

Irrigation characterization improved by the direct use of SMAP soil moisture anomalies within a data assimilation system

Yonghwan Kwon^{1,2}, Sujay V. Kumar¹, Mahdi Navari^{1,2}, David M. Mocko^{1,3}, Eric M. Kemp^{1,4}, Jerry W. Wegiel^{1,3}, James V. Geiger¹ and Rajat Bindlish¹

¹Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA.

²Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA.

³Science Applications International Corporation, Reston, VA, USA.

⁴Science Systems and Applications, Inc., Lanham, MD, USA.

Email:

yonghwan.kwon@nasa.gov; yhkwon@umd.edu

Received xxxxxx

Accepted for publication xxxxxx

Published xxxxxx

Abstract

Prior soil moisture data assimilation (DA) efforts to incorporate human management features such as agricultural irrigation has only shown limited success. This is partly due to the fact that observational rescaling approaches for bias correction used in soil moisture DA systems are less effective when unmodeled processes such as irrigation are the dominant source of systematic biases. In this article, we demonstrate an alternative approach, i.e. anomaly correction for overcoming this limitation. Unlike the rescaling approaches, the proposed method does not scale remote sensing soil moisture retrievals to the model climatology, but it extracts the temporal variability information from the retrievals. The study demonstrates this approach through the assimilation of soil moisture retrievals from the Soil Moisture Active Passive (SMAP) mission into the Noah land surface model. The results demonstrate that DA using the anomaly correction method can better capture the effect of irrigation on soil moisture in agricultural areas while providing comparable performance to the DA integrations using rescaling approaches in non-irrigated areas. These findings emphasize the need to reduce inconsistencies between remote sensing and the models so that assimilation methods can employ information from remote sensing more directly to develop representations of unmodeled processes such as irrigation.

Keywords: soil moisture, irrigation, data assimilation, anomaly correction, cumulative distribution function (CDF) matching, Land Information System (LIS), SMAP

1. Introduction

Soil moisture is an important environmental variable in that it modulates fluxes of water, energy, and carbon at the land-atmosphere interface (e.g. Graf *et al* 2014, Green *et al* 2019, Hirschi *et al* 2011, Trugman *et al* 2018). Therefore, accurate

knowledge of the spatial and temporal distribution of soil moisture is critical for many environmental applications such as weather and climate forecasting (e.g. Koster *et al* 2010, Wu and Dickinson 2004), agricultural water resources management (e.g. Jalilvand *et al* 2019, Shin and Jung 2014), and drought and flood prediction (e.g. Hong and Kalnay 2000, Wanders *et al* 2014, Yuan *et al* 2011). Incorporating remote sensing retrievals of soil moisture into land surface models through data assimilation (DA) is often considered a viable approach to developing observation-informed, spatially-distributed estimates of soil moisture (Heathman *et al* 2003, Reichle *et al* 2002a). Though there have been numerous efforts on assimilating soil moisture retrievals from different satellite measurement platforms (e.g. Blyverket *et al* 2019, De Lannoy *et al* 2019, Kumar *et al* 2009, 2014, Reichle *et al* 2007, Seo *et al* 2021), most of these studies report relatively minor improvements from DA, particularly when high quality precipitation information is used to drive the model simulations. This is partly because the focus of most soil moisture DA studies is on improving short-term anomalies, as opposed to improving biases, which are often the dominant source of soil moisture errors (Reichle *et al* 2004). Further, these soil moisture DA studies and the evaluation metrics used in them are mostly focused on comparing to in-situ measurements, essentially ignoring the utility of remote sensing data to represent the significant spatial heterogeneity of soil moisture processes. Recent studies (e.g. Nearing *et al* 2018) have also shown that the coarse-scale remote sensing measurements may not have sufficient information about point-scale processes, making such comparisons to in-situ data inconclusive. On the other hand, a significant potential utility of satellite soil moisture observations should be in capturing the impact of processes that are influenced by factors outside precipitation or natural variability, which are ubiquitous on the land surface. For example, irrigation is one of the important and pervasive human activities across the world and has a direct effect on soil moisture in agricultural areas (Douglas *et al* 2009, Ozdogan *et al* 2010, Zhang *et al* 2017), yet is difficult to capture in models. The vast majority of the current soil moisture DA studies have not examined the usefulness of DA in the context of capturing such unmodeled processes (Kumar *et al* 2015).

Because agricultural irrigation is often driven by subjective practices, irrigation formulation in land surface models (LSMs) are poorly represented (Abolafia-Rosenzweig *et al* 2019, Gibson *et al* 2017, Lawston *et al* 2017). Recent studies (e.g. Brocca *et al* 2018, Lawston *et al* 2017) have demonstrated that microwave-based satellite remote sensing, particularly the Soil Moisture Active Passive (SMAP) satellite, can capture the effect of irrigation on soil moisture. Though assimilation of satellite soil moisture retrievals should offer the possibility of incorporating the irrigation signals into

model estimates, technical limitations of DA methods have been shown to be restrictive in this context (Kumar *et al* 2015).

Specifically, most soil moisture DA systems are forced to employ observational rescaling strategies due to large differences in soil moisture representation in models and remote sensing retrievals (Koster *et al* 2009). Rescaling approaches such as Cumulative Distribution Function (CDF) matching (Reichle and Koster 2004) are used to convert the remote sensing retrievals into the LSM's climatology prior to DA. While this approach is effective in eliminating systematic differences in all statistical moments between the modeled and remote sensing retrievals, they have been shown to cause statistical errors when the observations contain signals that are not modeled (Kumar *et al* 2015). They have also been shown to be significant sources of information loss in DA (Nearing *et al* 2018). Further, the CDF-matching is ineffective when unmodeled processes such as irrigation are the dominant source of systematic errors between LSMs and observations, as demonstrated in Kumar *et al* (2015). That is, CDF-matching, by design, results in most irrigation impacts being excluded prior to data assimilation. Similar limitations also apply to other rescaling methods such as variance matching, least squares regression-based rescaling, and triple collocation analysis-based rescaling, that have been proposed in the literature (Yilmaz and Crow 2013). Reducing the reliance on rescaling methods is essential to improving the capability of DA systems to represent such processes.

Observation rescaling methods can be avoided if the representative and interpretation differences between modeled and remote sensing soil moisture are reduced. This is not easy to do given that significant revisions in model parameters and conceptual formulations in land surface and retrieval models are needed. Alternatively, here we explore the utility of simpler a priori bias correction approach if the assumptions about the model and remote sensing systematic errors are relaxed within reason. Compared to the CDF-matching method, the newly applied bias correction method does not scale soil moisture retrievals to the model climatology, but instead extracts the soil moisture anomaly information (i.e. temporal variability) from the retrievals and uses it directly. Hereafter, we refer to the newly applied bias correction method as “anomaly correction” throughout this paper. Compared to CDF-matching, this approach better preserves the short-term anomaly information from the remote sensing retrievals. Our results demonstrate that employing the anomaly correction method can help the soil moisture DA framework to better preserve unmodeled processes such as irrigation from satellite-based soil moisture retrievals in agricultural areas while achieving comparable or better performance in estimating soil moisture than using the CDF-matching method. This study serves as a benchmark for future DA approaches to improve the information exploitation by directly incorporating remote sensing soil moisture retrievals.

2. Methods

2.1 Soil moisture data assimilation experimental setup

In this study, we conduct data assimilation experiments using the NASA Land Information System (LIS; Kumar *et al* 2006, 2008, Peters-Lidard *et al* 2007), which is a multiscale land surface modeling and data assimilation system for hydrologic applications. Among various state-of-the-art land surface models available in LIS, we employ the Noah LSM version 3.9 (Ek *et al* 2003) as a forward forecast model. The LSM integrations are conducted over the continental United States (CONUS) domain, which is composed of about 10-km resolution grid cells. We assimilate the SMAP Level 2 Radiometer Half-Orbit 36 km Equal-Area Scalable Earth (EASE)-Grid Soil Moisture product (SPL2SMP; O'Neill *et al* 2020a) into the model soil moisture simulation using an ensemble Kalman filter (EnKF) scheme (Evensen 1994, Reichle *et al* 2002b). To remove climatological biases in soil moisture between the Noah simulation and SMAP retrievals within DA, we use the CDF-matching and anomaly correction (see section 2.2 for more details). Outputs from the SMAP soil moisture DA experiments are evaluated using in situ soil moisture measurements from the International Soil Moisture Network (ISMN; Dorigo *et al* 2011, 2013). Experimental setup details including the atmospheric forcing, land parameters, model initialization, perturbation settings, and quality-control (QC) procedure are provided in Text S1.

2.2 Bias correction methods

The CDF-matching follows the approach of Reichle and Koster (2004), in which the satellite soil moisture retrievals are scaled to the model-simulated soil moisture climatology so that the CDFs of the scaled retrievals and the model estimates match. This CDF-matching approach corrects all the statistical moments of the distribution of observations regardless of its shape (Reichle and Koster 2004).

The newly proposed bias correction method (i.e. anomaly correction) follows two simple steps: (1) the climatology of the soil moisture retrievals is removed from the original SMAP data to obtain the SMAP soil moisture anomaly, and (2) the SMAP anomaly is then added to modeled soil moisture climatology, which is then assimilated into the LSM. The anomaly correction method assumes that the climatological soil moisture difference between models and observations is the main source of the systematic bias while the differences in higher moments such as standard deviation are small. Through the anomaly correction method, we expect that the climatological bias is removed, but the temporal variability of the original SMAP soil moisture retrievals is maintained so that the unmodeled irrigation signals can be captured by the model via assimilation.

Both the bias correction methods are performed for each model grid cell in the domain. The CDFs and climatologies of the model and observation are pre-generated for each grid cell using all available data from 2008 to 2020 (for the Noah LSM) and from 2015 to 2020 (for the SMAP retrievals), respectively. Based on the recommendations from prior studies (e.g. Kumar *et al* 2015) that show improved DA performance and reduced spurious statistical artifacts, CDFs are computed at a monthly scale. In the anomaly-based bias correction, the anomalies are computed at each grid cell by subtracting the long-term mean across the entire data record from each day's data.

3. Results and discussion

3.1 Rationality of the anomaly correction assumption

The key assumption in CDF-matching is that systematic differences in all statistical moments, such as mean, standard deviation, skewness, and kurtosis, of the model and observation distributions must be corrected. In contrast, the anomaly correction method assumes that correcting only the first moment of the systematic differences in soil moisture between models and observations is sufficient, and that maintaining the temporal variability of observations can help models to better capture the irrigation signals via assimilation.

Figure 1 shows the spatial distribution of the mean and standard deviation of surface soil moisture from the model (OL; without assimilation) and SMAP retrievals across 2015 to 2020. Both the model and SMAP exhibit a similar spatial pattern (i.e. relatively dry and wet soil moisture conditions in the western and central/eastern CONUS, respectively) of the temporal mean surface soil moisture (Figures 1(a) and 1(b)). However, compared to the SMAP retrievals, the Noah LSM-simulated soil moisture shows wetter and drier conditions in the western and central/eastern CONUS, respectively (Figure 1(c)). The standard deviation of surface soil moisture is relatively small in the dry western part of the study domain while it is larger in the wet central/eastern CONUS (Figures 1(d) and 1(e)). The spatial pattern of the differences in the standard deviation of soil moisture (Figure 1(f)) is similar to that of the mean soil moisture difference (Figure 1(c)), but its magnitude is less than $\pm 0.04 \text{ m}^3 \text{ m}^{-3}$ for 99% of grid cells in the CONUS domain (Figure 1(g)). These smaller differences in the standard deviation between the model and observation support the key assumption of the anomaly correction method that the biases in the mean is the main source of the systematic error. Note that similar checks must be conducted for every combination of remote sensing retrieval and land surface model before DA.

In the following sections, we demonstrate the feasibility of the anomaly correction method for improving the model estimates of soil moisture over the CDF-matching method via assimilation of the SMAP soil moisture retrievals while

preserving useful information about human-induced processes such as irrigation captured by the SMAP data, but not represented in the model.

3.2 Evaluations using in-situ soil moisture measurements

Figure 2 presents an evaluation of soil moisture DA performance by comparing against in-situ ISMN soil moisture measurements with the two a priori bias correction methods, that is, CDF-matching (DA_{CDF}) and anomaly correction (DA_{AC}). Note that the in-situ comparison is used primarily to evaluate the impact of the relaxed assumptions of

climatological differences with the anomaly correction. Figure 2 shows the unbiased root mean square errors (ubRMSE) and anomaly correlation coefficients (anomaly R) with the normalized differences in the ubRMSE and anomaly R between the DA and OL cases computed as $(DA - OL)/OL \times 100$. Negative values in the ubRMSE normalized difference and positive values in the anomaly R normalized difference indicate improvements because of DA. The percentage of model grid cells where the ubRMSE and anomaly R are improved (i.e. reduced ubRMSE and increased anomaly R) or degraded via assimilation of the SMAP soil moisture retrievals is also presented.

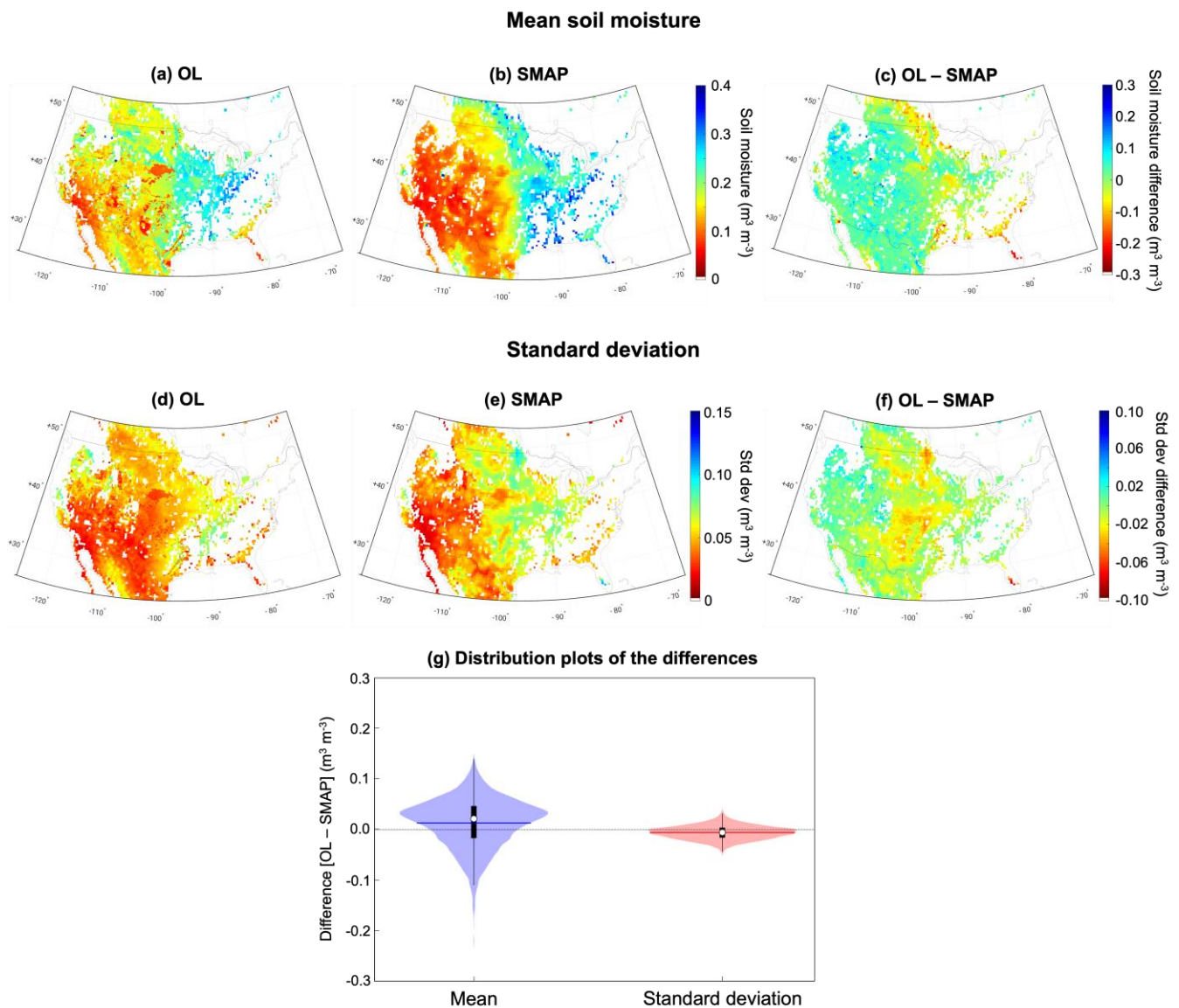


Figure 1. Temporal mean (uppermost panel) and standard deviation (middle panel) of surface soil moisture from the OL (a and d) and uncorrected (original) SMAP data (b and e). Differences in the (c) mean and (f) standard deviation between the OL and SMAP (i.e., OL – SMAP) are also presented. The lowermost panel (g) shows distribution plots of the differences where white dots and solid horizontal lines indicate the median and mean, respectively, and black boxes and vertical lines represent the interquartile range and 95% confidence interval, respectively.

Overall, compared to the OL run, both DA_{CDF} and DA_{AC} lead to small, but improved estimates of surface soil moisture in terms of the domain-averaged ubRMSE and anomaly R (Figure 2). DA_{AC} outperforms DA_{CDF} in reducing the domain-averaged ubRMSE (Figure 2(a)) while DA_{CDF} results in a higher domain-averaged anomaly R than DA_{AC} (Figure 2(b)). Figure 2(c) shows that a 0.35% and a 2.54% reduction in ubRMSE are achieved by DA_{CDF} and DA_{AC}, respectively, and a 4.04% and a 2.02% increase in anomaly R are achieved by DA_{CDF} and DA_{AC}, respectively. When the normalized improvements in ubRMSE and anomaly R are combined, DA_{AC} exhibits slightly better performance than DA_{CDF}. Regarding the number of improved model grid cells, the DA_{AC} case is superior to the DA case using the CDF-matching for both the ubRMSE and anomaly R (Figure 2(d)). That is, DA_{AC} reduces the ubRMSE and increases the anomaly R for 51.6% (27.7% with statistically significant changes) and 60.3% (25.0% with statistically significant changes) of the evaluated grid cells, respectively, while DA_{CDF} improves 46.1% (14.1% with statistically significant changes) and 58.3% (16.2% with statistically significant changes) of the evaluated model grid cells with respect to the ubRMSE and anomaly R, respectively. Note that additional DA experiments using the lumped CDFs calculated from all data points exhibit degraded performance, especially for anomaly R, compared to other DA

cases (Figure S2) due to spurious artifacts in the observations after scaling. This supports the use of the monthly-based CDFs in subsequent analyses.

These results are consistent with previous soil moisture DA studies that report small improvements (e.g. Blyverket *et al* 2019, De Lannoy *et al* 2019, Kumar *et al* 2014) in in-situ comparisons when high quality precipitation is used to drive the models. The domain-averaged improvements in the soil moisture estimates achieved from DA using both the a priori bias correction methods are marginal and are not statistically significant when compared against the ISMN soil moisture measurements. This is presumably attributed to the good performance of the OL integration driven by the high-quality meteorological forcing. That is, 55.2% and 43.5% of the evaluated model grid cells fall within the lower range of ubRMSE (less than 0.08 m³ m⁻³) and higher range of anomaly R (greater than 0.6), respectively. In the context of this paper, these results confirm that use of anomaly correction provides comparable DA performance to the CDF-matching approaches, in the evaluations against in-situ measurements, most of which are located in non-irrigated areas where soil moisture changes are dominated by natural variability (i.e. precipitation-driven alone). The utility of the anomaly correction method in capturing anthropogenic signals such as irrigation is demonstrated in the following sections.

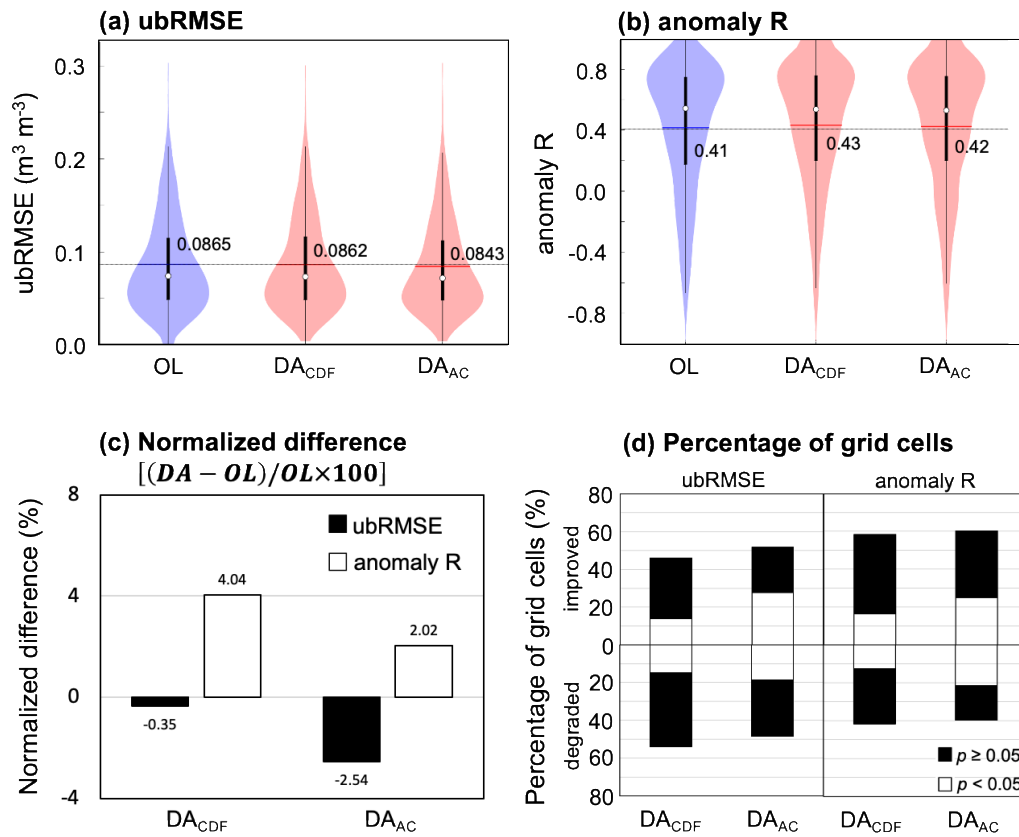


Figure 2. Soil moisture evaluation statistics: (a) ubRMSE; (b) anomaly R; (c) normalized differences in the ubRMSE and anomaly R [computed as $(DA - OL)/OL \times 100$] where a negative ubRMSE difference and a positive anomaly R difference indicate improvement in soil moisture estimates by DA; and (d) percentage of improved (i.e. reduced ubRMSE and increased anomaly R) and degraded grid cells by DA with ($p < 0.05$) and without ($p \geq 0.05$) statistically significant changes (p -values are obtained from a t -test). DA_{CDF} and DA_{AC} represent data assimilation cases using the CDF-matching and anomaly correction methods, respectively. In the distribution plots of (a) ubRMSE and (b) anomaly R, white dots and solid horizontal lines indicate the median and mean (the exact mean values are indicated in the figure), respectively, and black boxes and vertical lines represent the interquartile range and 95% confidence interval, respectively.

3.3 Capturing irrigation signals

In this section, we discuss the ability of the soil moisture DA framework to capture the effects of agricultural irrigation on soil moisture variability. Figure 3 shows the time-series of precipitation rate and surface soil moisture from the uncorrected (original) and corrected (after the CDF-matching or anomaly correction) SMAP, and the OL and DA cases for a single year (2016) at five example agricultural locations: California (marked with a yellow circle), Idaho (Snake River basin, marked with an orange triangle), Nebraska (marked with a red square), Texas (marked with green circle), and Georgia (marked with green triangle). These locations are selected from the Moderate Resolution Imaging

Spectroradiometer (MODIS)-based map (Salmon *et al* 2015) of the irrigated grid cell fraction (Figure 3(a)).

The example locations are in general characterized by the wet winter season and dry summer season based on the precipitation rate, and thus are areas where irrigation is the main water source for crops during the warm, dry summer season. The SMAP data show that wet soil moisture conditions are maintained during the dry summer season (Figure 3), which is attributed to irrigation practices. The irrigation signals, i.e. deviation of the soil moisture pattern from the precipitation pattern, are well captured by the SMAP retrievals especially in the California region (Figures 3(b) and 3(g)) where flood irrigation practices dominate, as discussed in Lawston *et al* (2017).

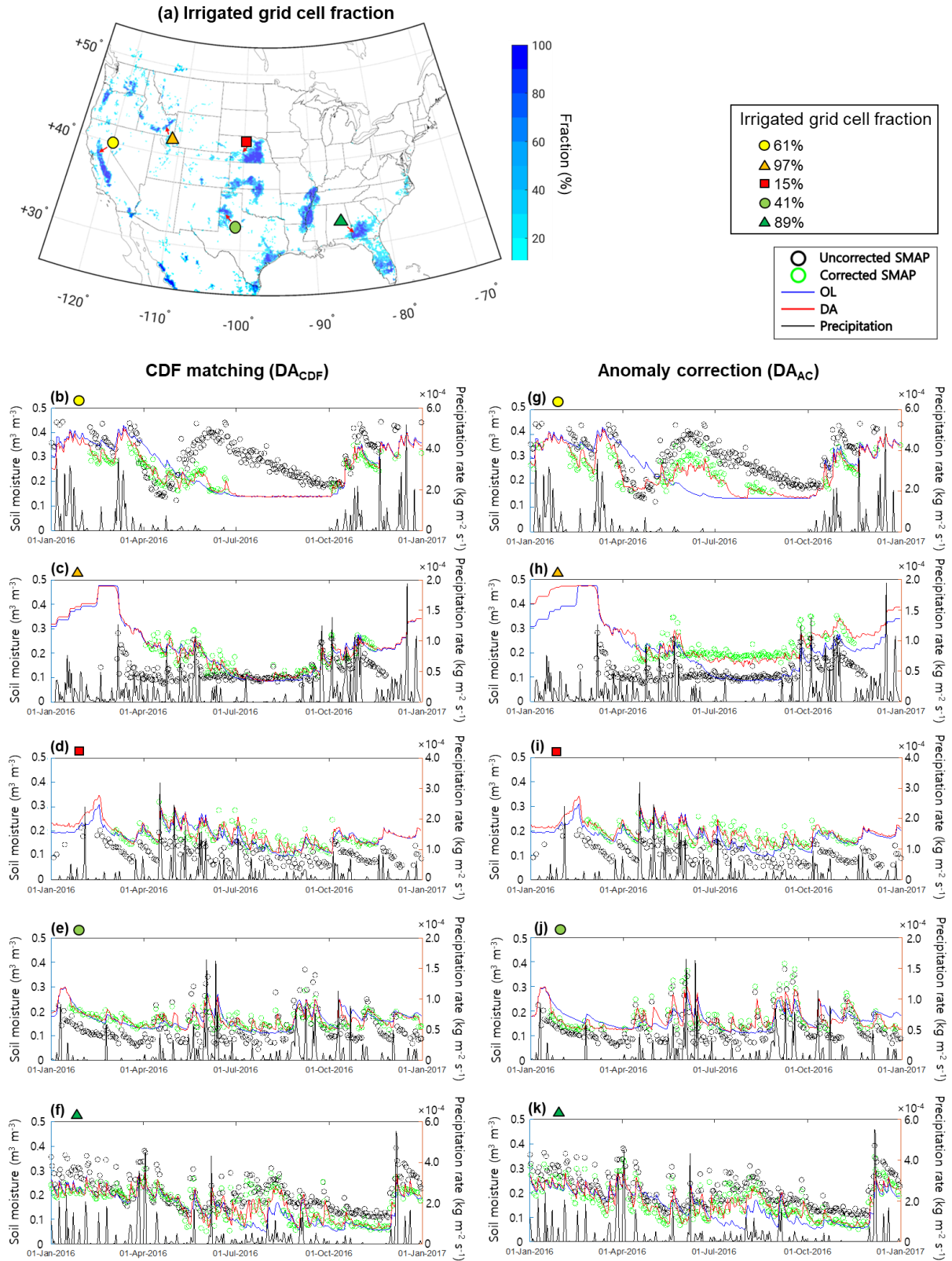


Figure 3. Time-series of precipitation and soil moisture from the OL, DA, and uncorrected (original) and corrected SMAP retrievals for example locations indicated in the (a) map of the irrigated grid cell fraction derived from MODIS data (Salmon *et al* 2015). The left and right columns present the DA results using the CDF-matching and anomaly correction, respectively, for the same locations and time period.

The bias-corrected SMAP soil moisture in Figure 3(b) shows that the irrigation signals, which are present in the uncorrected SMAP data, are not incorporated well after the CDF-matching. Therefore, assimilating the CDF-matched SMAP soil moisture data does not produce significantly higher soil moisture values during the summer months, which is presumably the time period when irrigation practices are active. On the other hand, the bias-corrected SMAP data after the anomaly correction is able to capture the effect of irrigation on soil moisture (Figure 3(g)). Consequently, merging the SMAP soil moisture anomaly information into the model via assimilation helps the model to represent the wet soil moisture features from the observations during the rain-free, dry summer season as obviously seen in the flood irrigation-dominated region (Figure 3(g)). Although only surface soil moisture is directly updated during the assimilation procedure, the impact is propagated to deeper soil layers by the model subsurface physics (Figure 4).

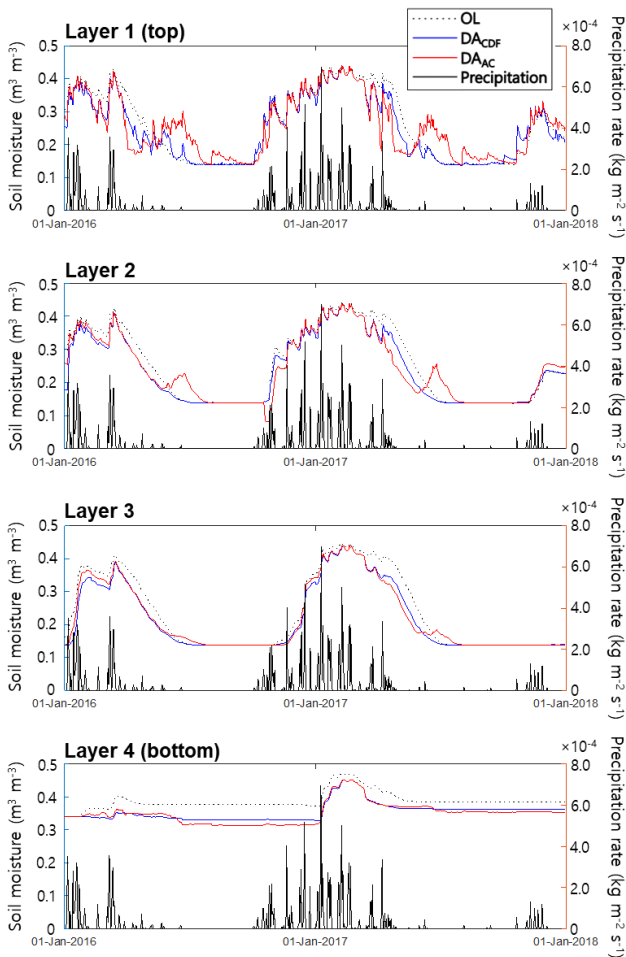


Figure 4. Time-series of precipitation and soil layer moisture content from the OL, DA_{CDF}, and DA_{AC} in the irrigated area in California.

Compared to the California region (Figure 3(g)), the obvious nature of the irrigation signal in the SMAP data is less evident at other locations, i.e. Idaho (Figures 3(c) and 3(h)), Nebraska (Figures 3(d) and 3(i)), Texas (Figures 3(e) and 3(j)), and Georgia (Figures 3(f) and 3(k)) regions, due to the simultaneous influence of irrigation and rainfall. Nevertheless, assimilating the SMAP data using the anomaly correction method (i.e. DA_{AC}) results in wetter soil moisture conditions than both the OL and DA_{CDF} during the dry summer season. Over the Snake River basin location, SMAP (uncorrected) observations show soil moisture levels maintained during April to September time period, presumably due to irrigation. While DA_{CDF} do not capture these features, DA_{AC} captures the persistence of the soil moisture at the same level seen in the observations.

3.4 Land surface temperature and latent heat flux estimates

Soil moisture conditions affect land surface temperature (LST) through various physical processes. Soil thermal diffusivity in general increases with wetter soil moisture, although their relationship is non-linear because both soil thermal conductivity and heat capacity increase with soil moisture (Al Nakshabandi and Kohnke 1965, Kwon and Koo 2017). Therefore, wetter soil moisture conditions may lead to more energy transfer from surface to subsurface soil layers during daytime. Increased soil moisture also increases the energy loss (i.e. latent heat flux to the atmosphere) from the land surface by enhancing evapotranspiration. Many studies (e.g. Chen and Dirmeyer 2019, Huang and Ullrich 2016, Thierry *et al* 2017) demonstrate that irrigation leads to strong evaporative cooling of the land surface. On the other hand, increased soil moisture darkens the land surface (i.e. decreases the surface albedo) and resultantly increases net radiation at the land surface. All these physical processes associated with soil moisture, especially soil moisture-atmosphere feedbacks, shape the probability distribution of LST (Berg *et al* 2014).

Here, we compute the Kullback-Leibler divergence (KLD; Kullback and Leibler 1951) by comparing the model-estimated LST distribution against that of the MODIS LST data (MOD11A1; Wan *et al* 2015) during the period of July to September (Figure 5). The KLD measures the dissimilarity between two different probability distributions, that is, a smaller KLD indicates that the probability density of the estimated LST is closer to that of the MODIS LST. This comparison is used as an independent evaluation of the ability of different soil moisture DA strategies for capturing irrigation impacts.

Figure 5 shows the difference in KLD between the DA_{AC} and DA_{CDF} cases. In the KLD difference map, negative values represent that the LST estimates are improved by DA_{AC} compared to the DA_{CDF} case. Overall, DA_{AC} is superior to DA_{CDF} with respect to the KLD in the irrigated areas (see

Figure 3(a)), especially in the central/eastern CONUS including Nebraska, Kansas, Oklahoma, Texas, lower Mississippi, and Georgia. Compared to other irrigated areas, a relatively small number of locations in California and Idaho exhibits clear differences in KLD between the DA cases, which is attributed to small changes in the LST estimates by DA with a limited number of SMAP data resulting from the QC procedures. These results provide additional evidence that the newly applied soil moisture bias correction method (i.e. anomaly correction) can be an effective alternative to the traditional method (i.e. CDF-matching) within the DA

framework to capture the irrigation signals in the soil moisture estimates via assimilation of the SMAP retrievals.

Improvements in evapotranspiration (ET) estimates over the irrigated areas are also observed (Figure 6), when the results are compared against the Atmosphere-Land Exchange Inverse (ALEXI; Anderson *et al* 2007) ET product. Except for the lower Mississippi region, the comparison to ALEXI indicates beneficial impacts from DA over irrigated areas in the snake river basin, parts of Nebraska and central valley of California. Further evaluations using ground-based measurements need to be conducted in future studies.

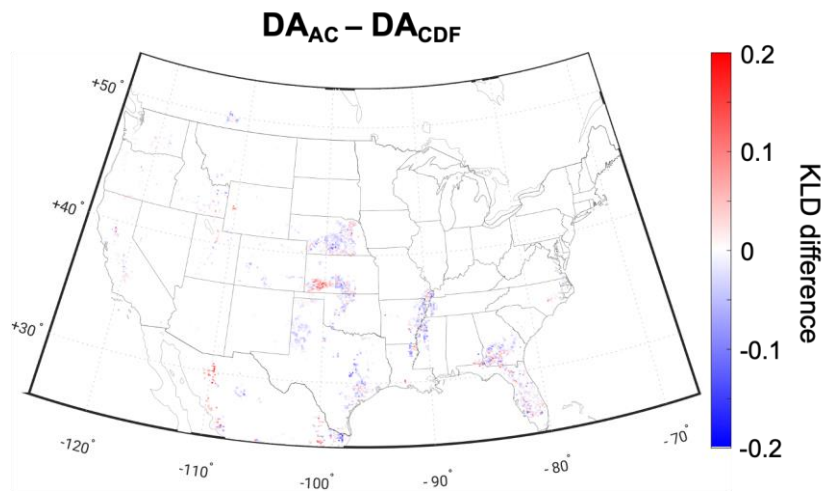


Figure 5. Difference in Kullback-Leibler divergence (KLD) of the model-estimated land surface temperature (LST) distribution from that of the MODIS LST between the DA_{AC} and DA_{CDF} cases. Note that the KLD difference is plotted over the irrigated areas (Figure 3(a)), and the values close to zero are excluded from the figure for clarity.

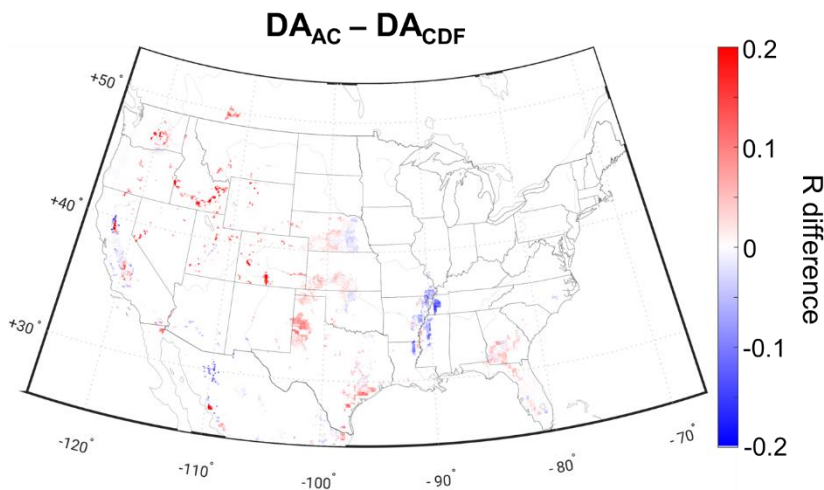


Figure 6. Latent heat flux correlation coefficient (R) difference between the DA_{AC} and DA_{CDF} cases. The R values are computed using the ALEXI latent heat flux product. Note that the latent heat flux R is plotted over the irrigated areas (Figure

3(a)). Positive values (red color) represent improvement in the latent heat flux estimates by using the anomaly correction method compared to the CDF-based DA.

4. Summary and conclusions

This study demonstrates a new bias correction method (i.e. anomaly correction) within the soil moisture data assimilation framework with the aim to improve model estimates of soil moisture in irrigated areas. The newly applied method is different from the commonly used CDF-matching in that it does not scale remotely-sensed soil moisture retrievals to the modeled soil moisture climatology. Instead, the anomaly correction approach directly uses the anomalies from the remote sensing retrievals. Thus, the anomaly correction method maintains the temporal variability of the soil moisture retrievals, which undergoes significant changes when the CDF-matching method is employed to correct systematic biases between models and observations. The difference in the soil moisture standard deviations between the model and observation used in this study (from SMAP) is much smaller than the difference in the mean. This supports the key assumption of the anomaly correction method, that is, the bias in the mean is the primary source of the systematic differences in soil moisture between models and observations, and thus correcting the bias only in the first moment is sufficient for performing soil moisture data assimilation.

The experimental results demonstrate that correcting biases using the anomaly correction method can ameliorate issues (i.e. loss of irrigation signals from observations by employing the CDF-matching-based bias correction method) reported from previous studies. The results show that DA employing the anomaly correction can better capture the effect of irrigation on soil moisture in agricultural areas. Further, the use of anomaly correction exhibits comparable or better performance in estimating soil moisture compared to the DA case using the CDF-matching method in the comparisons to in-situ measurements. These findings are encouraging and suggest the potential of the soil moisture DA framework to represent unmodeled processes in the soil moisture estimates at large spatial scales.

This paper demonstrates the potential added utility of directly assimilating information from remote sensing retrievals, using SMAP data. This more relaxed assumption in DA is possible because the systematic errors between SMAP and the LSM is primarily in the first moment of the statistical distribution. For retrievals from other sensors with larger errors in the variability of soil moisture (e.g. standard deviation), this approach is unlikely to be viable. It should be emphasized that further evaluations of our methodology using other soil moisture retrievals need to be conducted in future studies.

The results of this paper also highlight the need to improve the consistency across soil moisture retrievals and model estimates so that they can be used more directly in DA or end-use application environments. Currently, soil moisture remote sensing retrievals are produced in volumetric water content or as estimates of water saturation, and conversion between the two physical spaces require the knowledge of soil properties such as bulk density, which are often not reported along with retrieval products. Similarly, LSM-based soil moisture estimates are model-specific quantities, representing a measure of the wetness of the soil. To assimilate satellite soil moisture estimates directly, the inconsistencies between models and remote sensing retrievals must be addressed. Though the soil moisture community has largely sidestepped this issue by focusing on anomaly-based evaluations, studies have shown that they underestimate the errors (Dorigo *et al* 2010). The results of this study also show the promise of smaller information loss and larger utility from DA if more direct approaches to assimilating remote sensing retrievals can be developed.

Data availability statement

The NASA Land Information System (LIS) is available for public access at <https://github.com/NASA-LIS/LISF>. The Soil Moisture Active Passive (SMAP) soil moisture retrievals can be obtained from <https://doi.org/10.5067/F1TZ0CBN1F5N>. The International Soil Moisture Network (ISMN) in situ soil moisture measurements can be downloaded from https://www.geo.tuwien.ac.at/insitu/data_viewer/. The Moderate Resolution Imaging Spectroradiometer (MODIS) land surface temperature can be accessed at <https://e4ftl01.cr.usgs.gov/MOLT/MOD11A1.006/>. The data that support the findings of this study are available upon request from the authors.

Acknowledgments

This research was supported by the United States Air Force under grant no. F2BDAZ9263G101. Computational resources were provided by the NASA Center for Climate Simulation.

References

- Abolafia-Rosenzweig R, Livneh B, Small E E and Kumar S V 2019 Soil moisture data assimilation to estimate irrigation water use *J. Adv. Model. Earth Syst.* **11** 3670–3690 <https://doi.org/10.1029/2019MS001797>
- Al Nakshabandi G and Kohnke H 1965 Thermal conductivity and diffusivity of soils as related to moisture tension and other physical properties *Agr. Meteorol.* **2** 271–279 [https://doi.org/10.1016/0002-1571\(65\)90013-0](https://doi.org/10.1016/0002-1571(65)90013-0)

- Anderson M, Norman J, Mecikalski J, Otkin J and Kustas W 2007 A climatological study of evapotranspiration and moisture stress across the continental U.S. based on thermal remote sensing: I. Model formulation *J. Geophys. Res.* **112** D11112 <https://doi.org/10.1029/2006JD007506>
- Berg A, Lintner B R, Findell K L, Malyshev S, Loikith P C and Gentine P 2014 Impact of soil moisture–atmosphere interactions on surface temperature distribution *J. Climate* **27** 7976–7993 <https://doi.org/10.1175/JCLI-D-13-00591.1>
- Blyverket J, Hamer P D, Bertino L, Albergel C, Fairbairn D and Lahoz W A 2019 An evaluation of the EnKF vs. EnOI and the assimilation of SMAP, SMOS and ESA CCI soil moisture data over the contiguous US *Remote Sens.* **11** 478 <https://doi.org/10.3390/rs11050478>
- Brocca L, Tarpanelli A, Filippucci P, Dorigo W, Zaussinger F, Gruber A and Fernández-Prieto D 2018 How much water is used for irrigation? A new approach exploiting coarse resolution satellite soil moisture products *Int. J. Appl. Earth Obs.* **73** 752–766 <https://doi.org/10.1016/j.jag.2018.08.023>
- Chen L and Dirmeyer P A 2019 Global observed and modelled impacts of irrigation on surface temperature *Int. J. Climatol.* **39** 2587–2600 <https://doi.org/10.1002/joc.5973>
- De Lannoy G J M, de Rosnay P and Reichle R H 2019 Soil moisture data assimilation. In Q Duan, F Pappenberger, A Wood, H Cloke and J Schaake (Eds.) *Handbook of hydrometeorological ensemble forecasting* (Berlin/Heidelberg, Germany: Springer) https://doi.org/10.1007/978-3-642-39925-1_32
- Dorigo W A, Scipal K, Parinussa R M, Liu Y Y, Wagner W, de Jeu R A M and Naeimi V 2010 Error characterisation of global active and passive microwave soil moisture datasets *Hydrol. Earth Syst. Sc.* **14** 2605–2616 <https://doi.org/10.5194/hess-14-2605-2010>
- Dorigo W A, Wagner W, Hohensinn R, Hahn S, Paulik C, Xaver A, Gruber A, Drusch M, Mecklenburg S, van Oevelen P, Robock A and Jackson T. 2011 The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements *Hydrol. Earth Syst. Sc.* **15** 1675–1698 <https://doi.org/10.5194/hess-15-1675-2011>
- Dorigo W A, Xaver A, Vreugdenhil M, Gruber A, Hegyiová A, Sanchis-Dufau A D, Zamojski D, Cordes C, Wagner W and Drusch M 2013 Global automated quality control of in situ soil moisture data from the international soil moisture network *Vadose Zone J.* **12** vzj2012.0097 <https://doi.org/10.2136/vzj2012.0097>
- Douglas E M, Beltrán-Przekurat A, Niyogi D, Pielke Sr. R A and Vörösmarty C 2009 The impact of agricultural intensification and irrigation on land-atmosphere interactions and Indian monsoon precipitation – A mesoscale modeling perspective *Global Planet. Change* **67** 117–128 <https://doi.org/10.1016/j.gloplacha.2008.12.007>
- Ek M B, Mitchell K E, Lin Y, Rogers E, Grunmann P, Koren V, Gayno G and Tarpley J D 2003 Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model *J. Geophys. Res.* **108** 8851 <https://doi.org/10.1029/2002JD003296>
- 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>
- Gibson J, Wang T E, Gates J, Grassini P, Yang H and Eisenhauer D 2017 A case study of field-scale maize irrigation patterns in western Nebraska: implications for water managers and recommendations for hyper-resolution land surface modeling *Hydrol. Earth Syst. Sc.* **21** 1051–1062 <https://doi.org/10.5194/hess-21-1051-2017>
- Graf A, Bogen H R, Drüe C, Hardelauf H, Pütz T, Heinemann G and Vereecken H 2014 Spatiotemporal relations between water budget components and soil water content in a forested tributary catchment *Water Resour. Res.* **50** 4837–4857 <https://doi.org/10.1002/2013WR014516>
- Green J K, Seneviratne S I, Berg A M, Findell K L, Hagemann S, Lawrence D M and Gentine P 2019 Large influence of soil moisture on long-term terrestrial carbon uptake *Nature* **565** 476–479 <https://doi.org/10.1038/s41586-018-0848-x>
- Heathman G C, Starks P J, Ahuja L R and Jackson T J 2003 Assimilation of surface soil moisture to estimate profile soil water content *J. Hydrol.* **279** 1–17 [https://doi.org/10.1016/S0022-1694\(03\)00088-X](https://doi.org/10.1016/S0022-1694(03)00088-X)
- Hirschi M, Seneviratne S I, Alexandrov V, Boberg F, Boroneant C, Christensen O B, Formayer H, Orlowsky B and Stepanek P 2011 Observational evidence for soil-moisture impact on hot extremes in southeastern Europe *Nat. Geosci.* **4** 17–21 <https://doi.org/10.1038/ngeo1032>
- Hong S Y and Kalnay E 2000 Role of sea surface temperature and soil-moisture feedback in the 1998 Oklahoma-Texas drought *Nature* **408** 842–844 <https://doi.org/10.1038/35048548>
- Huang X and Ullrich P A 2016 Irrigation impacts on California’s climate with the variable-resolution CESM *J. Adv. Model. Earth Sy.* **8** 1151–1163 <https://doi.org/10.1002/2016MS000656>
- Jalilvand E, Tajrishy M, Hashemi S A G Z and Brocca L 2019 Quantification of irrigation water using remote sensing of soil moisture in a semi-arid region *Remote Sens. Environ.* **231** 111226 <https://doi.org/10.1016/j.rse.2019.111226>
- Koster R D, Guo Z, Yang R, Dirmeyer P A, Mitchell K and Puma M J 2009 On the Nature of Soil Moisture in Land Surface Models *J. Climate* **22** 4322–4335 <https://doi.org/10.1175/2009JCLI2832.1>
- Koster R D, Mahamanama S P P, Yamada T J, Balsamo G, Berg A A, Boissarie M, Dirmeyer P A, Doblas-Reyes F J, Drewitt G, Gordon C T, Guo Z, Jeong J H, Lawrence D M, Lee W S, Li Z, Luo L, Malyshev S, Merryfield W J, Seneviratne S I, Stanelle T, van den Hurk B J J M and Wood E F 2010 Contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment. *Geophys. Res. Lett.* **37** L02402 <http://dx.doi.org/10.1029/2009GL041677>
- Kullback S and Leibler R A 1951 On information and sufficiency *Ann. Math. Stat.* **22** 79–86 <https://doi.org/10.1214/aoms/1177729694>
- Kumar S V, Reichle R H, Koster R D, Crow W T and Peters-Lidard C D 2009 Role of subsurface physics in the assimilation of surface soil moisture observations *J. Hydrometeorol.* **10** 1534–1547 <https://doi.org/10.1175/2009JHM1134.1>
- Kumar S V, Peters-Lidard C D, Santanello J A, Reichle R H, Draper C S, Koster R D, Nearing G and Jasinski M F 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. Sc.* **19** 4463–4478 <https://doi.org/10.5194/hess-19-4463-2015>
- Kumar S V, Peters-Lidard C D, Mocko D, Reichle R, Liu Y, Arsenault K R, Xia Y, Ek M, Riggs G, Livneh B and Cosh M 2014 Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation *J. Hydrometeorol.* **15** 2446–2469 <https://doi.org/10.1175/JHM-D-13-0132.1>

- Kumar S V, Peters-Lidard C D, Tian Y, Houser P R, Geiger J, Olden S, Lighty L, Eastman J L, Doty B, Dirmeyer P, Adams J, Mitchell K, Wood E F and Sheffield J 2006 Land information system: An interoperable framework for high resolution land surface modeling *Environ. Modell. Softw.* **21** 1402–1415 <https://doi.org/10.1016/j.envsoft.2005.07.004>
- Kumar S, Peters-Lidard C, Tian Y, Reichle R H, Geiger J, Alonge C, Eylander J and Houser P 2008 An integrated hydrologic modeling and data assimilation framework *Computer* **41** 52–59 <https://doi.org/10.1109/MC.2008.475>
- Kwon Y and Koo B K 2017 Development and evaluation of the Soil and Water Temperature Model (SWTM) for rural catchments. *J. Hydrol.* **553** 457–470 <http://dx.doi.org/10.1016/j.jhydrol.2017.08.017>
- Lawston P M, Santanello Jr J A and Kumar S V 2017 Irrigation signals detected from SMAP soil moisture retrievals *Geophys. Res. Lett.* **44** 11,860–11,867 <https://doi.org/10.1002/2017GL075733>
- Lawston P M, Santanello Jr J A, Zaitchik B F and Rodell M 2015 Impact of Irrigation Methods on Land Surface Model Spinup and Initialization of WRF Forecasts *J. Hydrometeorol.* **16** 1135–1154 <https://doi.org/10.1175/JHM-D-14-0203.1>
- Nearing G, Yatheendradas S, Crow W, Zhan X, Liu J and Chen F 2018 The efficiency of data assimilation *Water Resour. Res.* **54** 6374–6392 <https://doi.org/10.1029/2017WR020991>
- O'Neill P E, Chan S, Njoku E G, Jackson T, Bindlish R and Chaubell J 2020a SMAP L2 Radiometer Half-Orbit 36 km EASE-Grid Soil Moisture (Boulder, Colorado USA: NASA National Snow and Ice Data Center Distributed Active Archive Center) <https://doi.org/10.5067/F1T20CBN1F5N>
- Ozdogan M, Rodell M, Beaudoin H K and Toll D L 2010 Simulating the effects of irrigation over the United States in a land surface model based on satellite-derived agricultural data. *J. Hydrometeorol.* **11** 171–184 <https://doi.org/10.1175/2009JHM1116.1>
- Peters-Lidard C D, Houser P R, Tian Y, Kumar S V, Geiger J, Olden S, Lighty L, Doty B, Dirmeyer P, Adams J, Mitchell K, Wood E F and Sheffield J 2007 High-performance Earth system modeling with NASA/GSFC's Land Information System *Innovations Syst. Softw. Eng.* **3** 157–165 <https://doi.org/10.1007/s11334-007-0028-x>
- Reichle R H and Koster R D 2004 Bias reduction in short records of satellite soil moisture *Geophys. Res. Lett.* **31** L19501 <https://doi.org/10.1029/2004GL020938>
- Reichle R, Koster R D, Dong J and Berg A A 2004 Global soil moisture from satellite observations, land surface models, and ground data: Implications for data assimilation *J. Hydrometeorol.* **5** 430–442 [https://doi.org/10.1175/1525-7541\(2004\)005<0430:GSMFSO>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0430:GSMFSO>2.0.CO;2)
- Reichle R H, Koster R D, Liu P, Mahanama S P P, Njoku E G and Owe M 2007 Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR) *J. Geophys. Res.* **112** D09108 <https://doi.org/10.1029/2006JD008033>
- Reichle R H, McLaughlin D B and Entekhabi D 2002a Hydrologic data assimilation with the ensemble Kalman filter *Mon. Weather Rev.* **130** 103–114 [https://doi.org/10.1175/1520-0493\(2002\)130<0103:HDAWTE>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2)
- Reichle R H, Walker J P, Koster R D and Houser P R 2002b Extended versus ensemble Kalman filtering for land data assimilation *J. Hydrometeorol.* **3** 728–740 [https://doi.org/10.1175/1525-7541\(2002\)003<0728:EVEKFF>2.0.CO;2](https://doi.org/10.1175/1525-7541(2002)003<0728:EVEKFF>2.0.CO;2)
- Salmon J M, Friedl M A, Frohling S, Wisser D and Douglas E M 2015 Global rain-fed, irrigated, and paddy croplands: A new high resolution map derived from remote sensing, crop inventories and climate data *Int. J. Appl. Earth Obs.* **38** 321–334 <https://doi.org/10.1016/j.jag.2015.01.014>
- Seo E, Lee M -I and Reichle R H 2021 Assimilation of SMAP and ASCAT soil moisture retrievals into the JULES land surface model using the local Ensemble Transform Kalman Filter *Remote Sens. Environ.* **253** 112222 <https://doi.org/10.1016/j.rse.2020.112222>
- Shin Y and Jung Y 2014 Development of irrigation water management model for reducing drought severity using remotely sensed soil moisture footprints *J. Irrig. Drain. Eng.* **140** 04014021 [https://doi.org/10.1061/\(ASCE\)IR.1943-4774.0000736](https://doi.org/10.1061/(ASCE)IR.1943-4774.0000736)
- Thiery W, Davin E L, Lawrence D M, Hirsch A L, Hauser M and Seneviratne S I 2017 Present-day irrigation mitigates heat extremes *J. Geophys. Res.-Atmos.* **122** 1403–1422 <https://doi.org/10.1002/2016JD025740>
- Trugman A T, Medvigy D, Mankin J S and Anderegg W R L 2018 Soil moisture stress as a major driver of carbon cycle uncertainty *Geophys. Res. Lett.* **45** 6495–6503 <https://doi.org/10.1029/2018GL078131>
- Wan Z, Hook S and Hulley G 2015 MOD11A1 MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V006. 2015, distributed by NASA EOSDIS Land Processes DAAC <https://doi.org/10.5067/MODIS/MOD11A1.006>. Accessed 2021-08-05.
- Wanders N, Karssen D, de Roo A, de Jong S M and Bierkens M F P 2014 The suitability of remotely sensed soil moisture for improving operational flood forecasting *Hydrol. Earth Syst. Sc.* **18** 2343–2357 <https://doi.org/10.5194/hess-18-2343-2014>
- Wu W and Dickinson R E 2004 Time scales of layered soil moisture memory in the context of land-atmosphere interaction *J. Climate* **17** 2752–2764 [https://doi.org/10.1175/1520-0442\(2004\)017<2752:TSOLSM>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<2752:TSOLSM>2.0.CO;2)
- Yilmaz M T and Crow W T 2013 The optimality of potential rescaling approaches in land data assimilation *J. Hydrometeorol.* **14** 650–660 <https://doi.org/10.1175/JHM-D-12-052.1>
- Yuan X, Wood E F, Luo L and Pan M 2011 A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction *Geophys. Res. Lett.* **38** L13402 <https://doi.org/10.1029/2011GL047792>
- Zhang G, Liu C, Xiao C, Xie R, Ming B, Hou P, Liu G, Xu W, Shen D, Wang K and Li S 2017 Optimizing water use efficiency and economic return of super high yield spring maize under drip irrigation and plastic mulching in arid areas of China *Field Crop. Res.* **211** 137–146 <https://doi.org/10.1016/j.fcr.2017.05.026>