

# Near-Real-Time Forecast of Satellite-Based Soil Moisture Using Long Short-Term Memory with an Adaptive Data Integration Kernel

KUAI FANG AND CHAOPENG SHEN

*Department of Civil and Environmental Engineering, The Pennsylvania State University, University Park, Pennsylvania*

(Manuscript received 1 August 2019, in final form 16 December 2019)

## ABSTRACT

Nowcasts, or near-real-time (NRT) forecasts, of soil moisture based on the Soil Moisture Active and Passive (SMAP) mission could provide substantial value for a range of applications including hazards monitoring and agricultural planning. To provide such a NRT forecast with high fidelity, we enhanced a time series deep learning architecture, long short-term memory (LSTM), with a novel data integration (DI) kernel to assimilate the most recent SMAP observations as soon as they become available. The kernel is adaptive in that it can accommodate irregular observational schedules. Testing over the CONUS, this NRT forecast product showcases predictions with unprecedented accuracy when evaluated against subsequent SMAP retrievals. It showed smaller error than NRT forecasts reported in the literature, especially at longer forecast latency. The comparative advantage was due to LSTM's structural improvements, as well as its ability to utilize more input variables and more training data. The DI-LSTM was compared to the original LSTM model that runs without data integration, referred to as the projection model here. We found that the DI procedure removed the autocorrelated effects of forcing errors and errors due to processes not represented in the inputs, for example, irrigation and floodplain/lake inundation, as well as mismatches due to unseen forcing conditions. The effects of this purely data-driven DI kernel are discussed for the first time in the geosciences. Furthermore, this work presents an upper-bound estimate for the random component of the SMAP retrieval error.

## 1. Introduction

Surface soil moisture plays an important role in the water, carbon, and energy cycles by directly coupling atmospheric processes to land surface states. Soil moisture is critical for many applications, for example, irrigation planning, weather forecasts (Koster 2004), monitoring drought (Sheffield and Wood 2008), flood potential assessment (Norbiato et al. 2008), and landslide prediction (Ray et al. 2010; Brocca et al. 2012). Accurate near-real-time (NRT) forecasts of soil moisture have substantial societal value.

In the past decade, our measurement capability of surface soil moisture has been significantly improved by several satellite missions, including the Advanced Microwave Scanning Radiometer (AMSR) (Njoku et al. 2003), the Advanced Scatterometer (ASCAT) (Wagner et al. 1999), the Soil Moisture and Ocean Salinity (SMOS) (Kerr et al. 2010), and the Soil Moisture Active Passive (SMAP) (Entekhabi et al. 2010), among others. These spaceborne missions provide a global

measurement of soil moisture with typically  $\sim 2$ – $3$  days of revisit time. Due to the inherent characteristics of L-band microwave remote sensing, this revisit time is not expected to be reduced. Notwithstanding their great value, such temporal gaps may limit its use in applications demanding NRT soil moisture estimates.

Alternatively but not employed in this study, land surface models (LSMs) and data assimilation (DA) techniques such as ensemble Kalman filtering (EnKF) (Evensen 1994, 2003) could be used to generate NRT soil moisture forecasts or nowcasts. With the help of an observational operator, an estimate of observational uncertainty, and an ensemble of model runs that allows the estimation of a covariance matrix between simulated states, EnKF uses the observations to update model internal states. The assimilation of data is beneficial in that it will rectify model or forcing errors. With the updated states, the model can make better forecasts.

Several issues remain when we use DA to improve soil moisture forecasts. Because DA works through the lens of a dynamical system model (most often a process-based model), its effects critically depend on the structures of the LSM and a number of delicate techniques

---

*Corresponding author:* Chaopeng Shen, cshen@enr.psu.edu

DOI: 10.1175/JHM-D-19-0169.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) ([www.ametsoc.org/PUBSReuseLicenses](https://www.ametsoc.org/PUBSReuseLicenses)).

and user choices, for example, assimilation frequency, variables to be updated, and data preprocessing. For example, DA requires the observation to be unbiased with respect to the model. However, for soil moisture, satellite observation and LSM simulations often exhibit quite different mean values and variability. Often, as a preprocessing step of DA, satellite soil moisture products are locally rescaled and shifted to match the model climatology (Reichle 2004). It has been reported that such bias correction practices tend to exclude signals that disagree with model hypotheses (Kumar et al. 2015) and hence remove independent information provided by observations (Kolassa et al. 2017). In addition, because choices need to be made regarding which states to include in the covariance matrix, the DA scheme needs to be tailored and extensively tested for each different observational variable and LSM.

Previously, Koster et al. (2017, hereafter K17) introduced a data-driven method that produced NRT soil moisture forecasts based on SMAP data, with impressive and state-of-the-art results then. In that approach, soil moisture loss from evapotranspiration and drainage is defined as a function of soil moisture state, and the shape of the soil moisture loss function can be regarded as piecewise linear, to be estimated locally at each pixel. However, it is not clear if such a linear model could fully describe the process of soil moisture change, and, especially, provide strong performance for forecasts with a few days of latency.

Deep learning (DL) is well known for its ability to learn nonlinear mapping relationships and model dynamical systems (Shen 2018; LeCun et al. 2015; Schmidhuber 2015). In our previous work (Fang et al. 2017), we employed long short-term memory (LSTM) in a time series DL model to predict surface soil moisture based on climatic forcings and physiographic attributes such as soil texture and terrain slope. We showed that the LSTM model could reproduce SMAP soil moisture with unprecedented fidelity compared to conventional methods (Fang et al. 2017). Evaluated against in situ data, a SMAP-trained LSTM model could also add value to the long-term predictions and capture long-term trends (Fang et al. 2019a). However, previously our approach did not have the capability to assimilate NRT observations. Observations were used merely as the target in the training period, and not during forward simulations (“inference” in the machine learning terminology), that is, we did not exploit the value of recent observations to improve forecasts. Although forwarding lagged targets is a common practice in LSTM applications, for example, see Karpathy and Fei-Fei (2015), it is nontrivial to inject SMAP observations due to its irregular temporal gaps, which are

unavoidable for remote sensing, as well as many in situ measurements. We refer to such a model without data injection during a forward simulation as the *projection* model, in contrast to the *forecast* model that injects the latest observations during its prediction.

The overall objectives of this paper are 1) to introduce an adaptive and easy-to-implement kernel for the LSTM model to achieve NRT forecasts for soil moisture with high fidelity to SMAP, amenable to irregular observations, and 2) to interpret the differences between the projection and the forecast model and shed light on how hydrologic processes could impact soil moisture predictions. With the proposed approach, not only do we avoid the need for an LSM, we also avoid making explicit choices about how to use the observations. We call our procedure data integration (DI). DI achieves the goal of improving forecast using observations, but, compared to DA, it does not rely on running a forward model and then correcting its states. This model could serve as an operational large-scale soil moisture forecasting tool that can benefit downstream applications.

## 2. Methods

As an overview, a LSTM network with a data integration kernel is trained for NRT forecast of the SMAP L3 product, using atmospheric forcing time series, geographical attributes, and SMAP observations with a certain amount of latency (time lag between observation and prediction steps) as inputs.

### a. Input and target data

For the LSTM model, the training target is the SMAP L3 passive radiometer product (L3\_SM\_P) (Entekhabi et al. 2010). The SMAP mission intends to retrieve top 5-cm soil moisture by passive observations of surface brightness temperature on the L-band microwave, and the L3 product combines available swaths on a daily basis. The spatial resolution is 36 km on Equal-Area Scalable Earth Grid (EASE-Grid). The input data consist of climatic forcing time series and static physiographic attributes. Climatic forcings were extracted from North American Land Data Assimilation System phase 2 (NLDAS-2) (Xia et al. 2015), and included precipitation, temperature, radiation, humidity, and wind speed. Physiographic attributes contain soil properties extracted from the World Soil Information (ISRIC-WISE) database (Batjes 1995), including sand, silt, and clay percentages, bulk density, and soil water capacity, as well as land cover attributes provided by SMAP auxiliary data, including mountainous terrain, ice, surface roughness, urban areas, water bodies, land cover classes, and vegetation density. All of the inputs were regridded to EASE-Grid based on area weighting.

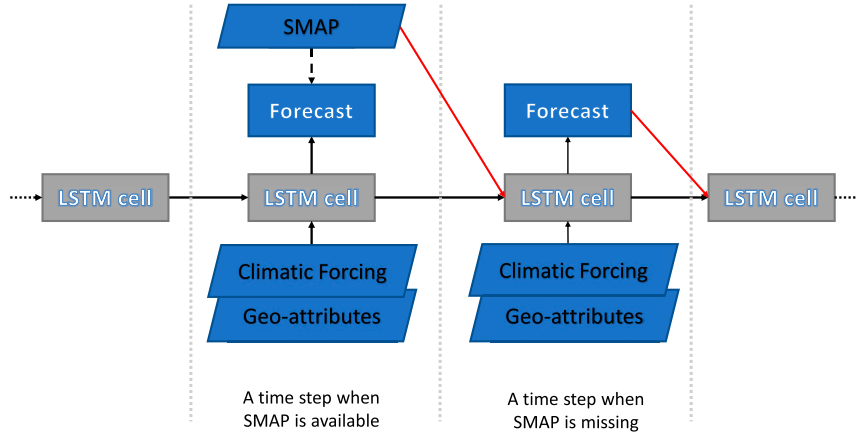


FIG. 1. A flow diagram of the data integration kernel. The solid lines stand for information passing forward, and the dashed lines stand for backward propagation. The LSTM cells have the same weights for all time steps. “Geo-attributes” stands for geographically distributed physiographic attributes.

*b. LSTM with an adaptive data integration kernel*

LSTM is a type of recurrent neural network (RNN), which makes use of sequential information. The “vanilla” RNN suffered from the “vanishing gradient” issue—the gradient values in the network shrink exponentially through time steps and prevent it from learning long-term dependencies. To deal with this issue, LSTM introduces a memory mechanism, where “memory states” units and “gates” are trained to decide when and what to remember or forget. The forward pass of LSTM is written as

$$\text{(input transfer)} x^{(t)} = \text{ReLU}(W_{xx} \mathbf{x}_0^{(t)} + b_{xx}), \quad (1)$$

$$\begin{aligned} \text{(input node)} g^{(t)} = \tanh & \left[ W_{gx} \mathcal{D}(x^{(t)}) \right. \\ & \left. + W_{gh} \mathcal{D}(h^{(t-1)}) + b_g \right], \end{aligned} \quad (2)$$

$$\text{(input gate)} i^{(t)} = \sigma [W_{ix} \mathcal{D}(x^{(t)}) + W_{ih} \mathcal{D}(h^{(t-1)}) + b_i], \quad (3)$$

$$\text{(forget gate)} f^{(t)} = \sigma [W_{fx} \mathcal{D}(x^{(t)}) + W_{fh} \mathcal{D}(h^{(t-1)}) + b_f], \quad (4)$$

$$\text{(output gate)} o^{(t)} = \sigma [W_{ox} \mathcal{D}(x^{(t)}) + W_{oh} \mathcal{D}(h^{(t-1)}) + b_o], \quad (5)$$

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

$$\text{(hidden gate)} h^{(t)} = \tanh(s^{(t)}) \odot o^{(t)}, \quad \text{and} \quad (7)$$

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

where the superscript refers to the time step;  $\mathbf{x}_0$  is the vector of raw inputs;  $z$  is observation, and  $y$  is network output;  $h$  is the hidden state;  $s$  is memory cells, which are designed to hold long-term memory; ReLU is the rectified linear unit;  $\sigma$  is the activation function;  $\mathcal{D}$  is the dropout operator (Zaremba et al. 2015);  $\odot$  refers to the point-wise multiplication;  $W$  is network weight; and  $b$  is bias.

In the original design of LSTM, at no time were recent observations employed in the predictions. Here we would like to append recent observations as a data-injection term to the other inputs described above. However, due to the prevalence of missing SMAP data (about 2/3 time steps for SMAP L3 data), a special, “closed-loop” procedure was needed to provide the data-injection term when no observations were available (Fig. 1). In this implementation, at time step  $t - 1$ , the network would produce a prediction for time step  $t$ , which will, at the next time step, serve as the default data-injection term, unless actual observations exist:

$$\begin{aligned} \text{(observation integration)} \mathbf{x}_0^t \\ = \begin{cases} [\mathbf{X}^{(t)}, z^{(t-1)}], & \text{if } z^{(t-1)} \text{ exist} \\ [\mathbf{X}^{(t)}, y^{(t-1)}], & \text{otherwise} \end{cases}, \end{aligned} \quad (9)$$

where  $\mathbf{X}$  is the vector of input data including climatic forcings and physiographic attributes, and  $z^{(t-1)}$  and  $y^{(t-1)}$  are observed and predicted soil moisture from the last time step, respectively.

This closed-loop is different from directly supplying lagged observations as an input to LSTM. A major difference is that if we directly supply lagged observations,

all training data can be prepared before training starts, yet this cannot be done for the closed-loop one. For the latter, the network essentially provides itself with training data for the data-injection term when no new information is available. This happens at runtime (both training and testing) so that each absent observation may influence the training data of the next time step and so on. To enable it, a time loop is implemented in the “forward” function, and a condition of whether observation exists or not was tested in the loop. The closed-loop kernel was out of necessity: without such a kernel, the LSTM algorithm will crash when missing data or “NaN” are fed to it as inputs. Interpolation is not suitable because values from future time steps are not available for interpolation for forecast tasks. Hence we need a forward extrapolator, and the LSTM algorithm itself is the most suitable one. This setup means that, during the first few epochs of the training, the network would produce poor predictions that would lead to incorrect gap filling data. However, as the prediction network improves during training, it also improves its gap filling capability, and the training would converge.

As mentioned earlier, here we refer to the model with the data integration kernel as the “forecast model” and the one without it as the “projection model.” The data integration kernel is “adaptive” as it can accommodate irregular observational schedules. It is also near-real time in the sense that all observations, as soon as they become available, are subsequently employed in the prediction of future time steps. Typically, SMAP data are disseminated with 1-day latency. This model was implemented using PyTorch (Paszke et al. 2017). As in Fang et al. (2017), the input and hidden size of LSTM were set to 256, and dropout rate is 0.5. The network was optimized for 500 epochs using the AdaDelta algorithm (Zeiler 2012), which adaptively updates the learning rate.

### c. Model training and evaluation

The training period was from 1 April 2015 to 31 March 2016, and the testing period was from 1 April 2016 to 31 March 2018. Considering computational efficiency, we collected one pixel from every  $2 \times 2$  pixels to form the training data, resulting in a 1/4 coverage of the continental United States (CONUS). Four statistical metrics, including the time-averaged difference (bias), root-mean-square error (RMSE), RMSE calculated after removing bias (ubRMSE), and Pearson correlation ( $R$ ) were calculated between model-predicted soil moisture and SMAP.

For the forecast model, we reported the error metrics for different numbers of forecast days. For example, for any forecast time step  $t_f$ , we found its most recently

available SMAP observation in the past, at  $t_o$ , and referred to it as a  $t_f - t_o$  day forecast. As the SMAP satellite returns every  $\sim 2\text{--}3$  days and would not report soil moisture when the ground is frozen, forecasts within 3 days cover 84.5% of the time window on the CONUS (41.0%, 31.2%, and 12.1% for 1-, 2-, and 3-day forecasts, respectively). However, we can only calculate error metrics when SMAP observations are available, which is 41.2% of the total time. Within those times, the proportions of 1-, 2-, and 3-day forecasts are 23.9%, 47.0%, and 25.1%, respectively. In addition, to illustrate the benefit of data integration, we calculated the difference in RMSE and  $R$  between the forecast and projection model.

We trained a separate model to compare with the model in K17, which was trained with 5 months of data (1 May 2015–30 September 2015) and only precipitation as the input. This model was tested on 1 May 2016–30 September 2016. We would like to provide different levels of comparisons to understand the respective benefits that were offered by (i) the structural advantage of LSTM over the linear functions, (ii) more training data, and (iii) more climate forcing variables as inputs. To this end, we retrained the DI-LSTM models using identical training periods as K17, and, separately, with precipitation only and with all forcing data.

## 3. Results and discussion

### a. Overall performance of forecast model

Overall, the forecast model was capable of predicting the SMAP L3 soil moisture product with unprecedented accuracy. We only present the error metrics of 1-, 2-, and 3-day forecasts as they fill most of the temporal gaps of the SMAP product. As Fig. 2a shows, the 1-day forecast product is highly consistent with the SMAP L3 product, producing a median RMSE of 0.022 and a median Pearson correlation of 0.92. For 2- and 3-day forecasts, the performance decayed slightly to a median RMSE of 0.024 and a median correlation of 0.90, but there were no obvious differences between the 2- and 3-day forecasts. The forecast models also produced autocorrelations that were much closer to SMAP than those from a land surface model (appendix A and Fig. A1). Comparing this version of the model with the results from K17, our approach reduced the error by roughly 20%. In general, there are many pixels with RMSE values less than 0.01 and few pixels above 0.04 (Fig. 3a). In comparison, the patch under Lake Michigan is 0.03 in our work and 0.06 in K17 (Fig. B1).

If trained on the same period as K17, with only precipitation as the forcing, DI-LSTM models still showed advantages over K17's scheme, and the differences were

## Error metrics of projection and forecast model

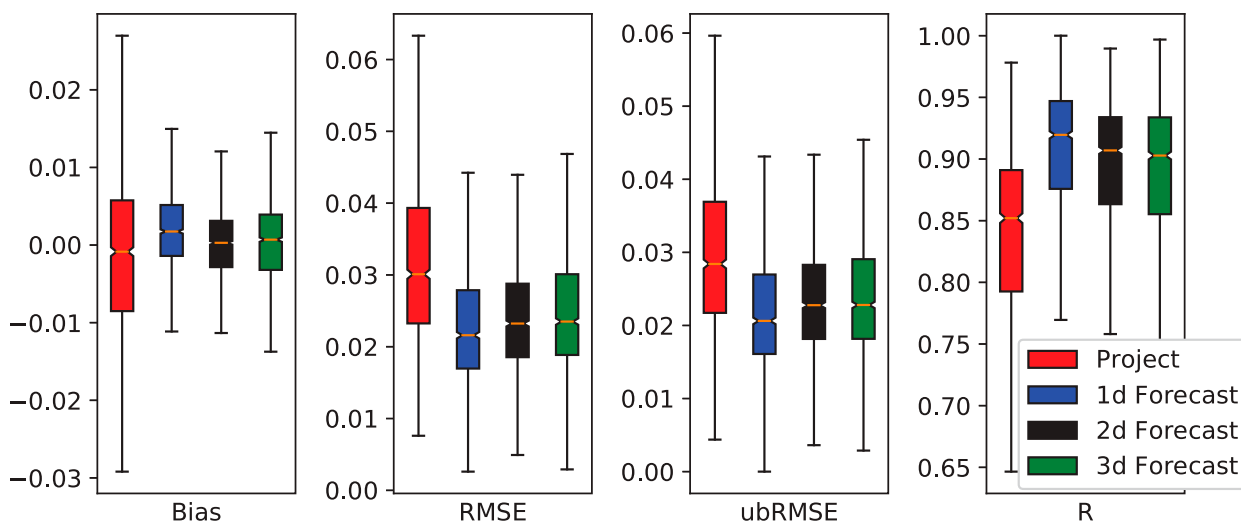


FIG. 2. Error metrics of soil moisture prediction from the projection model, and 1-, 2-, and 3-day forecast models.

more noticeable for the 2- and 3-day forecasts than the 1-day forecast. Evaluated using the mask from K17, the mean RMSE of DI-LSTM increased from 0.022 (with full forcings and 1-yr training data) to 0.025 for a 1-day forecast, which is roughly a 7% improvement from K17 (with a RMSE of 0.027). For 2- and 3-day forecasts, DI-LSTM with precipitation only had an average RMSE of 0.027 and 0.028, which are 13% and 18% smaller than K17 (0.032 and 0.034). Detailed pixel-level comparisons with K17 are presented in appendix B, Figs. B1 and B2. Comparing the DI-LSTM models trained with different forcings and with different lengths of training data, we learned that the advantages of the full DI-LSTM model, especially for 2- and 3-day forecasts, were contributed by all of the following: (i) the structural advantage of LSTM to model nonlinear processes and evolve it in time with memory, (ii) the ability to easily integrate more forcing fields other than precipitation, and (iii) the ability to improve with longer training data (appendix B).

Despite the overall high fidelity of forecasting product, there are regions of notably larger error, consistent with the expected SMAP error patterns (Fig. 3b). The eastern CONUS generally has larger errors compared to the west, which is due to larger annual precipitation and consequently larger moisture variability. The error map highlights the northeast CONUS along the Appalachian range, which resulted from the lower quality of SMAP data there due to the high vegetation water content (O'Neill et al. 2015), and higher fraction of time with frozen soil. We also see sporadic low- $R$  pixels along the southeast coast. Some other

notable regions of larger error include central north CONUS (Minnesota and southern Wisconsin where tens of thousands of lakes exist), and to the west of the lower Mississippi (eastern side of Missouri and Arkansas, where we find agricultural lands occupying wide floodplains).

The remaining RMSE with the forecast model contains three components: forcing error, model structural error, and random SMAP noise. The model structural error refers to difficult-to-predict hydrologic processes such as lakes, irrigation, and floodplain inundation. Because the DI procedure prevented autocorrelated effects of the first two components from accumulating, the remaining RMSE should be dominated by the noise component. As powerful as LSTM is, it cannot predict random noise during the test period. Hence the map in Fig. 3 constitutes an upper bound estimate of the random and nonsystematic component of the SMAP data's error.

The magnitude of forecast error is significantly smaller than earlier evaluation of SMAP products against in situ data, even the unbiased RMSE (Colliander et al. 2017). This suggests a large portion ( $\sim 30\%$ ) of the difference between SMAP and in situ data is due to systematic discrepancies. For example, a well-known discrepancy is the mismatch in sensing depths between SMAP and in situ data (Rondinelli et al. 2015). There may be other autocorrelated retrieval errors that are not as well known. We should not interpret the unbiased error against the in situ data as entirely random instrument noise. If it is possible to remedy those systematic errors using either enhanced

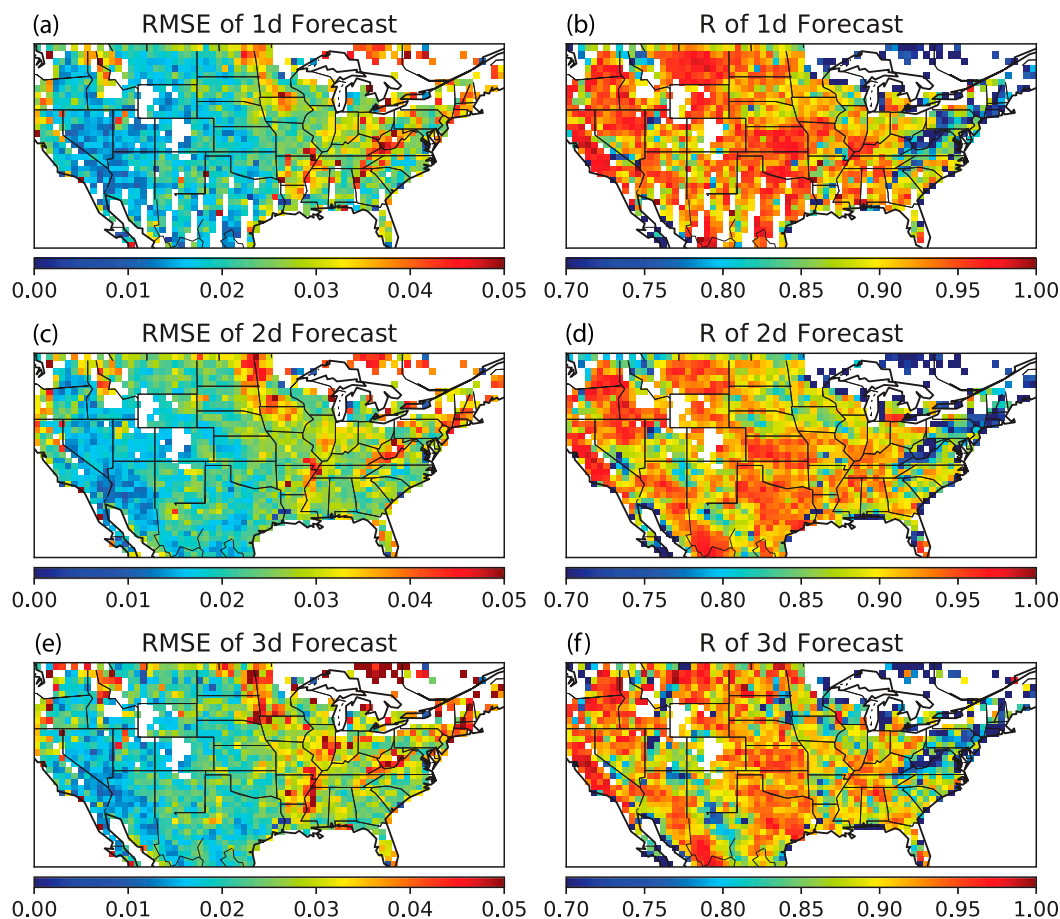


FIG. 3. Performance of soil moisture forecasts evaluated against SMAP L3 product. Shown are (a),(c),(e) the RMSE for 1-, 2-, and 3-day forecasts, respectively, and (b),(d),(f)  $R$ .

retrieving algorithms, the difference could be much smaller.

#### *b. Improvement over the projection model*

Comparing the projection model (without DI) to the forecast model (with DI), the data integration process rectified forcing errors and clearly improved the performance (Fig. 2). From the projection to the forecast model (merged forecast of all leading days), median RMSE decreased from 0.030 to 0.022, and median  $R$  increased from 0.85 to 0.9. More notably, the median magnitude of bias reduced from 0.009 to 0.003. Compared to data assimilation using LSMs, the LSTM has a very generic structure for modeling dynamical systems. The proposed DI kernel led to improvement on almost all pixels (>96%) over the CONUS. However, its spatial gradient (Fig. 4) suggests that different reasons are behind this improvement.

The most obvious effect of DI is to remove error autocorrelation, that is, it prevents errors from accumulating or persisting. For example, for several selected

pixels with relatively large differences between the projection and forecast models, the time series plots (Fig. 5) clearly show that most of the projection model errors were autocorrelated, that is, the difference between the projection and SMAP persisted for many days after its initial occurrence. With pixel A, the difference starting in April 2017 did not vanish until the beginning of July. For pixel B, a difference that occurred in October 2016 lasted until January 2017. For pixel F, the projection model overestimated soil moisture starting at the beginning of 2017, and a very similar difference was carried through April. In contrast, DI-LSTM seldom deviated more than one consecutive day from SMAP.

The improvement due to DI (Fig. 4) was substantial over many agricultural regions over the CONUS, including the southern parts of Illinois and the boundary of Iowa and Missouri (an example is pixel C in Fig. 5), North Dakota (pixel B in Fig. 5), the Central Valley, and the Mississippi Embayment (an example is pixel E in Fig. 5). The Central Valley region in California saw >30% reduction in RMSE, although the influence

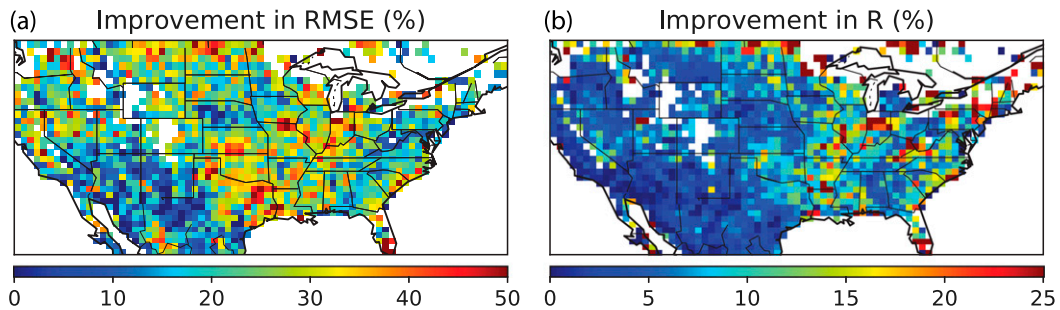


FIG. 4. Improvements from projection to forecast models. (a) RMSE and (b)  $R$  improvements were calculated as  $\{[RMSE(\text{projection}) - RMSE(\text{forecast})]/RMSE(\text{projection})\} \times 100\%$  and  $\{[R(\text{forecast}) - R(\text{projection})]/R(\text{projection})\} \times 100\%$ .

on  $R$  is not obvious there. The northern half of North Dakota is covered by spring wheat fields. The Mississippi Embayment region has extensive cropping of rice, soybean, and cotton. These patterns suggest that irrigation may be an important factor causing projection model error, which agrees with previous studies showing that SMAP could detect irrigation (Kumar et al. 2015; Lawston et al. 2017). As irrigation is absent from the input, the projection model could not have anticipated the added water input and hence underestimated soil moisture, as shown with pixel B in Fig. 5 in 2017. However, the forecast model greatly reduced that error after assimilating recent SMAP observations. In fact, as Fig. 6 shows, for different types of crops, the improvement from data integration had distinct climatology that generally agrees with the crop's growing schedule: the improvement is the most significant over summer, spring, and winter for corn, spring wheat, and winter wheat, respectively, while the improvement for corn is the weakest. In contrast to Felfelani et al.'s (2018) work, which assimilated SMAP into the Community Land Model via a 1D EnKF to constrain irrigation, the LSTM did not have an irrigation module and required the DI to reduce error.

One can compare the green-colored regions surrounding pixel C in Fig. 5b and the corn fraction map in Fig. 6b to see that the DI benefit was small for corn-dominated regions. This pattern agrees with the general understanding that a large fraction of the corn fields in the central plains are rainfed. Thus, the projection model already knew the water input (rainfall) and performed quite well in this region (Fig. 3). It is also worth mentioning that the Dakotas were not generally regarded as heavily irrigated regions (Kumar et al. 2015). However, the spring wheat has quite a different seasonal cycle in water use. There were also some discrepancies in the amount of irrigation in the Dakotas between

survey-based and MODIS-derived estimates (Ozdogan and Gutman 2008).

In addition to rice and other crops, the Mississippi Embayment (pixel E in Fig. 5) also contains extensive woody wetlands, large floodplains, and many river confluences into the Mississippi, all contributing to large open water surface areas. The surface areas, although not large enough to occupy SMAP pixels, could have impacted SMAP signals. Over adjacent fields classified as rice, significant DI benefits were obvious not only in May and June (due to rice irrigation), but more so in November and December (Fig. 6a). The DI benefits in November and December were likely attributable to the changes in floodplain inundation. As shown in the example of pixel E in Fig. 5, the projection model tended to overestimate soil moisture during this period. November and December are months where the Mississippi River stage is typically low but rising. Our interpretation is that, during low water periods, the inundation extent was lower than average. As the river stage was unknown to LSTM, the projection model could not resolve the reason for this decline and thus overestimated soil moisture. On the flip side, the projection model tended to underestimate during January–April of the year, which is the high-water stage. After river stage stabilized and started to change more slowly, the DI benefits became smaller. Similarly, relatively large DI benefits can be found in Wisconsin and Minnesota (two states to the west of Lake Michigan), where there are “thousands of lakes.” Presumably, the projection model had difficulty capturing riverine and lake inundation processes, while DI prevented errors from accumulating.

Data integration could also fix issues due to forcing conditions unseen in the training period. This issue is inevitable when the training data are still being accumulated. For example, for a pixel in northern Texas, if the projection model was trained on 2017, we notice underestimation of soil moisture during 2015 when

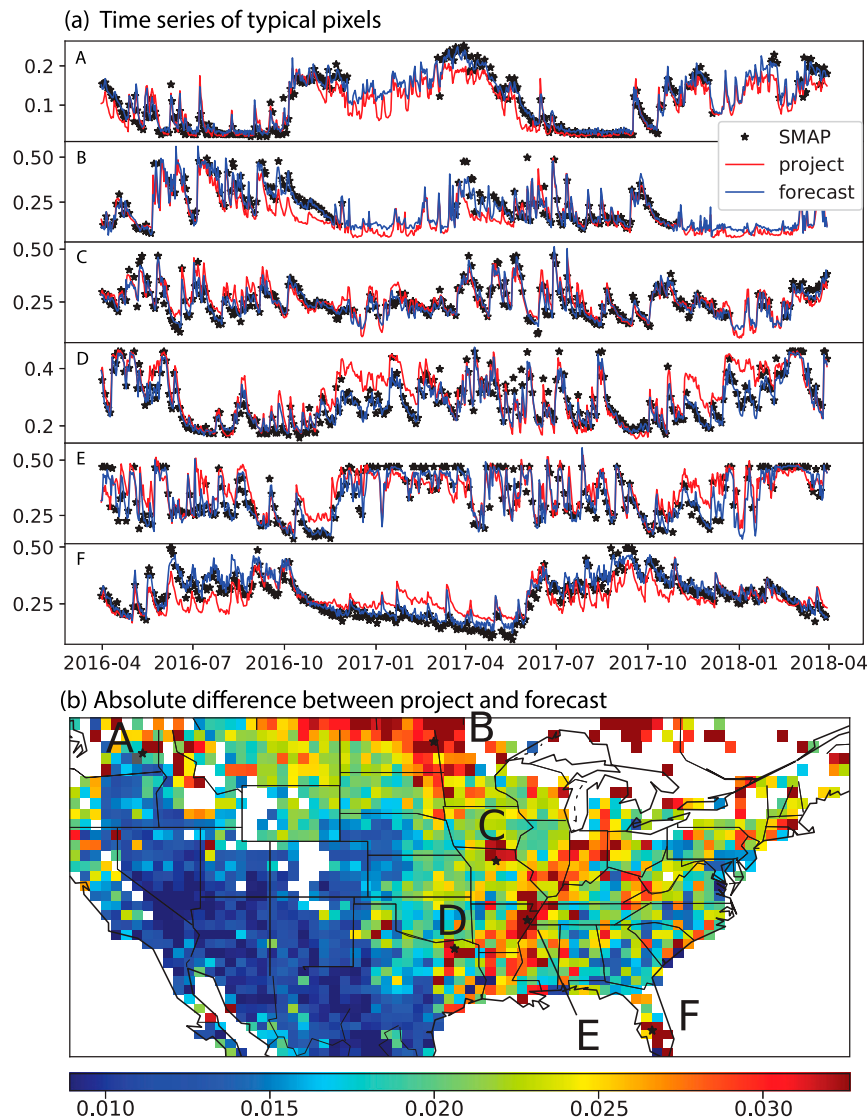


FIG. 5. (a) Soil moisture time series for pixels that have large differences between the projection and forecast models. (b) Locations of the chosen pixels.

extreme precipitation occurred (Fig. 7). Data integration can effectively correct this issue. It is worth mentioning that this issue could be mitigated by training on a comprehensive dataset or using weighting techniques in the loss function to emphasize the extreme events, which could be a future pursuit.

In summary, the projection model cannot predict processes that are not described by the inputs, for example, irrigation and riverine/lake inundation. Also, the performance of the projection model would decrease in unseen events. Data integration is most effective in removing the autocorrelated effects of forcing errors. With our DI framework, observation data can significantly reduce the model bias for

such situations and elevate the accuracy to an unprecedented level.

### c. Further discussion

While its first application in hydrology appears to be as recent as 2017 (Fang et al. 2017), LSTM has already been utilized in a number of prediction tasks, for example, streamflow (Kratzert et al. 2018), lake water temperature (Jia et al. 2019), pan evaporation (Majhi et al. 2020), water table depth (J. Zhang et al. 2018), and sewer outflow (D. Zhang et al. 2018), proving its versatility and general applicability. However, forecast applications have been more difficult due to often irregular observations. Here we show that, with a small amount of



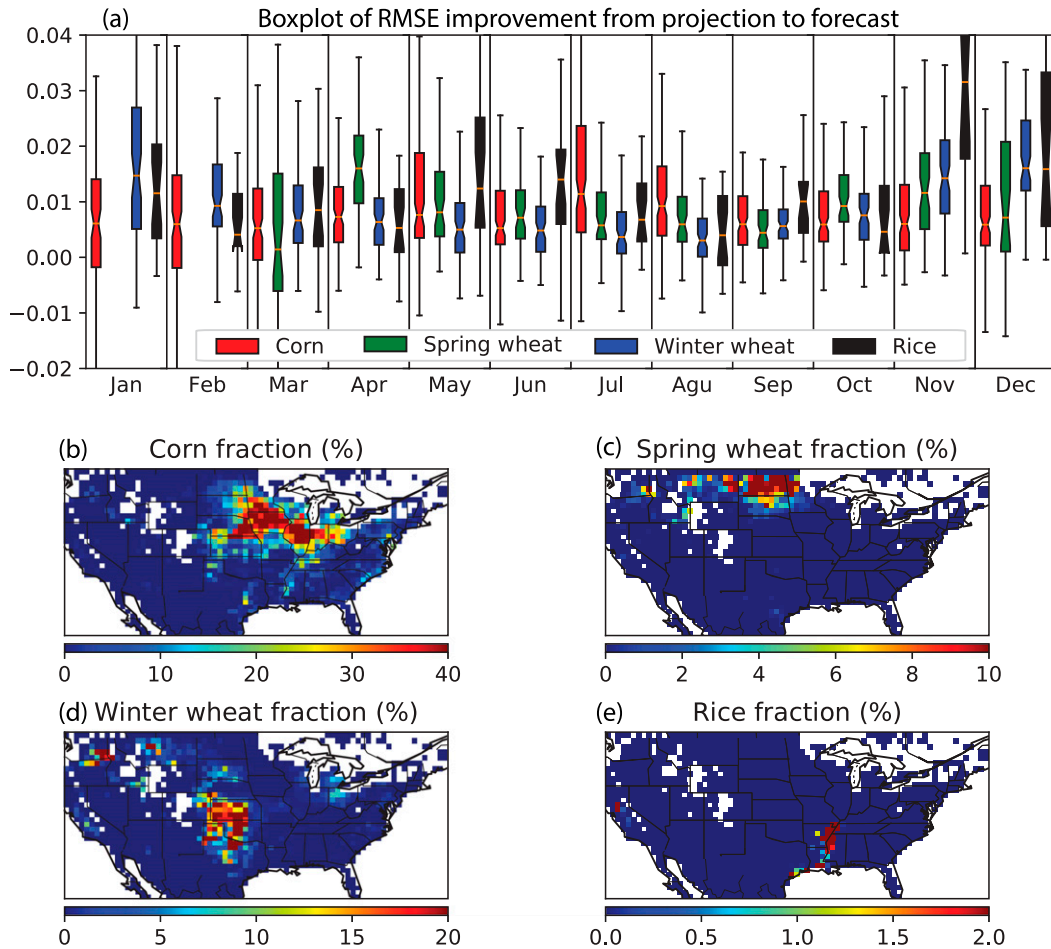


FIG. 6. (a) Box plot of DI benefits [RMSE(projection) – RMSE(forecast)] of different crops for each month. (b)–(e) Area fractions of corn, spring wheat, winter wheat, and rice, respectively.

modification [Eq. (9)], DL can be adaptive and can approach forecast problems and deliver superb performance. The network itself served as a forward extrapolator that provided training for itself when no observation was available, and it converged as the training proceeded. The adaptive nature of the proposed kernel gives us great convenience in dealing

with intermittent observations. If the projection model mimics a hydrologic model, then the forecast model approximates both the hydrologic process and the data injection procedure, which is a completely different problem than projection alone. Many applications could benefit from this DI framework. As the computational cost of a forward run of DL model is

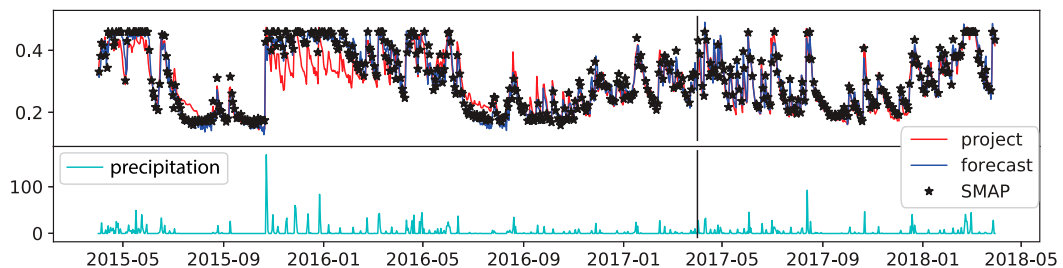


FIG. 7. Soil moisture time series from the projection and forecast models for a pixel in northern Texas. Here we trained the model from 1 Apr 2017 to 31 Mar 2018, and tested it from 1 Apr 2015 to 31 Mar 2017. (top) The soil moisture time series from the projection model, the forecast model, and SMAP. (bottom) Precipitation.

trivial, our model will bring the latency to the same level as meteorology datasets. It can also fill the gaps due to satellite revisit time with a small performance penalty.

Could the projection model have done a better job at describing the missing processes? We believe it could, if it was given better information and improved configuration of network structure. For example, if we further provide the time series of the Mississippi River stage and some information that characterizes the stage–inundation relationship, it could allow the network to better capture the impacts of riverine inundation on SMAP readings. For another example, if more information is provided regarding the crops and irrigation schedule, the corresponding errors would reduce. However, DI would still generate better results in these scenarios. Another point to further examine is the impacts of training regional models and a global model. K17 trained pixel-by-pixel models whereas our model is trained simultaneously with training data from all pixels, which potentially imposed stronger constraints and suppressed overfitting. We hypothesize that such training procedure allows the LSTM to extract commonality across regions and generalize the true soil moisture dynamics as modulated by different landscapes. This behavior could be why the present model showed smaller error metrics.

Ensemble model-based data assimilation and DL-based data integration deliver the forecast capability with very different mechanisms. With DA, one needs to make many decisions regarding the process-based model, a suitable DA and bias correction scheme, the observation operator, variables to include in the covariance matrix, and so on. With the proposed DI scheme, there is no need for a separate forward model and it does not require a specific assimilation scheme. One focuses on posing the problem, that is, which inputs are relevant to the outputs and providing the dataset, and allowing the neural network to discover the mathematical structures that connect them. Such automation is the soul of DL and one of the main reasons behind its recent rise in popularity, along with accurate predictions. This simplicity is bound to make socially relevant forecasts more accessible to a wider variety of users. In contrast, the advantage of DA is that the observations can be used to update other unobserved physical variables, as discussed previously. Currently, our DI scheme cannot perform this important task. It remains difficult to place physical significance on most the network-internal states.

LSTM contains hidden states and cell states, called  $h$  and  $s$  in Eq. (1), respectively, which serve as the internal memory and states of the model. Along with uncertainty

estimates (Fang et al. 2019b; Kendall and Gal 2017) and knowledge of observation variance, it is certainly possible to update these states using a more explicit approach that is similar in spirit to EnKF. It should also be possible to introduce physics in the deep network (Shen et al. 2018; Karpatne et al. 2017; Zhu et al. 2019). Such an effort is out of the scope of this paper. Here we elected to employ a simple scheme where LSTM decides how to best use the recent observations. We are quite certain that LSTM utilizes the assimilated observations to update both  $h$  and  $s$ , which, again, cannot be normally be interpreted as physical variables. It would be difficult to avoid a performance penalty when we add concurrent objectives or manual changes to the network structure. Nevertheless, we think different approaches should be attempted and compared in the future so that the community can learn their respective strengths.

Compared to the models in K17, DL can easily utilize forcing variables in the inputs due to its automation. Moreover, we showed that the performance of LSTM increased with additional training data (Fig. B2). DL is designed to thrive in a big data environment. Previous work in artificial intelligence found it is difficult to assess the potential limit in performance. For example, for image recognition tasks, Sun et al. (2017) found that a vision network's performance continued to improve even when the number of training images was increased to 300 million. Simpler functions may cease to improve (or become “knowledge-saturated”) much earlier, in our case perhaps after being trained with a few months of data.

In this work we mainly focused on the accuracy of hydrologic forecasts and neglected the error due to weather forecasts. The first reason is that SMAP data are provided with 1-day latency and thus accurate weather forcings could be available. The other reason is that the errors due to weather forecasts have already been shown by K17.

#### 4. Conclusions

In this paper we described a time series DL model with an adaptive data integration kernel to produce nowcasts and near-real-time forecasts for SMAP-based soil moisture. Adding to the war chest of hydrologic DL, this is the first time, to the best of our knowledge, a stepwise data integration kernel for DL has been reported in the hydrologic literature. The results showed unprecedented fidelity to the SMAP product, higher than SMAP design accuracy and those reported in the literature. The proposed DI kernel can be adopted by other operational forecast missions. Compared to data assimilation techniques, this approach avoids making choices such as bias

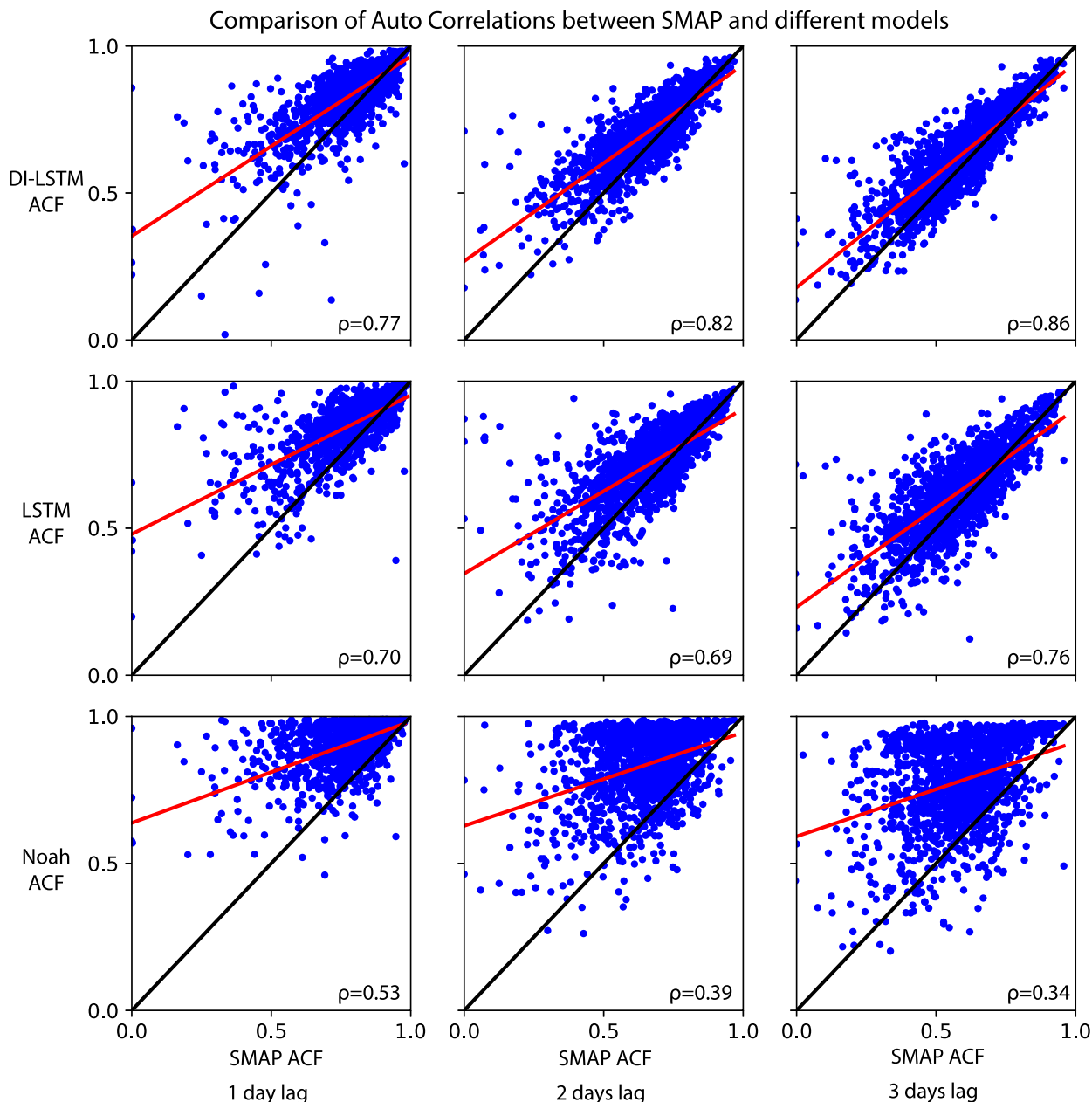


FIG. A1. Comparison of autocorrelation (ACF) between SMAP and simulated soil moisture from DI-LSTM, LSTM and the land surface model Noah (from NLDAS). Each point is a SMAP pixel. For each pixel, we aggregated data points with 1-, 2-, and 3-day lags between SMAP revisits, respectively, and calculated autocorrelations at these lags as the SMAP ACF. The model ACF was calculated on the same time steps as SMAP. We could see at 1-day lag the ACF has the highest range.

correction, and also circumvents structural assumptions in land surface models (it does not use a forward model), although it cannot modify unobserved physical variables. The scheme is simple to implement and can adapt to irregular observational schedule. We showed the scheme's advantage to linear functions used in the literature, especially for longer-term forecasts. The advantage was not only due to structural improvement but also due to its

ability to conveniently utilize longer training data and additional forcing variables.

The data integration sheds insights into the soil moisture dynamics and the sources of soil moisture product uncertainty over the CONUS. The procedure was effective at removing the autocorrelated effects of forcing errors or processes that could not be captured by a rainfall-soil moisture model, such as irrigation

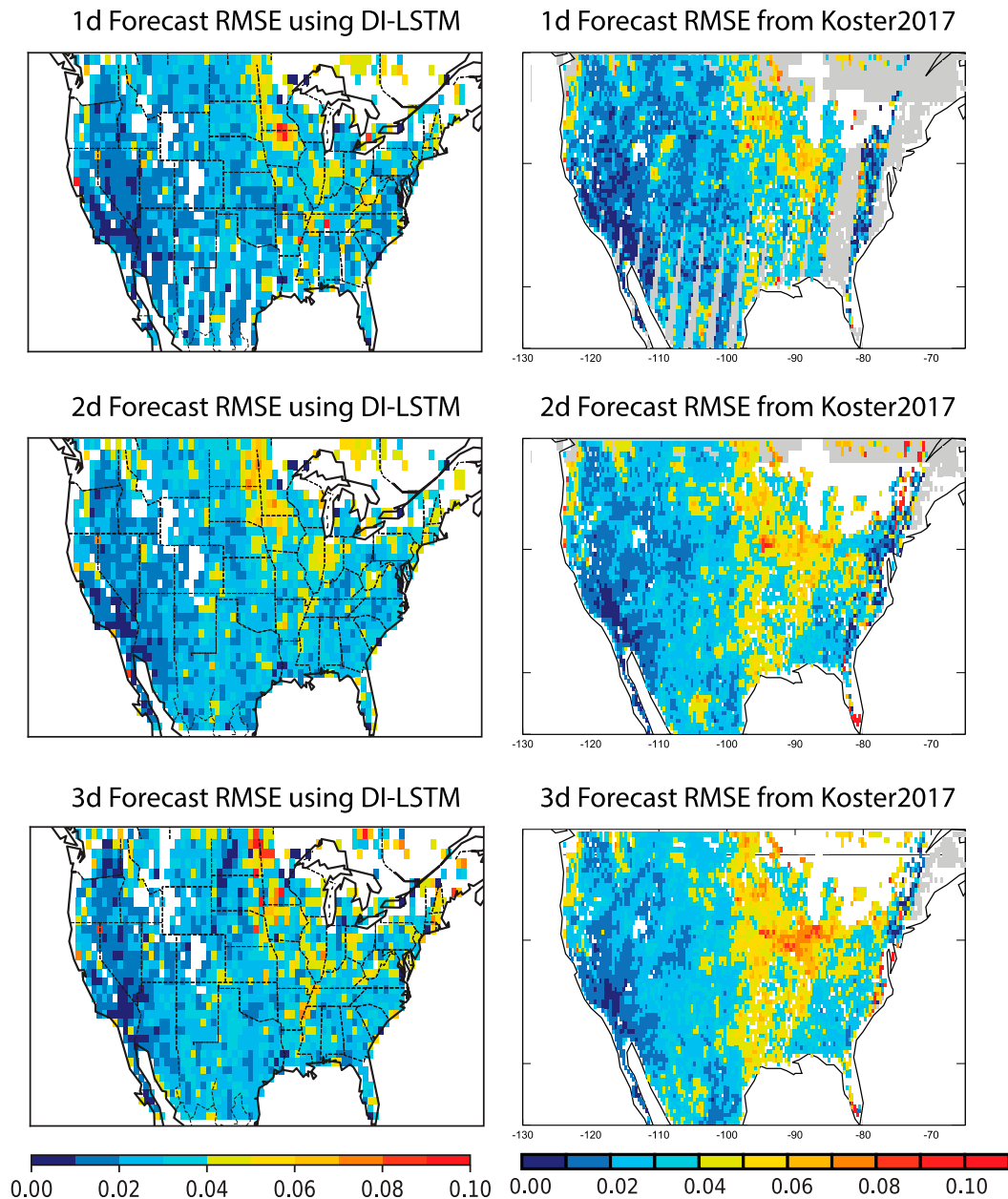


FIG. B1. RMSE from DI-LSTM in comparison with Fig. 3 in K17. The DI-LSTM was trained with the same data as K17 (only precipitation, from May 2015 to October 2015) and tested on the same time window (from May 2016 to October 2016).

and riverine/lake inundation. Also, it improved the predictive capability for extreme hydrologic events. The regions where we notice large DI benefits contained those with substantial irrigation, and the timing of the improvement was also in agreement with crops' growing schedule. We provided an upper-bound estimate of the random component of SMAP's observation error. As this estimate is 30% smaller than in situ comparisons, SMAP accuracy could be potentially improved if

enhanced retrieval algorithms could address some of the systematic deviations from in situ data. The forecast model error had similar spatial patterns to the recommended quality of the SMAP product, indicating that the remaining forecast errors are dominated by inherent noise with the SMAP measurement.

*Acknowledgments.* This study used publicly available data. Forcing and SMAP data can be downloaded from

RMSE of forecast models using different training set

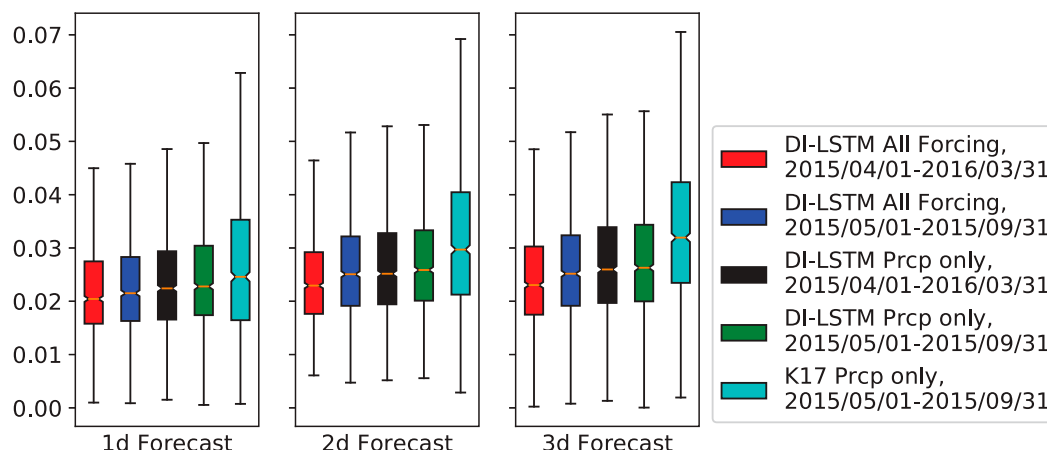


FIG. B2. RMSE of models trained on different datasets and tested from 1 May 2016 to 10 Oct 2016, as in K17. The training periods for the models are listed in the legend. Those four models were trained using 1) all forcing data (red; including precipitation, temperature, radiation, humidity, and wind speed) from April 2015 to April 2016. This model was used to produce the main results in this paper. 2) All forcing data from May 2015 to October 2015 (blue); 3) only precipitation from April 2015 to April 2016 (black); and 4) only precipitation from May 2015 to October 2015 (green), which is identical to K17. The K17 result is shown in the cyan box for comparison.

the NLDAS and SMAP websites, respectively. The code used in this paper is open-sourced at <https://github.com/mhpi/hydroDL>. This work was supported by partially by Office of Biological and Environmental Research of the U.S. Department of Energy under contract DE-SC0016605. KF was also partially supported by National Science Foundation under Grant 1832294, and partially by [Google.org](https://www.google.org) through the AI Impact Challenge. We thank Dr. Randall Koster, a reviewer of the paper, for sharing the detailed results of his previous work for comparisons. We appreciate the reviewers, whose comments helped us improve the paper.

APPENDIX A

Autocorrelation Comparison

How well did the LSTM capture the temporal autocorrelations in the data? Figure A1 shows the autocorrelation function (ACF) estimated from available SMAP data in comparison to that calculated from the model simulations. All models tended to overestimate ACF as opposed to SMAP. The DI-LSTM had the closest ACFs to SMAP, followed by LSTM, while the Noah tended to have quite higher ACFs than SMAP at all lags. Both DI-LSTM and LSTM also had higher correlation with SMAP in ACF than Noah, meaning that they captured the spatial variability in SMAP moisture recession dynamics well. The recession dynamics, on

the other hand, differed quite significantly between Noah and SMAP.

APPENDIX B

Detailed Comparisons with K17

The metrics reported in section 3 appeared stronger than those in K17, but the periods of simulations and the inputs were not identical. To learn about the respective benefits of the LSTM architecture and different input and training data, we trained models with the same inputs and training period as K17. We also compared it to the performance of two other models trained with (i) all forcing, May 2015–October 2015, and (ii) only precipitation, April 2015–April 2016.

The DI-LSTM generally produced smaller RMSE and similar spatial pattern as K17, due to the nature of SMAP error (Fig. B1). On Fig. B1, all settings were the same as K17 and only precipitation was used as inputs. Hence the benefits entirely derived from the LSTM architecture and its more complex memory capability that supports time integration. The difference is more noticeable for the agricultural Midwestern regions, especially for 2- and 3-day forecasts. LSTM had a more gentle performance decline for 2- and 3-day forecasts.

It appears that the effectiveness of including forcing variables other than precipitation was stronger than using longer training periods (Fig. B2). With only

5 months of training data, but with all forcing variables (blue box in Fig. B2), the test RMSE was only slightly higher than the full DI-LSTM (red box in Fig. B2), and lower than other models. Nevertheless, using longer training data, either with precipitation only or with all forcing data, helped to reduce the error.

The fact that the advantages of LSTM were stronger for 2- and 3-day forecasts suggests that LSTM is structurally better at keeping track of system memory and evolving the states over time. LSTM appeared to be better at modeling the moisture recession process than K17's scheme. On the other hand, model skills were obtained by easily incorporating temperature, wind, etc. in LSTM, which may not be straightforward in K17's scheme.

#### REFERENCES

- Batjes, N. H., 1995: A homogenized soil data file for global environmental research: A subset of FAO, ISRIC, and NRCS profiles, version 1.0. Tech. Rep. 95/10b, ISRIC, 43 pp., [https://www.isric.org/sites/default/files/isric\\_report\\_1995\\_10b.pdf](https://www.isric.org/sites/default/files/isric_report_1995_10b.pdf).
- Brocca, L., F. Ponziani, T. Moramarco, F. Melone, N. Berni, and W. Wagner, 2012: Improving landslide forecasting using ASCAT-derived soil moisture data: A case study of the Torgiovannetto landslide in central Italy. *Remote Sens.*, **4**, 1232–1244, <https://doi.org/10.3390/rs4051232>.
- Colliander, A., and Coauthors, 2017: Validation of SMAP surface soil moisture products with core validation sites. *Remote Sens. Environ.*, **191**, 215–231, <https://doi.org/10.1016/j.rse.2017.01.021>.
- Entekhabi, D., and Coauthors, 2010: The Soil Moisture Active Passive (SMAP) mission. *Proc. IEEE*, **98**, 704–716, <https://doi.org/10.1109/JPROC.2010.2043918>.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10143–10162, <https://doi.org/10.1029/94JC00572>.
- , 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, <https://doi.org/10.1007/s10236-003-0036-9>.
- Fang, K., C. Shen, D. Kifer, and X. Yang, 2017: Prolongation of SMAP to spatiotemporally seamless coverage of continental U.S. using a deep learning neural network. *Geophys. Res. Lett.*, **44**, 11 030–11 039, <https://doi.org/10.1002/2017GL075619>.
- , M. Pan, and C. Shen, 2019a: The value of SMAP for long-term soil moisture estimation with the help of deep learning. *IEEE Trans. Geosci. Remote Sens.*, **57**, 2221–2233, <https://doi.org/10.1109/TGRS.2018.2872131>.
- , C. Shen, and D. Kifer, 2019b: Evaluating aleatoric and epistemic uncertainties of time series deep learning models for soil moisture predictions. *36th ICML Workshop on Climate Change: How Can AI Help?* Long Beach, CA, ICML, 1–3, <https://arxiv.org/abs/1906.04595>.
- Fellfani, F., Y. Pokhrel, K. Guan, and D. M. Lawrence, 2018: Utilizing SMAP soil moisture data to constrain irrigation in the community land model. *Geophys. Res. Lett.*, **45**, 12 892–12 902, <https://doi.org/10.1029/2018GL080870>.
- Jia, X., J. Willard, A. Karpatne, J. Read, J. Zwart, M. Steinbach, and V. Kumar, 2019: Physics guided RNNs for modeling dynamical systems: A case study in simulating lake temperature profiles. *Proc. of the 2019 SIAM Int. Conf. on Data Mining*, Philadelphia, PA, Society for Industrial and Applied Mathematics, 558–566, <https://doi.org/10.1137/1.9781611975673.63>.
- Karpaty, A., and L. Fei-Fei, 2015: Deep visual-semantic alignments for generating image descriptions. *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Boston, MA, IEEE, 3128–3137, <https://doi.org/10.1109/CVPR.2015.7298932>.
- Karpatne, A., and Coauthors, 2017: Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Trans. Knowl. Data Eng.*, **29**, 2318–2331, <https://doi.org/10.1109/TKDE.2017.2720168>.
- Kendall, A., and Y. Gal, 2017: What uncertainties do we need in Bayesian deep learning for computer vision? Preprints, *31st Conf. on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, NIPS, 12 pp, <http://arxiv.org/abs/1703.04977>.
- Kerr, Y. H., and Coauthors, 2010: The SMOS mission: New tool for monitoring key elements of the global water cycle. *Proc. IEEE*, **98**, 666–687, <https://doi.org/10.1109/JPROC.2010.2043032>.
- Kolassa, J., and Coauthors, 2017: Data assimilation to extract soil moisture information from SMAP observations. *Remote Sens.*, **9**, 1179, <https://doi.org/10.3390/rs9111179>.
- Koster, R. D., 2004: Regions of strong coupling between soil moisture and precipitation. *Science*, **305**, 1138–1140, <https://doi.org/10.1126/science.1100217>.
- , R. H. Reichle, and S. P. P. Mahanama, 2017: A data-driven approach for daily real-time estimates and forecasts of near-surface soil moisture. *J. Hydrometeorol.*, **18**, 837–843, <https://doi.org/10.1175/JHM-D-16-0285.1>.
- Kratzert, F., D. Klotz, C. Brenner, K. Schulz, and M. Herrnegger, 2018: Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrol. Earth Syst. Sci.*, **22**, 6005–6022, <https://doi.org/10.5194/hess-22-6005-2018>.
- Kumar, S. V., C. D. Peters-Lidard, J. A. Santanello, R. H. Reichle, C. S. Draper, R. D. Koster, G. Nearing, and M. F. Jasinski, 2015: Evaluating the utility of satellite soil moisture retrievals over irrigated areas and the ability of land data assimilation methods to correct for unmodeled processes. *Hydrol. Earth Syst. Sci.*, **19**, 4463–4478, <https://doi.org/10.5194/hess-19-4463-2015>.
- Lawston, P. M., J. A. Santanello, and S. V. Kumar, 2017: Irrigation signals detected from SMAP soil moisture retrievals. *Geophys. Res. Lett.*, **44**, 11 860–11 867, <https://doi.org/10.1002/2017GL075733>.
- LeCun, Y., Y. Bengio, and G. Hinton, 2015: Deep learning. *Nature*, **521**, 436–444, <https://doi.org/10.1038/nature14539>.
- Majhi, B., D. Naidu, A. P. Mishra, and S. C. Satapathy, 2020: Improved prediction of daily pan evaporation using Deep-LSTM model. *Neural Comput. Appl.*, <https://doi.org/10.1007/S00521-019-04127-7>, in press.
- Njoku, E. G., T. J. Jackson, V. Lakshmi, T. K. Chan, and S. V. Nghiem, 2003: Soil moisture retrieval from AMSR-E. *IEEE Trans. Geosci. Remote Sens.*, **41**, 215–229, <https://doi.org/10.1109/TGRS.2002.808243>.
- Norbiato, D., M. Borga, S. Degli Esposti, E. Gaume, and S. Anquetin, 2008: Flash flood warning based on rainfall thresholds and soil moisture conditions: An assessment for gauged and ungauged basins. *J. Hydrol.*, **362**, 274–290, <https://doi.org/10.1016/j.jhydrol.2008.08.023>.
- O'Neill, P., S. Chan, E. Njoku, T. Jackson, and R. Bindlish, 2015: Algorithm theoretical basis document level 2 & 3 soil moisture

- (passive) data products, revision B. Tech. Rep. JPL D-66480, 80 pp., [http://smap.jpl.nasa.gov/system/internal\\_resources/details/original/316\\_L2\\_SM\\_P\\_ATBD\\_v7\\_Sep2015.pdf](http://smap.jpl.nasa.gov/system/internal_resources/details/original/316_L2_SM_P_ATBD_v7_Sep2015.pdf).
- Ozdogan, M., and G. Gutman, 2008: A new methodology to map irrigated areas using multi-temporal modis and ancillary data: An application example in the continental US. *Remote Sens. Environ.*, **112**, 3520–3537, <https://doi.org/10.1016/j.rse.2008.04.010>.
- Paszke, A., and Coauthors, 2017: Automatic differentiation in PyTorch. *31st Conf. on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, NIPS, 4 pp.
- Ray, R. L., J. M. Jacobs, and M. H. Cosh, 2010: Landslide susceptibility mapping using downscaled AMSR-E soil moisture: A case study from Cleveland Corral, California, US. *Remote Sens. Environ.*, **114**, 2624–2636, <https://doi.org/10.1016/j.rse.2010.05.033>.
- Reichle, R. H., 2004: Bias reduction in short records of satellite soil moisture. *Geophys. Res. Lett.*, **31**, L19501, <https://doi.org/10.1029/2004GL020938>.
- Rondinelli, W. J., B. K. Hornbuckle, J. C. Patton, M. H. Cosh, V. A. Walker, B. D. Carr, and S. D. Logsdon, 2015: Different rates of soil drying after rainfall are observed by the SMOS satellite and the South Fork in situ soil moisture network. *J. Hydrometeor.*, **16**, 889–903, <https://doi.org/10.1175/JHM-D-14-0137.1>.
- Schmidhuber, J., 2015: Deep learning in neural networks: An overview. *Neural Networks*, **61**, 85–117, <https://doi.org/10.1016/j.neunet.2014.09.003>.
- Sheffield, J., and E. F. Wood, 2008: Global trends and variability in soil moisture and drought characteristics, 1950–2000, from observation-driven simulations of the terrestrial hydrologic cycle. *J. Climate*, **21**, 432–458, <https://doi.org/10.1175/2007JCLI1822.1>.
- Shen, C., 2018: A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resour. Res.*, **54**, 8558–8593, <https://doi.org/10.1029/2018WR022643>.
- , and Coauthors, 2018: HESS opinions: Incubating deep-learning-powered hydrologic science advances as a community. *Hydrol. Earth Syst. Sci.*, **22**, 5639–5656, <https://doi.org/10.5194/hess-22-5639-2018>.
- Sun, C., A. Shrivastava, S. Singh, and A. Gupta, 2017: Revisiting unreasonable effectiveness of data in deep learning era. *Proc. IEEE Conf. on Computer Vision*, Venice, Italy, IEEE, 843–852, <https://doi.org/10.1109/ICCV.2017.97>.
- Wagner, W., G. Lemoine, and H. Rott, 1999: A method for estimating soil moisture from ERS scatterometer and soil data. *Remote Sens. Environ.*, **70**, 191–207, [https://doi.org/10.1016/S0034-4257\(99\)00036-X](https://doi.org/10.1016/S0034-4257(99)00036-X).
- Xia, Y., M. B. Ek, Y. Wu, T. Ford, and S. M. Quiring, 2015: Comparison of NLDAS-2 simulated and NASMD observed daily soil moisture. Part I: Comparison and analysis. *J. Hydrometeor.*, **16**, 1962–1980, <https://doi.org/10.1175/JHM-D-14-0096.1>.
- Zaremba, W., I. Sutskever, and O. Vinyals, 2015: Recurrent neural network regularization. *Int. Conf. on Learning Representations 2015*, San Diego, CA, ICLR, 8 pp., <https://arxiv.org/abs/1409.2329>.
- Zeiler, M. D., 2012: Adadelta: An adaptive learning rate method. arXiv, 6 pp., <https://arxiv.org/abs/1212.5701>.
- Zhang, D., G. Lindholm, and H. Ratnaweera, 2018: Use long short-term memory to enhance Internet of Things for combined sewer overflow monitoring. *J. Hydrol.*, **556**, 409–418, <https://doi.org/10.1016/j.jhydrol.2017.11.018>.
- Zhang, J., Y. Zhu, X. Zhang, M. Ye, and J. Yang, 2018: Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. *J. Hydrol.*, **561**, 918–929, <https://doi.org/10.1016/j.jhydrol.2018.04.065>.
- Zhu, Y., N. Zabararas, P.-S. Koutsourelakis, and P. Perdikaris, 2019: Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data. *J. Comput. Phys.*, **394**, 56–81, <https://doi.org/10.1016/j.jcp.2019.05.024>.