Earth Syst. Sci. Data, 17, 4587–4611, 2025
https://doi.org/10.5194/essd-17-4587-2025

Open Access Earth System Science Data

A 1 km soil moisture dataset over eastern CONUS generated by assimilating SMAP data into the Noah-MP land surface model

Sheng-Lun Tai, Zhao Yang, Brian Gaudet, Koichi Sakaguchi, Larry Berg, Colleen Kaul, Yun Qian, Ye Liu, and Jerome Fast

Pacific Northwest National Laboratory, Richland, WA 99352, USA

Correspondence: Sheng-Lun Tai (sheng-lun.tai@pnnl.gov)

Received: 17 December 2024 – Discussion started: 28 January 2025
Revised: 25 June 2025 – Accepted: 8 July 2025 – Published: 19 September 2025

Abstract. An improved fine-scale soil moisture (SM) dataset at 1 km grid spacing, covering much of the eastern continental US, was generated by assimilating 9 km Soil Moisture Active Passive (SMAP) SM data into the v4.0.1 Noah-MP land surface model. With 12 ensemble members, the assimilation was carried out using the ensemble Kalman filter algorithm within NASA's Land Information System. The SM analysis for 2016 was fully validated against in situ observations from four different networks and compared with four other existing datasets. Results indicate that this SM analysis surpasses other datasets in top-layer SM distribution, including a machine-learning-based product, despite all SM estimates being less heterogeneous than observed. The analysis of anomalous errors suggests that large similarity in intrinsic errors is likely due to overlapping data sources among the selected SM datasets. More detailed evaluations were performed over two geographic areas. The observations collected by the Atmospheric Radiation Measurement facility in Oklahoma suggest that soil temperature and surface heat fluxes are concurrently simulated with good accuracy. Investigation into the 2016 southeastern US drought response further indicates drier conditions and higher evapotranspiration estimates compared to GLEAMv4.1. Notably, large errors are associated with grids having clay soil textures, underscoring the need for refined model treatments for specific soil types to further improve SM estimates. The dataset is publicly available on Zenodo at https://doi.org/10.5281/zenodo.14370563 (Tai et al., 2024).

1 Introduction

Soil moisture (SM) is a critical component in the complex interactions between the land surface and the atmosphere, influencing a range of processes that are vital for weather and climate dynamics. More specifically, it plays a significant role in regulating surface energy fluxes by controlling the partitioning of incoming solar radiation into sensible and latent heat fluxes, thereby impacting atmospheric stability, boundary layer dynamics, and the initiation of convective systems (Dirmeyer et al. 2016; Ek and Holtslag, 2004; Betts, 2002; Taylor et al., 2011).

In addition to groundwater, precipitation falling onto ground surface contributes to the SM availability. Conversely, variations in SM heterogeneity can also influence the spatial and temporal distribution of precipitation through its effects on evapotranspiration rates and the atmospheric moisture and energy budgets (Katul et al., 2012; Hsu and Dirmeyer, 2023). Hence, the feedback loop between SM and precipitation is crucial for understanding and predicting regional hydrological cycles, droughts, and flood events (Koster et al., 2004; Dirmeyer et al., 2016). Furthermore, SM conditions can impact weather extremes such as heatwaves by modulating the surface energy balance and the efficiency of heat exchange between the land surface and the atmosphere (Seneviratne et al., 2010). These interactions occur across various spatial and temporal scales, underscoring the need to accurately capture the spatial and temporal variabilities of SM distribution.

A variety of sensors such as time domain reflectometry (TDR) sensors, capacitance probes, and neutron probes have been used for in situ (ground-based) SM measurements. These measurements provide direct assessments of SM content at specific locations with high temporal resolution and accuracy in the soil column and are most useful for validating remote sensing data and calibrating hydrological models (Robock et al., 2000; Rasheed et al., 2022). However, their relatively sparse distribution hinders their applicability for characterizing realistic local to regional SM variability in broader regions despite efforts made to expand soil moisture observation networks (Diamond et al., 2013; Schaefer et al., 2007; Hawdon et al., 2014; Dorigo et al., 2021a; McPherson et al., 2007; Wang et al., 2023). Conversely, remote sensing satellites such as the Soil Moisture Active Passive (SMAP) mission, Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), Soil Moisture and Ocean Salinity (SMOS), and Sentinel-1 (Entekhabi et al., 2010; Njoku et al., 2003; Kerr et al., 2001; Torres et al., 2012) provide nearly global coverage of soil moisture estimates measured by passive and active microwave sensors. Passive microwave sensors measure soil moisture based on microwave emissions from the Earth's surface, while active radar sensors use backscatter measurements to infer soil moisture levels (e.g., Kerr et al., 2001; Wagner et al., 2013). These satellite-based retrievals offer spatially extensive coverage and reasonable revisit times (1–3 d), contributing to large-scale hydrological and climate studies. Nevertheless, known uncertainties of satellite SM retrievals such as relatively coarse resolution [O (10 km)], limited accuracy (affected by vegetation, surface roughness, and temperature), shallow depth (only a depth of 0–5 cm is measured), and environmental interference (rain, cloud, and snow cover) have posed challenges for their contributions to represent local- to regional-scale SM distribution (e.g., Colliander et al., 2017).

Land surface models (LSMs) can simulate soil moisture conditions for any region by representing the interactions among the atmosphere, vegetation, and ground (Niu et al., 2011; Lawrence et al., 2019; Liang et al., 1994). Key processes such as precipitation, infiltration, lateral flow, evaporation, plant transpiration, and groundwater table variations are parameterized in LSMs. When precipitation occurs, water can infiltrate into the soil, accumulate, or run off, depending on soil characteristics and the rate of rainfall. Evaporation from the soil and transpiration from plants (collectively called evapotranspiration) reduce soil moisture, while infiltration and percolation move water downward through the soil profile. LSMs typically predict these processes to provide estimates of soil moisture at different depths over time. Various depths of soil layers can be configured to model the water movement between these layers in the soil column. A retrospective LSM simulation forced by observation-constrained surface atmospheric conditions (rainfall, temperature, wind, humidity, and radiation, etc.) and land and soil properties (leaf area index (LAI), albedo, land cover, soil texture, and permeability, etc.) is commonly used to reproduce the soil conditions. Despite the advantages, state-of-the-art LSMs still contain uncertain, incomplete, and/or unresolved physical processes that may introduce biases into the simulated land surface properties.

As a way to mitigate such modeling issues, data assimilation (DA) techniques such as the ensemble Kalman filter (EnKF), variational methods (e.g., 3DVar and 4DVar), and Bayesian approaches have been used to merge multiple sources of observational data (in situ measurements and satellite retrievals) with LSM simulations to optimize soil moisture simulations by improving initial conditions and parameter estimates, enhancing the accuracy of soil moisture predictions and hydrological forecasts (e.g., Reichle et al., 2002; Crow and Wood, 2003; Kumar et al., 2008; Chao et al., 2022; Martens et al., 2017). In any DA approach, the assimilation scheme must be coupled with an LSM. As such, the generated analysis consists of model states which are always physically balanced and can be directly used as the initial conditions of an LSM. Some additional advantages of utilizing DA techniques in generating high-resolution SM data include their flexibility in data resolution (output frequency, horizontal grid spacing, and vertical layers) and domain coverage, the possibility to incorporate any improvements in the coupled models and/or new observables, and the availability of the full suite of land surface properties relevant for studies of atmospheric boundary layer and hydraulic processes.

A variety of satellite soil moisture retrievals have been assimilated into different LSMs. For example, Draper et al. (2012) assimilated data measured by both an active microwave advanced scatterometer (ASCAT) and the passive AMSR-E into a catchment model. Liu et al. (2011) also assimilated ASCAT and AMSR-E, but the Noah LSM was chosen as the core model. Seo et al. (2021) conducted experiments over CONUS, which assimilates SMAP and ASCAT data into the Joint UK Land Environment Simulator (JULES) using the local ensemble transform Kalman filter (LETKF). They found that SMAP data are more beneficial than ASCAT in terms of improvement in soil moisture estimate. Mousa and Shu (2020) assessed the potential impacts of assimilation of SMAP, SMOS, and ASCAT on spatial representation in soil moisture over Africa and reported that SMAP has overall superior performance compared to SMOS and ASCAT. Earlier studies have specifically explored the impact of SMAP soil moisture data assimilation in terms of soil moisture estimates, hydrological modeling, and drought monitoring across different regions of the globe. For example, studies have shown promising results by assimilating SMAP soil moisture data into the Noah-MP land surface model (e.g., Rouf et al., 2021; Ahmad et al., 2022). The Noah-MP LSM is widely used in both research and operational systems (e.g., Ma et al., 2017; He et al., 2023; Johnson et al., 2023). Compared to its predecessor (Noah LSM), Noah-MP introduces substantial improvements that enhance realism, flexibility, and process representation. Those up-

dates include dynamic vegetation models, multi-layer snow and soil physics, stomatal resistance schemes, and canopy interception processes. Given the compatibility with coupled models, Noah-MP is integrated into state-of-the-art atmospheric models including the Weather Research and Forecasting (WRF) model, making it an excellent choice of LSM to study land–atmosphere coupling processes. Research by Rouf et al. (2021) discussed how the spatial resolution of SMAP SM data (36 km versus 9 km) and the grid spacing of analysis (12.5 and 0.5 km) would impact SM estimation over Oklahoma using the framework of NASA's Land Information System (LIS). They showed that the accuracy of SM analysis is enhanced when assimilating the 9 km SMAP data with 0.5 km LSM grid spacing. Likewise, Yin and Zhan (2020) showed a positive influence of soil moisture data assimilation coupled with Noah-MP simulations in the continental US (CONUS) that underscores the need for fine-scale soil moisture data to achieve an optimal result. Ahmad et al. (2022) further demonstrated the positive impact of SMAP DA on soil moisture estimate in South Asia along with sensitivities to SMAP data bias correction settings. In Chakraborty et al. (2024), an improved soil moisture distribution over India was obtained by incorporating SMAP soil moisture into the Indian Land Data Assimilation System (ILDAS).

Emerging higher-resolution (i.e., 1 km) soil moisture datasets such as SMAP-derived 1 km downscaled surface soil moisture data (Fang et al., 2022) and Sentinel-1 surface soil moisture data (Fan et al., 2025) could potentially provide finer-scale soil moisture information and may be incorporated into data assimilation processes. However, as described in Fang et al. (2022), the spatial coverage and availability of the downscaled SMAP dataset are notably reduced compared to the 9 km dataset. On the other hand, although there are multiple studies demonstrating the impacts of assimilation of Sentinel-1 data (e.g., Brocca et al., 2024; Filippucci et al., 2022; Foucras et al., 2020; Gao et al., 2017; Meyer et al., 2022), the experiments were all performed over smaller and more localized areas as opposed to the more extensive domains used in SMAP-based studies. This is primarily due to the contrasts in sensor characteristics between these two satellites. While Sentinel-1 measures at a higher spatial resolution ($\sim$ 1 km) than SMAP, it has relatively lower radiometry sensitivity, has a much longer revisit time (3 to 4 times), and requires more complex preprocessing. This means that Sentinel-1 may be less sensitive to subtle differences in soil moisture content than SMAP and would be unlikely to capture day-to-day variability. Hence, in general, SMAP data are more suitable for regional- and global-scale applications than Sentinel-1.

Building upon these studies, we aim to improve local to regional soil moisture distributions over much of the east CONUS region by assimilating the SMAP Level 3 (L3) 9 km soil moisture product into the 1 km grid spacing Noah-MP LSM. While earlier studies (e.g., Rouf et al., 2021) chose not to use higher-resolution precipitation forcing data, we use the 4 km NCEP Stage IV Quantitative Precipitation Estimate (QPE) data (Lin and Mitchell, 2005) as the LSM's precipitation forcing. The Stage IV product is a unique precipitation dataset since it takes advantage of both the weather radar and rain gauge observation networks over the CONUS to robustly reconstruct precipitation distribution. Moreover, instead of only focusing on the importance of SM DA to improve soil moisture estimates as other studies did, we also explore how SM DA may influence other simulated land surface properties on both seasonal and regional bases. We assess the performance of our dataset over the full study domain but also explore key regions in additional detail. Specifically, we evaluate the performance of our dataset in a known "hotspot" of land–atmosphere coupling using the dense in situ observations collected by the Oklahoma Mesonet and DOE's Atmospheric Radiation Measurement (ARM) facility in the Southern Great Plains (SGP), which were also used in the study of Rouf et al. (2021). In addition, we examine our dataset's characterization of the extreme drought conditions affecting the southeast US over the fall and winter of 2016. Comparisons are made to alternative SM datasets, including datasets generated through machine learning approaches, to better understand the value of DA incorporated with an LSM. The primary goal of this study is to demonstrate the development of a year-long soil moisture dataset for the eastern US, which can also be readily used in land–atmosphere coupled simulations by providing the essential boundary conditions needed for model initialization.

The remaining parts of this paper are organized as follows: the analysis domain and period are described in Sect. 2. The methodologies and datasets employed in this study are detailed in Sect. 3. The results of the impact of SM data assimilation and the evaluations of the generated SM estimate along with the other existing SM datasets are discussed in Sect. 4. Lastly, the summary and discussion are provided in Sect. 5.

2 Analysis domain and period

Our study domain encompasses a wide swath of the central and eastern CONUS (Fig. 1). The time period for the analysis covers the entire year of 2016 from 1 January through 31 December 2016. This analysis period was selected in order to complement land–atmosphere coupled simulations associated with the 2016 Holistic Interactions of Shallow Clouds, Aerosols, and Land Ecosystems (HI-SCALE) field campaign (Fast et al., 2018). The locations of in situ measurements from the United Sates Climate Reference Network (USCRN), Soil Climate Analysis Network (SCAN), Oklahoma Mesonet (OKMet), and ARM SGP are overlaid on the map in Fig. 1a. The soil texture and land cover maps are given in Fig. 1b and c, respectively. Table 1 summarizes the grid numbers and their percentages over the study domain for each classification of soil texture and land cover. The top three soil types (besides water) are silt loam (24.02 %), loam

(18.88 %), and sandy loam (15.7 %), whereas grassland, cropland, and cropland/natural vegetation mosaics are the top three land cover types accounting for 22.2 %, 19.64 %, and 10.2 %, respectively, of points in the domain.

3 Methodology and datasets

3.1 NASA Land Information System and Noah-MP land surface model

The NASA Land Information System (LIS) is an advanced modeling and data assimilation framework designed to better simulate land surface processes and improve our understanding of terrestrial hydrology, biogeochemistry, and climate interactions (Kumar et al., 2006; Peters-Lidard et al., 2007). LIS incorporates multiple hydrological models and LSMs as well as data assimilation techniques to optimize the representation of land surface processes. This model–observation integration enhances the accuracy and reliability of simulations by leveraging the strengths of different models and observational datasets. It is functionable in assimilating satellite-derived observations of soil moisture, vegetation dynamics, and other land surface variables to improve the initialization and calibration of model simulations. Its versatility and scalability make it suitable for both research and operational uses. Given the above, LIS is primarily used in this study to generate realistic representation in soil states through assimilation of SMAP soil moisture retrievals into the Noah-MP land surface model.

The version 4.0.1 Noah-MP LSM (Ek et al., 2003; Niu et al., 2011; Yang et al., 2011) was run within LIS to simulate the relevant land surface processes across the study domain. The Noah-MP model was run with a 0.01° by 0.01° horizontal grid spacing and using a 15 min time step. The specific model configurations utilized are detailed in Table 2. Each soil column within the study region is represented by four layers with depths of 10, 30, 60, and 100 cm below the ground surface. The surface soil moisture updates are transmitted to deeper layers according to model formulations in water diffusivity and hydraulic conductivity. More specifically, while moisture fluxes between successive layers control how water moves within each soil column, excess water above saturation in any layer will be transferred to the next unsaturated layer downward. The Noah-MP LSM can be driven by many sources of meteorological forcing data as desired. Note that external irrigation and groundwater extraction were not explicitly simulated in Noah-MP, and these processes might be important for certain locations (Yang et al., 2020, 2021).

3.2 Datasets

The datasets employed in this study include the forcing data that drive the Noah-MP LSM (Sect. 3.2.1–3.2.3), as well as multiple in situ observations (Sect. 3.2.4) used as the benchmarks for intercomparison among our SM estimates and the other existing SM datasets (Sect. 3.2.5–3.2.7).

3.2.1 Enhanced SMAP Level 3 soil moisture data

The Soil Moisture Active Passive (SMAP) mission uses passive (radiometer) L-band microwave remote sensing to estimate land surface soil moisture and freeze/thaw state (O'Neill et al., 2014). The L-band radiometry offers all-weather, diurnal sensing of the surface dielectric properties, which are a function of the near-surface soil moisture. SMAP has a 2 to 3 d revisit frequency and two overpasses (morning and afternoon) at local times of 06:00 and 18:00, respectively. One of the SMAP products, the enhanced SMAP Level 3 soil moisture product (SPL3SMP_E; O'Neill et al., 2023), is primarily used for assimilation in this study. It consists of daily estimates of global soil moisture within the top soil layer (∼ 5 cm depth) on a cylindrical 9 km equal-area scalable Earth grid (https://nsidc.org/data/spl3smp_e/versions/6, last access: 8 September 2025), spanning 31 March 2015 to present.

3.2.2 North America Land Data Assimilation System Phase 2 (NLDAS-2)

The NLDAS-2 (Xia et al., 2012) aims to provide high-resolution, near-real-time, and retrospective datasets that integrate land surface model outputs with observations to monitor and simulate land surface conditions across North America. It is available at hourly intervals and on a 12.5 km spatial grid from January 1979 to the present. A wide range of land surface variables such as soil moisture, soil temperature, snow cover, evapotranspiration, and runoff are provided. Meteorological forcing variables such as precipitation, temperature, wind speed, and solar radiation are also included. The NLDAS-2 is used in this study as the meteorological forcing data to drive the Noah-MP LSM.

3.2.3 NCEP Stage IV Quantitative Precipitation Estimate

The NCEP Stage IV Quantitative Precipitation Estimate (QPE) (Lin and Mitchell, 2005) is a high-resolution, quality-controlled dataset produced by the National Centers for Environmental Prediction (NCEP). It integrates precipitation data from multiple sources, including the Next-Generation Radar (NEXRAD) network rain gauges, and satellite observations, to provide accurate and detailed precipitation estimates across the contiguous United States. With a grid spacing of 4 km at hourly intervals, Stage IV QPE is widely used in meteorology, hydrology, and climate research for tasks such as weather forecasting, flood modeling, and studying precipitation trends. We replace the precipitation data in the NLDAS-2 with the Stage IV QPE data as they provide not only higher-resolution and more realistic precipitation forc-

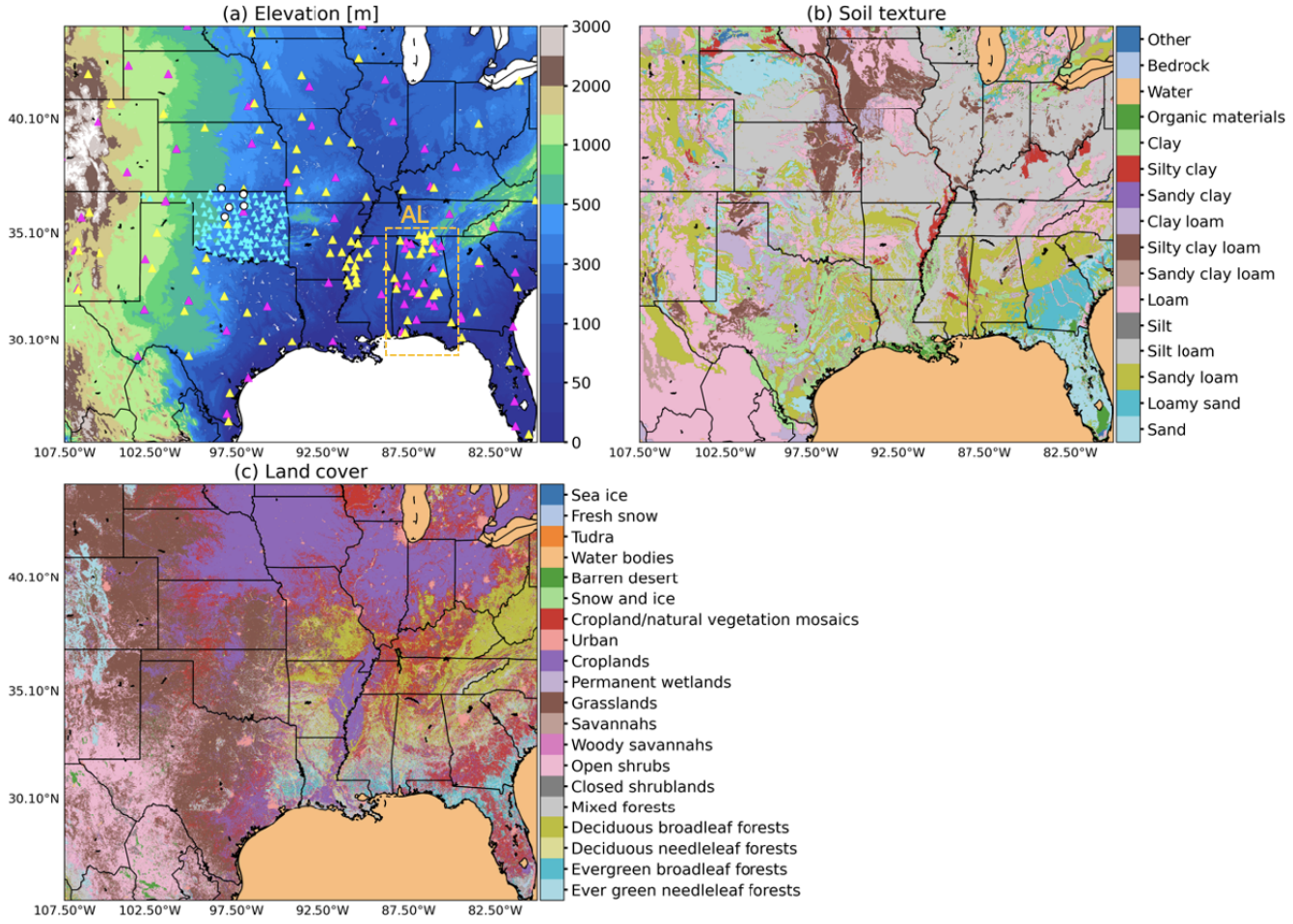

Figure 1. Maps illustrating the study domain over eastern CONUS. The yellow, magenta, and cyan triangles denote the stations of SCAN, USCRN, and OKMet observational networks, respectively. The white circles mark the locations of selected ARM SGP sites. The domain soil texture was categorized into 14 soil types **(c)** according to the NCEP/STATSGO+FAO classification. The domain land cover comprised 18 main types based on the MODIS-derived IGBP classification. The subdomain AL is denoted by the orange box (dashed line) in **(a)**.

ing over the CONUS region but also improved SM estimates in our test simulations. As an example, the comparison of instantaneous rain rate obtained from NLDAS-2 and Stage IV precipitation at 00:00 UTC on 30 August 2016 (Fig. S1 in the Supplement) demonstrates that the Stage IV data provide a more heterogeneous precipitation distribution than NLDAS-2 over the study domain.

3.2.4 In situ measurements

In situ soil moisture observations used in this study were obtained from the (1) US Climate Reference Network (USCRN), (2) Soil Climate Analysis Network (SCAN), (3) Oklahoma Mesonet (OKMet, McPherson et al., 2007), and (4) ARM SGP (Sisterson et al., 2016). The USCRN and SCAN data are acquired from the International Soil Moisture Network (Dorigo et al., 2021a). The four networks are selected as the benchmarks of our SM analysis due to either their relatively wide spatial coverages or preferred site locations. Besides atmospheric and environmental parameters such as air temperature, humidity, and wind conditions, both SCAN and USCRN stations are equipped with sensors that measure critical soil parameters, including soil moisture and temperature at depths of 5, 10, 20, 50, and 100 cm. The USCRN and SCAN are superior among available soil moisture networks as many of their stations (112 and 91 sites from USCRN and SCAN, respectively) are uniformly distributed over the study domain (Fig. 1). They are used to evaluate our SM analysis along with other existing SM datasets (Table 3). The OKMet (120 sites) and ARM SGP (6 sites) observations are adopted as their site locations are densely distributed (the average distance between any two stations is shorter than 30 km) over a portion of the Southern Great Plains (SGP) region, which is one of the hotspots with strong land–atmosphere coupling (e.g., Fast et al., 2018; Sakaguchi et al., 2022). In addition to SM, the soil temperature observations and the latent and sensible heat fluxes measured by the

Table 1. Summary of grid numbers and percentage of total grids for the soil texture/land cover types.

Soil texture			Land cover		
Class	No. of grids	Percentage of total grids [%]	Class	No. of grids	Percentage of total grids [%]
Sand	370 227	7.12	Evergreen needleleaf forests	120 816	2.32
Loamy sand	140 072	2.69	Evergreen broadleaf forests	66 193	1.27
Sandy loam	816 849	15.70	Deciduous needleleaf forests	344	0.007
Silt loam	1 249 730	24.02	Deciduous broadleaf forests	373 716	7.18
Loam	982 206	18.88	Mixed forests	332 114	6.38
Sandy clay loam	41 522	0.80	Closed shrublands	23 319	0.45
Silty clay loam	240 916	4.63	Open shrubs	514 667	9.89
Clay loam	187 450	3.60	Woody savannahs	100 390	1.93
Silty clay	63 818	1.23	Savannahs	27 746	0.53
Clay	205 676	3.95	Grasslands	1 154 805	22.20
Organic materials	39 598	0.76	Permanent wetlands	9591	0.18
Water	842 008	16.19	Cropland	1 021 681	19.64
Other	22 069	0.42	Urban	64 997	1.25
			Cropland/natural vegetation mosaics	521 029	10.02
			Snow and ice	33	0.0006
			Barren desert	28 692	0.55
			Water bodies	842 008	16.19

Table 2. Selected parameters, parameterizations, and forcing data used in the configured Noah-MP LSM.

LSM parameter/parameterization/forcing data	
Land cover	MODIS (IGBP-NCEP) (Friedl et al., 2002)
Elevation, slope, and aspect	SRTM30-v2.0 (Farr et al., 2007)
Greenness	National Centers for Environmental Prediction (Gutman and Ignatov, 1998)
Vegetation	Dynamic vegetation option
Maximum albedo	National Centers for Environmental Prediction (Robinson and Kukla, 1985)
Canopy stomatal resistance	Ball–Berry method (Ball et al., 1987)
Snow surface albedo	Canadian land surface scheme (Verseghy, 1991)
Runoff and groundwater	Simple groundwater model, SIMGM (Niu et al., 2007)
Surface-layer drag coefficient	General Monin–Obukhov similarity theory (Brutsaert, 1982)
Snow and soil temperature	Semi-implicit option
Partitioning of rain and snowfall	Jordan91 (Jordan, 1991)
Lower boundary of soil temperature	Noah native option
Supercooled liquid water and frozen soil permeability	NY06 (Niu et al., 2007)
Surface meteorological forcing	NLDAS-2 and Stage IV QPE (precipitation)

Soil Temperature and Moisture Profiles (STAMP) and Eddy Correlation Flux Measurement System (ECOR) deployed by the ARM SGP facility are also used to concurrently assess the simulated soil properties and surface heat fluxes. Note that soil moisture (temperature) measured at a depth of 5 cm below the ground surface was primarily used to compare with the model-estimated surface soil moisture (soil layer depth = 0 to 10 cm).

3.2.5 ERA5-Land reanalysis

ERA5-Land (Muñoz-Sabater et al., 2021) is a global reanalysis dataset that provides essential land variables with a grid spacing of 0.1° and is valid at hourly frequency, spanning January 1950 to the present. It is continuously produced by rerunning the land component (Tiled ECMWF Scheme for Surface Exchanges over Land incorporating land surface hydrology – H-TESSEL) of the ECMWF ERA5 climate reanalysis that sequentially assimilates available meteorological observations (Hersbach et al., 2020). Despite model uncertainties due in part to imperfect atmospheric forcing, unresolved physical processes, and lack of observational constraint, the spatiotemporal coverage of the ERA5-Land dataset has been advantageous in many land surface applications including flood or drought monitoring and fore-

Table 3. Soil moisture estimates analyzed in this study.

Soil moisture product	Grid spacing	Spatial coverage	Temporal resolution	Temporal coverage	References
SPL3SMP_E	9 km	Global	Daily	31 March 2015–present	O'Neill et al. (2023)
ERA5-Land	0.1°	Global	Hourly	1950–present	Muñoz-Sabater et al. (2021)
GLASS SM	1 km	Global	Daily	2000–2020	Zhang et al. (2023)
GLEAM v4.1	0.1°	Global	Daily	1980–2023	Miralles et al. (2025)
SMAPDA	1 km	East CONUS	6 hourly	2016	–

casting. It is thus employed in this study as SM reference data, providing more insights through the comparison.

3.2.6 Global Land Surface Satellite soil moisture (GLASS SM)

The global, daily 1 km GLASS soil moisture product (GLASS SM; Zhang et al., 2023) was derived using an ensemble learning model (eXtreme Gradient Boosting – XGBoost) that integrates multiple datasets as the machine learning (ML) model's inputs, including the remotely sensed Global Land Surface Satellite (GLASS) products (Liang et al., 2021), ERA5-Land reanalysis products (Muñoz-Sabater et al., 2021), and static auxiliary datasets (e.g., Multi-Error-Removed Improved-Terrain – MERIT – and Global gridded soil information – SoilGrids; Poggio et al., 2021). The ground-based soil moisture archived by the International Soil Moisture Network (ISMN) and the 0.25° grid spacing combined soil moisture data from the European Space Agency's Climate Change Initiative (ESA CCI; Dorigo et al., 2017) are collectively used as the target data for training in ML. The validations carried out for the GLASS SM product in Zhang et al. (2023) demonstrated its capability to capture temporal dynamics of measured soil moisture. Hence, given its novelty in the methodology and high spatial resolution (1 km), the GLASS SM data are used as one of the benchmarks in this study.

3.2.7 Global Land Evaporation Amsterdam Model (GLEAM)

GLEAM (Global Land Evaporation Amsterdam Model; Miralles et al., 2011) is a state-of-the-art dataset that provides global estimates of soil moisture, terrestrial evaporation (or evapotranspiration), and related hydrological components. GLEAM soil moisture data are derived from satellite observations and model simulations. It integrates a variety of satellite observations and meteorological data, such as soil moisture from microwave remote sensing, vegetation indices, and meteorological data on precipitation, air temperature, and radiation. Version 4.1 of GLEAM (Miralles et al., 2025) is used in our analysis, which is available at 0.1° resolution for the period of 1980 to 2023.

3.3 Open-loop and data assimilation simulations

The open-loop simulation (named "OL" hereafter) employs the configuration noted in Table 2 and spins up between 1 January 2011 and 31 March 2015. It refers to the integration of the Noah-MP land surface model without any assimilation of external observations. The long spin-up period (greater than 4 years) ensures that soil states reach the equilibrium state before conducting data assimilation (Cosgrove et al., 2003; Rodell et al., 2005). Since the SMAP SM data are only available from 31 March 2015 onwards, the DA simulation started at 00:00 UTC on 1 April 2015 and ended at 00:00 UTC on 1 January 2017. The ensemble Kalman filter (EnKF) assimilation algorithm implemented in the LIS is utilized to assimilate the SMAP SM retrievals into the Noah-MP-modeled estimates. The EnKF's sequential assimilation algorithms including two main steps (model propagation and data assimilation update) are coupled with model integration and executed recursively. Here, Noah-MP is the nonlinear forward model to advance the propagation step and generate the prognostic state vector forward in time. The update step occurs whenever any observations are valid, and the update of prognostic state variable can be described by the equation below:

$$\hat{x}^{\mathrm{a}}_{k+1} = \hat{x}^{\mathrm{b}}_{k+1} + K\left(y_{k+1} - H_{k+1}\left(\hat{x}^{\mathrm{b}}_{k+1}\right)\right), \quad (1)$$

where $\hat{x}^{\mathrm{a}}_{k+1}$ stands for the analyzed (updated) state of variable x at time step $k+1$. $\hat{x}^{\mathrm{b}}_{k+1}$ represents the background state of variable x integrated from time step k. The Kalman gain matrix K and the innovation vector $\left(y_{k+1} - H_{k+1}\left(\hat{x}^{\mathrm{b}}_{k+1}\right)\right)$ are required when updating the background state. Here, y_{k+1} denotes the observations valid at time step $k+1$, and H_{k+1} is the observation operator that applies conversion and interpolation in time and space to the model state variable in order to conform with the observable.

The ensemble simulations are required at each propagation step to provide an estimate of the model spread (uncertainty). Here, the NASA Land Data Toolkit (LDT; Arsenault et al., 2018) is used to initialize the ensemble simulations based on the OL simulation restart output file at 23:45 UTC on 31 March 2015. The initial conditions of those ensemble members are obtained by perturbing atmospheric forcing variables as listed in Table 4. Perturbation type is grouped

as either multiplicative (M), sampled from a lognormal distribution, or additive (A), which is sampled from a normal distribution.

According to the sensitivity study regarding the impact of ensemble size in Ahmad et al. (2022), the ensemble spread (measured by standard deviation across all members) may be flattened when the number of replicates increases beyond 15. We experimented with 12 and 24 ensemble members, and the result suggested minor differences in terms of soil moisture representation. Hence, the DA experiment we show here has an ensemble size of 12. The model and SMAP soil moisture retrieval error standard deviations are set as $0.04\,m^3\,m^{-3}$. Due to the existence of relative systematic differences between SMAP and modeled SM, the cumulative distribution function (CDF) matching technique (Reichle and Koster, 2004) is used for bias correction of the SMAP soil moisture retrievals using Noah-MP model data as the reference. Monthly CDFs of the SMAP soil moisture retrievals and the Noah-MP-simulated soil moisture were both generated using the NASA LDT and used to map the SMAP SM retrievals into the Noah-MP-modeled soil moisture space prior to assimilation. The reference period for the monthly matching is 2 years in total, ranging from 1 January 2015 to 31 December 2016. Since the SMAP SM data are representative of the top soil layer (~ 5 cm deep from surface), the topmost soil layer soil moisture is employed as the model state variable during assimilation. The DA simulation and its SM data are abbreviated as "SMAPDA" hereafter. A more detailed discussion regarding its performance for estimated SM is covered in Sect. 4.

3.4 Metrics for DA impact measuring and evaluation

3.4.1 Soil moisture analysis increment

To assess the impact of the SMAP SM data assimilation on the soil moisture estimates, we analyze the soil moisture analysis increments generated from the DA experiment (SMAPDA). The analysis increment refers to the difference between the analysis (optimized estimate of the state after DA) and the background forecast (model state before DA). It is a measure of how much the model state has been corrected (updated) by incorporating new observations, which is not only related to the deviation from model background to observation but also modulated by observation and model errors. In the EnKF approach, the model error varies in time and space and is estimated using the ensemble spread (standard deviation of ensemble simulations). We use the cumulative number and temporal mean of soil moisture analysis increments to indicate the spatial distribution of observational constraint by the SMAP_L3_E data and highlight the areas that experience overall wetting or drying due to the cycling of assimilation. Note that the SMAP_L3_E data were subset to hourly data and assimilated when they matched the model time step.

3.4.2 Evaluation against in situ measurements

The soil moisture estimates generated through different approaches are evaluated against in situ measurements using the metrics of anomaly correlation coefficient (ACC), root mean square error (RMSE), and bias, defined as follows.

$$\mathrm{ACC} = \frac{\sum_{i=1}^{n}\left(P'_i - \overline{P'}\right)\left(M'_i - \overline{M'}\right)}{\sqrt{\sum_{i=1}^{n}\left(P'_i - \overline{P'}\right)^2 \sum_{i=1}^{n}\left(M'_i - \overline{M'}\right)^2}} \tag{2}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(P_i - M_i)^2} \tag{3}$$

$$\mathrm{Bias} = \frac{1}{N}\sum_{i=1}^{n}(P_i - M_i), \tag{4}$$

P represents the estimated top-layer soil moisture, and M stands for corresponding in situ soil moisture measurement. N is the total number of selected samples. The ACC, ranging from -1 to 1, measures how well the temporal anomalies (departures from the monthly mean) of two time series (model estimates P' and observation M') match each other. Here in Eq. (2), it is essentially computed as the Pearson correlation coefficient using the estimated and observed soil moisture anomaly time series in each location. Since the soil moisture time series has strong seasonal cycles, the removal of the seasonal signal when computing ACC helps quantify the skill in capturing soil moisture temporal variations across all timescales. The ACC is commonly used to verify the impact of soil moisture data assimilation due to the necessity of isolating the seasonal cycle, which is highly consistent between open-loop and assimilation experiments (e.g., Kumar et al., 2009). While RMSE (Eq. 3) is used to measure the mean difference between the modeled and in situ SM, bias (Eq. 4) is computed as the overall deviation (including the signs) of the modeled SM from in situ SM observations. In addition, the standard deviation (SD) is also calculated for each SM dataset to quantify the spatial heterogeneity in SM across the given sites at different locations:

$$\mathrm{SD} = \sqrt{\frac{1}{N}\sum_{i=1}^{n}\left(S_i - \overline{S}\right)}. \tag{5}$$

Here, S_i refers to individual SM data points and $\overline{S}$ stands for the mean over the entire dataset.

4 Results

4.1 Increments from SMAP soil moisture data assimilation

To gauge how much observational information was effectively assimilated into the model, we examined the outputs of SM analysis increments at the top layer (5 cm depth). Figure 2 illustrates the maps of the cumulative number (hours)

Table 4. Description of parameters used in meteorological forcing perturbations for the ensemble simulations.

Perturbed meteorological forcing	Perturbation type	Standard deviation	Cross-correlations with perturbations			
			SW	LW	P	T_{air}
Shortwave radiation (SW)	M	$0.2\,W\,m^{-2}$	1.0	−0.3	−0.5	0.3
Longwave radiation (LW)	A	$30\,W\,m^{-2}$	−0.3	1.0	0.5	0.6
Precipitation (P)	M	0.5 mm	−0.5	0.5	1.0	−0.1
Near-surface air temperature (T_{air})	A	0.5 K	0.3	−0.6	−0.1	1.0

of SM analysis increments over each of the 3-month-long periods. Overall, the SMAP SM data assimilation is more effective in spring, summer, and fall (Fig. 2b, c, and d) as opposed to winter (January–February–March; Fig. 2a). The relatively small number of analysis increments shown in the Jan–Feb–Mar period (Fig. 2a) is likely due to the increased uncertainty in L-band microwave radiometer SM retrieval as a result of snow cover and frozen ground in the cold season (e.g., Liu et al., 2021). While analysis increments are distributed over the majority of domain, there are grids that received zero update, especially in the eastern part of the domain. In the default setting, the LIS would only assimilate observations where the SMAP data's retrieval quality is flagged as successful. Those zero-update pixels are most likely covered by dense vegetation. As such, the sensitivity of surface SM is usually distinctly reduced and thus flagged as unsuccessful retrievals. Nevertheless, despite generally less effective assimilation over this region, a few spots in Florida and partially Georgia and South Carolina show that most frequent updates from DA across the entire domain.

Figure 3 demonstrates the spatial distribution of mean SM analysis increments over the four seasons. The calculation of mean increment only includes samples with nonzero increments. While consistently positive increments are shown in Texas and northern Mexico throughout the year, seasonal variations are evident in portions of the Great Plains. For instance, in Kansas, more negative (positive) increments are seen for January–February–March/April–May–June (Fig. 3a and b) and July–August–September/October–November–December (Fig. 3c and d), respectively. This suggests that compared to SMAP observations, the model most likely has a consistent dry bias over part of Texas and the adjacent Mexican territory, and the biases are more variable temporally in other parts of the domain including the northern SGP. Possible causes of those DA increments are the model deficiencies and missing physical processes. For instance, SMAP detects increased soil moisture, which may be partially due to irrigation, whereas the Noah-MP LSM used in this study does not explicitly simulate irrigation (e.g., Felfelani et al., 2018; Lawston et al., 2017). In addition, practices like tilling or cover cropping affect surface moisture and are likely not captured by the model physics. Other missing/unrealistic model treatments in runoff schemes, dynamic groundwater level, seasonal varying vegetation, and root systems may also modulate the increment patterns. Moreover, biases in meteorological forcing (e.g., radiation, temperature, and winds) may also affect how evapotranspiration is estimated and thus the soil moisture. SMAP data assimilation often compensates for these errors, especially after dry spells or in transition seasons (spring and fall).

4.2 Comparison with existing surface SM datasets

To assess the performance of our SM analyses (SMAPDA and OL) along with other existing SM products, we conduct a comprehensive intercomparison among all derived datasets (Table 3) against a collection of in situ measurements from four observational networks (USCRN, SCAN, OKMet, and ARM SGP). Note that the assessments were conducted separately against the USCRN and the SCAN datasets despite both having well-distributed site locations over the study domain. This was carried out purposely in order to verify whether any inconsistency between their instruments and/or measurements may alternatively bias the validation results. The following subsections discuss the evaluation results referenced by using the observations from each network.

4.2.1 Evaluation using USCRN soil moisture observations

SM estimates from SMAPDA, OL, GLASS SM, ERA5-Land, GLEAMv4.1, and SMAP AM (the morning overpass of SMAP_L3_E) are first evaluated using the in situ observations from the USCRN (Fig. 1). The metrics described in Sect. 3.4.2 are computed accordingly. Since only SMAPDA and ERA5-Land consist of SM representations through the entire soil column, surface (top-layer) SM representations are primarily assessed here. To perform one-to-one comparisons with in situ data, for each SM product, the daily SM time series data at the grid cells closest to the observational site locations are extracted. The 2D histograms given in Fig. 4 visualize the differences between the observations and the estimates and depict the contrasts among the datasets. All scatter points are grouped by 50 bins (2D pixels), and the contours are smoothed using the Gaussian filter for an improved visualization. The more samples are concentrated along the diag-

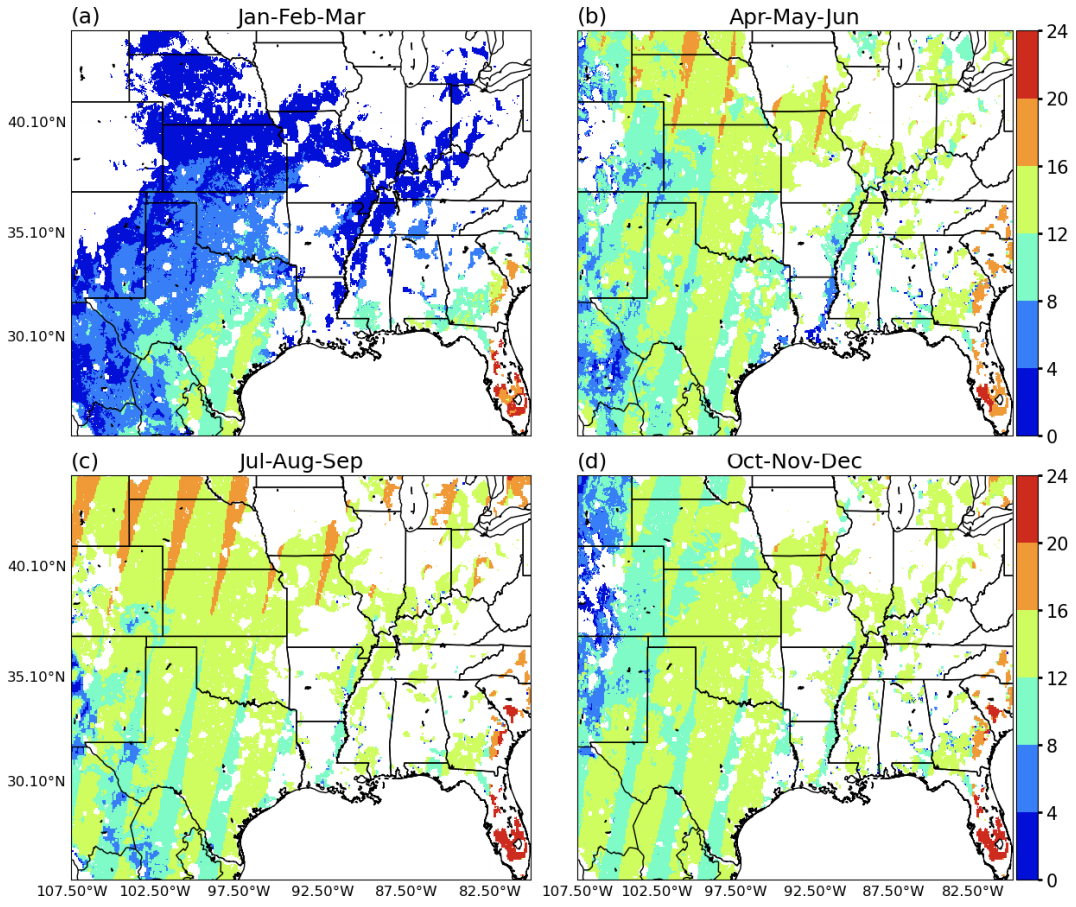

Figure 2. Maps of the cumulative number of DA SM increments computed for the periods of **(a)** January–February–March, **(b)** April–May–June, **(c)** July–August–September, and **(d)** October–November–December in 2016.

onal line, the better the estimate would be (placeholder for root zone SM evaluation).

The results indicate that SMAPDA has the highest ACC (~ 0.8) among all SM estimates. While it is slightly higher than OL in general, Fig. 5b shows that SMAPDA improves over OL at most of the sites, indicating that SM DA does optimize the SM dynamics despite overall minor differences in RMSE and bias. It also shows that SMAPDA's RMSE and bias (0.085 and $0.005\,m^3\,m^{-3}$) are slightly larger than what GLASS SM has (0.083 and $-0.004\,m^3\,m^{-3}$). Since the GLASS SM uses in situ observations including those from USCRN as the target when training the ML model (i.e., not independent), it is not surprising the GLASS SM magnitudes better align with the USCRN data in general. However, the GLASS SM has the smallest ACC (0.574) compared to all other estimates. This implies that the ML algorithm may not capture the temporal evolution of SM but only the instantaneous SM values due to the selected variables (i.e., absolute SM) in the cost function. At any rate, these two 1 km grid spacing products significantly outperform others. Constructed in 0.1° grid spacing, both ERA5-Land and GLEAMv4.1 (Fig. 4c and f) may partially suffer from a relatively coarser resolution in addition to differences in treatments of physical processes. As a result, their RMSEs are all greater than $0.1\,m^3\,m^{-3}$. Meanwhile, relatively larger biases are also computed (0.017 and $-0.006\,m^3\,m^{-3}$). The SMAP AM also has poor skill in SM estimation given its highly scattered samples in the 2D histogram despite relatively low bias. Although it has the second-highest ACC overall (0.625), it still underrepresents temporal variability at the site level when compared with SMAPDA (Fig. 5e).

We also note that the cut-in (smallest) values of surface SM vary notably across the SM products. For example, GLEAMv4.1 and SMAPDA have relatively larger cut-in values of ~ 0.05 to $0.06\,m^3\,m^{-3}$, whereas the ERA5-Land and SMAP AM values are valid above approximately $0.02\,m^3\,m^{-3}$. The GLASS SM has a negligible limit on the smallest SM value. The differences in these cut-in SM values may be associated with either the formulations of land surface models or the observational sensitivities and could at least partially affect how well each estimate agrees with the observation.

Figure 6a illustrates the disaggregated RMSEs for SMAPDA at each USCRN site. The RMSE differences be-

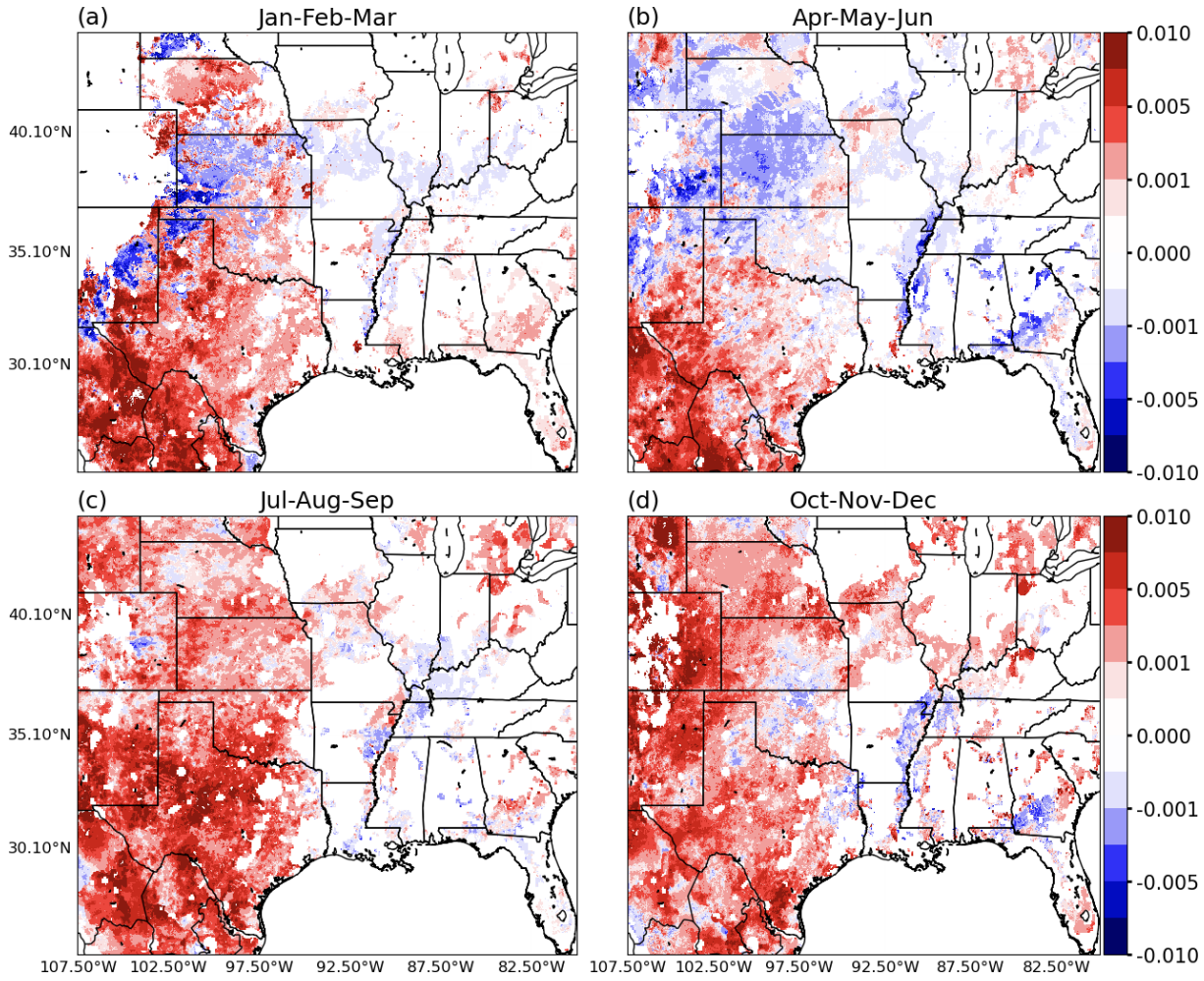

Figure 3. Similar to Fig. 2, but for mean SM increments.

tween SMAPDA and other estimates are given in Fig. 6b–f to better visualize relative performance. Not surprisingly, relatively minor differences between OL and SMAPDA are analyzed (Fig. 6b). While the RMSE differences are rather mixed between SMAPDA and GLASS SM (Fig. 6d), distinct and extensive increases in RMSEs are observed in the cases of ERA5-Land, SMAP AM, and GLEAMv4.1 (Fig. 6c, e, and f), especially for sites in the southeast US and coastal sites in Florida and Texas.

Likewise, the biases are displayed in Fig. 7. In all SM datasets, a wet bias is more evident in the southeastern US sites than others, whereas a dry bias is distinct across many sites in the northern and eastern Great Plains despite variability in their magnitudes. This consistent bias pattern implies that these SM estimates may share common sources of uncertainties in observational data and/or treatments in the models. Further model improvements may be carried out to focus on the correction of this common issue.

To further examine the potential errors in common among the six SM estimates, we calculated the RMSE anomaly for each dataset. The RMSE anomaly is obtained by subtracting annual mean RMSE from the daily time series of each estimate. It extracts intrinsic variation in SM errors from the original SM time series and thus facilitates bias-free intercomparison. A diverse variation among the datasets is shown in Fig. 8. Despite relatively large day-to-day variability in the SMAP AM time series than other datasets, the multiday variability in SMAP AM is similar to GLEAMv4.1. For example, both of them show much larger SM errors from January to April and relatively smaller errors present in late spring and summer. The errors increase when it transitions into late fall and early winter. There is also a high level of similarity between the ERA5-Land and GLASS SM time series. Despite minor discrepancies, compared to other datasets, they both show relatively smaller variation over the 1-year period with slightly larger errors in April, June, and July. These results are not surprising as SMAP data are one of the ingredients of GLEAMv4.1 (Miralles et al., 2025), whereas GLASS SM adopts ERA5-Land soil moisture as the SM input data in the ML model (Zhang et al., 2023). Figure 8 also indicates that

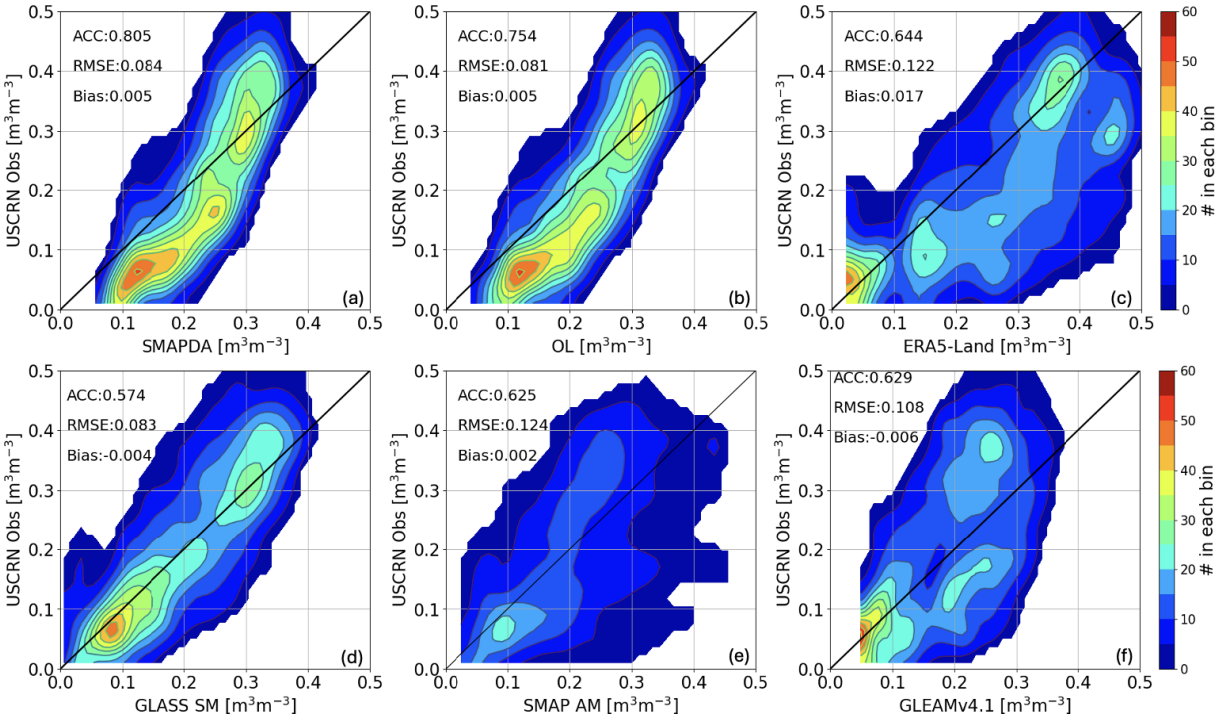

Figure 4. 2D histograms summarizing the evaluation results using the observational data measured by the USCRN. Panels **(a)** to **(f)** represent results of SMAPDA, OL, ERA5-Land, GLASS SM, SMAP AM, and GLEAMv4.1. A total of 50 bins are used to generate the 2D histograms. The anomaly correlation coefficient (CC), RMSE, and bias are given in the upper left corner of each panel.

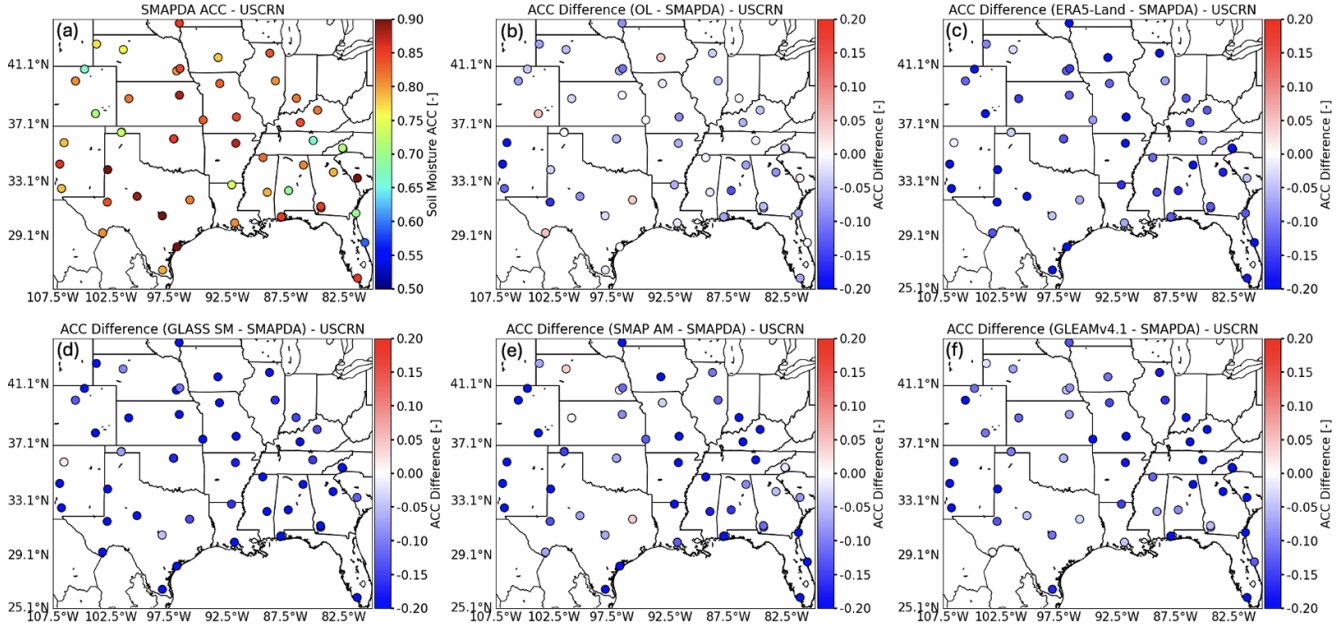

Figure 5. **(a)** Site-wise SM ACC computed for SMAPDA using the USCRN observations. The ACC differences subtracting SMAPDA results from **(b)** OL, **(c)** ERA5-Land, **(d)** GLASS SM, **(e)** SMAP AM, and **(f)** GLEAMv4.1 are also illustrated.

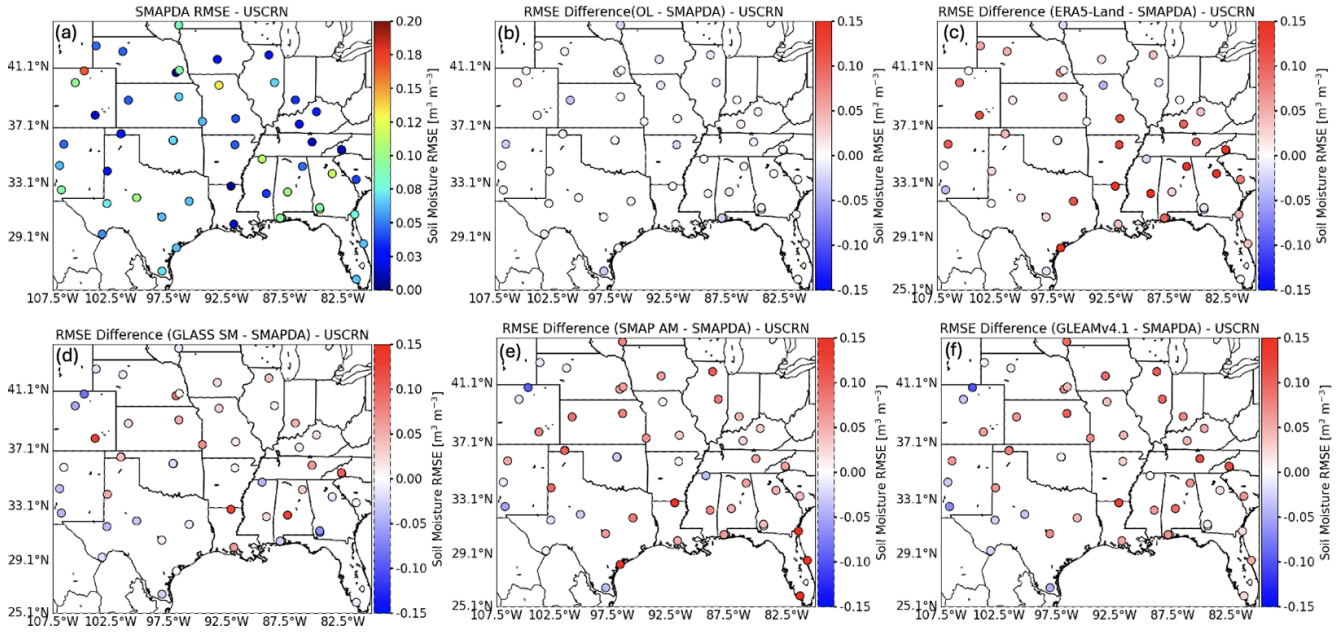

Figure 6. Similar to Fig. 5, but for SM RMSE.

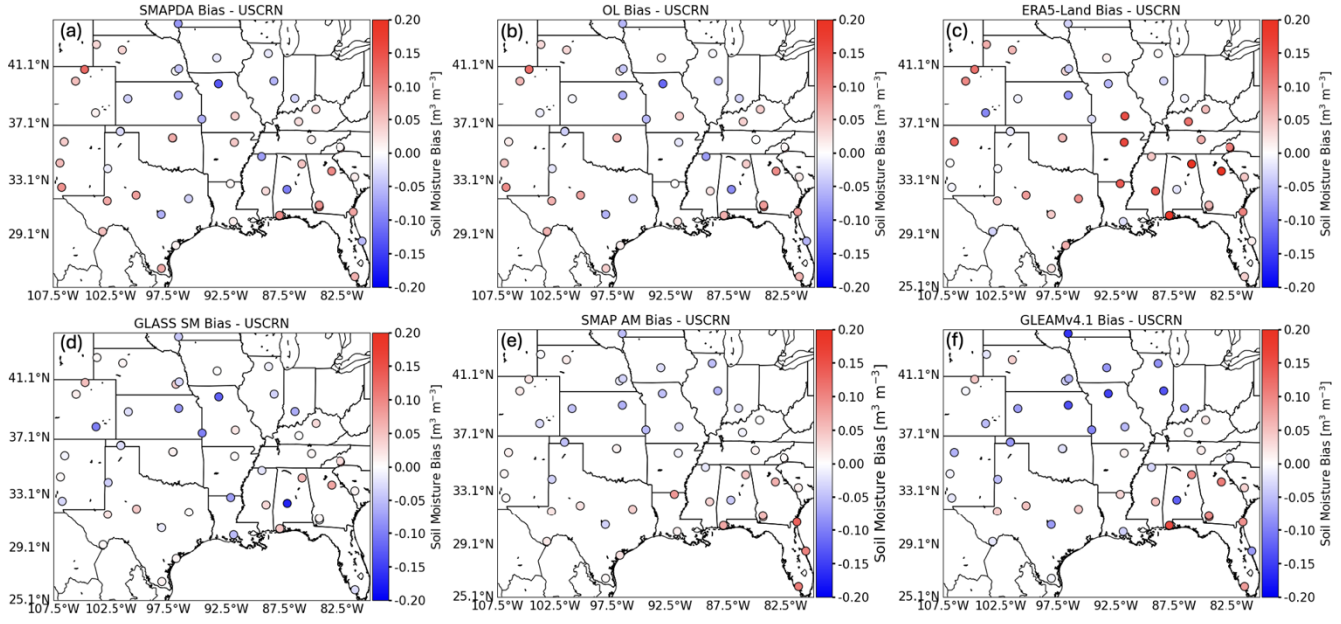

Figure 7. Site-wise bias computed using the USCRN observations. Results for **(a)** SMAPDA, **(b)** OL, **(c)** ERA5-Land, **(d)** GLASS SM, **(e)** SMAP AM, and **(f)** GLEAMv4.1 are illustrated.

SMAPDA demonstrates a unique trend, with the smallest errors before June and peak errors occurring in early July and late November.

4.2.2 Evaluation using SCAN soil moisture observations

Along the same line as discussed in Sect. 4.2.1, we examined the SM 2D histograms as referenced by the SCAN observations (Fig. S2). Overall, similar conclusions can be drawn from the comparisons, implying that the evaluation is robust with very little dependence on selected reference SM observations. For example, SMAPDA further improves ACC on top of OL (Fig. S1a and b), showcasing the positive impact on capturing SM variability in time. Moreover, SMAPDA and GLASS SM remain the top two performers among the SM products in terms of low RMSE (0.089 and

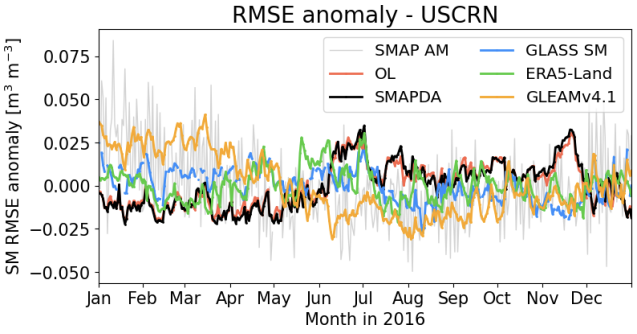

Figure 8. The SM RMSE anomaly time series computed against USCRN observations during 2016. Results of SMAP AM, OL, SMAPDA, ERA5-Land, GLEAMv4.1, and GLASS SM are denoted by colored lines as indicated in the legend.

0.095 $m^3\ m^{-3}$) and bias (0.001 and $-0.006\ m^3\ m^{-3}$). But again, despite better alignment with the diagonal in general, GLASS SM has much smaller ACC than SMAPDA. It also suggests that GLASS SM has more off-diagonal samples than the SMAPDA, likely due in part to the use of ERA5-Land as the initial SM guess for GLASS SM. Since ERA5-Land has relatively scattered samples in the 2D histogram (Fig. S2c) and the ML algorithm does not overfit by design (Zhang et al., 2023), some pixels may receive less correction than others. The estimate from GLEAMv4.1 (Fig. S2f) suffers from generally smaller SM estimates (capped around $\sim 0.38\ m^3\ m^{-3}$), which potentially causes severe underestimation. SMAP AM has the least bias among all estimates (Fig. S2e). However, it also owns many samples far off the diagonal, which lowers the overall skill scores. As a result, the RMSEs are all greater than $0.11\ m^3\ m^{-3}$, which is about 20 % more than what is computed for SMAPDA.

The site-wise ACC and RMSEs given in Figs. S3 and S4 confirm that SMAPDA is the top performer among the SM estimates in general as it shows consistency in producing more realistic temporal evolution and relatively small error across the SCAN sites. Excessive SM errors (positive differences) are found at several sites in the southeastern US when evaluating the ERA5-Land and SMAP AM. Despite small contrast in RMSE differences across the SCAN sites, the estimates from GLEAMv4.1 show extensively larger errors at more locations, leading to a nontrivial mean RMSE of $0.112\ m^3\ m^{-3}$. Figure S5 shows that, regardless of the in situ observations, the bias pattern of each SM estimate resembles those given in Fig. 7, suggesting that the analyzed biases should be rather robust and representative.

Figure S6 shows that the SM RMSE anomaly of SMAPDA is very similar to those of GLASS SM (light blue) and even SMAP AM (light gray) when assessed using the SCAN data. ERA5-Land and GLEAMv4.1 exhibit trends partially different from the other three estimates. Specifically, ERA5-Land (GLEAMv4.1) has relatively smaller (larger) errors than the other three estimates from January to April and tends to produce rather larger (smaller) errors from October to December. Despite the differences in these results and those from the comparison against USCRN (Fig. 8), similarities in RMSE anomalies among the analyses remain clear. This is most likely due to the various SM estimates using duplicate sources of SM data, even though different methods are employed to arrive at the final estimates.

4.2.3 Regional assessment over the Southern Great Plains

The Southern Great Plains (SGP), including Oklahoma, has been recognized as one of the hotspots for strong land–atmosphere coupling (LAC; Santanello et al., 2009; Tao et al., 2019). Earlier studies revealed the key physical processes that modulate the strength of LAC and how LAC influences the life cycle of convective clouds using model and observational datasets generated for this region. For instance, Fast et al. (2018) investigated the impact of SM spatial heterogeneity on simulated convective clouds near the ARM SGP site using a large eddy simulation model for a selected event during the 2016 HI-SCALE field campaign. They found that the scales of SM gradient in the model can significantly affect the presence of simulated cloud populations even with identical atmospheric conditions. Sakaguchi et al. (2022) further analyzed the LES model data produced by Fast et al. (2019) using spectral analysis and demonstrated that the SM spatial heterogeneity may strengthen secondary circulations and extend their spatial scales. Both studies concluded that a more realistic and high-resolution representation of SM is desired to better understand LAC at local to regional scales (~ 1 km and greater). This motivates us to examine how SMAPDA SM estimates perform in this region in comparison to other datasets, and the evaluations are carried out by leveraging highly concentrated observations measured by the OKMet (Fig. 1).

As shown in Fig. 9, the performance of GLASS SM degrades when evaluated against the OKMet data. The 2D histogram shows that most samples occur in the bins above the diagonal, meaning that GLASS SM (Fig. 9d) generally underestimates SM (mean bias: $-0.038\ m^3\ m^{-3}$), whereas when compared with data from USCRN and SCAN (Figs. 4d and S1d), GLASS SM has much better agreement with the observations. This is most likely due to the exclusion of OKMet data in the ML training process. Since ML is purely data-driven, the skill of ML-based SM estimates highly depends on the availability of in situ observations. Conversely, SMAPDA exhibits the lowest annual mean RMSE ($0.087\ m^3\ m^{-3}$) in general in comparison to the other four datasets (0.106, 0.107, 0.104, and $0.112\ m^3\ m^{-3}$ for ERA5-Land, GLASS SM, SMAP_AM, and GLEAMv4.1, respectively). The annual mean RMSE for SMAPDA stays close to the RMSEs obtained when compared to the USCRN and SCAN observations (0.084 and $0.0894\ m^3\ m^{-3}$, respectively). This demonstrates that a physically constrained

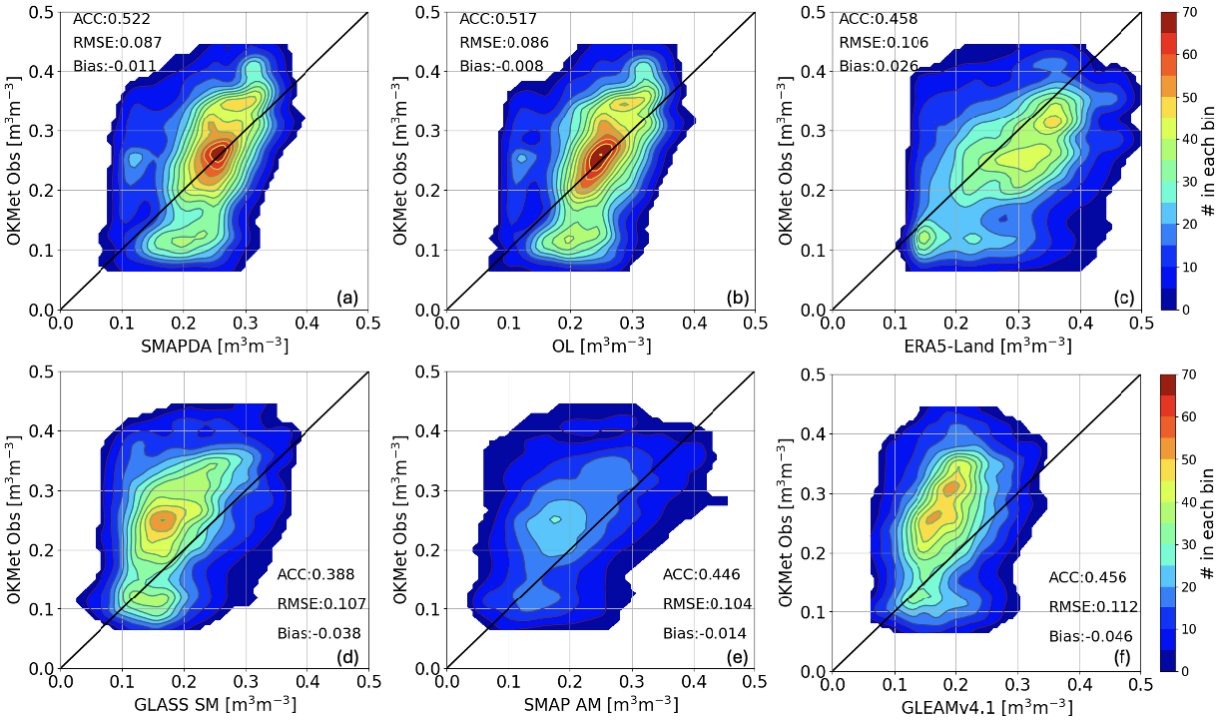

Figure 9. Similar to Figs. 4 and S1, but for results computed against the OKMet observations.

model tends to perform more consistently and mitigate soil moisture biases despite uncertain and neglected/unresolved physical processes in the model.

The relatively dense spatial distribution of the OKMet sites enables further investigation into the realism of estimated SM spatial heterogeneity. We computed the daily standard deviation (SD) across all OKMet sites for each SM estimate as a way to quantify the spatial SM heterogeneity (meaning how spread the SM values are in space). Figure 10 shows that observed SD (magenta) is mostly larger than what is estimated by any of the derived SM approaches over the year despite notable day-to-day variations. Even though SMAPDA and GLASS SM top the others in SM estimates based on the evaluations shown earlier, they both underestimate the SM spatial heterogeneity with an averaged SD of $\sim 0.6\,m^3\,m^{-3}$, which is about 25 % less than observed. GLEAMv4.1 and SMAP AM have even smaller SDs over the period. While ERA5-Land tends to have larger and more comparable variances as observed, it does not accurately distribute those SM values in space (Fig. S7).

In addition to soil moisture, the ARM SGP facility (through instruments of STAMP and ECOR) has collected soil temperature and surface heat flux (latent and sensible) measurements that are critical for land–atmosphere coupling research as these quantities directly modulate the strength of turbulent mixing in the atmospheric boundary layer. Here we primarily assess how SMAPDA represents SM, soil temperature (ST), latent heat flux (LHF), and sensible heat flux

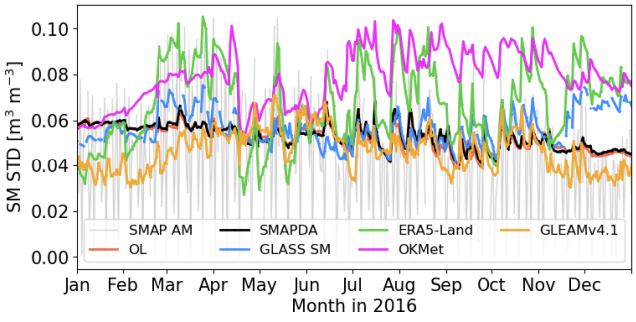

Figure 10. Daily time series of SM standard deviation across OKMet sites computed for each SM estimate.

(SHF) using the concurrent measurements collected across six ARM SGP sites (E31, E33, E37, E38, E39, and E41) as denoted in Fig. 1a. Note SM and ST data from the STAMP are not valid in January.

Results show that SMAPDA reproduces the observed monthly trends in SM and slightly overestimates SM (annual mean model–observation difference of $+\,0.04\,m^3\,m^{-3}$) with larger positive biases in winter months (Fig. 11a). The annual mean ST is warmer in SMAPDA than observed (difference: +0.84 K), which can be attributed to a relatively distinct warm bias in summer months (June–September) (Fig. 11b). While LHF has an annual mean difference of $-12.14\,W\,m^{-2}$ when compared to the observations (Fig. 11c), it is considered minor as annual LSM error can be approximately –

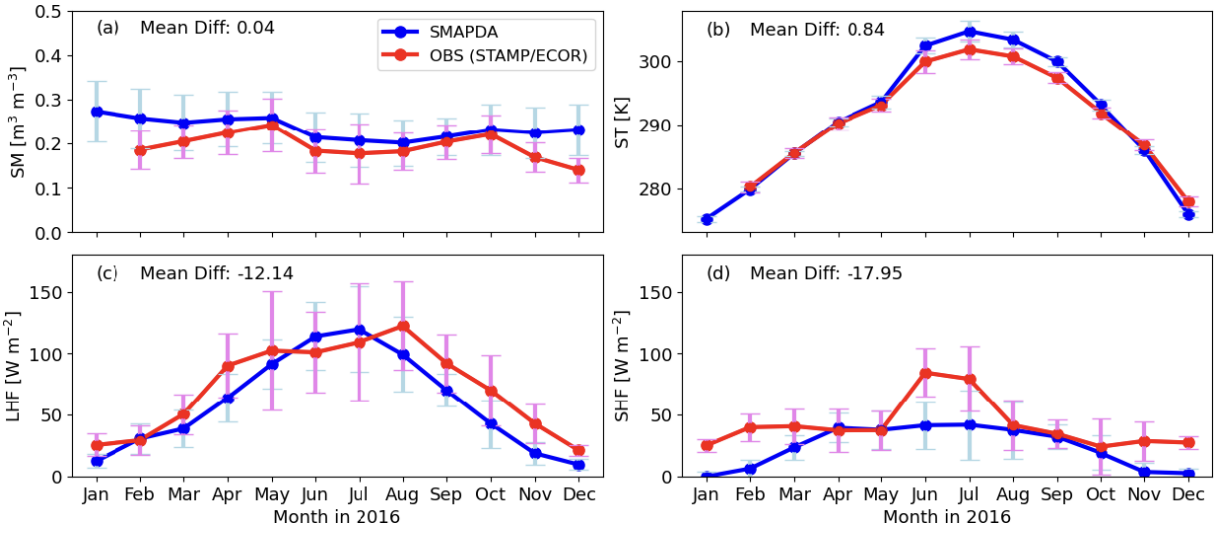

Figure 11. Time series of monthly mean **(a)** SM, **(b)** ST, **(c)** LHF, and **(d)** SHF. The blue (red) line with circles represents results obtained from SMAPDA simulations (ARM SGP observations (STAMP/ECOR)) as computed across six ARM SGP sites (Fig. 1a).

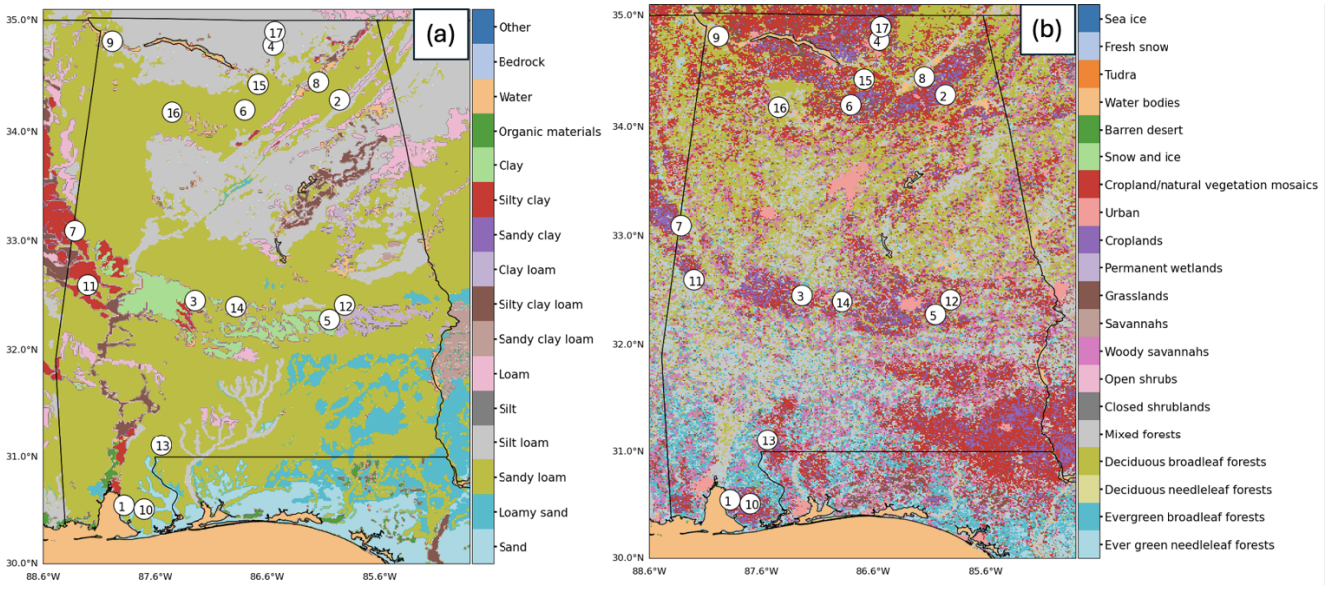

Figure 12. Zoomed-in maps of **(a)** soil and **(b)** land cover types over the AL subdomain as marked in Fig. 1a. The locations of 17 valid observational sites from USCRN and SCAN are denoted by the white circles with numbers overlaid in correspondence to panels in Fig. 13.

20 W m^{-2} based on earlier studies (ARM, 2014). In the case of SHF, SMAPDA tends to underestimate for most of the months (Fig. 11d). This is likely due to a consistent wet bias in SM throughout the year (Fig. 11a), leading to increased energy partitioning in latent heat and a corresponding reduction in sensible heat. Distinct positive biases even appear over summer months (June and July) despite higher simulated ST (Fig. 11b).

4.2.4 Regional analysis associated with 2016 drought in the southeastern US

The southeastern US experienced one of the most significant drought events in the region during the fall of 2016 (peaked in October and November) based on the historical record (Park Williams et al., 2017). It primarily affected parts of Georgia, Alabama, Tennessee, and the Carolinas. The drought reached extreme and exceptional levels, especially in northern Georgia and Alabama, where some areas experienced their driest conditions in history. A combination of factors, including below-average rainfall during the spring and summer months

and unusually high temperatures, led to increased evaporation and reduced soil moisture and thereby the drought conditions in fall. The drought severely impacted agriculture, leading to reduced crop yields, and contributed to widespread wildfires in the Appalachian region. The strained water resources also posed a great challenge to water availability for communities and industries. Hence, we aim to explore the representativeness of the SMAPDA SM estimate under the extreme drought conditions in the southeastern US.

The subdomain "AL", which covers Alabama (as denoted in Fig. 1a), was chosen for conducting the following analyses since the in situ measurements from USCRN and SCAN are relatively denser in Alabama than in other areas in the southeastern US (Fig. 1a). In addition to the spatial variability of soil types (Fig. 12a), the relatively large fraction of forest cover (Fig. 12b) can further complicate how soil moisture is distributed through hydraulic processes such as evapotranspiration (ET), interception, infiltration, runoff, groundwater recharge, and hydraulic redistribution due to the presence of root systems and tree canopies. Here, we selectively examine the relationship between SM and ET under the drought conditions by comparing the SMAPDA output with the GLEAMv4.1 data. GLEAMv4.1 was chosen for this specific comparison as it is the only product among all benchmark datasets that concurrently provides SM and observation-constrained ET representations.

To assess whether the two datasets (SMAPDA and GLEAMv4.1) represent the drought event, we first use the 5-year (2012–2016) soil moisture data from OL simulation to infer the monthly climatological mean and standard deviation at each pixel over the area. Based on the climatological baseline, the standardized soil moisture anomaly (SMA) can then be computed to better quantify the severity of drought conditions. According to Ontel et al. (2021), Jiménez-Donaire et al. (2020), and Tian et al. (2022), when SMA is between 0 and −1, the drought is considered a mild drought. Moderate drought occurs if SMA lies between −1 and −2. Lastly, when SMA falls below −2, a severe drought condition is defined. Both SMAPDA and GLEAMv4.1 suggest that moderate to severe drought conditions occurred in this region during the months of September, October, and November (Fig. 13). Due to the contrast in sample size, SMAPDA demonstrates more spatial heterogeneity in SMA compared to GLEAMv4.1. Otherwise, the result reconfirms the drought period as defined in other relevant studies.

The monthly statistics from both SMAPDA and GLEAMv4.1 are given in Fig. 14. Both estimates exhibit a decreasing trend in SM over the summer months (JJA) as well as a steeper decline in fall (SON) over the AL subdomain (Fig. 14a). Except in winter months (DJF) when the SM estimate is slightly larger in SMAPDA than in GLEAMv4.1, soil conditions produced by SMAPDA are consistently drier than GLEAMv4.1. Both datasets suggest increases in ET before June with similar magnitudes (Fig. 14c), mostly due to the seasonal increase in the solar

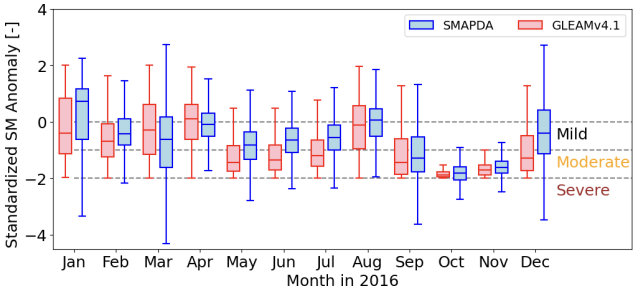

Figure 13. Box-and-whisker plot for monthly standardized SM anomaly (SSA) computed for SMAPDA and GLEAMv4.1 data in 2016 referenced by their own climatology (2012–2016). Dashed lines denote the thresholds for the defined drought conditions.

insolation as well as the leaf area. However, in summer (JJA), SMAPDA produces much larger ET than GLEAMv4.1 does, which leads to much drier soil conditions concurrently. This then facilitates the intensification of drought conditions in the fall, leading to further reduction in water availability through the soil columns, which significantly limits the amount of ET in contrast to GLEAMv4.1.

Although in situ ET observations are not available through the USCRN and SCAN measurements, the SM observations (Fig. 14b) suggest that while SMAPDA overall captures how SM evolves over time, GLEAMv4.1 gives much weaker responses to the SM drying process than SMAPDA. This ultimately produces an overall much larger wet bias in SM for GLEAMv4.1 than for SMAPDA in the fall. Whether data from all grid cells in the AL subdomain (Fig. 14a and c) or only the 17 grid cells nearest to the in situ measurements (Fig. 14b and d) are used, the trends in both SM and ET are very similar. This suggests that the evaluations illustrated in Fig. 14b and d are representative for the subdomain. As we investigate in more detail through comparison at each individual site (Fig. 15), we find that most of the large errors in SMAPDA's SM estimate can be attributed to sites' soil properties (Fig. 15), specifically where clay soil types are present (site #3 and 5: clay; site #7: silty clay). At those sites, the enhanced temporal variability in SM is distinct but underestimated by both SMAPDA and GLEAMv4.1. This suggests that both approaches are unable to capture the drastic changes in SM, likely due in part to the nature of clay soil texture. The conclusion regarding soil-texture-dependent errors seems to hold even when we extend the analysis to all sites (USCRN and SCAN) located in the study domain (Fig. S8). The relatively weaker dependency on land cover types illustrated in Fig. S9 confirms that the errors are much more sensitive to soil textures than land cover types. Overall, it shows that although the sample sizes vary among soil types, the soil moisture error remains relatively higher at sites with clay soil than other soil types. This echoes what was reported in Colliander et al. (2022), stressing that further model

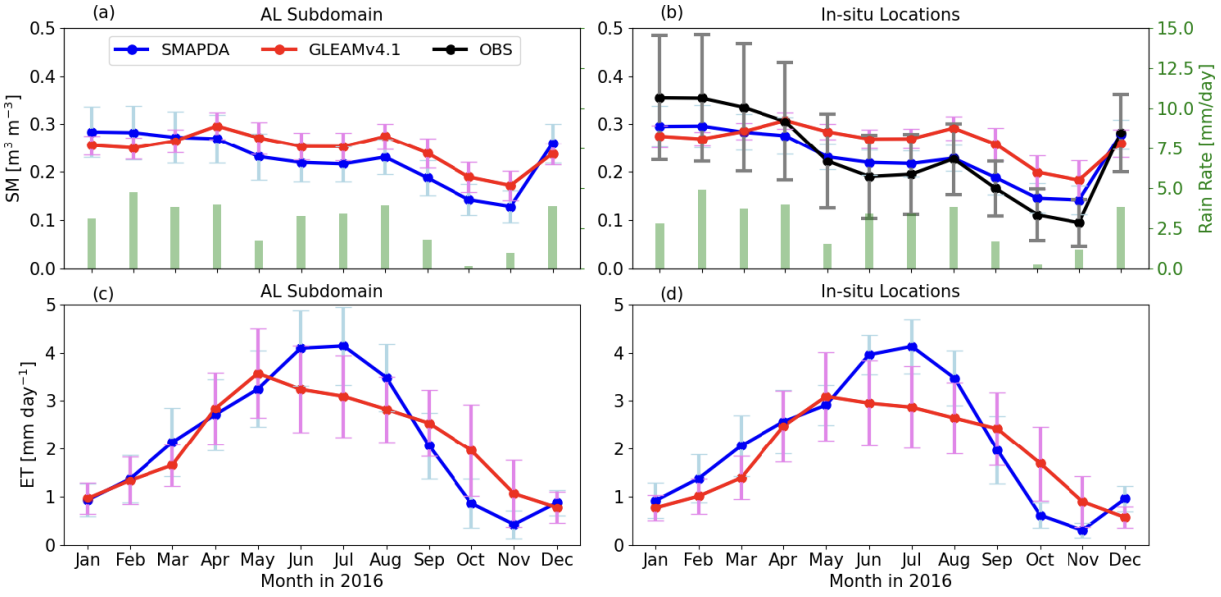

Figure 14. Monthly mean and standard deviation (denoted by error bars) of SM **(a, b)** and ET **(c, d)** over the AL subdomain and among 17 in situ observation locations as denoted in Fig. 12. The mean rain rates represented by green bars in **(a)** and **(b)** are computed from SMAPDA correspondingly.

refinements may be needed to improve treatments in resolving hydraulic processes for the variants of clay soil.

5 Data availability

The dataset generated and analyzed during the current study is available on Zenodo at https://doi.org/10.5281/zenodo.14370563 (Tai et al., 2024). The software package of the NASA Land Information System (LIS) can be downloaded through https://github.com/NASA-LIS/LISF (NASA-LIS, 2025; Kumar et al., 2006; Peters-Lidard et al., 2007). The enhanced SMAP Level 3 soil moisture product is accessible at https://doi.org/10.5067/M20OXIZHY3RJ (SPL3SMP_E; O'Neill et al., 2023). The NLDAS-2 data (Xia et al., 2012) are archived at https://disc.gsfc.nasa.gov/datasets/NLDAS_FORA0125_H_2.0/summary?keywords=NLDAS2.
The Stage IV QPE product (Lin and Mitchelll, 2005) is accessible at https://data.ucar.edu/dataset/ncep-emc-4km-gridded-data-grib-stage-iv-data. The USCRN and SCAN data are acquired from the International Soil Moisture Network (ISMN; https://ismn.earth/en/, Dorigo et al., 2021b). The Oklahoma Mesonet soil moisture observations (OKMet; McPherson et al., 2007) were acquired from the ARM discovery website: https://doi.org/10.5439/1027361 (Giangrande et al., 2020). For all other SM datasets please refer to Table 3. The STAMP (https://doi.org/10.5439/1238260, Kyrouac et al., 2016) and ECOR (https://doi.org/10.5439/1097546, Gaustad, 2003) data were sourced from the Atmospheric Radiation Measurement (ARM) user facility.

6 Summary and discussion

To facilitate an improved representation of local- to regional-scale SM distribution, we generated a high-resolution SM dataset at a 1 km grid spacing by assimilating the 9 km SMAP SM data into the Noah-MP land surface model. The dataset has spatial coverage over the east CONUS and a frequency of 6 h for the entire year of 2016. The SMAP SM data assimilation is accomplished under the framework of NASA's Land Information System using the EnKF algorithm. In the DA simulation, 12 ensemble members were initialized by perturbing the selected variables in meteorological forcing data (NLDAS-2 and Stage IV). The subset of daily SMAP SM overpasses is assimilated hourly when applicable. The generated SM estimate is comprehensively assessed by using the in situ SM observations collected in the networks of USCRN, SCAN, OKMet, and ARM SGP and compared with the performance of other existing SM datasets such as the morning overpass of SMAP (SPL3SMP_E) data, ERA5-Land, GLASS SM, and GLEAMv4.1.

Overall, the evaluation result suggests that the resulting soil moisture estimate through DA, which we refer to as SMAPDA, exhibits the top performance among the examined datasets. It improves SM temporal variability in most of the evaluated sites when compared with the estimate from the OL simulation (experiment without DA). While the SMAPDA and GLASS SM are considered the top two SM estimates based on the skill metrics computed against USCRN and SCAN observations (e.g., CCs are ~ 0.8 and ~ 0.7 and RMSEs are ~ 0.08 and $\sim 0.09\,m^3\,m^{-3}$, respectively), SMAPDA surpasses GLASS SM when vali-

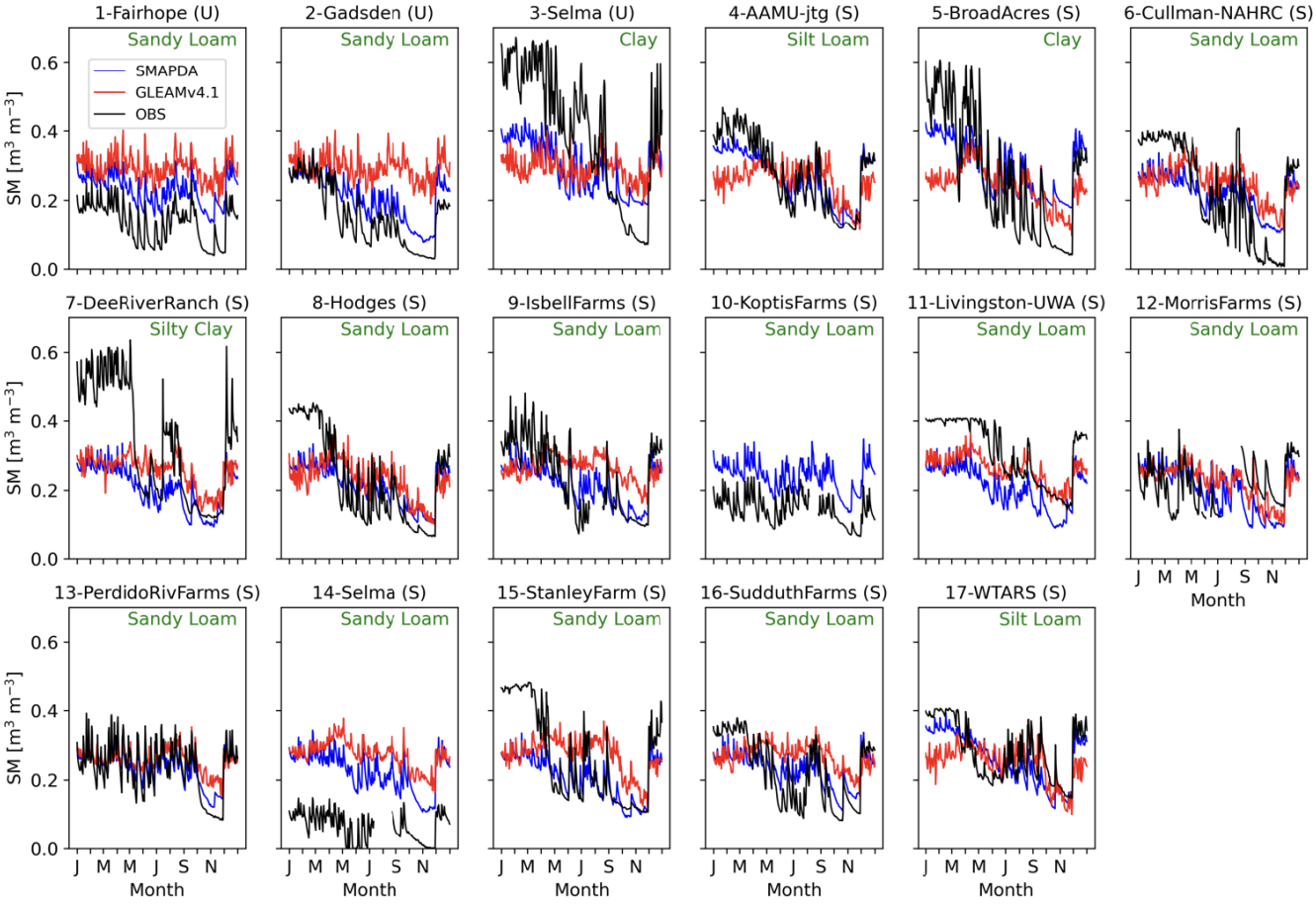

Figure 15. SM daily time series comparison at each in situ observation location. The numbers given in the title above each panel correspond to the locations as marked by the numbered white circles in Fig. 12. The site names and the corresponding observational networks (as indicated by either U (USCRN) or S (SCAN) in parentheses) are readable from the titles. Soil texture type is indicated by green text in the top right corner of each panel.

dated against OKMet data (independent observations for both SMAPDA and GLASS SM). Being a fully data-driven ML product, the GLASS SM achieves a better one-to-one alignment with the observations than SMAPDA when evaluated by the in situ data used in its training process (USCRN and SCAN). However, the relative accuracy of GLASS SM and SMAPDA is reverse when compared with the independent observations from OKMet, which implies the inclusion of physical constraints could be vital for a more consistent performance in SM estimate using the ML approach. From the analysis in anomalous errors, we show similar intrinsic errors among the selected SM datasets in some cases, which is most likely driven by overlapping data sources. Referenced to the OKMet observations, an investigation of the realism of estimated SM spatial heterogeneity indicates that all SM estimates, including the SMAPDA and GLASS SM, persistently underestimate the observed variances ($\sim 25\,\%$ less) across the sites over the study period. While ERA5-Land estimate shows larger and more comparable variances as observed, it does not accurately represent those SM values individually.

In addition to SM data, we showed that SMAPDA data reasonably represent ST and even surface heat fluxes when compared against the observations measured in ARM SGP sites. This suggests that the suite of the SMAPDA dataset is useful in characterizing land–atmosphere interactions. Moreover, it is also analyzed with respect to the response to a drought that occurred over the southeastern US during the fall of 2016. As one of the key components contributing to the drought, the reduction in SM is usually accompanied by increased evaporation in a water-limited scenario, which may potentially amplify and increase wildfire activity as well as stressing agricultural production until new precipitation. We explored the relationships between SM and ET with a focus on Alabama quantitatively, utilizing concurrent GLEAMv4.1 data as the reference. Results indicate that both datasets show declined SM in summer and fall, with SMAPDA consistently displaying drier soil conditions compared to GLEAMv4.1. ET trends from both datasets were relatively close until June but diverged in summer, with SMAPDA estimating higher ET, exacerbating the drought conditions. Data also high-

lighted that model discrepancies, particularly in clay-rich soils, suggest the need for refined treatments of hydraulic processes in models for accurate SM estimates.

A few uncertainties in our analysis are worth noting. For example, the evaluation result is most likely dependent on the data resolutions. Coarse-resolution SM estimates such as ERA5-Land, GLEAMv4.1, and SMAP AM suffer from insufficient representativeness of subgrid SM variability, underscoring the necessity of high resolution to better characterize highly heterogeneous SM distributions. The mismatch of spatial resolutions among different data sources likely introduces uncertainty in their intercomparison. More sophisticated upscale/interpolation algorithms may be employed to further mitigate the issue (Crow et al., 2012; Gruber et al., 2020; Quiring et al., 2016); however, as these methods introduce their own uncertainties, we opted not to use them for our evaluations. In addition, unresolved natural and anthropogenic processes such as surface and subsurface lateral flow (e.g., Yang et al., 2021), root water uptake and redistribution (e.g., Zeng, 2001), dynamic groundwater water table and capillary rise (e.g., Miguez-Macho and Fan, 2012), and irrigation (e.g., Yang et al., 2020) can potentially shift the SM estimates under various conditions. Along this line, the estimates in root zone SM would be worth validating to further constrain the overall performance. While our SM dataset encompassing much of the eastern CONUS is restricted to a 1-year period (2016), our results demonstrate a promising approach that can be applied to any local domain of interest with potentially longer analysis periods. This dataset could be used as lower boundary conditions to drive other meteorological model experiments that investigate the impact of land–atmosphere coupling on boundary layer properties and clouds. Lastly, there are many more ML algorithms, such as neural networks, random forests, and support vector machines, that have been applied to enhance the spatial and temporal resolution of soil moisture datasets and further improve accuracy in data-sparse regions (e.g., O and Orth, 2021; Han et al., 2023; Lei et al., 2022; Zhang et al., 2023). However, such approaches lack the inherent physical constraints of a data assimilation approach. Future studies may include more ML-based products in the assessment and discuss the impacts of physical constraints on estimated SM as suggested in this study.

Supplement. The supplement related to this article is available online at https://doi.org/10.5194/essd-17-4587-2025-supplement.

Author contributions. ST designed the model experiments, performed the simulations, and contributed to investigation, data curation, and visualization. ST, ZY, BG, KS, LB, CK, YQ, YL, and JF contributed to the scientific discussions. ST drafted the original paper, and all co-authors contributed to the revisions. JF, LB, and CK were responsible for conceptualization, supervision, and funding acquisition.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. This research was supported by the Atmospheric Science Research (ASR) program as part of the DOE Office of Biological and Environmental Research. Pacific Northwest National Laboratory is operated for the DOE by the Battelle Memorial Institute under contract DE-AC05-76RL01830. The modeling computation was performed on PNNL's Research Computing clusters. The authors would like to acknowledge all the technical support through the LIS GitHub (https://github.com/NASA-LIS/LISF/discussions, last access: 8 September 2025), particularly the responsible research scientist David Mocko.

Financial support. This research has been supported by Battelle (grant no. DE-AC05-76RL01830).

Review statement. This paper was edited by Yun Yang and reviewed by two anonymous referees.

References

Ahmad, J. A., Forman, B. A., and Kumar, S. V.: Soil moisture estimation in South Asia via assimilation of SMAP retrievals, Hydrol. Earth Syst. Sci., 26, 2221–2243, https://doi.org/10.5194/hess-26-2221-2022, 2022.

Arsenault, K. R., Kumar, S. V., Geiger, J. V., Wang, S., Kemp, E., Mocko, D. M., Beaudoing, H. K., Getirana, A., Navari, M., Li, B., Jacob, J., Wegiel, J., and Peters-Lidard, C. D.: The Land surface Data Toolkit (LDT v7.2) – a data fusion environment for land data assimilation systems, Geosci. Model Dev., 11, 3605–3621, https://doi.org/10.5194/gmd-11-3605-2018, 2018.

Ball, J. T., Woodrow, I. E., and Berry, J. A.: A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions, Prog. Photosynth. Res., 4, 221–224, https://doi.org/10.1007/978-94-017-0519-6_48, 1987.

Betts, A. K.: Surface diurnal cycle and boundary layer structure over Rondônia during the rainy season, J. Geophys. Res., 107, https://doi.org/10.1029/2001JD000356, 2002.

Brocca, L., Albergel, C., Balenzano, A., Barbagli, R., de Rosnay, P., Dorigo, W., Wagner, W., and Tarpanelli, A.: Exploring the actual spatial resolution of 1 km satellite

soil moisture products, Sci. Total Environ., 945, 174087, https://doi.org/10.1016/j.scitotenv.2024.174087, 2024.

Brutsaert, W.: Evaporation into the Atmosphere: Theory, History, and Applications, Springer, Dordrecht, https://doi.org/10.1007/978-94-017-1497-6, 1982.

Chakraborty, A., Saharia, M., Chakma, S., Kumar Pandey, D., Niranjan Kumar, K., Thakur, P. K., Kumar, S., and Getirana, A.: Improved soil moisture estimation and detection of irrigation signal by incorporating SMAP soil moisture into the Indian Land Data Assimilation System (ILDAS), J. Hydrol., 638, 131581, https://doi.org/10.1016/j.jhydrol.2024.131581, 2024.

Chao, L., Zhang, K., Wang, S., Gu, Z., Xu, J., and Bao, H.: Assimilation of surface soil moisture jointly retrieved by multiple microwave satellites into the WRFHydro model in ungauged regions: towards a robust flood simulation and forecasting, Environ. Modell. Softw., 154, 105421, https://doi.org/10.1016/j.envsoft.2022.105421, 2022.

Colliander, A., Jackson, T. J., Bindlish, R., Chan, S., Das, N. N., Kim, S. B., Cosh, M. H., and Njoku, E. G.: Validation of SMAP surface soil moisture products with core validation sites, Remote Sens. Environ., 191, 215–231, https://doi.org/10.1016/j.rse.2017.01.021, 2017.

Colliander, A., Kim, S. B., Chan, S., Das, N. N., Crow, W., Cosh, M. H., and Njoku, E. G.: Validation of soil moisture data products from the NASA SMAP Mission, IEEE J. Sel. Top. Appl.., 15, 364–392, https://doi.org/10.1109/JSTARS.2021.3124743, 2022.

Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., and Duan, Q.: Land surface model spin-up behavior in the North American Land Data Assimilation System (NLDAS), J. Geophys. Res.-Atmos., 108, https://doi.org/10.1029/2002JD003316, 2003.

Crow, W. T. and Wood, E. F.: The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using Ensemble Kalman filtering: a case study based on ESTAR measurements during SGP97, Adv. Water Resour., 26, 137–149, https://doi.org/10.1016/S0309-1708(02)00088-X, 2003.

Crow, W. T., Chen, F., Reichle, R. H., and Bindlish, R.: Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products, Rev. Geophys., 50, https://doi.org/10.1029/2011RG000372, 2012.

Diamond, H. J., Karl, T. R., Palecki, M. A., Baker, C. B., Bell, J. E., Leeper, R. D., Easterling, D. R., Lawrimore, J. H., Meyers, T. P., Helfert, M. R., Goodge, G., and Thorne, P. W.: U.S. Climate Reference Network after one decade of operations: Status and assessment, B. Am. Meteorol. Soc., https://doi.org/10.1175/BAMS-D-12-00170.1, 2013.

Dirmeyer, P. A., Halder, S., Bombardi, R., Norton, H. E., Fang, Y., and Quiring, S. M.: Confronting weather and climate models with observational data from soil moisture networks over the United States, J. Hydrometeorol., https://doi.org/10.1175/JHM-D-15-0196.1, 2016.

Dorigo, W., Gruber, A., De Jeu, R., Wagner, W., Stacke, T., Loew, A., Albergel, C., Brocca, L., Chung, D., Parinussa, R., and Kidd, R.: ESA CCI Soil Moisture for improved Earth system understanding: State-of-the-art and future directions, Remote Sens. Environ., 203, 185–215, https://doi.org/10.1016/j.rse.2017.07.001, 2017.

Dorigo, W., Himmelbauer, I., Aberer, D., Schremmer, L., Petrakovic, I., Zappa, L., Preimesberger, W., Xaver, A., Annor, F., Ardö, J., Baldocchi, D., Bitelli, M., Blöschl, G., Bogena, H., Brocca, L., Calvet, J.-C., Camarero, J. J., Capello, G., Choi, M., Cosh, M. C., van de Giesen, N., Hajdu, I., Ikonen, J., Jensen, K. H., Kanniah, K. D., de Kat, I., Kirchengast, G., Kumar Rai, P., Kyrouac, J., Larson, K., Liu, S., Loew, A., Moghaddam, M., Martínez Fernández, J., Mattar Bader, C., Morbidelli, R., Musial, J. P., Osenga, E., Palecki, M. A., Pellarin, T., Petropoulos, G. P., Pfeil, I., Powers, J., Robock, A., Rüdiger, C., Rummel, U., Strobel, M., Su, Z., Sullivan, R., Tagesson, T., Varlagin, A., Vreugdenhil, M., Walker, J., Wen, J., Wenger, F., Wigneron, J. P., Woods, M., Yang, K., Zeng, Y., Zhang, X., Zreda, M., Dietrich, S., Gruber, A., van Oevelen, P., Wagner, W., Scipal, K., Drusch, M., and Sabia, R.: The International Soil Moisture Network: serving Earth system science for over a decade, Hydrol. Earth Syst. Sci., 25, 5749–5804, https://doi.org/10.5194/hess-25-5749-2021, 2021a.

Chen, Y., Feng, X., and Fu, B.: An improved global remote-sensing-based surface soil moisture (RSSSM) dataset covering 2003–2018, Earth Syst. Sci. Data, 13, 1–31, https://doi.org/10.5194/essd-13-1-2021, 2021b (data available at: https://ismn.earth/en/, last access: 6 May 2025).

Draper, C. S., Reichle, R. H., De Lannoy, G. J. M., and Liu, Q.: Assimilation of passive and active microwave soil moisture retrievals, Geophys. Res. Lett., 39, https://doi.org/10.1029/2011GL050655, 2012.

Ek, M. B. and Holtslag, A. A. M.: Influence of soil moisture on boundary layer cloud development, J. Hydrometeorol., https://doi.org/10.1175/1525-7541(2004)005<0086:IOSMOB>2.0.CO;2, 2004.

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., and Tarpley, J. D.: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, J. Geophys. Res.-Atmos., 108, https://doi.org/10.1029/2002JD003296, 2003.

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., Entin, J. K., Goodman, S. D., Jackson, T. J., Johnson, J., Kimball, J., Piepmeier, J. R., Koster, R., Martin, N., McDonald, K., Moghaddam, M., Moran, S., Reichle, R., Shi, J. C., Spencer, M. W., Thurman, S. W., Tsang, L., and Van Zyl, J.: The Soil Moisture Active Passive (SMAP) Mission, Proc. IEEE, 98, 704–716, https://doi.org/10.1109/JPROC.2010.2043918, 2010.

Fan, D., Liu, Y. Y., Ye, N., Peng, J., Su, C.-H., Long, D., and Jiang, L.: A Sentinel-1 SAR-based global 1 km resolution soil moisture data product: Algorithm and preliminary assessment, Remote Sens. Environ., 318, 114579, https://doi.org/10.1016/j.rse.2024.114579, 2025.

Fang, B., Lakshmi, V., Cosh, M., Liu, P.-W., Bindlish, R., and Jackson, T. J.: A global 1 km downscaled SMAP soil moisture product based on thermal inertia theory, Vadose Zone J., 21, e20182, https://doi.org/10.1002/vzj2.20182, 2022.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., and Alsdorf, D.: The Shuttle Radar Topography Mission, Rev. Geophys., 45, https://doi.org/10.1029/2005RG000183, 2007.

Fast, J. D., Berg, L. K., Zhang, Y., Schmid, B., Shilling, J. E., Springston, S. R., Tomlinson, J. M., Pekour, M. S., Dubey,

M. K., Parworth, C., Mei, F., Watson, T. B., Goldberger, L., Aiken, A. C., and Shippert, T.: Overview of the HI-SCALE Field Campaign: A new perspective on shallow convective clouds, B. Am. Meteorol. Soc., https://doi.org/10.1175/BAMS-D-18-0030.1, 2018.

Fast, J. D., Berg, L. K., Newsom, R. K., Shaw, W. J., Xin, Y., McFarlane, S. A., Gustafson, W. I. Jr., Hardesty, J., Hutchinson, K., Larson, V. E., Smith, J. N., Turner, D. D., Ward, M. D., and Nelson, D. D.: Large-eddy simulation of urban boundary layer turbulence and scalar transport under realistic and homogeneous surface conditions, J. Adv. Model. Earth Sy., 11, 1933–1955, https://doi.org/10.1029/2018MS001504, 2019.

Felfelani, F., Pokhrel, Y., Guan, K., and Lawrence, D. M.: Utilizing SMAP soil moisture data to constrain irrigation in the Community Land Model, Geophys. Res. Lett., 45, 12892–12902, https://doi.org/10.1029/2018GL080870, 2018.

Filippucci, P., Brocca, L., Quast, R., Ciabatta, L., Saltalippi, C., Wagner, W., and Tarpanelli, A.: High-resolution (1 km) satellite rainfall estimation from SM2RAIN applied to Sentinel-1: Po River basin as a case study, Hydrol. Earth Syst. Sci., 26, 2481–2497, https://doi.org/10.5194/hess-26-2481-2022, 2022.

Foucras, M., Zribi, M., Albergel, C., Baghdadi, N., Calvet, J.-C., and Pellarin, T.: Estimating 500-m resolution soil moisture using Sentinel-1 and optical data synergy, Water, 12, 866, https://doi.org/10.3390/w12030866, 2020.

Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., Woodcock, C. E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., and Schaaf, C.: Global land cover mapping from MODIS: Algorithms and early results, Remote Sens. Environ., 83, 287–302, https://doi.org/10.1016/S0034-4257(02)00078-0, 2002.

Gao, Q., Zribi, M., Escorihuela, M. J., and Baghdadi, N.: Synergetic use of Sentinel-1 and Sentinel-2 data for soil moisture mapping at 100 m resolution, Sensors, 17, 1966, https://doi.org/10.3390/s17091966, 2017.

Gaustad, K.: Eddy Correlation Flux Measurement System (ECOR), U.S. Department of Energy Atmospheric Radiation Measurement (ARM) Climate Research Facility [data set], https://doi.org/10.5439/1097546, 2003.

Giangrande, S., Wang, D., and Gregory, L.: Oklahoma Mesonet Soil Moisture (OKMSOIL), 1998-01-01 to 2020-10-22, Southern Great Plains (SGP) External Data (satellites and others) (X1), Atmospheric Radiation Measurement (ARM) user facility, ARM Data Center [data set], https://doi.org/10.5439/1027361, 2020.

Gruber, A., Scanlon, T., Dorigo, W., and Wagner, W.: Validation practices for satellite soil moisture retrievals: What are (the) errors?, Remote Sens. Environ., 244, 111806, https://doi.org/10.1016/j.rse.2020.111806, 2020.

Gutman, G. and Ignatov, A.: The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models, Int. J. Remote Sens., 19, 1533–1543, https://doi.org/10.1080/014311698215333, 1998.

Han, Q., Zeng, Y., Zhang, L., Wang, C., Prikaziuk, E., Niu, Z., and Su, B.: Global long-term daily 1 km surface soil moisture dataset with physics-informed machine learning, Sci. Data, 10, 101, https://doi.org/10.1038/s41597-023-02011-7, 2023.

Hawdon, A., McJannet, D., and Wallace, J.: Calibration and correction procedures for cosmic-ray neutron soil moisture probes located across Australia, Water Resour. Res., 50, 5029–5043, https://doi.org/10.1002/2013WR015138, 2014.

He, C., Niu, G.-Y., Yang, Z.-L., Ek, M., Meng, C., Xia, Y., Zhou, T., Zhang, X., Fan, Y., and Su, H.: Enhancing the Community Noah-MP Land Model Capabilities for Earth Sciences and Applications, B. Am. Meteorol. Soc., https://doi.org/10.1175/BAMS-D-23-0249.1, 2023.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Fleming, S., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Q. J. Roy. Meteor. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hsu, H. and Dirmeyer, P. A.: Soil moisture–evaporation coupling shifts into new gears under increasing CO_2, Nat. Commun., 14, 1162, https://doi.org/10.1038/s41467-023-36794-5, 2023.

Jiménez-Donaire, M. P., Tarquis, A., and Giráldez, J. V.: Evaluation of a combined drought indicator and its potential for agricultural drought prediction in southern Spain, Nat. Hazards Earth Syst. Sci., 20, 21–33, https://doi.org/10.5194/nhess-20-21-2020, 2020.

Johnson, J. M., Frame, J. M., Ceraolo, S., Wood, A. W., Arsenault, K. R., Cosgrove, B., and Sampson, K.: Comprehensive analysis of the NOAA National Water Model: A call for heterogeneous formulations and diagnostic model selection, J. Geophys. Res.-Atmos., 128, e2023JD038534, https://doi.org/10.1029/2023JD038534, 2023.

Jordan, R.: A one-dimensional temperature model for a snow cover: Technical documentation for SNTHERM.89, U.S. Army Cold Regions Research and Engineering Laboratory, CRREL Report 91-16, 1991.

Katul, G. G., Oren, R., Manzoni, S., Higgins, C., and Parlange, M. B.: Evapotranspiration: A process driving mass transport and energy exchange in the soil–plant–atmosphere–climate system, Rev. Geophys., 50, https://doi.org/10.1029/2011RG000366, 2012.

Kerr, Y. H., Waldteufel, P., Wigneron, J.-P., Martinuzzi, J., Font, J., and Berger, M.: Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission, IEEE T. Geosci. Remote, 39, 1729–1735, https://doi.org/10.1109/36.942551, 2001.

Koster, R. D., Suarez, M., Ducharne, A., Stieglitz, M., and Kumar, P.: Regions of strong coupling between soil moisture and precipitation, Science, 305, 1138–1140, https://doi.org/10.1126/science.1100217, 2004.

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., and Wood, E. F.: 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, 2006.

Kumar, S. V., Reichle, R. H., Peters-Lidard, C. D., Koster, R. D., Zhan, X., Crow, W. T., Eylander, J. B., and Houser, P. R.: A land surface data assimilation framework using the Land Information System: Description and applications, Adv. Water Resour., 31,

1419–1432, https://doi.org/10.1016/j.advwatres.2008.01.013, 2008.

Kumar, S. V., Reichle, R. H., Koster, R. D., Crow, W. T., and Peters-Lidard, C. D.: Role of subsurface physics in the assimilation of surface soil moisture observations, J. Hydrometeorol., https://doi.org/10.1175/2009JHM1134.1, 2009.

Kyrouac, J., Cook, D., and Sisterson, D.: Soil Temperature and Moisture Profile (STAMP), Atmospheric Radiation Measurement (ARM) user facility, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA [data set], https://doi.org/10.5439/1238260, 2016.

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G. B., Collier, N., Ghimire, B., van Kampenhout, L., Kennedy, D., Kluzek, E., Knox, R. G., Lawrence, P. J., Li, H., Li, F., Lombardozzi, D., Lu, Y., Pandey, A., Lawrence, P. J., Ricciuto, D. M., Sacks, W. J., Shi, M., Wieder, W. R., Xu, C., Ali, A. A., Badger, A. M., Bisht, G., Broxton, P., Brunke, M. A., Burns, S. P., Buzan, J., Clark, M., Craig, A., Dahlin, K., Drewniak, B., Fisher, J. B., Flanner, M. G., Fox, A. M., Gentine, P., Hoffman, F., Keppel-Aleks, G., Knox, N. M., Kumar, S., Lenaerts, J. T. M., Lienert, S., Lipscomb, W. H., Pelletier, J. D., Perket, J., Randerson, J. T., Rollins, A. W., Shafer, S. L., Shuman, J. K., Tang, J., Tian, B., Touré-Niang, A. T., Vitt, F., Wong, A., Morrison, H., and Zeng, X.: The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty, J. Adv. Model. Earth Sy., 11, 4245–4287, https://doi.org/10.1029/2018MS001583, 2019.

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

Lei, F., Senyurek, V., Kurum, M., Gurbuz, A. C., Boyd, D., Moorhead, R., Crow, W. T., and Eroglu, O.: Quasi-global machine learning-based soil moisture estimates at high spatio-temporal scales using CYGNSS and SMAP observations, Remote Sens. Environ., 276, 113041, https://doi.org/10.1016/j.rse.2022.113041, 2022.

Liang, S., Zheng, T., Liu, R., Fang, H., Tsay, S.-C., and Running, S.: The Global Land Surface Satellite (GLASS) product suite, B. Am. Meteorol. Soc., https://doi.org/10.1175/BAMS-D-18-0341.1, 2021.

Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, J. Geophys. Res.-Atmos., 99, 14415–14428, https://doi.org/10.1029/94JD00483, 1994.

Lin, Y., and Mitchell, K. E.: The NCEP Stage II/IV hourly precipitation analyses: development and applications, in: Proc. 19th Conf. Hydrology, American Meteorological Society, San Diego, CA, 9–13 January 2005, https://ams.confex.com/ams/pdfpapers/83847.pdf (last access: 8 September 2025), 2005 (data available at: https://data.ucar.edu/dataset/ncep-emc-4km-gridded-data-grib-stage-iv-data(last access: 8 September 2025).

Liu, Y. Y., Parinussa, R. M., Dorigo, W. A., De Jeu, R. A. M., Wagner, W., van Dijk, A. I. J. M., McCabe, M. F., and Evans, J. P.: Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals, Hydrol. Earth Syst. Sci., 15, 425–436, https://doi.org/10.5194/hess-15-425-2011, 2011.

Liu, P.-W., Jackson, T. J., Bindlish, R., Cosh, M. H., McKee, L., and Slater, A. G.: Assessing disaggregated SMAP soil moisture products in the United States, IEEE J. Sel. Top. Appl., 14, 2577–2592, https://doi.org/10.1109/JSTARS.2021.3056001, 2021.

Ma, N., Niu, G.-Y., Xia, Y., Cai, X., Zhang, Y., Ma, Y., and Fang, Y.: A systematic evaluation of Noah-MP in simulating land-atmosphere energy, water, and carbon exchanges over the continental United States, J. Geophys. Res.-Atmos., 122, 12245–12268, https://doi.org/10.1002/2017JD027597, 2017.

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A., and Verhoest, N. E. C.: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, Geosci. Model Dev., 10, 1903–1925, https://doi.org/10.5194/gmd-10-1903-2017, 2017.

McPherson, R. A., Fiebrich, C. A., Crawford, K. C., Kilby, J. R., Grimsley, D. L., Martinez, J. E., Basara, J. B., Illston, B. G., Morris, D. A., and Wolfinbarger, J. M.: Statewide monitoring of the mesoscale environment: A technical update on the Oklahoma Mesonet, J. Atmos. Ocean. Tech., 24, 301–321, https://doi.org/10.1175/JTECH1976.1, 2007.

Meyer, R., Zhang, W., Kragh, S. J., Andreasen, M., Jensen, K. H., Fensholt, R., Stisen, S., and Looms, M. C.: Exploring the combined use of SMAP and Sentinel-1 data for downscaling soil moisture beyond the 1 km scale, Hydrol. Earth Syst. Sci., 26, 3337–3357, https://doi.org/10.5194/hess-26-3337-2022, 2022.

Miguez-Macho, G. and Fan, Y.: The role of groundwater in the Amazon water cycle: 1. Influence on seasonal streamflow, flooding and wetlands, J. Geophys. Res.-Atmos., 117, https://doi.org/10.1029/2012JD017539, 2012.

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, Hydrol. Earth Syst. Sci., 15, 453–469, https://doi.org/10.5194/hess-15-453-2011, 2011.

Miralles, D. G., Martens, B., Lievens, H., van der Schalie, R., Jiménez, C., Dorigo, W. A., and Verhoest, N. E. C.: GLEAM4: global land evaporation and soil moisture dataset at 0.1° resolution from 1980 to near present, Sci. Data, 12, 416, https://doi.org/10.1038/s41597-025-04610-y, 2025.

Mousa, B. G. and Shu, H.: Spatial evaluation and assimilation of SMAP, SMOS, and ASCAT satellite soil moisture products over Africa using statistical techniques, Earth Space Sci., 7, e2019EA000841, https://doi.org/10.1029/2019EA000841, 2020.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth Syst. Sci. Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

NASA-LIS: LISF, https://github.com/NASA-LIS/LISF, GitHub [code], last access: 15 September 2025.

Niu, G.-Y., Yang, Z.-L., Dickinson, R. E., Gulden, L. E., and Su, H.: Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data, J. Geophys. Res.-Atmos., 112, https://doi.org/10.1029/2006JD007522, 2007.

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi,

D., Rosero, E., Tewari, M., and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, J. Geophys. Res.-Atmos., 116, https://doi.org/10.1029/2010JD015139, 2011.

Njoku, E. G., Jackson, T. J., Lakshmi, V., Chan, T. K., and Nghiem, S. V.: Soil moisture retrieval from AMSR-E, IEEE T. Geosci. Remote, 41, 215–229, https://doi.org/10.1109/TGRS.2002.808243, 2003.

O, S. and Orth, R.: Global soil moisture data derived through machine learning trained with in-situ measurements, Sci. Data, 8, 170, https://doi.org/10.1038/s41597-021-00964-1, 2021.

O'Neill, P., Chan, S., Njoku, E., Jackson, T., and Bindlish, R.: Soil Moisture Active Passive (SMAP) Algorithm Theoretical Basis Document: Level 2 & 3 Soil Moisture (Passive) Data Products, Jet Propulsion Laboratory, California Institute of Technology, Tech. Rep., https://smap.jpl.nasa.gov/system/internal_resources/details/original/489_ATBD_L2-3_SM_P_v5.pdf (last access: 8 September 2025), 2014.

O'Neill, P. E., Chan, S. K., Njoku, E. G., Jackson, T. J., Bindlish, R., Chaubell, J., and Colliander, A.: SMAP Enhanced L3 Radiometer Global and Polar Grid Daily 9 km EASE-Grid Soil Moisture (SPL3SMP_E), Version 6, NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, Colorado, USA [data set], https://doi.org/10.5067/M20OXIZHY3RJ, 2023.

Ontel, I., Irimescu, A., Boldeanu, G., Mihailescu, D., Angearu, C.-V., Nertan, A., Craciunescu, V., and Negreanu, S.: Assessment of soil moisture anomaly sensitivity to detect drought spatio-temporal variability in Romania, Sensors, 21, 8371, https://doi.org/10.3390/s21248371, 2021.

Park Williams, A., Cook, B. I., Smerdon, J. E., Bishop, D. A., Seager, R., and Mankin, J. S.: The 2016 Southeastern U.S. drought: An extreme departure from centennial wetting and cooling, J. Geophys. Res.-Atmos., 122, 10888–10905, https://doi.org/10.1002/2017JD027523, 2017.

Peters-Lidard, C. D., Houser, P. R., Tian, Y., Kumar, S. V., Geiger, J., Olden, S., Lighty, L., Eastman, J. L., Dirmeyer, P., Adams, J., Mitchell, K., and Wood, E. F.: High-performance Earth system modeling with NASA/GSFC's Land Information System, Innov. Syst. Softw. Eng., 3, 157–165, https://doi.org/10.1007/s11334-007-0028-x, 2007.

Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., and Rossiter, D.: SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty, SOIL, 7, 217–240, https://doi.org/10.5194/soil-7-217-2021, 2021.

Quiring, S. M., Ford, T. W., Wang, J. K., Khong, A., Harris, E., Lindgren, T., Goldberg, D. W., and Li, Z.: The North American Soil Moisture Database: Development and applications, B. Am. Meteorol. Soc., https://doi.org/10.1175/BAMS-D-13-00263.1, 2016.

Rasheed, M. W., Mahmood, S. A., Hafeez, M., Rehman, T., and Ashfaq, A.: Soil moisture measuring techniques and factors affecting the moisture dynamics: A comprehensive review, Sustainability, 14, 11538, https://doi.org/10.3390/su141811538, 2022.

Reichle, R. H. and Koster, R. D.: Bias reduction in short records of satellite soil moisture, Geophys. Res. Lett., 31, https://doi.org/10.1029/2004GL020938, 2004.

Reichle, R. H., McLaughlin, D. B., and Entekhabi, D.: 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, 2002.

Robinson, D. A. and Kukla, G.: Maximum surface albedo of seasonally snow-covered lands in the Northern Hemisphere, J. Clim. Appl. Meteorol., 24, 402–411, https://doi.org/10.1175/1520-0450(1985)024<0402:MSAOSS>2.0.CO;2, 1985.

Robock, A., Vinnikov, K. Y., Srinivasan, G., Entin, J. K., Hollinger, S. E., Speranskaya, N. A., Liu, S., and Namkhai, A.: The Global Soil Moisture Data Bank, B. Am. Meteorol. Soc., 81, 1281–1299, https://doi.org/10.1175/1520-0477(2000)081<1281:TGSMDB>2.3.CO;2, 2000.

Rodell, M., Houser, P. R., Berg, A. A., and Famiglietti, J. S.: Evaluation of 10 methods for initializing a land surface model, J. Hydrometeorol., 6, 146–155, https://doi.org/10.1175/JHM414.1, 2005.

Rouf, T., Girotto, M., Houser, P., and Maggioni, V.: Assimilating satellite-based soil moisture observations in a land surface model: The effect of spatial resolution, J. Hydrol. X, 13, 100105, https://doi.org/10.1016/j.hydroa.2021.100105, 2021.

Sakaguchi, K., Yang, Z., Qian, Y., Berg, L. K., Lin, W., and Fast, J. D.: Determining spatial scales of soil moisture–cloud coupling pathways using semi-idealized simulations, J. Geophys. Res.-Atmos., 127, e2021JD035282, https://doi.org/10.1029/2021JD035282, 2022.

Santanello, J. A., Peters-Lidard, C. D., Kumar, S. V., Alonge, C., and Tao, W.-K.: A modeling and observational framework for diagnosing local land–atmosphere coupling on diurnal time scales, J. Hydrometeorol., 10, 577–599, https://doi.org/10.1175/2009JHM1066.1, 2009.

Schaefer, G. L., Cosh, M. H., and Jackson, T. J.: The USDA Natural Resources Conservation Service Soil Climate Analysis Network (SCAN), J. Atmos. Ocean. Tech., 24, 2073–2077, https://doi.org/10.1175/2007JTECHA930.1, 2007.

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., and Teuling, A. J.: Investigating soil moisture–climate interactions in a changing climate: A review, Earth-Sci. Rev., 99, 125–161, https://doi.org/10.1016/j.earscirev.2010.02.004, 2010.

Seo, E., Lee, M.-I., Reichle, R. H., and Koster, R. D.: 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, 2021.

Sisterson, D. L., Peppler, R. A., Cress, T. S., Lamb, P. J., and Turner, D. D.: The ARM Southern Great Plains (SGP) site, Meteorol. Monogr., 57, 6.1–6.14, https://doi.org/10.1175/AMSMONOGRAPHS-D-16-0004.1, 2016.

Tai, S.-L., Yang, Z., Gaudet, B., Sakaguchi, K., Berg, L., Kaul, C. M., Qian, Y., Liu, Y., and Fast, J.: A 1 km soil moisture data over eastern continental U.S. generated through assimilating SMAP data into the Noah-MP land surface model, Zenodo [data set], https://doi.org/10.5281/zenodo.14370563, 2024.

Tao, C., Zhang, Y., Tang, S., Tang, Q., Ma, H.-Y., Xie, S., and Zhang, M.: Regional moisture budget and land-atmosphere coupling over the U.S. Southern Great Plains inferred from the ARM

long-term observations, J. Geophys. Res.-Atmos., 124, 10091–10108, https://doi.org/10.1029/2019JD030585, 2019.

Taylor, C. M., Gounou, A., Guichard, F., Harris, P. P., Ellis, R. J., Couvreux, F., and De Kauwe, M.: Frequency of Sahelian storm initiation enhanced over mesoscale soil-moisture patterns, Nat. Geosci., 4, 430–433, https://doi.org/10.1038/ngeo1173, 2011.

Tian, L., Zhang, B., and Wu, P.: A global drought dataset of standardized moisture anomaly index incorporating snow dynamics (SZI_{snow}) and its application in identifying large-scale drought events, Earth Syst. Sci. Data, 14, 2259–2278, https://doi.org/10.5194/essd-14-2259-2022, 2022.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Navas Traver, I., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'Abbate, M., Pietropaolo, A., Huchler, M., and Rostan, F.: GMES Sentinel1 mission, Remote Sens. Environ., 120, 9–24, https://doi.org/10.1016/j.rse.2011.05.028, 2012.

Verseghy, D. L.: CLASS – A Canadian land surface scheme for GCMs. I. Soil model, Int. J. Climatol., 11, 111–133, https://doi.org/10.1002/joc.3370110202, 1991.

Wagner, T. J., Turner, D. D., Berg, L. K., and Krueger, S. K.: Ground-based remote retrievals of cumulus entrainment rates, J. Atmos. Ocean. Tech., 30, 1460–1471, https://doi.org/10.1175/JTECH-D-12-00187.1, 2013.

Wang, C., Gu, X., Zhou, X., Yang, J., Yu, T., Tao, Z., Gao, H., Liu, Q., Zhan, Y., Wei, X., Li, J., Zhang, L., Li, L., Li, B., Feng, Z., Wang, X., Fu, R., Zheng, X., Wang, C., Sun, Y., Li, B., and Dong, W.: Chinese Soil Moisture Observation Network and time series data set for high-resolution satellite applications, Sci. Data, 10, 424, https://doi.org/10.1038/s41597-023-02234-8, 2023.

Williams, A. P., Cook, B. I., Smerdon, J. E., Bishop, D. A., Seager, R., and Mankin, J. S.: The 2016 Southeastern U.S. Drought: An Extreme Departure From Centennial Wetting and Cooling, J. Geophys. Res.-Atmos., 122, 10888–10905, https://doi.org/10.1002/2017JD027523, 2017.

Xia, Y., Mitchell, K. E., Ek, M. B., Sheffield, J., Cosgrove, B. A., Wood, E. F., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D., Koren, V., Duan, Q., Lohmann, D., Fan, Y., and Mocko, D.: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products, J. Geophys. Res.-Atmos., 117, D03109, https://doi.org/10.1029/2011JD016048, 2012 (data available at: https://disc.gsfc.nasa.gov/datasets/NLDAS_FORA0125_H_2.0/summary?keywords=NLDAS2, last access: 8 September 2025).

Yang, Z., Qian, Y., Liu, Y., Berg, L. K., Gustafson, W. I., Feng, Z., Sakaguchi, K., Fast, J. D., Tai, S.-L., Huang, M., and Xiao, H.: Understanding irrigation impacts on low-level jets over the Great Plains, Clim. Dynam., 55, 925–943, https://doi.org/10.1007/s00382-020-05301-7, 2020.

Yang, Z., Huang, M., Berg, L. K., Qian, Y., Gustafson, W. I., Fang, Y., Liu, Y., Fast, J. D., Sakaguchi, K., and Tai, S.-L.: Impact of lateral flow on surface water and energy budgets over the Southern Great Plains – A modeling study, J. Geophys. Res.-Atmos., 126, e2020JD033659, https://doi.org/10.1029/2020JD033659, 2021.

Yang, Z.-L., Niu, G.-Y., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Longuevergne, L., Manning, K., Niyogi, D., Tewari, M., and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins, J. Geophys. Res.-Atmos., 116, D12110, https://doi.org/10.1029/2010JD015140, 2011.

Yin, J. and Zhan, X.: Scale impact of soil moisture observations to Noah-MP land surface model simulations, Remote Sens., 12, 1169, https://doi.org/10.3390/rs12071169, 2020.

Zeng, X.: Global vegetation root distribution for land modeling, J. Hydrometeorol., 2, 525–530, https://doi.org/10.1175/1525-7541(2001)002<0525:GVRDFL>2.0.CO;2, 2001.

Zhang, Y., Liang, S., Ma, H., He, T., Wang, Q., Li, B., Xu, J., Zhang, G., Liu, X., and Xiong, C.: Generation of global 1 km daily soil moisture product from 2000 to 2020 using ensemble learning, Earth Syst. Sci. Data, 15, 2055–2079, https://doi.org/10.5194/essd-15-2055-2023, 2023.