Soil moisture/evapotranspiration over-coupling and L-band brightness temperature assimilation: sources and forecast implications

Wade T. Crow[1], Concepcion Arroyo Gomez[1], Joaquín Muñoz Sabater[2], Thomas Holmes[3], Christopher R. Hain[4], Fangni Lei[5], Jianzhi Dong[1], Joseph G. Alfieri[1] and Martha C. Anderson[1]

[1] USDA ARS Hydrology and Remote Sensing Laboratory, Beltsville, Maryland, USA

[2] European Center for Medium Range Weather Forecasts, Reading, UK

[3] NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

[4] NASA Marshall Space Flight Center, Huntsville, Alabama, USA

[5] Geosystems Research Institute, Mississippi State University, Starkville, Mississippi, USA

Corresponding Author: Wade Crow (wade.crow@usda.gov)

Abstract: The assimilation of L-band surface brightness temperature (Tb) into the land surface model (LSM) component of a numerical weather prediction (NWP) system is generally expected to improve the quality of summertime 2-m air temperature (T2m) forecasts during water-limited surface conditions. However, recent retrospective results from the European Centre for Medium-Range Weather Forecasts (ECMWF) suggest that the assimilation of L-band Tb from the European Space Agency's (ESA) Soil Moisture Ocean Salinity (SMOS) mission may, under certain circumstances, degrade the accuracy of growing-season 24-h T2m forecasts within the central United States. To diagnose the source of this degradation, we evaluate ECMWF soil moisture (SM) and evapotranspiration (ET) forecasts using both in situ and remotely sensing resources. Results demonstrate that the assimilation of SMOS Tb broadly improves the ECMWF SM analysis in the central United States while simultaneously degrading the quality of 24-h ET forecasts. Based on a recently derived map of true global SM/ET coupling and a synthetic fraternal twin data assimilation experiment, we argue that the spatial and temporal characteristics of ECMWF SM analyses and ET forecast errors are consistent with

Early Online Release: This preliminary version has been accepted for publication in *Journal of Hydrometeorology,* may be fully cited, and has been assigned DOI 10.1175/JHM-D-20-0088.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

the hypothesis that the ECMWF LSM over-couples SM and ET and, as a result, is unable to effectively convert an improved SM analysis into enhanced ET and T2m forecasts. We demonstrate that this over-coupling is likely linked to the systematic underestimation of root-zone soil water storage capacity by LSMs within the United States Corn Belt region.

1. Introduction

During the growing season, soil moisture (SM) typically controls the partitioning of available energy between sensible and latent heat flux at the soil-atmosphere interface and thereby influences the energetic relationship between the land surface and the lower atmosphere. Furthermore, SM time series contain significant temporal persistence that can be exploited to forecast this relationship out in time. Therefore, the realistic initialization of SM states in the land surface model (LSM) component of a numerical weather prediction (NWP) system should, in theory, contribute to the skill of near-surface summer air temperature forecasts. However, this potential is not yet realized in operational weather prediction systems. Instead, SM values in operational NWP systems are often updated in a non-physical manner to minimize differences between observed and analyzed near-surface air temperature and relative humidity (Drusch and Viterbo 2007).

The shortcomings of this approach have spurred interest in the assimilation of SM information into operational NWP systems (Liu et al. 2012). Since ground-based observations of SM are seldomly available in near-real-time, NWP centers have instead focused on the development of data assimilation (DA) techniques to merge near-surface SM information acquired from satellite-based observations into their LSMs (Dharssi et al. 2011, Muñoz-Sabater et al. 2015; 2019, Carrera et al. 2015; 2019, Zheng et al. 2018). This approach combines best-possible estimates of land surface states based on available observations and short-range atmospheric

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

forecasts provided by the NWP system. In this regard, the European Space Agency (ESA) Soil Moisture and Ocean Salinity (SMOS) mission (Kerr et al. 2012), specifically designed to measure surface SM and ocean salinity from space, provides a unique opportunity to assimilate L-band microwave brightness temperature (Tb) observations that are highly sensitive to surface SM levels (Muñoz-Sabater et al. 2015). The assimilation of SMOS Tb should provide a more realistic representation of initial SM conditions, and subsequently, improved atmospheric forecasts in areas of significant land-atmosphere coupling.

Despite this potential, recent results have suggested that the assimilation of SMOS Tb can, under certain circumstances, degrade 2-m air temperature forecasts (Muñoz-Sabater et al. 2019, Carrera et al. 2019). Figure 1, based on results published previously in Muñoz-Sabater et al. (2019), illustrates this for 2012 and 2013 summer forecasts obtained from a retrospective analysis by the European Center for Medium-Range Weather Forecasts (ECMWF) NWP system over the central United States. The figure plots differences in root-mean-square error (RMSE) for 24-h forecasts (corresponding to ~18:00 local solar time in the central United States) of 2-m air temperature (T2m) for three separate DA cases: i) a control (CTRL) case based on the operational ECMWF approach of assimilating T2m and 2-m relative humidity (RH2m) observations to update SM states, ii) a new experimental (EXPR) case based on the assimilation of only L-band SMOS Tb and iii) a baseline open loop (OL) case of no land data assimilation. See below and Muñoz-Sabater et al. (2019) for further case details.

Red shading in Figure 1 indicates areas where the EXPR case has increased RMSE in 24-h T2m forecasts relative to either the CTRL (Figure 1a) or OL (Figure 1b) baseline cases. The increased 24-h T2m forecast RMSE (relative to the CTRL case) found along the eastern seaboard of the United States in Figure 1b is not wholly unexpected. The presence of significant

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

forest cover in this region reduces the amount of SM information present in SMOS Tb observations. In addition, the regional prevalence of energy-limited surface conditions reduces the value of SM for improving surface energy flux and, subsequently, T2m forecasts. As a result, it is not surprising that the assimilation of T2m and RH2m observations (in the CTRL case) is a more effective assimilation strategy in this region.

In contrast, the degradation of 24-h T2m forecast skill in the EXPR case over the north-central United States is more concerning. This region contains relatively little forest cover and commonly exhibits water-limited summertime surface conditions. Therefore, SMOS Tb observations should contain significant amounts of SM information, and this information should, in turn, improve ECMWF's ability to track surface energy fluxes and issue reliable 24-h T2m forecasts. This is especially true for comparisons against an OL case that is unaided by any data assimilation (Figure 1a). Bias results (not shown) reveal that elevated EXPR T2m RMSE values in this region are generally associated with a positive T2m bias.

Consequently, EXPR T2m forecast degradation in the central United States suggests a breakdown (somewhere) in the beneficial sequential chain linking: i) successful SMOS L-band Tb assimilation, ii) improved SM analyses, iii) improved short-term evapotranspiration (ET) forecasts and iv) improved short-term T2m forecasts. Our goal here is to systematically examine individual links in this chain and clarify if, and how, T2m forecast skill is squandered in the EXPR case.

ET forecasts at the center of this conceptual chain provide a critical link between SM analyses and forecasted T2m. However, the accuracy of ET forecasts is difficult to evaluate over large geographic regions. Recent work has illustrated that thermal infrared (TIR) remote sensing can be used to accurately constrain LSM representation of surface water and energy balance

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

processes – see, e.g., Han et al. (2015). Therefore, in addition to our conventional use of sparse, ground-based SM and ET observations to examine the SM-ET-T2m forecast chain, we also utilize ET retrievals acquired from TIR remote sensing and the Atmosphere-Land Exchange Inverse (ALEXI) model (Anderson et al. 2007; 2011) to continuously characterize the accuracy of ECMWF ET forecasts within a regional-scale domain. If successful, this application of large-scale, satellite-based ET retrievals as a diagnostic tool would represent an important advance in our ability to track the impact of SM analysis errors on NWP forecasts of the lower atmosphere.

Section 2 describes the ECMWF forecasts, ALEXI ET retrievals and ground-based SM and ET observations utilized in our analysis. Results are presented in Section 3 and discussed in Section 4 with the aid of synthetic fraternal twin synthetic experiments generated using a simplified soil water balance model. Finally, key paper conclusions are summarized in Section 5.

2. Data and methods

a. ECMWF data assimilation experiments

Launched in late 2009, ESA's SMOS project is the first satellite mission designed specifically to provide global retrievals of surface (0-5 cm) SM and sea-surface salinity. Still functioning as of mid-2020, the SMOS sensor passively measures microwave radiation emitted by the Earth's surface within the L-band portion of the electromagnetic spectrum (1.4 GHz) using an interferometric radiometer (Kerr et al. 2012). At this frequency, microwave Tb is modestly affected by both vegetation cover and the atmosphere and relatively more sensitive to surface SM conditions than higher frequency C- and X-band observations available from older passive microwave satellite missions. The SMOS instrument acquires individual L-band Tb retrievals at a spatial resolution of about 40 km and with a repeat time of every 2-3 days (at the equator).

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

ECMWF has conducted a series of hindcasting DA experiments to gauge the impact of assimilating SMOS Tb into their operational weather forecasting system (Muñoz-Sabater et al. 2015; 2019). Our focus here is on experiments conducted during the 2012– 2013 boreal summer and described in detail by Muñoz-Sabater et al. (2019). As discussed above, these experiments are based on comparisons between a "control" (CTRL) case that assimilates only screen-level meteorological variables (T2m and RH2m) versus an "experimental" (EXPR) case that assimilates only SMOS L-band Tb. An "open loop" (OL) case lacking any land data assimilation is also considered as a baseline. In all three cases, the LSM is the improved Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land (HTESSEL) used operationally by ECMWF (Balsamo et al. 2009) within the ECMWF Integrated Forecasting System (IFS).

All ECMWF data assimilation experiments were based on a 12-hour assimilation window in which all available observations of T2m and RH2m (for the CTRL case) and SMOS Tb (for the EXPR case) were collected and assimilated to update HTESSEL soil moisture states. For the CTRL case, the assimilation system assigned error standard deviations of 1 [°K] and 4 [%] for T2m and R2H observations, respectively. For the EXPR case, a variable SMOS Tb error standard deviation was assigned depending on the radiometric accuracy of the assimilated SMOS Tb observation. Updated states of soil moisture at 00 UTC were then used to launch the 24-hour T2m and ET forecasts examined here. For further details, see Muñoz-Sabater et al. (2019).

All ECMWF forecasts and analyses were interpolated to a spatial resolution of 0.25° (from their original non-regular grid at a horizontal resolution of approximately 40-km). Unless otherwise noted, forecasts were issued at 00 UTC with a lead time of 24 hours. Therefore, ET forecasts

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

[MJ m^{-2} d^{-1}] reflect the accumulation of forecasted flux between 00 and 23:59 UTC. Likewise, 24-h T2m forecasts reflect predictions of 2-m air temperature [°K] at 00 UTC – corresponding to ~18:00 local solar time in the central United States

All presented SM results are based on a DA analysis that reflects the best-available estimate of current soil moisture conditions based on all prior information. Specifically, SM analyses represent volumetric soil moisture [m^3 m^{-3}] content at 00 UTC for three vertical HTESSEL soil layers (0-7 cm, 7-28 cm, and 28-100 cm). Our period of interest is the 2012 and 2013 growing seasons (1 May to 30 September). Unfortunately, 2012 ET and SM OL fields were lost during the cyclical purging of experimental results at ECMWF. Therefore 2012 results shown below are limited to EXPR versus CTRL comparisons.

b. Satellite retrieval of daily ET

As introduced above, ALEXI is a diagnostic thermal infrared (TIR) model that calculates surface energy fluxes using the two-source energy balance (TSEB) approach of Norman et al. (1995). It models the land surface as a composite of soil and vegetation cover and couples the TSEB with an atmospheric boundary layer model to capture land–atmosphere feedback on T2m (Anderson et al. 2007; 2011). The land-surface representation in the ALEXI model partitions TIR retrievals of surface radiometric temperature (T_{RAD}) into its soil and canopy temperature components (T_s and T_c) assuming that $f(\theta)$ represents the apparent vegetation cover fraction at sensor view angle θ:

$$T_{RAD}\,(\theta) = \left[f(\theta){T_c}^4 + [1 - f(\theta)]{T_s}^4\right]^{1/4}. \qquad (1)$$

For a homogeneous vegetation canopy with a given leaf area index (LAI) and spherical leaf angle distribution, $f(\theta)$ is approximated as:

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

$$f(\theta) = 1 - \exp\left(\frac{-0.5\Omega(\theta)LAI}{cos\theta}\right). \tag{2}$$

where Ω is a vegetation clumping factor at view angle θ used to characterize non-random leaf area distributions (Anderson et al., 2005). Based on remote-sensing estimates of T_{RAD}, LAI, and radiative forcing, ALEXI solves for the soil (subscript 's') and the canopy (subscript 'c') energy budget terms individually and calculates composite (soil plus canopy) net radiation (*RN*), sensible (*H*), latent heat (*λE*) and soil heat (*G*) fluxes as:

$$RN = H + \lambda E + G \quad \begin{cases} RN = RN_c + RN_s \\ H = H_c + H_s \\ \lambda E = \lambda E_c + \lambda E_s \end{cases} \tag{3}$$

during cloud-free days. During cloudy days, fluxes are estimated by temporal smoothing and gap-filling the ratio of ET to solar radiation obtained on clear days and then multiplying this ratio by daily solar insolation values.

For this study, time series of morning T_{RAD} are obtained from the TIR channel (11 µm) on the Geostationary Operational Environmental Satellites (GOES) and LAI information from the Moderate Resolution Imaging Spectrometer (MODIS). The ALEXI model has been used to retrieve continuous daily ET since 2001 over the United States (Anderson et al. 2007, Hain et al. 2011). Here, we extracted daily (00 to 23:59 UTC) 0.25° ALEXI ET estimates [MJ m^{-2} d^{-1}] acquired during the 2012 and 2013 growing seasons (1 May to 30 September).

c. Ground-based SM observations

ECMWF surface-layer (0- to 7-cm) SM analyses were evaluated using observations acquired at a 5-cm measurement depth from the USDA Soil Climate Analysis Network (SCAN) and NOAA United States Climate Reference Network (USCRN). All SCAN and USCRN sites passing a basic quality check were considered (see below for details).

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

In addition, ECMWF root-zone layer (0- to 1-m) SM analyses were evaluated at selected USDA SCAN sites in the central United States. These analyses were based on the weighted averaging of SM estimates for the top three HTESSEL vertical soil layers (i.e., 0-7 cm, 7-28 cm and 28-100 cm). Corresponding USDA SCAN 1-m averages were based on the weighted averaging of SM observations available at ~5-, 10-, 20-, 50- and 100-cm depths assuming constant soil moisture within vertical soil layers (defined using boundaries corresponding to the mid-points between measurements obtained at successive depths). See Figure 2 for all site locations.

For all USCRN and SCAN observations (regardless of depth), temporal measurement gaps of less than 6 hours in SM measurements were bilinearly interpolated. The resulting hourly SM time series were then sub-sampled to acquire daily estimates of SM at 00 UTC. Days containing gaps larger than 6 hours were masked, and at least 100 valid daily SM measurements were required (in total) during the 2012 and 2013 growing seasons (1 May to 30 September) for a given site to be considered. Point-scale ground observations were assumed to represent an entire 0.25° grid cell. To identify non-representative sites, a minimum correlation of 0.30 [-] was required between USCRN/SCAN daily and (both) EXPR- and CTRL-case SM analyses for a given SM measurement site to be considered.

d. Ground-based ET observations

In addition to ALEXI ET retrievals, surface energy flux observations acquired at AmeriFlux network sites within the central United States (Table 1) were used to evaluate the quality of ECMWF 24-h ET forecasts. At these sites, all valid summertime 30-minute ET observations were multiplied by 48 and averaged within each day to obtain a daily (00 to 23:59 UTC) ET total [MJ m^{-2} d^{-1}]. At least 36 valid half-hourly observations per day were required for a given day to be considered, and we enforced a minimum threshold requirement of at least 25 daily

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

data pairs per year between ECMWF forecasts and ground observations. Flux tower sites not meeting this availability threshold, or providing discontinuous and/or non-realistic time series, were not considered. Note that certain tower sites met these thresholds for only one year of our two-year analysis. For the case of highly clustered sites within a single 0.25° grid cell, AmeriFlux observations from multiple towers were averaged into a single daily ET time series (Table 1).

In addition to the 17 AmeriFlux sites/clusters listed in Table 1, ground-based ET data were collected within the South Fork Watershed of the Iowa River at a Joint Experiment for Crop Assessment and Monitoring (JECAM) site maintained by the USDA Agricultural Research Service. During the 2012 and 2013 growing seasons (1 May to 30 September), 30-minute eddy covariance flux estimates were obtained from neighboring corn and soybean fields. Fluxes from these two sites were averaged based on weights consistent with local corn and soybean land cover fractions and summed into 00 to 23:59 UTC daily averages prior to their comparison against collocated 0.25° ECMWF ET forecasts.

Despite our best-effort attempts to maximize the spatial support of the ground-based ET measurements, it is inevitable that residual spatial representativeness errors will be present when flux tower observations are used as a point-of-reference for 0.25° ECMWF ET forecasts. The impact of these errors is discussed below.

e. SM/ET coupling assessment

Due to the impact of random retrieval error, it is generally difficult to assess SM/ET coupling strength using remote sensing products. Left uncorrected, random retrieval errors in SM and ET remote sensing products will spuriously bias observation-based coupling estimates low and compromise their value as an absolute benchmark for LSMs (Findell et al. 2015). To address

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

this issue, Crow et al. (2015) proposed a triple-collocation (TC) approach that uses multiple independent estimates of both SM and ET to calculate unbiased estimates of the true Spearman rank coefficient of determination (bounded as [0,1]) between SM and ET - even in the presence of significant random retrieval error in individual SM and ET products.

Lei et al. (2018) refined the approach of Crow et al. (2015) and applied it globally to weekly SM and ET products from a variety of global remote sensing and LSM sources. Specifically, they applied remotely sensed SM products acquired from the C-band Advanced SCATerometer (ASCAT) using the Vienna University of Technology (TU-Wien) change-detection algorithm (Naeimi et al. 2009) and passive microwave SM retrievals taken from the ESA Climate Change Initiative (CCI) Soil Moisture (v3.2) product (Dorigo et al. 2018). Remote sensing ET products were generated by applying the ALEXI model to both TIR- (Hain and Anderson 2017) and microwave-based land surface temperature retrievals (Holmes et al. 2015). LSM-based SM and ET products used to complete the required SM and ET triplets were obtained from offline LSM output provided by the Global Land Data Assimilation System (Rodell et al. 2004).

Based on these products, Lei et al. (2018) constructed a global map of benchmark SM/ET coupling strength (i.e., the true Spearman rank coefficient of determination between weekly SM and ET values). The exact 0.25°-resolution SM/ET coupling strength values utilized here were derived by applying the Lei et al. (2018) approach to SM and ET products collected during the 2012 and 2013 growing seasons.

Provided the error assumptions underlying the application of TC are satisfied (i.e., estimation errors are orthogonal and mutually independent), this assessment can be considered robust and independent of the specific datasets used to create it (Crow et al. 2015, Lei et al. 2018). Therefore, it provides an absolute point-of-reference for evaluating (correlation-based) SM/ET

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

coupling strength estimates provided by HTESSEL. Since it has been shown to represent the most realistic soil moisture conditions, HTESSEL is evaluated based on EXPR case results generated within the ECMWF IFS system. Nevertheless, several limitations in this approach should be acknowledged. First, due to the lack of global root-zone SM products available from remote sensing, this benchmark is based on ET coupling with surface (0-5 cm), and not root-zone (0-1 m), SM products. Second, like all TC assessments, the approach converges slowly in time. Therefore, two growing seasons of data (i.e., 2012 and 2013) represent a relatively short period for its application. Finally, the approach requires a minimum threshold of skill to be present in the SM and ET products it utilizes. Areas where this threshold is not met, due, e.g., to the loss of skill in surface SM retrievals under dense vegetation cover, must be masked.

3. Results

As noted above, the assimilation of SMOS surface SM (in the EXPR DA case) does not uniformly improve the accuracy of 24-h forecasts of T2m relative to the baseline CTRL case of assimilating T2m and RH2m or the OL case of no land data assimilation at all. Our primary goal here is explaining the source of this degradation within the central United States (Figure 1).

a. 00 UTC SM analyses

To start, it is important to confirm that SMOS Tb data assimilation improves the HTESSEL SM analysis at multiple soil depths. To this end, Figure 2 summarizes EXPR temporal correlation (R) differences, versus both the OL and CTRL baseline cases, for surface- (top row; 0- to 5-cm) and root-zone (bottom row; 0- to 1-m) 00 UTC SM analyses. All temporal correlations are sampled against benchmark SM observations acquired at USDA SCAN and NOAA USCRN sites (see Section 2.c). Prior to their assimilation in the EXPR case, SMOS Tb observations were linearly rescaled to match the climatological mean and standard deviation of the Tb values

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

estimated by applying a microwave forward model to surface state estimates provided by the ERA-Interim reanalysis (de Rosnay et al. 2019). This rescaling ensures that the assimilation of SMOS Tb cannot correct stable bias in HTESSEL SM estimates (used to generate the reanalysis) and, therefore, cannot significantly improve RMSE in cases where such bias is the major component of RMSE (Crow et al. 2005). Therefore, Figure 2 focuses on relative improvements in temporal *R* to summarize overall EXPR SM performance. Note that assessments of product-to-product *R* differences (for example, determining if EXPR or CTRL SM correlates better with true SM) are relatively insensitive to spatial representative errors (Dong et al. 2019, 2020).

At both depths (0-5 cm and 0-1 m), the EXPR case consistently improves the precision (i.e., correlation versus a high-quality reference) of SM analyses relative to the CTRL and OL baseline cases. Such improvement is particularly strong versus the OL case of no land data assimilation. Due to the inability of SM DA to correct bias (see above), RMSE results (not shown) are relatively more mixed. Nevertheless, the EXPR DA case still generally reduces surface SM RMSE across a large swath of the central United States and has, at worst, a neutral impact on root-zone SM RMSE.

Therefore, Figure 2 suggests that the degradation of EXPR T2m forecasts in the central United States in Figure 1 cannot be tied to a comparable degradation in the EXPR SM analysis. Instead, the SMOS Tb DA system functions as expected with regards to its net positive impact on the precision of ECMWF SM analyses. The relatively short temporal period of our analysis, combined with the highly autocorrelated nature of SM times series (particularly in the root-zone), prevents us from establishing the statistical significance of most precision improvements in Figure 2. However, these result are broadly consistent with a number of prior studies that

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

demonstrated the positive impact of L-band Tb (or SM) assimilation on the accuracy of LSM surface- and root-zone SM estimates (Muñoz-Sabater et al. 2019, Reichle et al. 2017; 2019, Blankenship et al. 2016, Mladenova et al. 2019, Carrera et al. 2015; 2019).

b. 24-h ET forecasts

Given that the EXPR case appears to enhance SM analysis precision (Figure 2), it becomes important to examine ET forecasts as the next link in the SM-ET-T2m forecast chain and a potential source of T2m forecast degradation within the central United States (see Figure 1). To this end, the background images in Figure 3 describe temporal *R* (top row) and RMSE (bottom row) differences between 24-h EXPR ET forecasts versus both the OL (left column) and CTRL (right column) baseline cases for the case of utilizing ALEXI ET retrievals as the reference benchmark. Note that while SMOS Tb assimilation (i.e., the EXPR case) often makes ECMWF ET forecasts more precise and accurate (i.e., improves *R* and RMSE fit to independent ALEXI ET retrievals), consistent degradation relative to both the CTRL and OL baseline cases is found over an area of the central United States that corresponds roughly to the region of degraded T2m forecasts in Figure 1. This implies that the net degradation in EXPR T2m forecasts seen in Figure 1 is linked to a comparable degradation in ET forecasts. That is, the beneficial chain linking improved SM analyses, ET forecasts, and T2m forecasts appears to break down at the interface between SM and ET.

As with the case of T2m forecasts in Figure 1, the net degradation in ET forecast accuracy is larger versus the CTRL baseline than against the OL case. Because of the beneficial impact of assimilating T2m and RH2m observations on surface flux forecasts, the CTRL case is a more accurate baseline and thus relatively harder to improve upon. In contrast, EXPR versus OL differences reflect only the (relatively smaller) net impact of assimilating SMOS Tb.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

While ALEXI ET retrievals used as a benchmark in Figure 3 are not error free, random errors in ALEXI daily ET estimates should not preferentially favor any of the forecast cases. Consequently, comparison against ALEXI ET retrievals provide a reliable assessment of relative accuracy (or precision) differences across multiple DA cases. In addition, ALEXI-based assessments of relative ET precision/accuracy are generally consistent with analogous assessments based on sparse, ground-based flux tower observations. Note the approximate correspondence in Figure 3 between the color shading of the background (derived using ALEXI as the ET benchmark) and the symbol fill colors (derived using sparse flux-tower listed in Table 1 as the ET benchmark). The agreement between these two independent assessments lends credibility to the conclusion that, within a broad swath of the central United States, the assimilation of SMOS Tb (in the EXPR DA case) degrades the accuracy of ECMWF short-term ET, and subsequently T2m forecasts, relative to both the CTRL and OL baseline cases (Figure 3). As discussed above, this degradation occurs despite the apparent improvement of the EXPR SM analysis relative to both baseline cases (Figure 2).

Figure 3 also reveals that, within the central United States, the CTRL case provides a far better fit to ALEXI ET than the OL case – note how ET degradation in the EXPR case becomes much more apparent when measured against the superior CTRL baseline (see the second column of Figure 3). Since the CTRL case is based on the use of T2m and RH2m observations to constrain ET, this improvement implies that the ECMWF IFS is correctly linking ET and T2m – which is consistent with the conclusion that the relationship between SM and ET represents the weak link in ECMWF IFS' representation of the SM-ET-T2m chain.

c. SM/ET temporal coupling

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

Taken as a whole, Figures 2-3 suggest that something in the way HTESSEL relates summertime SM to ET within the central United States prevents ET forecasts from realizing benefits derived from an improved SM analysis. This degradation in ET, in turn, appears responsible for the T2m forecast degradation seen in Figure 1.

Figure 4 explores this possibility by replotting the background of Figure 3c (i.e., the change in 24-h ET forecast RMSE between the EXPR and OL cases) and comparing it to a map of estimated bias in HTESSEL's representation of SM/ET temporal coupling – as calculated using the TC approach in Lei et al. (2018). As described in Section 2.e, the Lei et al. (2018) approach is noteworthy in that it corrects for the spurious low bias present in remote sensing-based estimates of SM/ET coupling due to the presence of independent random error afflicting estimates of SM and ET derived from various modelling and remote sensing sources. Therefore, it provides a robust estimate of absolute SM/ET coupling strength that is insensitive to the specific set of SM and ET products used to derive it (Crow et al. 2015). It can therefore be directly compared to LSM-based estimates of SM/ET coupling strength to identify LSM coupling-strength biases.

Areas where the assimilation of SMOS Tb degrades the accuracy of 24-h ET forecasts - see positive (blue) values in Figure 4a - generally correspond to regions where HTESSEL over-couples SM and ET - see positive (blue) values in Figure 4b. This suggests that SM/ET over-coupling is linked to the inability of the EXPR case to convert favorable EXPR SM results (Figure 2) into improved EXPR ET and T2m forecasts (Figures 1 and 3). It should also be noted that a general tendency towards LSM SM/ET over-coupling is also consistent with previous studies of LSM land-atmosphere coupling strength – see, e.g., Dirmeyer et al. (2018), Ukkola et al. (2016) and Lei et al. (2018).

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

4. Discussion

The specific mechanism linking HTESSEL SM/ET over-coupling (see Figure 4b) with the degradation of both ET and T2m EXPR forecasts is not immediately obvious. In this section, we will utilize a set of synthetic twin data assimilation experiments to clarify this mechanism and explain conditional biases present in ET and SM time series results at three central United States locations (A, B, and C; labelled in Figure 4a) where EXPR ET degradation is particularly strong.

a. Synthetic fraternal twin experiments

Figure 3 demonstrates that assimilation of SMOS Tb often degrades ET forecasts in the central United States despite having a consistently beneficial impact on the precision of SM estimates (Figure 2). Here we utilize a set of synthetic twin data assimilation experiments to resolve this apparent paradox. These experiments are based on the synthetic generation of "true" and "observed" SM states using a dynamic model and the re-assimilation of these synthetic observations back into the original dynamic model (after it has been degraded by synthetic modelling error). We will additionally differentiate the models applied in the observation-generation and assimilation steps by systematically introducing differences with respect to the assumed strength of SM/ET coupling (see above). Therefore, these synthetic twin experiments are "fraternal" in the sense that the assimilation model systematically differs from the base model used to generate the synthetic observations. Such experiments provide a well-controlled testbed for examining the impact of systematic modelling errors on data assimilation performance.

To this end, we will employ a simple model (and an assumption of statistically stationary climate) to describe the temporal evolution of SM as:

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

$$\mathrm{SM}_{t+1} = \exp(-\alpha)\mathrm{SM}_t - \beta_t + P_t \quad (4)$$

where P_t [mm] is time-varying precipitation; $\exp(-\alpha)\mathrm{SM}_t$ [mm] is a loss term assumed to be proportional to SM; β_t [mm] is a representation of random time-varying loss that is not linked to SM and α is a unit-less constant. Both loss terms in (4) are assumed due to ET - which can therefore be expressed via water balance principles as:

$$\mathrm{ET}_t = [1 - \exp(-\alpha)]\mathrm{SM}_t + \beta_t. \quad (5)$$

It is easily confirmed that the coupling strength between SM and ET (i.e., the partial derivative of (5) with respect to SM) is a monotonically increasing function of α. Therefore, hereinafter, α is used as a (nonlinear) unit-less proxy for SM/ET coupling strength.

Using the modelling system in (4-5), we conducted a series of synthetic fraternal twin experiments whereby synthetic "truth" estimates of SM were: i) generated via (4), ii) degraded through the introduction of synthetic random error and iii) then re-assimilated back into (4) using a Kalman Filter (KF) following the degradation of the P_t time series via random additive noise. A large set of such experiments was then produced where both true and assumed values of α were systematically varied (see axes on Figure 5). As such, these experiments illustrate the impact of assimilating SM observations into a model that systematically misrepresents the strength of SM/ET coupling (i.e., the magnitude of α). See Appendix A for additional methodological details on these experiments.

Our representation of this conditional bias in Figure 5 is based on the binary classification of true SM conditions as either "wet" or "dry" (i.e., less than or greater than the median value of the entire true SM times series). Conditional bias manifests itself as a difference between these opposing wet and dry time periods (i.e., column-wise differences in Figure 5 for a given row).

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

For presentation purposes, a single, long-term SM value has been removed from each individual synthetic result prior to plotting. Note that this has no impact on the magnitude of conditional biases.

Prior to DA, the inaccurate specification of α leads to conditional SM and ET biases in the OL case (see column versus column differences for the top two rows of Figure 5). Naturally, these biases are largest for cases where the assimilation model misrepresents SM/ET coupling (i.e., the off-diagonal portions of sub-plots in Figure 5 where assumed α does not match true α). However, the misspecification of α leads to contrasting signs in SM and ET conditional biases. That is, under conditions where the OL *underestimates* ET, excess moisture accumulates in the soil, leading to an *overestimation* of SM (and vice versa).

This sign contrast has important consequences for SM data assimilation. Since our simple model always assumes SM and ET are positively correlated via (5), efforts to correct time-varying errors in SM will tend to move ET in the wrong direction. Therefore, ET conditional bias is generally worsened by the correction of SM via DA in models that poorly describe SM/ET coupling. To see this, compare off-diagonal ET results for the OL case in the second row of Figure 5 to off-diagonal results for the KF case shown in the bottom row of Figure 5. This amplification of conditional bias during DA is generally stronger for the case of over-coupling (captured in the bottom-right corner of plots in Figure 5) than under-coupling (captured in the top-left corner). This break in symmetry occurs because the impact of SM errors on ET is relatively small when SM and ET are under-coupled. This allows the under-coupled SM/ET case to circumvent the negative inter-play between SM and ET conditional biases seen in the over-coupled case. Therefore, from the perspective of estimating ET using SM DA, over-coupling SM and ET is relatively more dangerous than analogous under-coupling.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

Also, note that ET degradation occurs despite the relatively robust removal of conditional SM bias present in the OL SM results by SM DA (compare the top row and the third rows in Figure 5). That is, the amplification of conditional bias by DA is only evident in ET estimates and is not reflected in the SM analysis.

b. Link to ECMWF forecast cases

Fraternal synthetic twin experiments summarized in Figure 5 illustrate that systematic errors in SM/ET coupling can lead to ET conditional biases that are exacerbated by the subsequent assimilation of SM observations. While these results are generated using a simplistic SM model, there is a substantial amount of overlap between synthetic twin DA results in Figure 5 and earlier real-data results presented in Figures 1-4.

To start, the observed ability of SMOS Tb DA to consistently improve the precision of SM analyses (see Figure 2) is consistent with the improvement of SM in the synthetic twin case (compare the first and third rows of Figure 5) – even for cases where SM/ET coupling is poorly characterized by the assimilation model. At the same time, synthetic results in Figure 5 illustrate how SM/ET over-coupling can produce a DA analysis where degraded ET forecasts and enhanced SM analyses simultaneously co-exist – thus explaining the apparent paradox noted above in the real-data EXPR SM and ET results over the central United States. The presence of SM/ET over-coupling in the central United States is also implied by comparisons between HTESSEL SM/ET coupling strengths and the independent SM/ET coupling strength assessment provided by Lei et al. (2018) (see Figure 4b).

Insight from the synthetic experiments in Figure 5 can also be used to explain SM and ET time series results (see Figure 6) extracted at labelled locations in Figure 4a. To start, it should be stressed that the model underlying the synthetic results is based on the simplistic assumption

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

that SM and ET are linearly related – see (5). However, both in nature and in HTESSEL physics, such coupling exists only for relatively dry SM conditions consistent with water-limited surface energy fluxes. Therefore, only the "dry-case" synthetic results (captured in the right-hand column of Figure 5) are likely to be directly relevant for interpreting time series results in Figure 6. Therefore, we will focus on the impact of SM/ET over-coupling during generally dry mid- to late-summer conditions.

During this period, all three sites in Figure 6 show a sharp decline in 1-m SM levels (see the bottom row of Figure 6). Because surface energy fluxes in the Corn Belt are commonly water-limited during the summer, this drying leads to a reduction in ET for the OL case (see the top row of Figure 6). However, since HTESSEL generally over-couples summertime SM and ET in the region (see Figure 4b), the resulting reduction in ET is excessive and induces a spurious reduction into OL ET results relative to the independent ALEXI ET benchmark (see the OL ET results along the top row of Figure 6). This reduction causes excess SM to progressively accumulate at all three sites during the late summer due to water balance considerations. As a result, late-summer SMOS Tb assimilation tends to remove soil water in the EXPR DA case (note the gap between OL and EXPR SM results that develops during this period in Figure 6). While this removal of water generally improves the HTESSEL SM analysis (see Figure 2), it also degrades ET forecasts relative to the OL baseline (Figure 3 and Figure 6) which, in turn, negatively impacts summertime T2m forecasts (Figure 1).

Note that these (real-data) dynamics are entirely consistent with earlier "dry" case synthetic results in Figure 5 for the over-coupled assimilation case (shown in the bottom-right of each plot along the right column of Figure 5). That is, during dry late summer conditions, over-coupling SM and ET leads to a simultaneous positive SM conditional bias (see bottom-right

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

portion of Figure 5b) and negative ET conditional bias in OL results (see bottom-right portion of Figure 5d). When SM DA is performed, the positive conditional SM bias is generally corrected (see bottom-right portion of Figure 5f); however, the negative conditional ET bias is exacerbated (compare the bottom-right portions of Figure 5d and 5h). Therefore, time series results in Figure 6 are consistent with expectations concerning the assimilation of SM (or Tb) information into a land model that over-couples SM and ET.

In addition, given the expected link between lower ET and higher T2m, the underestimation of growing season ET for the EXPR case in Figure 6 is consistent with the noted tendency for EXPR T2m RMSE results to be elevated by a positive T2m forecast bias in the central United States (see discussion of Figure 1 in Section 1). This also qualitatively agrees with independent results in Carrera et al. (2019) who noted that - in their conceptually similar Canadian Land Data Assimilation system - L-band Tb assimilation tends to introduce a negative bias into summertime 2-m dew point temperature forecasts within the central United States. Such a dry bias in near surface conditions is a natural consequence of underestimating surface ET.

c. Role of root-zone capacity

Given the apparent importance of SM/ET coupling strength bias on ECMWF EXPR ET and T2m forecasts, it is worthwhile to consider various candidate sources for this bias. One clue is the spatial correspondence between the region of degraded ET forecasts for the EXPR case relative to the CTRL baseline and the regional extent of the United States Corn Belt region (see Figure 3b).

Due to the depth and high organic content of its soils, the Corn Belt is generally characterized by very high values of root-zone soil water holding capacity. This capacity is exploited by the rapid vertical development of corn and soybean rooting systems that commonly extend below 1-

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

m in depth by late-summer (Ordonez et al. 2018, Abendroth et al. 2011). However, HTESSEL lumps all cultivated land under a single “crop” land cover type and assigns 96% of root volume for this land cover type into the top 1-m of the soil column – see Table 8.4 in ECMWF (2018). This suggests that real corn and soybean crops commonly have access to deeper (i.e., > 1-m) soil water storage than assumed by HTESSEL, and actual conditions exhibit less sensitivity (relative to HTESSEL) to temporal fluctuations in shallower SM values. Therefore, a low bias in root-zone water holding capacity (arrived at via mischaracterization of either soil type or rooting depth) will be associated with a high bias in HTESSEL SM/ET coupling strength, and the HTESSEL OL case can reasonably be expected to underestimate the (considerable) ability of the real Corn Belt system to buffer temporal periods of drying (Williams et al. 2016).

For the CTRL case, any such bias in root-zone capacity is mitigated by a DA analysis that systematically adds water during dry late summer (note the wetting of the CTRL case versus the OL baseline in Figure 6b) conditions to increase ET and match screen-level T2m and RH2m observations. For the NW Iowa and NE Kansas sites in Figure 6, such re-wetting of the soil column compensates for the late-summer underestimation of root-zone storage capacity in the model and generally maintains CTRL ET levels at or near independent ALEXI ET retrievals. In effect, the CTRL case adds water to the top-1-m of the soil column (and bolsters ET) to compensate for HTESSEL’s inability to capture the root extraction of soil water below 1-m. However, this compensating mechanism is not present in the OL case - causing a low bias in late-summer ET (Figure 6). This OL tendency is only exacerbated by SMOS Tb assimilation (in the EXPR case) due to the impact of SM/ET over-coupling (see earlier discussion in Section 4.b).

d. Alternative explanations

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

Above we argue that ECMWF T2m forecast errors are linked to SM/ET over-coupling in HTESSEL which, in turn, is associated with a low bias in assumed root-zone soil water holding capacity. However, since our case is admittedly circumstantial, the misrepresentation of other key processes within the United States Corn Belt region should also be considered.

1) NEGLECT OF C4 CROPS

In addition to large soil water holding capacities, a second defining characteristic of the Corn Belt is the preponderance of C4 crop cover (i.e., corn) and the inability of most LSMs to appropriately distinguish between C3 and C4 crops. The neglect of highly nonlinear C4 crop water stress processes has been shown to be a major limitation of existing LSMs (Verhoef and Egea 2014) and can cause systematic errors in the representation of SM/ET coupling strength - even in the case where root-zone water holding capacities are properly specified. However, an underestimation of non-linearity in the relationship between SM and ET does not appear to explain key ET conditional biases noted earlier in the Corn Belt for the OL case. For instance, if HTESSEL truly neglects nonlinearity in the grid-scale relationship between SM and ET (due to its neglect of C4 crops), then its OL case will produce too little ET during wet springtime conditions and too much ET during dry late-summer conditions (relative to a more nonlinear model that abruptly transitions between very high and very low ET conditions within a narrow root-zone SM window). This tendency is effectively the opposite of that seen in Figure 6 where, relative to the ALEXI ET baseline, the HTESSEL OL has too much ET in the spring and too little in the late summer.

In addition, the abrupt shut-off of ET in the nonlinear case would likely lead to higher late-summer SM than the linear case (where ET continues as a significant soil water loss mechanism down to much lower SM levels). Therefore, an excessively linear SM/ET case would likely

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

produce a low bias in late-summer SM conditions – whereas a comparison of EXPR and OL results in Figure 6 suggests the opposite (i.e. a positive late-summer SM bias in the HTESSEL OL case). One potential explanation for this is that the highly nonlinear evaporative stress relationship governing C4 crop ET response at a plot-scale (~10-m) is effectively linearized when applied to a coarse-scale grid containing large amounts of sub-grid SM spatial variability (Crow and Wood 1999). Therefore, the relatively linear HTESSEL evaporative stress relationship may, in the end, be more appropriate at the coarse grid-scale (~40-km) utilized in the ECMWF forecast system.

2) NEGLECT OF TILE DRAINAGE

A third defining characteristic of the United States Corn Belt (in addition to deep soil and C4 crop cover) is the widespread installation of tile drains to compensate for poor natural drainage from the soil column. These drains represent a key sink of root-zone soil water in the region that is typically neglected by LSMs (Hain et al. 2015, Yang et al. 2017). Therefore, it is reasonable to suggest the neglect of tile drainage in HTESSEL may produce a large-scale bias in HTESSEL OL ET and SM estimates. In fact, SM OL time series results in the bottom row of Figure 6 are generally consistent with this possibility. Note that SMOS Tb assimilation in the EXPR DA case tends to remove summertime SM from the OL case – which is consistent with the hypothesis that the HTESSEL OL case overestimates summertime SM due to its neglect of tile drainage losses. However, it is reasonable to expect that the neglect of tile drainage would also lead to excessive ET – since tile drainage increases the loss of spring SM storage and hastens the development of water-limited ET conditions later in the summer. This expected ET signal is not seen in OL ET results presented in the top row of Figure 6. To the contrary, the OL

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

case appears to underestimate ET in the late summer - which is difficult to rectify with the neglect of tile drainage from a water balance perspective.

3) NEGLECT OF IRRIGATION

Finally, while agriculture in the Corn Belt is generally rain-fed, the neglect of irrigation could potentially explain the observed underestimation of summertime ET for the HTESSEL OL case in Figure 6. However, the neglect of irrigation would also be associated with the underestimation of SM (particularly during the late summer) and an increase of SM (versus the OL case) in the EXPR DA case – a tendency that contradicts SM results in the bottom row of Figure 6.

In addition, the single area of the Corn Belt with extensive irrigation (eastern Nebraska; Green et al. 2018) is also the single Corn Belt sub-region where the EXPR case improves 24-h ET forecasts relative to the OL case (see the red-shaded area to the northwest of point "C" in Figure 4a). This suggests that unrepresented irrigation is not a plausible reason for the general degradation of EXPR ET forecasts across the Corn Belt. In fact, the presence of significant irrigation in eastern Nebraska seems to enhance the relative performance of the EXPR case since SMOS Tb assimilation provides an opportunity to compensate ET forecasts for irrigation water inputs that are missed in the OL case. Note that such compensation is generally consistent with previous assessments that L-band microwave observations (or SM retrievals based on these observations) can detect the presence of irrigation (Lawston et al. 2017).

Table 2 briefly summarizes the expected impact of missing (and/or mis-parameterized) land surface processes discussed above on late-summer SM and ET biases in HTESSEL OL output and compares these anticipated biases to actual biases found in Figure 6. While multiple processes operating within the United States Corn Belt are potentially neglected and/or poorly

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

represented by the HTESSEL OL case, only our original hypothesis of SM/ET over-coupling due to the underestimation of root-zone soil water holding capacity is fully consistent with the sign of observed late-summer SM and ET HTESSEL OL biases.

5. Summary and conclusions

It is commonly assumed that the improved representation of land surface states via DA will directly translate into better estimates of surface water and energy fluxes. This reasoning has formed the basis for intensive efforts to enhance NWP via the assimilation of microwave brightness temperature (Tb) observations (or surface soil moisture retrievals derived from such observations) into LSMs. While some success has been reported in these efforts (Muñoz-Sabater et al. 2019, Carrera et al. 2019), it is important to critically diagnose cases where expected improvements have not materialized. Here, we focus on the specific degradation of 24-h T2m forecasts within the central United States produced by the ECMWF forecast system during an experimental retrospective analysis assimilating SMOS L-band Tb ((Muñoz-Sabater et al. 2019).

An area of degraded 24-h T2m forecasts (Figure 1) in the central United States corresponds to a region where SMOS Tb assimilation improves surface and root-zone SM analyses (Figure 2), degrades ET forecasts (Figures 3) and the HTESSEL LSM over-couples SM and ET (Figure 4b) relative to the independent coupling benchmark provided by Lei et al. (2018). Using a synthetic twin analysis (Figure 5), we demonstrate that this third observation (i.e., SM/ET over-coupling) effectively explains the first two. In particular, the over-coupling of SM and ET can induce conditional biases into SM and ET estimates that are consistent with those found in the real-data OL results. In addition, the sign contrast in OL SM and ET conditional biases ensures that ET biases are exacerbated (rather that mitigated) by L-band Tb (or SM) DA. Therefore, SMOS Tb

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

DA (in the EXPR case) corrects surface and root-zone SM but simultaneously intensifies an existing conditional bias in ET. Based on this mechanism, the EXPR case DA systematically underpredicts ET during the middle to late summer (Figures 4 and 6), which, in turn, degrades T2m forecasts relative to both the CTRL and OL baseline cases (Figure 1).

Given that the area of degraded ET and T2m forecasts correspond well to the spatial extent of the United States Corn Belt, an agricultural source for SM/ET over-coupling (and associated ET and T2m degradation of the EXPR DA case) appears likely. The Corn Belt region is characterized by deep and organically rich soils and, as a result, very large root-zone soil water holding capacities. LSMs often underappreciate this capacity. In fact, the systematic under-estimation of root-zone soil water holding capacity by HTESSEL is generally consistent with the temporal and spatial characteristics of observed ET and SM conditional biases (see Section 4.c). Other agricultural characteristics of the Corn Belt region that are potentially neglected by the HTESSEL OL case (i.e., C4 crop cover, tile drainage and irrigation) are shown to be less likely causes of the bias due to their inability to explain the observed time and space characteristics of conditional biases present in OL SM and ET results (see Section 4.d and Table 2).

Alternative DA re-scaling techniques (capable of correcting for the presence of seasonally varying relative bias between HTESSEL and SMOS SM estimates) may improve EXPR DA results (Yilmaz et al. 2016). However, such a solution is arguably ad hoc and does not address the underlying SM/ET coupling strength bias present in the LSM. Instead, direct modifications to HTESSEL appear necessary for a robust solution. To this end, ECMWF is currently testing an HTESSEL implementation that utilizes a more extensive soil column (up to 8-m in depth

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

with 10 soil layers) capable of accommodating much deeper crop rooting depths. Results presented here are supportive of this approach.

While our focus here is solely on the central United States, ECMWF EXPR DA results were also degraded relative to the OL and CTRL cases over agricultural areas of central California, eastern Australia, the Sahel and the Eurasian wheat belt – see Figure 9 in Muñoz-Sabater et al. (2019). This implies that results presented here are relevant for multiple agricultural regions worldwide. Future research will explore this possibility.

Overall, results highlight the need to consider systematic aspects of LSMs before assuming the correction of random error in land surface states will directly translate into improved estimates of surface water and energy fluxes. Specifically, we highlight that systematic coupling errors can produce cases where conditional flux biases are reinforced (rather than mitigated) by DA. While past research has demonstrated that improperly parameterized DA systems can degrade model and state predictions (Reichle et al. 2008), this analysis illustrates that this danger extends to the case of a high-quality DA implementation for an LSM with systematic errors in its representation of state/flux coupling. Therefore, in a broader sense, this work highlights the danger of assuming that all LSM flux errors – regardless of their source – can be corrected by DA state correction. Instead, our results suggest that broader approaches considering the effects of both random and systematic errors sources must be used before land DA can consistently contribute additional value to NWP. In this regard, on-going improvements in the availability of remotely sensed ET retrievals (Holmes et al. 2018) and the improved accuracy of remote-sensing-based estimates of SM/ET coupling strength (Lei et al. 2018) provide valuable large-scale baselines for improving LSM representation of state/flux coupling strength.

Acknowledgments

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

Research was supported by NASA Aqua-Terra-Suomi research award 80HQTR18TO117 "Integrating multi-platform satellite soil moisture and evapotranspiration retrievals to constrain land surface water and energy balance coupling."

Data availability statement

Ameriflux data used here is available at https://ameriflux.lbl.gov/data/how-to-uploaddownload-data/. USDA SCAN and NOAA USCRN data is available at https://www.wcc.nrcs.usda.gov/scan/ and https://data.nodc.noaa.gov. All other datasets will be made available upon request.

Appendix A – Synthetic fraternal twin experiments

Synthetic twin experiments presented in Section 4.a and Figure 5 are based on the following steps. First, (4) is integrated forward in time for 150000 daily times for the case where: $SM_0 = 0$; β_t is sampled from a uniform distribution bounded between 0 and 10 mm; the parameter α is set to an arbitrary "true" value, and P_t is non-zero on 20% of days and sampled from the uniformed distribution bounded between 0 to 50 mm on rainy days. The results of this integration are assumed to represent a set of "true" SM_t observations.

Second, these "true" SM_t values are degraded via the introduction of mean-zero, additive, random, Gaussian noise with a variance of 25 mm^2 to represent observation certainty (i.e., the classical R in Kalman Filtering equations). Likewise, the precipitation time series P_t is degraded by mean-zero, random, Gaussian noise with a variance of 25 mm^2 to represent model forecast uncertainty (i.e., the classical Q in the Kalman filtering equations).

Third, the degraded observation are assimilated back into an integration of (4) driven by the degraded precipitation time series and using an assumed value of α. Assimilation is based on

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

applying a Kalman Filter (KF) and the same Q and R parameters given above. In addition, following typical practice in soil moisture data assimilation, the degraded SM_t time series is de-biased with respect to a temporal integration of (4) using the degraded P_t time series and the assumed value of α. This is done to minimize systematic errors arising from the misspecification of α and allow the KF to focus solely on the correction of random errors.

Finally, conditional bias is calculated in the KF analysis results relative to the true SM_t time series calculated in the first step. To construct the two-dimensional fields plotted in Figure 5, the entire procedure is systematically repeated for a range of true and assumed values of α. Plotted results in Figure 5 are averages obtained across 10000 separate experimental iterations.

As discussed in the main text, the term "fraternal twin" is used because the assimilation model and the "true" model simulation diverge due to the use of different α values. However, the KF assimilation system is considered optimal in the sense that it utilizes the correct values of Q and R (i.e., the error statistics that the Kalman Filter assumes to merge model estimates with observations are the exact statistics used to degrade the model and the observations in the synthetic experiment). This issue does become slightly ambiguous, however, due to the introduction of systematic error via the misspecification of α in the assimilation model. Therefore, one could argue that Q should be inflated in the Kalman filter implementation to capture the impact of *both* random error (explicitly introduced in the synthetic experiment) and this additional (implicit) source of systematic error. However, re-generating Figure 5 using Q inflation factors between 2 and 10 [-] had no qualitative impact on results.

Work cited

Abendroth, L.J., R.W. Elmore, M.J. Boyer, and S.K. Marlay. 2011: Corn growth and development. PMR 1009, Iowa State University Extension, Ames, IA.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

Anderson, M.C., J.M. Norman, W.P. Kustas, F. Li, J.H. Prueger, and J.M Mecikalski. 2005: Effects of vegetation clumping on two-source model estimates of surface energy fluxes from an agricultural landscape during SMACEX. *J. Hydrometeorol*. **6**, 892-909.

Anderson, M.C., J.M. Norman, J.R. Mecikalski, J.A. Otkin, and W.P. Kustas. 2007: A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 1. Model formulation, *Journal of Geophysical Research*, **112,** D10117, https://doi.org/10.1029/2006JD007506.

Anderson, MC., C. Hain, B. Wardlow, A. Pimstein, J.R. Mecikalski, and W.P. Kustas, 2011: Evaluation of drought indices based on thermal remote sensing of evapotranspiration over the continental United States, *Journal of Climate*, **24**, 8, 2025–2044, https://doi.org/10.1175/2010JCLI3812.1.

Balsamo, G., P. Viterbo, A.C.M. Beljaars, B. van den Hurk, M. Hirschi, A. Betts, and K. Scipal, K., 2009: A revised hydrology for the ECMWF model: verification from field site to terrestrial water storage and impact in the Integrated Forecast System, *Journal of Hydrometeorology*, **10**, 623–643, https://doi.org/10.1175/2008JHM1068.1.

Blankenship, C.B., J.L. Case, B.T. Zavodsky, and W.L. Crosson, 2016: Assimilation of SMOS Retrievals in the Land Information System. *IEEE Trans Geosci Remote Sens*., **54**, 11, 6320–6332, https://doi.org/10.1109/TGRS.2016.2579604.

Carrera, M.L., S. Bélair, and B. Bilodeau, 2015: The Canadian Land Data Assimilation System (CaLDAS): Description and synthetic evaluation study, *J. Hydrometeor.*, **16**, 1293–1314, https://doi.org/10.1175/JHM-D-14-0089.1.

Carrera, M.L., B. Bilodeau, S. Bélair, M. Abrahamowicz, A. Russell, and X. Wang, 2019: Assimilation of passive L-band microwave brightness temperatures in the Canadian Land Data Assimilation System: Impacts on short-range warm season numerical weather prediction. *J. Hydrometeor.*, **20**, 1053–1079, https://doi.org/10.1175/JHM-D-18-0133.1.

Crow, W.T. and E.F. Wood, 2002: Impact of soil moisture aggregation on surface energy flux prediction during SGP97, *Geophysical Research Letters*, **29**, 1, https://doi.org/10.1029/2001GL013796.

Crow, W.T., R.D. Koster, R.H. Reichle, and H. Sharif, 2005: Relevance of time-varying and time-invariant retrieval error sources on the utility of spaceborne soil moisture products, *Geophysical Research Letters.* **32**, L24405, https://doi.org/10.1029/2005GL024889.

Crow, W.T., E. Han, D. Ryu, C.R. Hain, and M.C. Anderson, 2017: Estimating annual water storage variations in medium-scale (10^3 to 10^4 km^2) basins using microwave-based soil moisture retrievals, *Hydrologic and Earth System Sciences*, **21**, 1849-1862, https://doi.org/10.5194/hess-21-1849-2017.

Crow, W.T., F. Lei, C. Hain, M.C. Anderson, R.L. Scott, D. Billesbach, and T. Arkebauer, 2015: Robust estimates of soil moisture and latent heat flux coupling strength obtained from triple collocation, *Geophysical Research Letters*, **42**, 8415-8423, https://doi.org/10.1002/2015GL065929.

De Rosnay, R., J. Muñoz-Sabater, C. Albergel, L. Isaksen, M. Drusch, J.P. Wigneron, 2019: SMOS brightness temperatures forward modelling, bias correction and long-term monitoring at ECMWF, *Remote Sens. Env*., **237**, 111424, https://doi.org/10.1016/j.rse.2019.111424.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

Dharssi, I., K.J. Bovis, B. Macpherson and C.P. Jones, 2011: Operational assimilation of ASCAT surface soil wetness at the Met Office, *Hydrol. Earth Syst. Sci*., **15**, 2729–2746, https://doi.org/10.5194/hess-15-2729-2011.

Dirmeyer, P. A., and co-authors, 2018: Verification of land-atmosphere coupling in forecast models, reanalyses, and land surface models using flux site observations, *Journal of Hydrometeorology*, **19**, 2, 375–392, https://doi.org/10.1175/JHM-D-17-0152.1.

Dong, J., W.T. Crow, R. Reichle, Q. Liu, F. Lei, and M. Cosh, 2019: Global assessment of added value in the SMAP Level 4 soil moisture product relative to its baseline land surface model, *Geophysical Research Letters*, **46**, 6604-6613. https://doi.org/10.1029/2019GL083398.

Dong, J., W.T. Crow, R. Reichle, K.J. Tobin, M.H. Cosh, D.D. Bosh, P.J. Starks, M. Seyfried, and C. Holifield, 2020: Comparison of microwave remote sensing and land surface modeling for surface soil moisture climatology estimation, in press, *Remote Sensing of Environment*.

Dorigo, W., and co-authors. 2018: ESA soil moisture climate change initiative (Soil_Moisture_cci): Version 03.2 data collection, *Centre for Environmental Data Analysis*, https://doi.org/10.5285/d2eea061026240eb8a2f9cc64a691338.

Drusch, M., and P. Viterbo, 2007: Assimilation of screen-level variables in ECMWF's Integrated Forecast System: A study on the impact on the forecast quality and analyzed soil moisture, *Mon. Weather Rev.*, **135**, 300–314, https://doi.org/10.1175/MWR3309.1.

ECMWF, 2018: IFS documentation – Cy45r1, operational implementation 5 June 2018, Part IV: physical process, https://drive.google.com/file/d/1beTmxnWx51hX-DvRMJtmshEybhB9xDDG/view.

Findell, K., P. Gentine, B. Lintner, and B. Guillod, 2015: Data length requirements for observational estimates of land–atmosphere coupling strength, *Journal of Hydrometeorology*, **16**, 4, 1615-1635, https://doi.org/10.1175/JHM-D-14-0131.1.

Green, T.R., H. Kipka, O. David, and G.S. McMaster, 2018: Where is the USA Corn Belt, and how is it changing?, *Science of the Total Environment*, **618**, 1613-1618, https://doi.org/10.1016/j.scitotenv.2017.09.325.

Hain, C.R., W.T. Crow, J.R. Mecikalski, M.C. Anderson, and T.R.H. Holmes. 2011: An intercomparison of available soil Moisture estimates from thermal infrared and passive microwave remote sensing and land surface modeling, *Journal of Geophysical Research-Atmospheres*, **116**, D15107, https://doi.org /10.1029/2011JD015633.

Hain, C.R., W.T. Crow, M.C. Anderson, and M.T. Yilmaz. 2015: Diagnosing neglected soil moisture source/sink processes via a thermal infrared-based two-source energy balance model, *Journal of Hydrometeorology*, **16**, 1070-1086, https://doi.org/10.1175/JHM-D-14-0017.1.

Hain, C. R., and M. C. Anderson. 2017: Estimating morning change in land surface temperature from MODIS day/night observations: Applications for surface energy balance modeling, *Geophysical Research Letters*, **44**, 9723–9733, https://doi.org/10.1002/2017GL074952.

Han, E., W.T. Crow, C.R. Hain, and M.C. Anderson, 2015: On the use of a water balance to evaluate inter-annual terrestrial ET variability, *Journal of Hydrometeorology*, **16**, https://doi.org/10.1175/JHM-D-14-0175.1.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

Holmes, T. R. H., Crow, W. T., Hain, C. R., Anderson, M., and W. P. Kustas. 2015: Diurnal temperature cycle as observed by thermal infrared and microwave radiometers, *Remote Sensing of Environment*, **158**, 110–125. https://doi.org/10.1016/j.rse.2014.10.031.

Holmes, T.H., C.R. Hain, W.T. Crow, M.C. Anderson, and W. Kustas, 2018: Microwave implementation of two-source energy balance approach for estimating evapotranspiration, *Hydrologic and Earth System Sciences*, **22**, 1351-1369, https://doi.org/10.5194/hess-22-1351-2018.

Kerr, Y.H., and co-authors, 2012: The SMOS soil moisture retrieval algorithm, *IEEE Transactions on Geoscience and Remote Sensing*, **50**, 5, 1384–1403, https://org.oi/doi:10.1109/TGRS.2012.2184548.

Lawston, P. M., J.A. Santanello, and S.V. Kumar, 2017: Irrigation signals detected from SMAP soil moisture retrievals, *Geophysical Research Letters*, **44**, 11860–11867, https://doi.org/10.1002/2017GL075733.

Lei, F., W.T Crow, T. Holmes, C. Hain, and M. Anderson, 2018: Global investigation of soil moisture and latent heat flux coupling strength, *Water Resources Research*, **54**, 8196-8215, https://doi.org/10.1029/2018WR023469.

Liu, Y., 2012: Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, *Hydrol. Earth Syst. Sci*., **16**, 3863–3887, https://doi.org/10.5194/hess-16-3863-2012.

Mladenova, I.E., J.D. Bolten, W.T. Crow, N. Sazib, M.H. Cosh, C.J. Tucker, and C. Reynolds. 2019: Evaluating the operational application of SMAP for global agricultural drought monitoring, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **12**, 9, 3387-3397, https://doi.org/10.1109/jstars.2019.2923555.

Muñoz-Sabater, J. 2015: Incorporation of passive microwave brightness temperatures in the ECMWF soil moisture analysis, *Remote Sensing*, **7**, 5, 5758–5784, https://doi.org/10.3390/rs70505758.

Muñoz- Sabater, J., and co-authors. 2019: Assimilation of SMOS brightness temperatures in the ECMWF Integrated Forecasting System, *Q J R Meteorol Soc*., **145**, 2524-2548, https://doi.org/10.1002/qj.3577.

Naeimi, V., Scipal, K., Bartalis, Z., Hasenauer, S., and W. Wagner. 2009: An improved soil moisture retrieval algorithm for ERS and METOP scatterometer observations, *IEEE Transactions on Geoscience and Remote Sensing*, **47**, 7, 1999–2013, https://doi.org/10.1109/TGRS.2008.2011617.

Norman, J.M., W.P. Kustas, K.S. Humes, 1995: A two-source approach for estimating soil and vegetation energy fluxes from observations of directional radiometric surface temperature, *Agric. For. Meteorol.,* **77**, 263-293, https://doi.org/10.1016/0168-1923(95)02265-Y.

Ordonez, R., M. Castellano, J. Hatfield, M. Helmers, M. Licht, M. Liebman, R. Dietzel, R. Martinez-Feria, J. Iqbal, L. Puntel, C. Cordova, K. Togliatti, E. Wright, S.V. Archontoulis, 2018: Corn and soybean root front velocity and maximum depth in Iowa, USA. *Field Crops Research*, **215**, 122-131, https://doi.org/10.1016/j.fcr.2017.09.003.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

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

Reichle, R.H. and co-authors, 2017: Assessment of the SMAP Level-4 Surface and Root-Zone Soil Moisture Product using in situ measurements, *Journal of Hydrometeorology*, **18**, 10, 2621-2645, https://doi.org/10.1175/JHM-D-17-0063.1.

Reichle, R.H. and co-authors, 2019: Version 4 of the SMAP Level-4 soil moisture algorithm and data product, *Journal of Advances in Modeling Earth Systems*, **11**, 106-3130. https://doi.org/10.1029/2019MS001729.

Rodell, M., and co-authors. 2004: The global land data assimilation system, *Bulletin of the American Meteorological Society*, **85**, 3, 381–394, https://doi.org/10.1175/BAMS-85-3-381.

Schnitkey, G., 2013: Concentration of corn and soybean production in the U.S., *Farmdoc Daily*, **3**, 130, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign.

Ukkola, A.M., and co-authors, 2016: Land surface models systematically overestimate the intensity, duration and magnitude of seasonal-scale drought, *Environmental Research Letters*, **11**, 104012, https://doi.org/10.1088/1748-9326/11/10/104012.

Verhoef, A., and G. Egea., 2014: Modeling plant transpiration under limited soil water: Comparison of different plant and soil hydraulic parameterizations and preliminary implications for their use in land surface models, *Agricultural and Forest Meteorology*, **191**, 22–32, https://doi.org/10.1016/j.agrformet.2014.02.009.

Williams A., and co-authors., 2016: Soil water holding capacity mitigates downside risk and volatility in US rainfed maize: Time to invest in soil organic matter? *PloS ONE*, **11**, 8, e0160974, https://doi.org/10.1371/journal.pone.0160974.

Yang, Y., M. Anderson, F. Gao, C. Hain, W. Kustas, T. Meyers, W.T. Crow, R. Finocchiaro, J. Otkin, L. Sun, and Y. Yang., 2017: Impact of tile drainage on evapotranspiration in South Dakota, USA based on high spatiotemporal resolution evapotranspiration time series From a multisatellite data fusion system, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **10**, 6, 2250-2264, https://doi.org/10.1109/JSTARS.2017.2680411.

Yilmaz, M.T., M.T. Crow, and D. Ryu, 2016: Impact of model relative accuracy in framework of rescaling observations in hydrological data assimilation studies, *Journal of Hydrometeorology*, **17**, 2245–2257, https://doi.org/10.1175/JHM-D-15-0206.1.

Zheng, W., X. Zhan, J. Liu, and M. Ek, 2018: A preliminary assessment of the impact of assimilating satellite soil Moisture data products on the NCEP Global Forecast System, *Advances in Meteorology*, 7363194, https://doi.org/10.1155/2018/7363194.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

Table 1. List of Ameriflux stations utilized in the analysis. Multiple site IDs listed under a single cluster number were averaged into a single time series prior to comparison with 0.25° ECMWF ET forecast grids.

Cluster #	Site ID	Latitude	Longitude	Elevation (m)	Land Cover
1	IB1	41.86	-88.22	226	Crop
	IB2	41.84	-88.24	226	Crop
2	RO1	44.71	-93.01	290	Crop
	RO2	44.73	-93.09	292	Crop
3	NE1	44.16	-96.48	361	Crop
	NE2	44.16	-96.47	362	Crop
	NE3	44.18	-96.44	362	Crop
4	ARM	36.61	-97.49	314	Crop
5	CRT	41.63	-83.35	180	Crop
6	KFS	39.06	-95.19	310	Grassland
7	KLS	38.77	-97.57	373	Grassland
8	GLE	41.37	-106.24	3197	Evergreen Forest
9	WHS	31.74	-110.05	1360	Shrubland
10	CPK	41.07	-106.12	2750	Evergreen Forest
11	NR1	40.03	-105.54	3050	Evergreen Forest

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

12	MOZ	38.74	-92.20	219	Deciduous Forest
13	PFA	45.95	-90.27	470	Mixed Forest
14	WCR	45.81	-90.08	520	Deciduous Forest
15	SYV	46.24	-89.35	540	Mixed Forest
16	MMS	39.32	-86.41	275	Deciduous Forest
17	UMB	45.56	-84.70	239	Deciduous Forest

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

Table 2. Summary of signs in observed and expected late-summer SM and ET biases. The positive sign for the "observed" SM OL bias is inferred from the tendency for SMOS Tb assimilation (i.e., the EXPR case) to remove soil water from late-summer OL results in Figure 6. Likewise, the negative sign for "observed" ET OL bias is based on late-summer comparisons between OL and ALEXI ET time series in Figure 6.

	SM OL bias (late summer)	**ET OL bias (late summer)**
***Observed* (see Section 4.c and Figure 6):**		
	Positive	Negative
***Anticipated impacts of model errors* (see Section 4.d):**		
Too little soil water cap.	Positive	Negative
C4 crops neglected	Negative	Positive
Tile-drainage neglected	Positive	Positive
Irrigation neglected	Negative	Negative

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

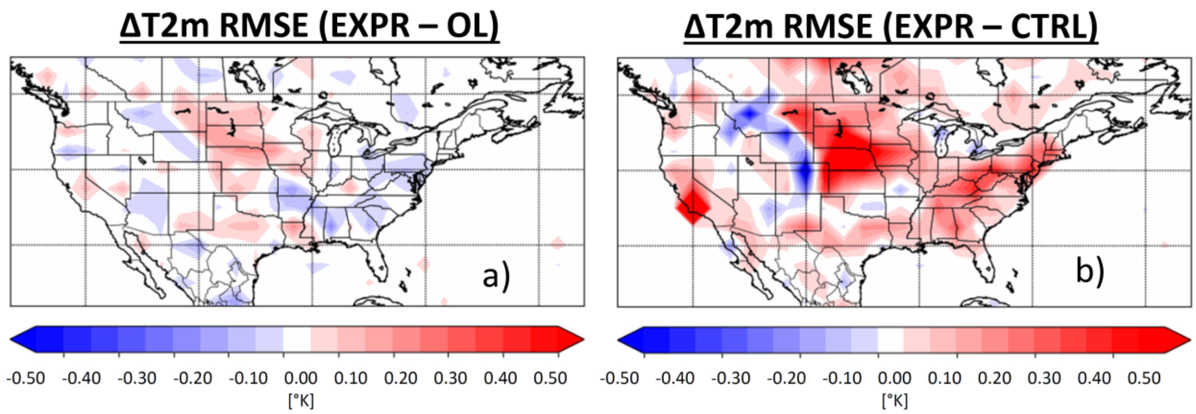

Figure 1. Change in EXPR T2m RMSE relative to the a) OL and b) CTRL cases for 24-h T2m forecasts. RMSE [°K] results are sampled across the 2012 and 2013 growing seasons (1 May to 30 September). Red shading indicates areas where SMOS Tb assimilation degrades T2m forecast skill relative to either the OL or CTRL baselines. Results taken from Muñoz-Sabater et al. (2019).

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

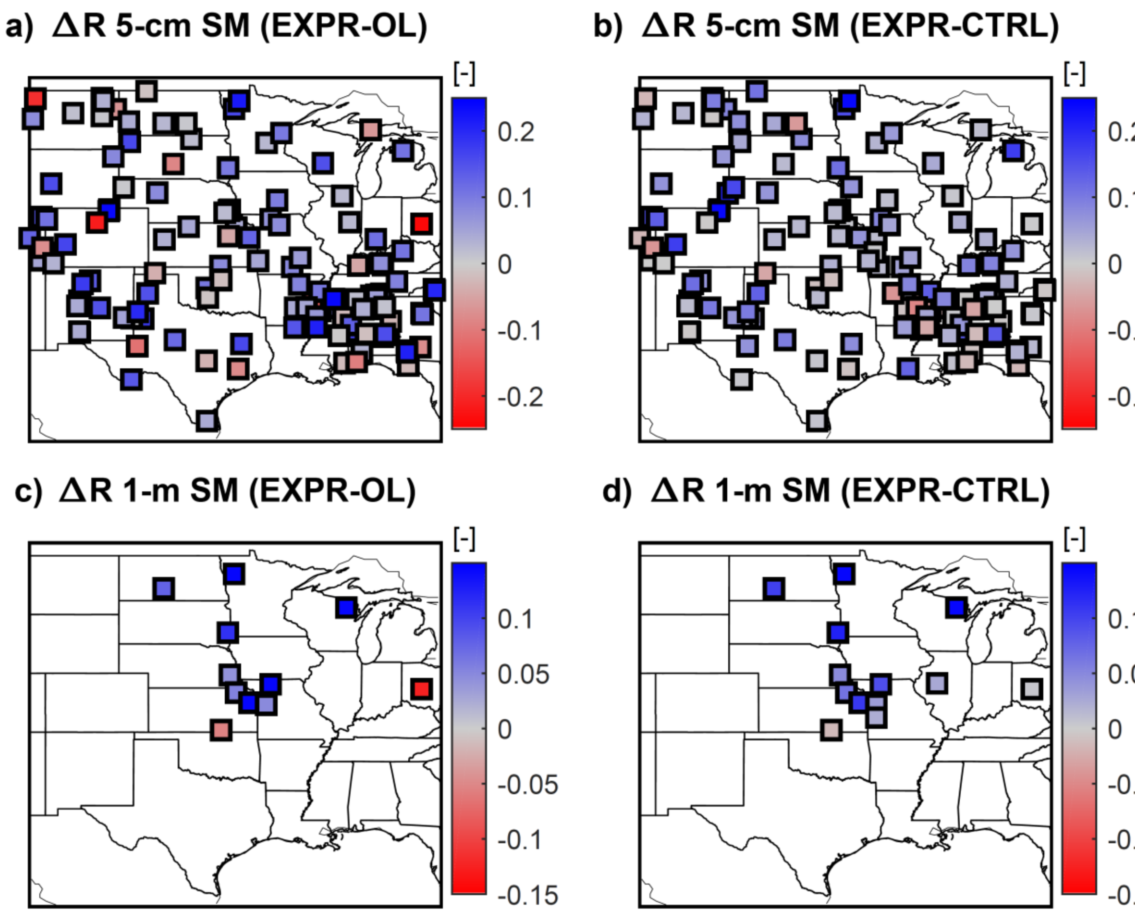

Figure 2. For USDA SCAN and NOAA USCRN ground sites, EXPR surface- (top row) and root-zone (bottom row) SM analysis correlation differences (Δ*R*) versus both the OL (left column) and CTRL (right column) baselines (00 UTC analyses). As discussed in the main text, EXPR-CTRL comparisons are for the 2012 and 2013 growing seasons while EXPR-OL comparisons are for the 2013 growing season only. The reduction of site density for the root-zone analysis reflects the limited availability of adequate profile soil moisture observations to obtain accurate surface to 1-m estimates.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

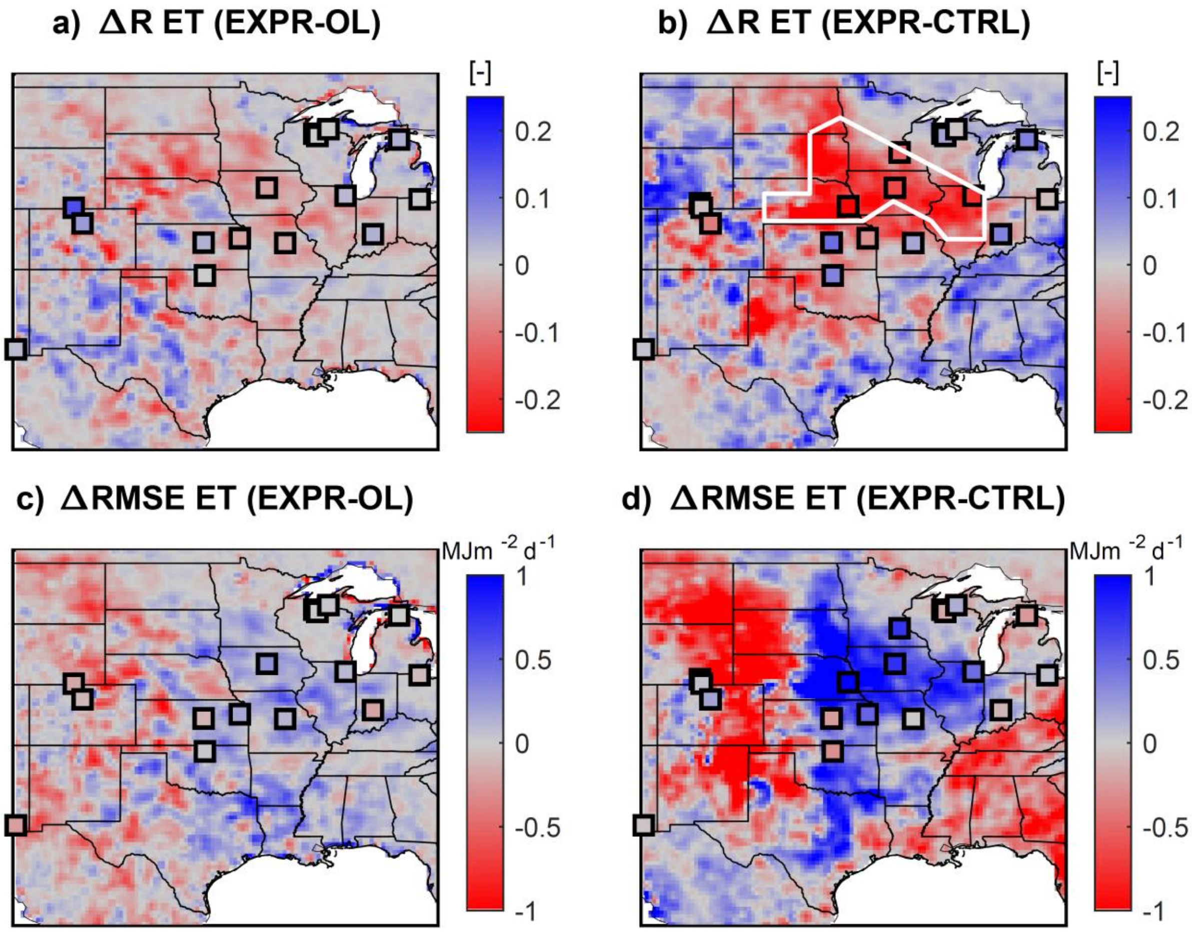

Figure 3. Change in EXPR ET 24-h forecast accuracy versus both the OL (left column) and CTRL (right column) baseline cases for temporal *R* (top row) and RMSE (bottom row) evaluation metrics. Background and symbol fill color shading reflect metric differences sampled against ALEXI ET retrievals and flux-tower ET observations, respectively. Plotted EXPR-CTRL differences (right column) are for the 2012 and 2013 growing seasons. EXPR-OL differences (left column) are for the 2013 growing season only. The white outline in part b) approximates the United States "Corn Belt" region (Schnitkey, 2014). All maps have been smoothed via a 2 x 2 moving-average filter applied to the original 0.25°-resolution image.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

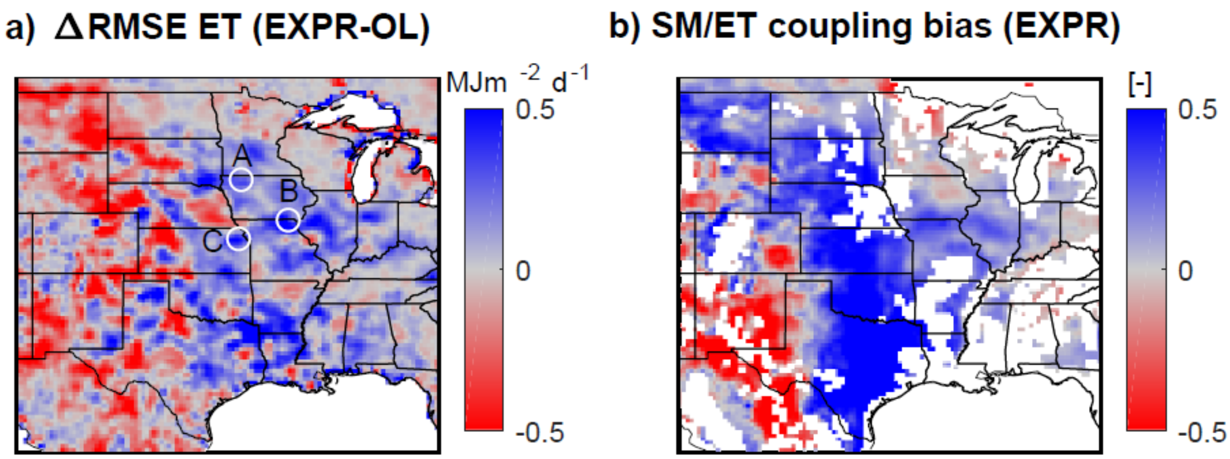

Figure 4. a) Replotting of the background in Figure 3c (i.e., the change in RMSE versus the ALEXI ET baseline between the EXPR and OL cases) with labeled locations (A, B and C) of sites examined later in Figure 6. b) HTESSEL SM/ET coupling bias (expressed as the Spearman rank coefficient of determination between weekly variables) versus the SM/ET coupling baseline provided in Lei et al. (2018). White areas in b) reflect regions where the approach in Lei et al. (2018) could not be reliably applied due to the low accuracy (or inadequate availability) of remotely sensed SM retrievals. Both maps have been smoothed via a 2 x 2 moving-average filter applied to the original 0.25°-resolution image.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

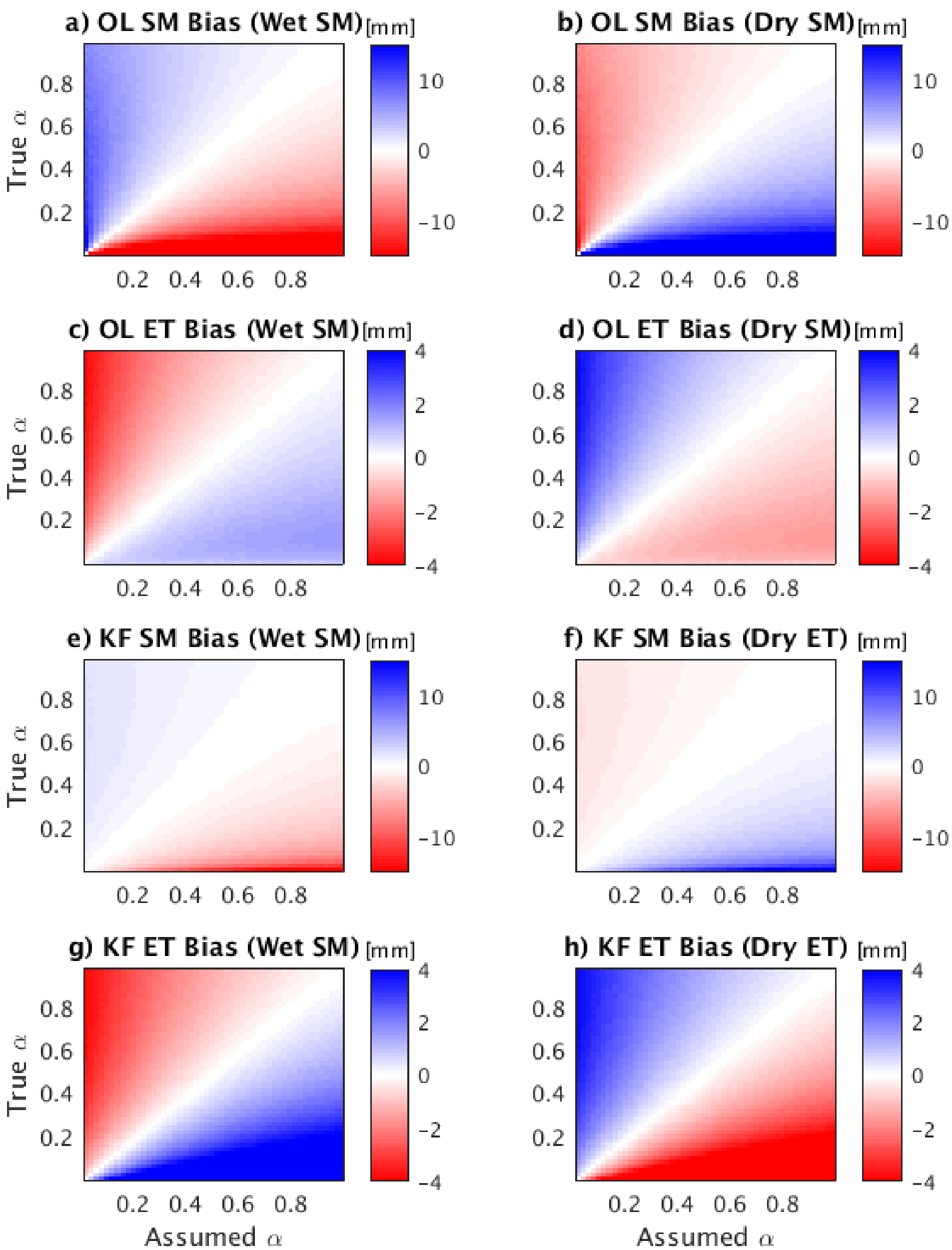

Figure 5. Daily OL SM (top row), OL ET (second row), KF SM (third row), and KF ET (bottom row) biases conditioned on true SM into "wet" (left column) and "dry" (right column) classifications. For each case, results are systematically generated for a range of true and assumed cases of SM/ET coupling strength (i.e., α). Open loop (OL) and Kalman Filter (KF) results correspond to before and after SM assimilation, respectively**.** Note that, in contrast to real-data results, ET is expressed in depth [mm] units.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.

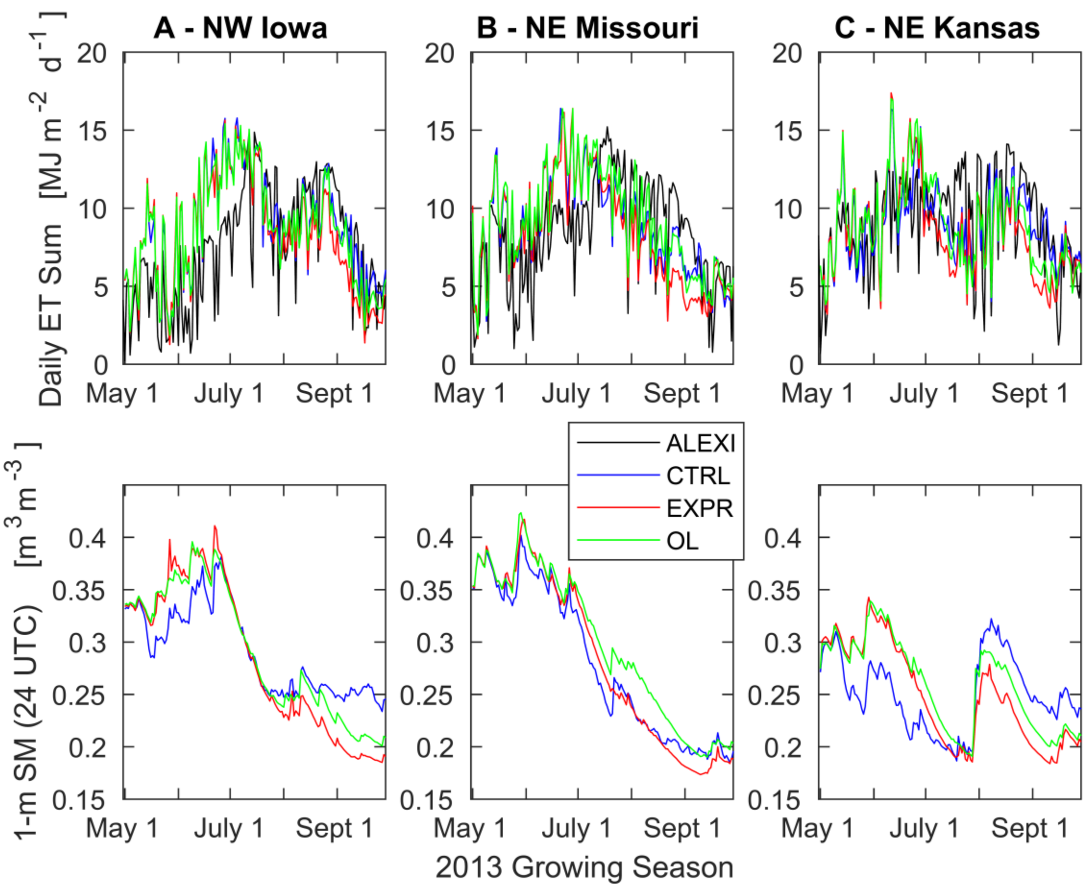

Figure 6. 2013 growing season time series of 24-h ET forecasts and 1-m SM analyses (00 UTC) for the CTRL, EXPR and OL DA cases (plus ALEXI ET retrievals) at sites (from left to right) in NW Iowa (A), NE Missouri (B) and NE Kansas (C) - see Figure 4a for exact site locations. Note that ALEXI does not provide SM estimates.

Accepted for publication in *Journal of Hydrometeorology*. DOI 10.1175/JHM-D-20-0088.1.