods of missing SMAP data during the winter freeze. During this period, LSTM shows high oscillations. During hindcast, *in situ* data could be missing in the winter as well. When there is snow on the ground, SMAP's retrieval quality flag would indicate that a retrieval was attempted, but the quality cannot be guaranteed. The data quality might not be reliable during when snow is present. Also, the mission does not report soil moisture when it suspects that the ground is frozen.

Similar to the surface soil moisture hindcast, root-zone moisture hindcast also inherits the systematic discrepancies between SMAP and *in situ*. At Little Washita, the soil moisture lows of *in situ* data are lower than that of the SMAP's. We see a similar pattern for the hindcast period (Fig. 6). At Little River, LSTM and SMAP tend to overestimate the peaks. At South Fork, both the LSTM and SMAP tend to underestimate soil moisture peaks. In contrast, Noah could dry down too much as well as overestimate the peaks, in fact, more prominently than the surface soil moisture.

B. Evaluation Over the Sparse Network

We evaluated LSTM and SMAP against CRN *in situ* network over CONUS. As there is a scale disparity between a SMAP pixel and point-scale CRN measurements, we expect their differences to be significantly larger than those between SMAP and core sites. Indeed, the median ubRMSE between SMAP and surface *in situ* (CRN) data as well as between LSTM hindcast and *in situ* are now around 0.05, which is

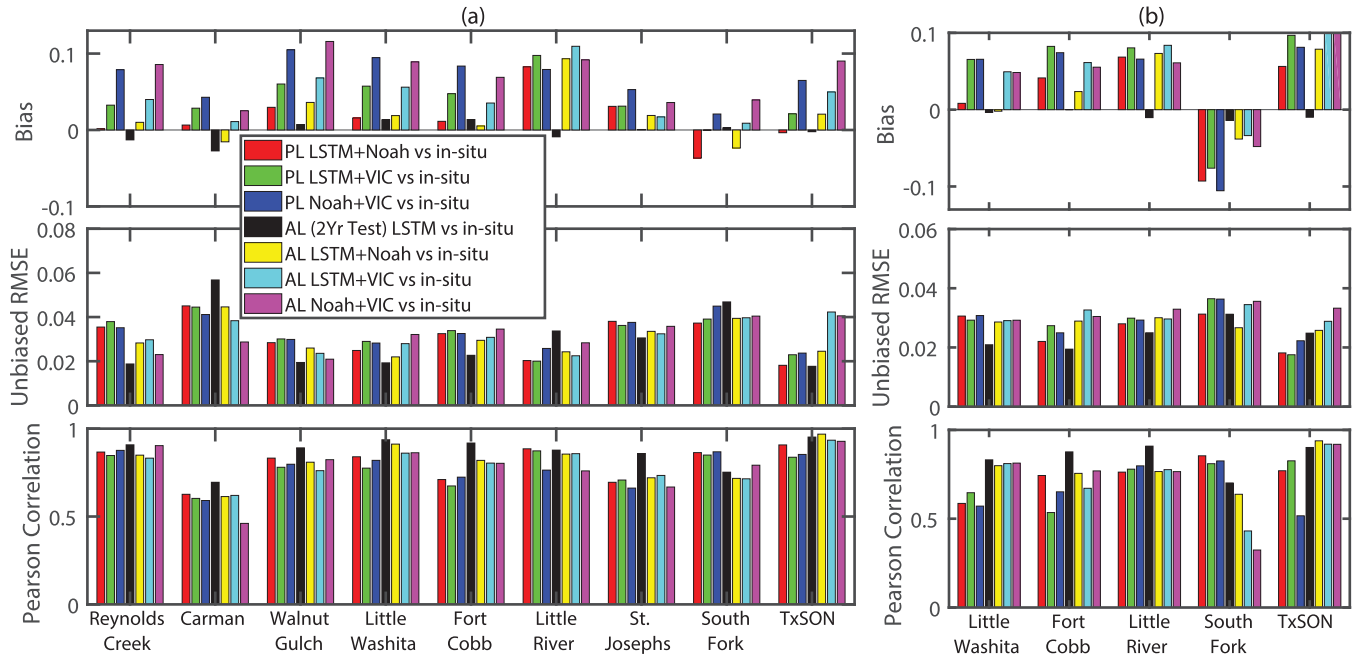


Fig. 4. Same as Figure 3, but here we compare the LSTM+Noah (called Comb in Figure 3) with Noah+VIC, which seems weaker than Noah+LSTM. We also show the test metrics between LSTM and SMAP when it is trained in one year and tested in another (AL 1 Yr Test).

notably higher than that at the core sites [Fig. 7(a)]. The correlation with the surface soil moisture is slightly lower than that at the core sites. At some sites, where LSTM has high correlations, during the training period, R decreased during the test period, as evidenced by the lower 75-percentile box edge for the hindcast period. Despite this decrease, the median LSTM R values for all the CRN sites is above 0.72 for the surface hindcast, which is comparable to the median of SMAP. For the root-zone data, LSTM only declines slightly from training to hindcast period, and they are both limited by SMAP's relatively low R .

Noah, in general, compares more favorably with CRN than LSTM and SMAP, again attesting to its good performance. However, Comb's R is higher than that of Noah's alone for the surface sites during both the training or the hindcast period. CRN probes are installed in triplicate redundancy at 5, 10, 20, 50, and 100 cm depths (sometimes only 5 and 10 cm are available). Thus, CRN sites have the same sensing depth discrepancy from SMAP and in turn LSTM. However, integrating information from SMAP data may nonetheless help correct the model errors. For the RZSM, it seems Noah has more better-performing sites, as it has a higher 75-percentile R .

The CRN sites can demonstrate spatial patterns of LSTM and SMAP performances and help us understand where the correlations have dropped from the training to the test periods [Fig. 8(a)]. LSTM surface hindcast resembles the spatial pattern as SMAP. However, we found several pockets of sites where LSTM is noticeably lower than that of SMAP's. Test R for some northeastern sites [eastern ellipsoid in Fig. 8(b)] have decreased significantly, i.e., below 0.5. These sites are located in cold and heavily vegetated regions, where SMAP does not work properly due to large vegetation water content

and snow. Thus, LSTM could not find useful relationships in this region. Another pocket of notable decrease as compared to SMAP lies at the boundaries between Utah, Colorado, and Wyoming (the western ellipsoid). Of these three, the two western sites are located on the Rocky mountains, while the easternmost site is on a plain east to the mountains. We think the large discrepancy here might be due to the poor quality of forcings in this region. For CONUS, NLDAS-2 precipitation integrates Climate Prediction Center gauge data and hourly radar data [49]. This region has the lowest weather station density and yet a large spatial variability in rainfall due to orographic effects. All of these factors also challenge the radar precipitation sensing. Therefore, hindcast quality here is likely adversely impacted by lower forcing quality.

C. Evaluation of Interannual Trends

Because SMAP has only three years of data, we employed both the CRN and the core sites to investigate the extent to which LSTM can reproduce the interannual trends with RZSM. While we expect large differences between hindcast and CRN data due to scale disparity, their interannual trends should be more comparable. Many of these sites showed only small trends (with an absolute value of less than 0.5% per year), and LSTM mostly predicts similarly small trend magnitudes (Fig. 9). When the trend magnitudes are this small in value, the statistical significance of the slopes are low. For the sites that have larger interannual trends, LSTM also predicts larger trends. R between LSTM hindcast and *in situ* (both CRN and core sites) trends is 0.88 for the sites that have large trends ($>0.5\%$ annually), indicating that SMAP prolongation by LSTM is a promising approach that can capture the interannual variability in RZSM. It should be

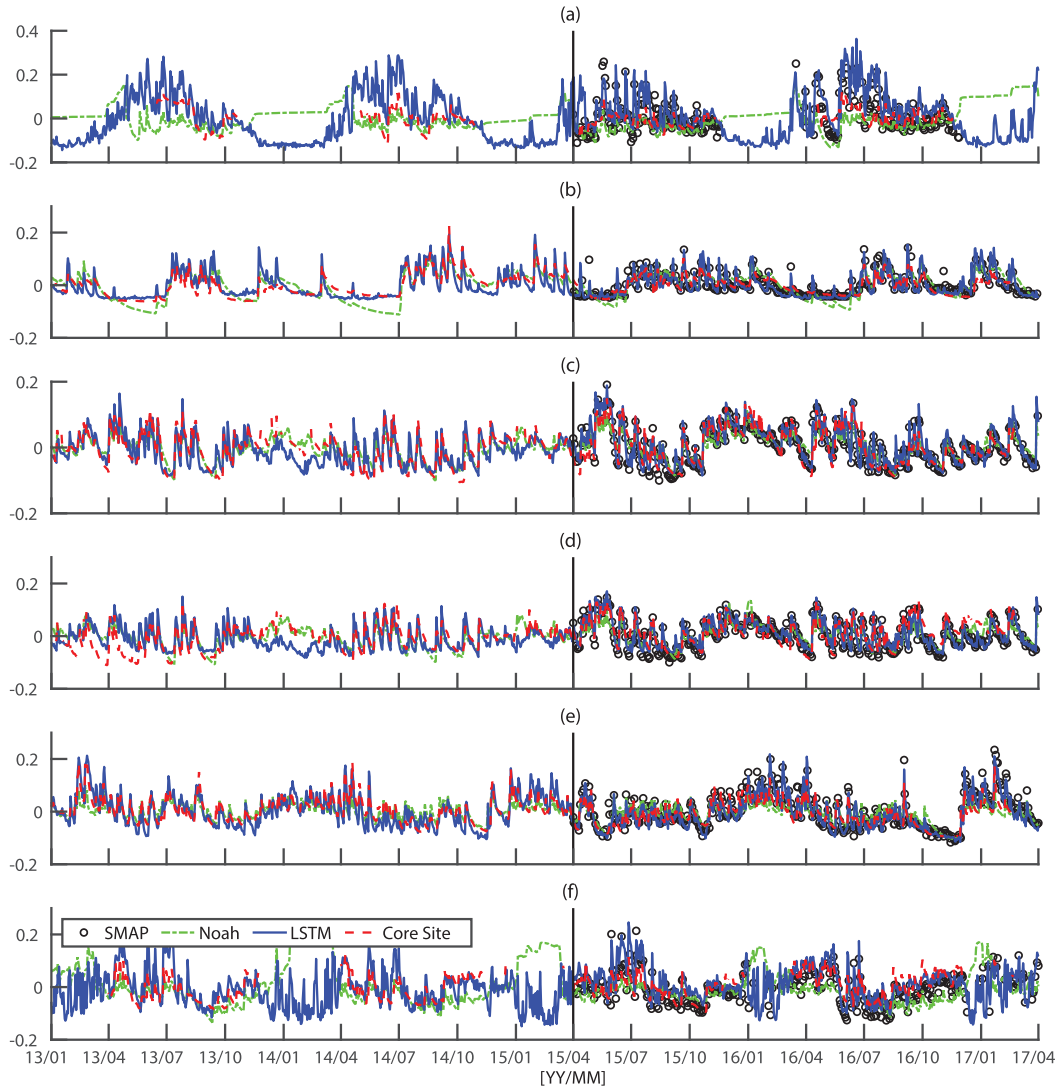


Fig. 5. Time series comparison of LSTM hindcast, SMAP, and *in situ* soil moisture anomalies at six different sites. The means were subtracted from all the series. LSTM hindcast inherits the patterns of discrepancies between SMAP and *in situ* data, e.g., faster drydowns for SMAP. At Carman, SMAP shows high-frequency oscillations, which disagree significantly with the *in situ* data. LSTM shows the same oscillations. At St. Josephs, we note some periods of missing SMAP and *in situ* data due to frozen ground or snow. (a) Carman 09013601. (b) Walnut Gulch 16013604. (c) Little Washita 16023603. (d) Fort Cobb 16033603. (e) Little River 16043604. (f) St. Josephs 16063603.

recalled that the model was only trained using three years of data. The increase in hindcast length does not appear to strongly influence the results.

D. Further Discussion and Future Research

Although we noted some issues, such as winter soil moisture, the LSTM results are presented “as is” (without any further corrections) for the sake of thoroughly exploring the strengths and limitations of the LSTM-based long-term predictions. When SMAP itself poorly matches the *in situ* dynamics, the LSTM hindcast also degrades. The inaccuracies in SMAP might amplify, which were found to be mild. As per other studies, a main difference between SMAP and *in situ* data appears to be related to the different sensing depths. SMAP appears to be capturing shallower soil moisture than the *in situ* probes. In contrast, the process-based model Noah seems

to simulate similar depths concerning the *in situ* data, which has led to Noah’s better performance at some sites. It also raises the question as to how SMAP data should be calibrated against the core site data. A corrective step accounting for the variable sensing depths seems necessary for the incremental improvement of SMAP’s retrieval algorithm.

In contrast to other model-data integration methods, such as DA and calibration, the power of LSTM is to learn the relationship between inputs and soil moisture without explicit assumptions concerning the underlying process. Thus, LSTM can largely avoid systematic (parametric or model structural) errors due to such assumptions. Our results show the learned knowledge is permanent and does not decay with time, when judged by the learning target. As previously mentioned, LSTM cannot correct stochastic forcing errors as DA could. For example, if a rainfall event was missing in the forcing,

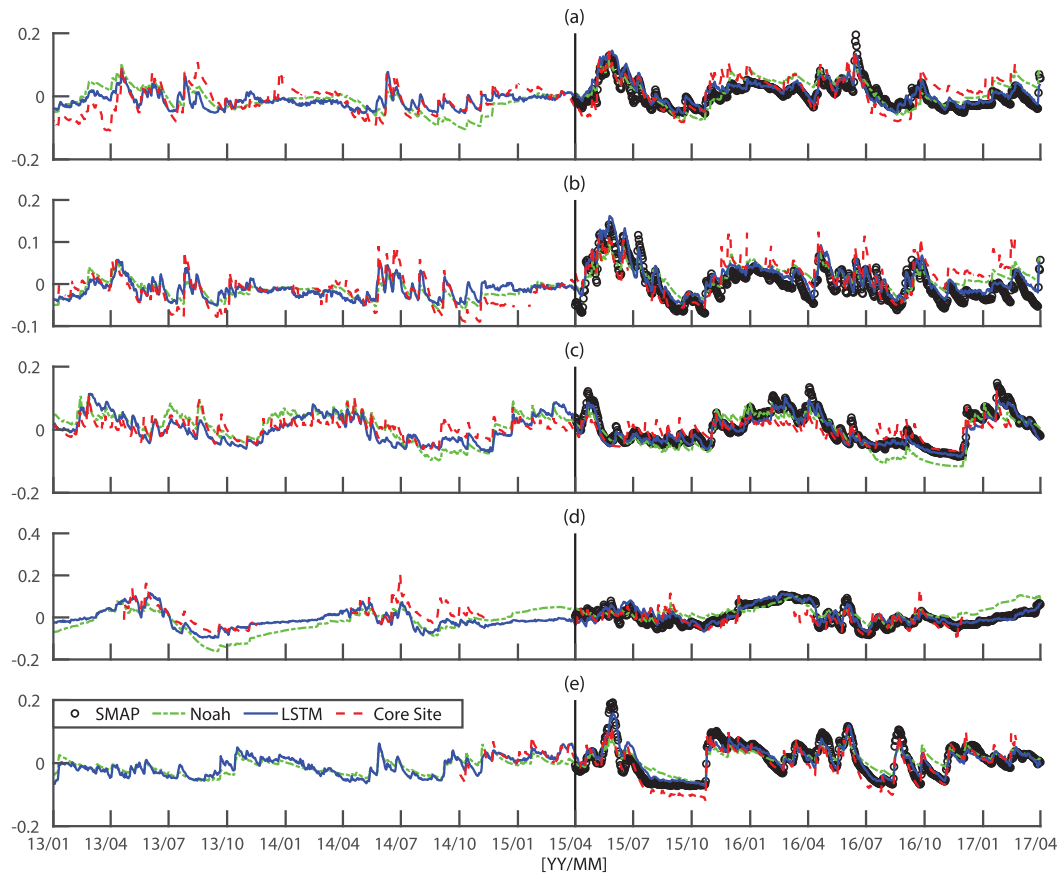


Fig. 6. Time series comparison of LSTM hindcast, SMAP, and *in situ* RZSM on five core validation sites. (a) Little Washita 16020917. (b) Fort Cobb 16030911. (c) Little River 16040904. (d) South Fork 16070905. (e) TxSON 48010911.

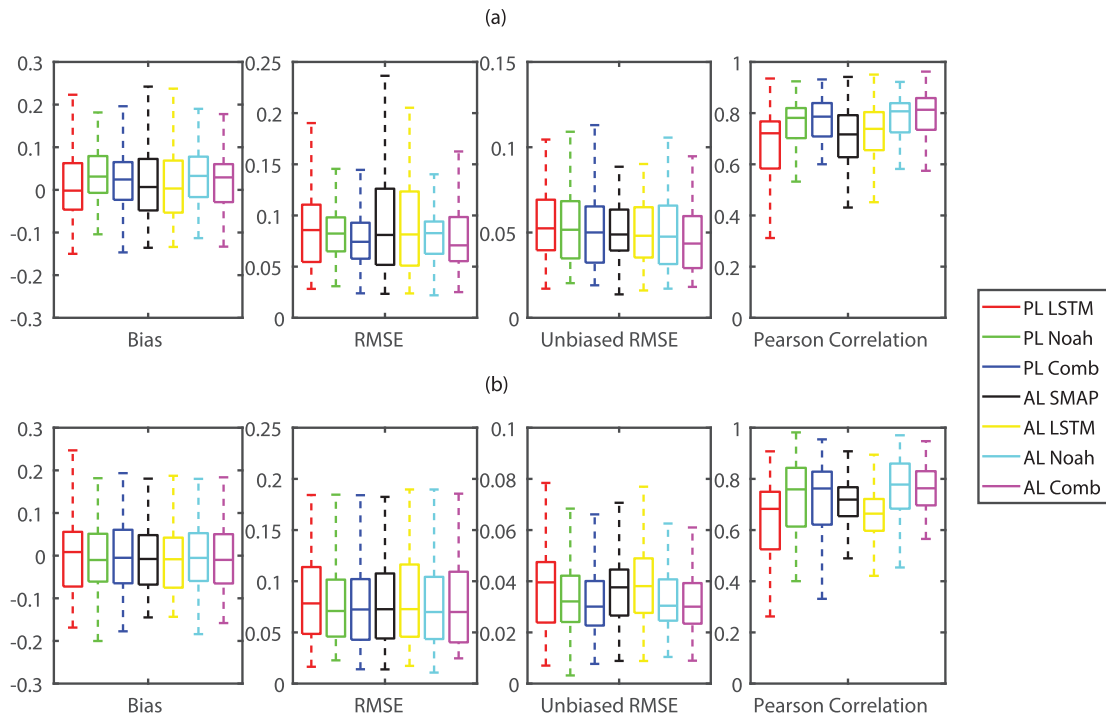


Fig. 7. RMSE, bias, unbiased-RMSE, and Pearson correlation of (a) surface and (b) RZSM between CRN sites, LSTM, and SMAP.

LSTM would be unable to predict the corresponding soil moisture rise. However, while DA can correct for stochastic forcing errors, its effects wane with time, leading to little

impact on long-term projections. The model states would drift away from the corrected states some time after data ingestion [50]. On the other hand, calibration only addresses

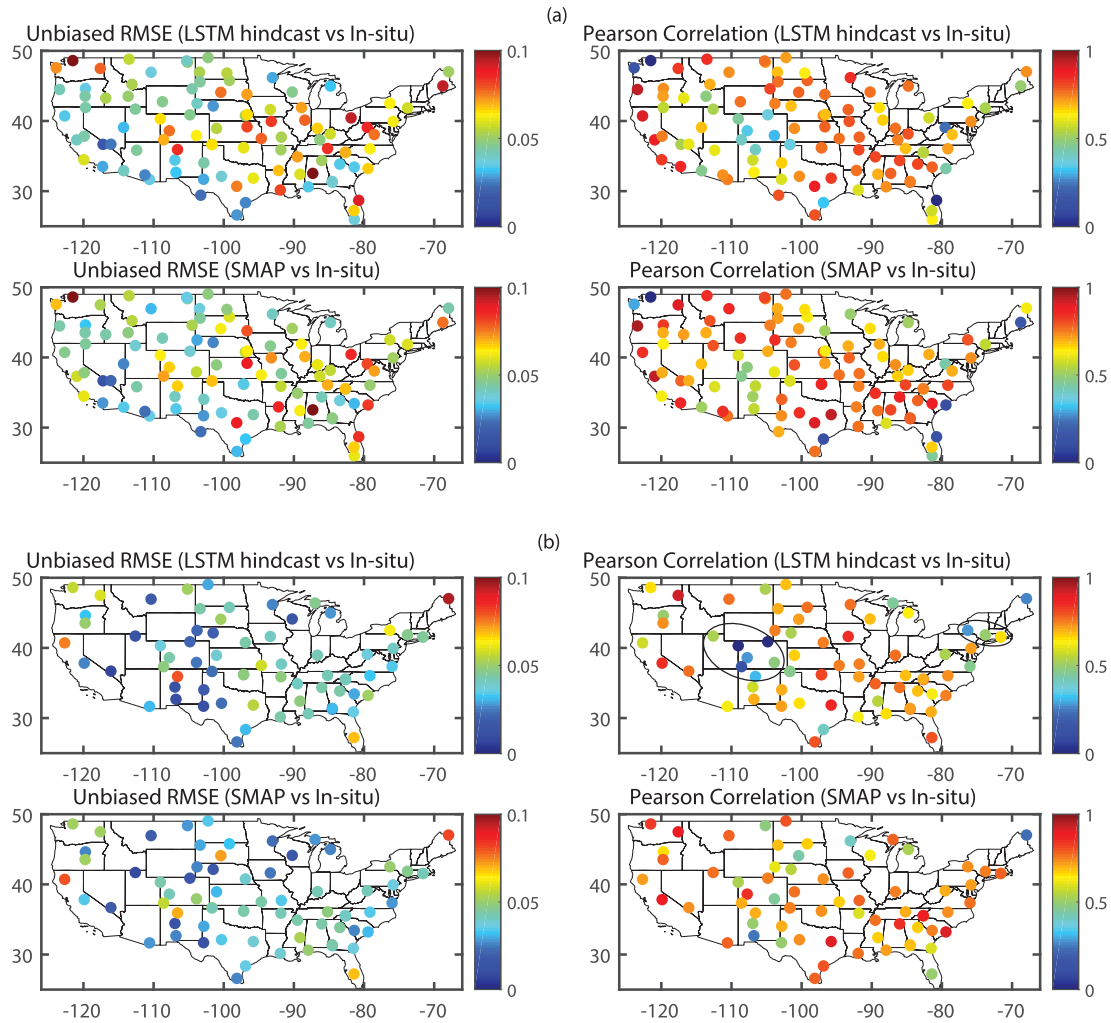


Fig. 8. Map of unbiased-RMSE and Pearson correlation of (a) surface and (b) RZSM between *in situ*, LSTM, and SMAP for periods before and after SMAP launch, on CRN *in situ* network. LSTM generally has the same spatial pattern as CRN, but there are a few instances of a notable decline in R , as annotated on the figure.

parametric errors. Too much calibration might in fact distort the parameters' values and model dynamics to compensate for the structural inadequacies. Therefore, all these techniques have their respective advantages and disadvantages, and might be applied synergistically to address different sources of errors.

As discussed in Section I, extending SMAP records back in time (hindcast) with high fidelity to SMAP observations might provide considerable value to myriad applications. For example, a high-fidelity data-driven historical reconstruction of SMAP can help benchmarking or calibrating the SMAP retrieval algorithms. For another example, future research may examine combining data merging techniques [20] with DL to exploit stochastic signal measured by older satellites. In Liu *et al.*'s [20] work, we cannot merge SMAP and older satellites because they had no overlaps. However, the LSTM hindcast may help create that overlap to enable CDF matching and separation of seasonality and random errors, among other outcomes.

Here, we only showed a simple average between LSTM and Noah to demonstrate the value of integrating information from

SMAP via a DL model. This simple approach is obviously not the optimal way to exploit the value of SMAP data. Currently, uncertainty estimation methods for DL models are still in the developing stage. As these methods become more matured, we can more adaptively combine the models, given our knowledge about their location-dependent and time-dependent accuracy. Further, the LSTM-extended data can be injected into LSM through either DA or model calibration to improve the model's internal dynamics other than soil moisture. On the other hand, we can construct more informative training data for winter periods, perhaps by combining observations with simulations by process-based models, to fill the gaps due to frozen soil. We envision the process-based model will continue to be a valuable constraint for temporal extrapolation as well.

The good performance of LSTM suggests that it learns the soil moisture evolution dynamics, as modulated by the land surface characteristics rather than a simple relationship between climate forcings and soil moisture. For forecast purposes, we have the additional information from present and near-past observations, which are expected to help improve

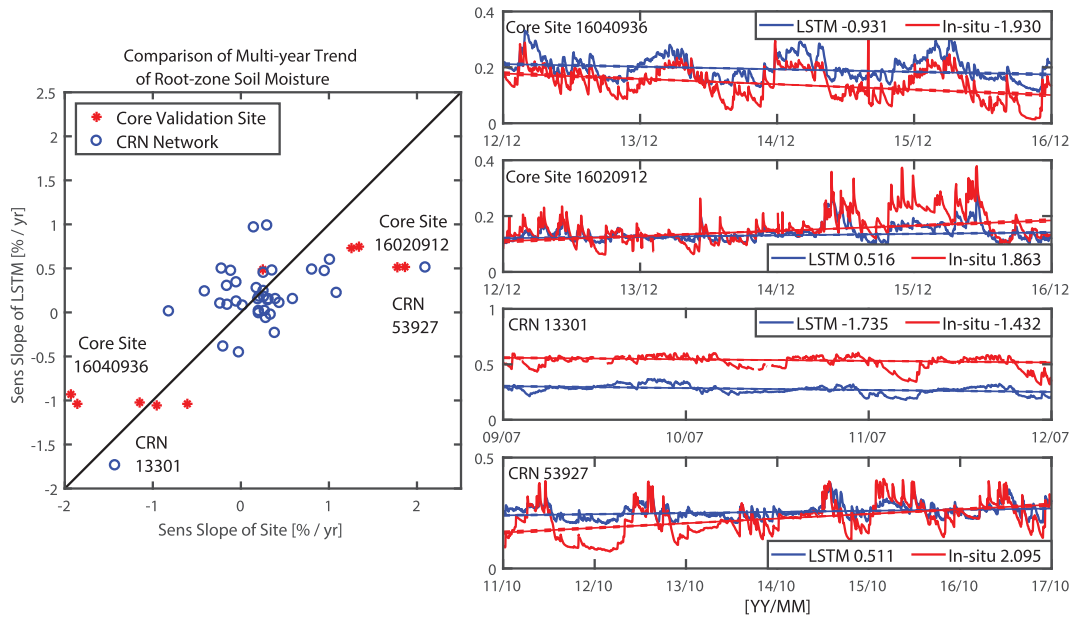


Fig. 9. Comparison of long-term RZSM trends between core validation sites, CRN sites, and LSTM hindcast. The correlation between the slopes of core site and LSTM is 0.89. The correlation between CRN and LSTM is 0.53 for all sites and 0.82 for sites with trends greater than 0.5% per year. (Right) Time series of several selected sites. (Left) Slopes are annotated on the legend with a unit of percent change per year.

forecast skills. The forecast model will require a slight reformulation of our network training procedure, where the current and past soil moisture observations are included as inputs. It must be noted that forecasting skills will depend heavily on the skill of the weather forecasts. We saw that the hindcast quality appears to deteriorate for sites that are surrounded by lower weather station density.

IV. CONCLUSION

Our results indicate that the long-term LSTM predictions have mostly retained the quality of the SMAP data. LSTM has a solid performance in regions where SMAP has good skills. For the root-zone product, the hindcast is able to capture and reproduce multiyear trends correctly, even if there are only three years of training data. LSTM's performance is limited by SMAP as a performance ceiling and is outperformed by Noah at some sites. A major factor is that SMAP has a different sensing depth in comparison to the probes measuring at 5 cm depth.

More importantly, despite the limitations, averaging LSTM and LSM predictions often produces better predictions. Even at sites where this is not true, Comb's performance is typically very close to Noah's. As a comparison, we have also shown that randomly combining models, e.g., Noah+VIC, do not always produce better results. Thus, SMAP extended by LSTM has added value and helped correct errors in Noah. There could be many novel uses of this approach, as described in the further discussion section.

In this application, the LSTM model aimed at reproducing a model-free, seamless replica for the information contained in SMAP. LSTM predictions are entirely data driven. As such, we expect that as SMAP data accrue and better retrieval algorithms are implemented, the DL models will also improve. Therefore, there is substantial value in the LSTM model as an

ensemble model member for the reconstruction of the past or the prediction of future trends. Meanwhile, we confirm the LSM Noah as a valuable contributor of information in ensemble predictions.

As LSTM was trained with SMAP as the target, it inherited systematic discrepancies between SMAP and *in situ* data. When the discrepancy is large, the error seems to be mildly amplified, leading to an increase in uBRMSE, which is around 0.01. This amplification is attributed to less detectable patterns to learn and more noise in the SMAP data, which can contaminate the learning process. Hindcasts have larger errors during winter for cold regions. SMAP data can be missing and less reliable under snow and frozen soil conditions, which reduces the available training data.

ACKNOWLEDGMENT

The authors thank NVIDIA Corporation for donating a Titan XP graphics card which accelerated this research.

REFERENCES

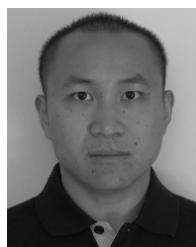
- [1] V. A. Orchard and F. Cook, "Relationship between soil respiration and soil moisture," *Soil Biol. Biochem.*, vol. 15, no. 4, pp. 447–453, Jan. 1983. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/003807178390010X>
- [2] D. Norbiato, M. Borga, S. D. Esposti, E. Gaume, and S. Anquetin, "Flash flood warning based on rainfall thresholds and soil moisture conditions: An assessment for gauged and ungauged basins," *J. Hydrol.*, vol. 362, nos. 3–4, pp. 274–290, 2008.
- [3] R. D. Koster *et al.*, "Regions of strong coupling between soil moisture and precipitation," *Science*, vol. 305, no. 5687, pp. 1138–1140, 2004, doi: [10.1126/science.1100217](https://doi.org/10.1126/science.1100217).
- [4] E. G. Njoku, T. J. Jackson, V. Lakshmi, T. K. Chan, and S. V. Nghiem, "Soil moisture retrieval from AMSR-E," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 2, pp. 215–229, Feb. 2003. [Online]. Available: <http://ieeexplore.ieee.org/document/1196040/>
- [5] W. Wagner, G. Lemoine, and H. Rott, "A method for estimating soil moisture from ERS scatterometer and soil data," *Remote Sens. Environ.*, vol. 70, no. 2, pp. 191–207, Nov. 1999. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S003442579900036X>

- [6] Y. H. Kerr *et al.*, "The SMOS mission: New tool for monitoring key elements of the global water cycle," *Proc. IEEE*, vol. 98, no. 5, pp. 666–687, May 2010. [Online]. Available: <http://ieeexplore.ieee.org/document/5446359/>
- [7] D. Entekhabi *et al.*, "The Soil Moisture Active Passive (SMAP) mission," *Proc. IEEE*, vol. 98, no. 5, pp. 704–716, May 2010.
- [8] W. Dorigo *et al.*, "ESA CCI soil moisture for improved Earth system understanding: State-of-the art and future directions," *Remote Sens. Environ.*, vol. 203, pp. 185–215, Dec. 2017. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0034425717303061>
- [9] R. Oliva *et al.*, "SMOS radio frequency interference scenario: Status and actions taken to improve the RFI environment in the 1400–1427-MHz passive band," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1427–1439, May 2012.
- [10] R. Oliva *et al.*, "Status of Radio Frequency Interference (RFI) in the 1400–1427 MHz passive band based on six years of SMOS mission," *Remote Sens. Environ.*, vol. 180, pp. 64–75, Jul. 2016. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425716300141>
- [11] J. R. Piepmeier *et al.*, "Radio-frequency interference mitigation for the soil moisture active passive microwave radiometer," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 761–775, Jan. 2013.
- [12] A. Colliander *et al.*, "Validation of SMAP surface soil moisture products with core validation sites," *Remote Sens. Environ.*, vol. 191, pp. 215–231, Mar. 2017.
- [13] T. Jackson *et al.* (2017). *Soil Moisture Active Passive (SMAP) Project Calibration and Validation for the L2/3_SM_P Version 3 Data Products*. [Online]. Available: <https://nsidc.org/data/smap/technical-references>
- [14] R. H. Reichle *et al.*, "Assessment of the SMAP Level-4 surface and root-zone soil moisture product using *in situ* measurements," *J. Hydrometeorol.*, vol. 18, no. 10, pp. 2621–2645, Oct. 2017, doi: [10.1175/JHM-D-17-0063.1](https://doi.org/10.1175/JHM-D-17-0063.1)
- [15] M. Pan, X. Cai, N. W. Chaney, D. Entekhabi, and E. F. Wood, "An initial assessment of SMAP soil moisture retrievals using high-resolution model simulations and *in situ* observations," *Geophys. Res. Lett.*, vol. 43, no. 18, pp. 9662–9668, 2016.
- [16] L. Karthikeyan, M. Pan, N. Wanders, D. N. Kumar, and E. F. Wood, "Four decades of microwave satellite soil moisture observations: Part 2. Product validation and inter-satellite comparisons," *Adv. Water Resour.*, vol. 109, pp. 236–252, Nov. 2017, doi: [10.1016/j.advwatres.2017.09.010](https://doi.org/10.1016/j.advwatres.2017.09.010)
- [17] M. Pan and E. F. Wood, "Impact of accuracy, spatial availability, and revisit time of satellite-derived surface soil moisture in a multiscale ensemble data assimilation system," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 3, no. 1, pp. 49–56, Mar. 2010.
- [18] R. D. Koster *et al.*, "Improved hydrological simulation using SMAP data: Relative impacts of model calibration and data assimilation," *J. Hydrometeorol.*, vol. 19, no. 4, pp. 727–741, Apr. 2018, doi: [10.1175/JHM-D-17-0228.1](https://doi.org/10.1175/JHM-D-17-0228.1)
- [19] Q. Liu *et al.*, "The contributions of precipitation and soil moisture observations to the skill of soil moisture estimates in a land data assimilation system," *J. Hydrometeorol.*, vol. 12, no. 5, pp. 750–765, Oct. 2011, doi: [10.1175/JHM-D-10-05000.1](https://doi.org/10.1175/JHM-D-10-05000.1)
- [20] Y. Y. Liu *et al.*, "Trend-preserving blending of passive and active microwave soil moisture retrievals," *Remote Sens. Environ.*, vol. 123, pp. 280–297, Aug. 2012. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425712001332#>
- [21] A. Gruber, W. A. Dorigo, W. Crow, and W. Wagner, "Triple collocation-based merging of satellite soil moisture retrievals," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 6780–6792, Dec. 2017. [Online]. Available: <http://ieeexplore.ieee.org/document/8046118/>
- [22] K. Fang, C. Shen, D. Kifer, and X. Yang, "Prolongation of SMAP to spatiotemporally seamless coverage of continental U.S. using a deep learning neural network," *Geophys. Res. Lett.*, vol. 44, no. 21, pp. 11030–11039, 2017.
- [23] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015. [Online]. Available: <http://www.nature.com/doifinder/10.1038/nature14539>
- [24] C. Shen *et al.*, "HESS Opinions: Deep learning as a promising avenue toward knowledge discovery in water sciences," *Hydrol. Earth Syst. Sci. Discuss.*, vol. 25, pp. 1–21, Apr. 2018. [Online]. Available: <https://www.hydrol-earth-syst-sci-discuss.net/hess-2018-168/>
- [25] G. J. M. De Lannoy and R. H. Reichle, "Assimilation of SMOS brightness temperatures or soil moisture retrievals into a land surface model," *Hydrol. Earth Syst. Sci.*, vol. 20, no. 12, pp. 4895–4911, Dec. 2016. [Online]. Available: <http://www.hydrol-earth-syst-sci.net/20/4895/2016/>
- [26] T. Jackson *et al.*, "Science data calibration and validation plan," Jet Propuls. Lab., Pasadena, CA, USA, Tech. Rep. JPL D-52544, 2012. [Online]. Available: <https://smap.jpl.nasa.gov/science/validation/>
- [27] R. D. Koster, M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar, "A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure," *J. Geophys. Res.*, vol. 105, no. D20, pp. 24809–24822, 2000, doi: [10.1029/2000JD900327](https://doi.org/10.1029/2000JD900327)
- [28] A. Ducharne, R. D. Koster, M. J. Suarez, M. Stieglitz, and P. Kumar, "A catchment-based approach to modeling land surface processes in a general circulation model: 2. Parameter estimation and model demonstration," *J. Geophys. Res.*, vol. 105, no. 24, pp. 24823–24838, 2000.
- [29] Y. Xia *et al.*, "Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow," *J. Geophys. Res. Atmospheres*, vol. 117, no. 3, pp. D03110-1–D03110-23, 2012.
- [30] N. H. Batjes, "A homogenized soil data file for global environmental research: A subset of FAO, ISRIC, and NRCS profiles (Version 1.0)," ISRIC, Wageningen, The Netherlands, Working Paper 95/10b, Jul. 1995.
- [31] M. B. Ek *et al.*, "Implementation of Noah land surface model advances in the national centers for environmental prediction operational mesoscale Eta model," *J. Geophys. Res., Atmospheres*, vol. 108, no. D22, pp. 8851-1–8851-16, 2003, doi: [10.1029/2002JD003296](https://doi.org/10.1029/2002JD003296)
- [32] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. ICASSP*, May 2013, pp. 6645–6649.
- [33] F. Beaufays. (2015). *The Neural Networks Behind Google Voice Transcription*. [Online]. Available: <https://research.googleblog.com/2015/08/the-neural-networks-behind-google-voice.html>
- [34] Z. C. Lipton, J. Berkowitz, and C. Elkan. (2015). "A critical review of recurrent neural networks for sequence learning." [Online]. Available: <https://arxiv.org/abs/1506.00019>
- [35] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014. [Online]. Available: <http://jmlr.org/papers/v15/srivastava14a.html>
- [36] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. (Jul. 2012). "Improving neural networks by preventing co-adaptation of feature detectors." [Online]. Available: <https://arxiv.org/abs/1207.0580>
- [37] Y. Gal and Z. Ghahramani. (Dec. 2015). "A theoretically grounded application of dropout in recurrent neural networks." [Online]. Available: <https://arxiv.org/abs/1512.05287>
- [38] S. Semeniuta, A. Severyn, and E. Barth. (Mar. 2016). "Recurrent dropout without memory loss." [Online]. Available: <https://arxiv.org/abs/1603.05118>
- [39] N. Leonard, S. Waghmare, Y. Wang, and J.-H. Kim. (2015). *RNN: Recurrent Library for Torch*. [Online]. Available: <https://github.com/Element-Research/rnn>
- [40] M. D. Zeiler. (Dec. 2012). "ADADELTA: An adaptive learning rate method." [Online]. Available: <https://arxiv.org/abs/1212.5701>
- [41] J. S. Famiglietti, D. Ryu, A. A. Berg, M. Rodell, and T. J. Jackson, "Field observations of soil moisture variability across scales," *Water Resour. Res.*, vol. 44, no. 1, pp. W01423-1–W01423-16, 2008.
- [42] J. E. Bell *et al.*, "U.S. climate reference network soil moisture and temperature observations," *J. Hydrometeorol.*, vol. 14, no. 3, pp. 977–988, 2013, doi: [10.1175/JHM-D-12-0146.1](https://doi.org/10.1175/JHM-D-12-0146.1)
- [43] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky, "Bayesian model averaging: A tutorial," *Stat. Sci.*, vol. 14, no. 4, pp. 382–401, 1999.
- [44] W. T. Crow, C.-H. Su, D. Ryu, and M. T. Yilmaz, "Optimal averaging of soil moisture predictions from ensemble land surface model simulations," *Water Resour. Res.*, vol. 51, no. 11, pp. 9273–9289, Nov. 2015, doi: [10.1002/2015WR016944](https://doi.org/10.1002/2015WR016944)
- [45] R. Reichle, R. D. Koster, G. De Lannoy, W. Crow, and J. Kimball. (2014). *Algorithm Theoretical Basis Documents*. [Online]. Available: <https://nsidc.org/data/smap/technical-references>
- [46] D. R. Helsel and R. M. Hirsch, "Statistical methods in water resources," in *Techniques of Water-Resources Investigations of the United States Geological Survey. Book 4, Hydrologic Analysis and Interpretation*. Reston, VA, USA: U.S. Geological Survey, 2002, ch. 12, pp. 323–352. [Online]. Available: <http://water.usgs.gov/pubs/twri/twri4a3/>
- [47] Y. Xia *et al.*, "Evaluation of multi-model simulated soil moisture in NLDAS-2," *J. Hydrol.*, vol. 512, pp. 107–125, May 2014, doi: [10.1016/j.jhydrol.2014.02.027](https://doi.org/10.1016/j.jhydrol.2014.02.027)

- [48] W. J. Rindinelli *et al.*, “Different rates of soil drying after rainfall are observed by the smos satellite and the south fork *in situ* soil moisture network,” *J. Hydrometeorol.*, vol. 16, no. 2, pp. 889–903, Apr. 2015, doi: [10.1175/JHM-D-14-0137.1](https://doi.org/10.1175/JHM-D-14-0137.1).
- [49] NLDAS-2. (2018). *NLDAS-2 Forcing Dataset Information*. [Online]. Available: <https://ldas.gsfc.nasa.gov/nldas/NLDAS2forcing.php#AppendixC>
- [50] G. J. M. De Lannoy, R. H. Reichle, P. R. Houser, V. R. N. Pauwels, and N. E. C. Verhoest, “Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter,” *Water Resour. Res.*, vol. 43, no. 9, Sep. 2007, doi: [10.1029/2006WR005449](https://doi.org/10.1029/2006WR005449).



Kuai Fang received the B.S. degree in geography from Beijing Normal University, Beijing, China, in 2012. He is currently pursuing the Ph.D. degree in water resource and hydrology with the Department of Civil Environmental Engineering, Pennsylvania State University, University Park, PA, USA. His Ph.D. dissertation focused on combining deep learning and remote sensing product to advance understanding and predictive capability of hydrologic functioning.



Ming Pan received the B.E. degree in hydraulic engineering from Tsinghua University, Beijing, China, in 2000, the M.A. and Ph.D. degrees in civil and environmental engineering from Princeton University, Princeton, NJ, USA, in 2002 and 2006, respectively.

He was a Post-Doctoral Researcher at the Massachusetts Institute of Technology, Cambridge, MA, USA, from 2006 to 2007. He was an Associate Research Scientist at Princeton University, where he is currently a Research Scholar with the Department

of Civil and Environmental Engineering, and involved in hydrology and remote sensing.

Dr. Pan is a member of the American Geophysical Union, the American Meteorological Society, and the European Geosciences Union.



Chaopeng Shen received the Ph.D. degree in environmental engineering from Michigan State University, East Lansing, MI, USA, in 2009.

He developed the hydrologic model process-based adaptive watershed simulator. He was a Post-Doctoral Research Associate with the Lawrence Berkeley National Laboratory, Berkeley, CA, USA, from 2011 to 2012. He is currently an Assistant Professor in civil engineering at Pennsylvania State University, University Park, PA, USA. His research interests are in hydrology

including interactions between water and other subsystems, catchment-floodplain-lake systems and scaling issues, process-based hydrologic modeling, multiscale modeling, hydrologic data mining, and hydrologic deep learning.

Dr. Shen is currently an Associate Editor of the *Water Resources Research*.