Impact of Design Factors for ESA CCI Satellite Soil Moisture Data Assimilation over Europe

Zdenko Heyvaert [a,b] Samuel Scherrer,[a,b] Michel Bechtold,[a] Alexander Gruber,[b] Wouter Dorigo,[b] Sujay Kumar,[c] and Gabriëlle De Lannoy[a]

[a] *Department of Earth and Environmental Sciences, Katholieke Universiteit Leuven, Heverlee, Belgium*
[b] *Department of Geodesy and Geoinformation, Technische Universität Wien, Vienna, Austria*
[c] *Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, Maryland*

ABSTRACT: In this study, soil moisture retrievals of the combined active–passive ESA Climate Change Initiative (CCI) soil moisture product are assimilated into the Noah-MP land surface model over Europe using a one-dimensional ensemble Kalman filter and an 18-yr study period. The performance of the data assimilation (DA) system is evaluated by comparing it with a model-only experiment (at in situ sites) and by assessing statistics of innovations and increments as DA diagnostics (over the entire domain). For both assessments, we explore the impact of three design choices, resulting in the following insights. 1) The magnitude of the assumed observation errors strongly affects the skill improvements evaluated against in situ stations and internal diagnostics. 2) Choosing between climatological or monthly cumulative distribution function matching as the observation bias correction method only has a marginal effect on the in situ skill of the DA system. However, the internal diagnostics suggest a more robust system parameterization if the observations are rescaled monthly. 3) The choice of atmospheric reanalysis dataset to force the land surface model affects the model-only skill and the DA skill improvements. The model-only skill is higher with input from the MERRA-2 than with input from the ERA5 reanalysis, resulting in larger DA skill improvements for the latter. Additionally, we show that the added value of the DA strongly depends on the quality of the satellite retrievals and land cover, with the most substantial soil moisture skill improvements occurring over croplands and skill degradation occurring over densely forested areas.

KEYWORDS: Soil moisture; Satellite observations; Data assimilation; Land surface model

1. Introduction

Soil moisture has been recognized as an essential climate variable by the Global Climate Observing System (GCOS) due to its important role in hydrological processes and its impact on global weather and climate through atmospheric feedbacks (Seneviratne et al. 2010; Mahfouf 2010). As the water source for evapotranspiration over land and an essential variable for vegetation growth, it is involved in the water, energy, and carbon cycles driving many land surface and atmospheric processes. The monitoring of soil moisture is thus essential in, for example, assessing possible drought development. This motivates the need for the development of long-term and large-scale data records of both surface and root-zone soil moisture.

Soil moisture can be derived from satellite-based brightness temperature or backscatter in the microwave spectrum (Kerr et al. 2010; Paloscia et al. 2013; Entekhabi et al. 2014). However, these microwave sensors have limited vertical penetration

depth (top few centimeters of the soil), and the resulting products contain both temporal and spatial discontinuities. Alternatively, land surface models (LSMs) can simulate surface and root-zone soil moisture to obtain physically consistent estimates at all times and locations (Srinivasan et al. 2000). Data assimilation (DA) uses observations to constrain LSM estimates and enables us to obtain superior estimates of soil moisture by combining the advantages of both sources for soil moisture estimation. Numerous studies have shown the benefit of satellite-based soil moisture DA, either using backscatter or brightness temperatures (e.g., Loew et al. 2009; De Lannoy and Reichle 2016a,b; Lievens et al. 2017; Reichle et al. 2019) or soil moisture retrievals (e.g., Reichle and Koster 2005; Draper et al. 2012; Lievens et al. 2015; Kumar et al. 2014, 2019). The latter approach does not require an observation operator that converts the observations into the prognostic model space, leaving out an additional level of complexity to the DA system. However, it lacks the advantage of being physically consistent and inevitably suffers from biases between the retrieval and model estimates of soil moisture (Reichle and Koster 2004). In this paper, we assimilate soil moisture retrievals from the European Space Agency Climate Change Initiative Soil Moisture (ESA CCI SM; Dorigo et al. 2017) product. In this product, soil moisture retrievals from various active and passive microwave instruments have been harmonized and bias corrected to create a

---

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JHM-D-22-0141.s1.

Corresponding author: Zdenko Heyvaert, zdenko.heyvaert@kuleuven.be

consistent merged long-term (over 30 years) soil moisture climate data record (Gruber et al. 2019).

Earlier studies have shown a positive impact of assimilating the ESA CCI SM product, e.g., for streamflow modeling in China (Liu et al. 2018) and estimating historical root-zone soil moisture records over India (Pal and Maity 2021). Blyverket et al. (2019) assimilated the ESA CCI SM, SMAP, and SMOS data over the contiguous United States using an ensemble optimal interpolation (EnOI) method to compare their respective performances. They found that out of the three resulting analyses, the one with assimilated ESA CCI SM data showed the smallest improvements in correlations with in situ soil moisture data. Over Europe, Albergel et al. (2017) found an improvement in the representation of above-ground biomass, gross primary production, and evapotranspiration by jointly assimilating the ESA CCI SM product and GEOV1 leaf area index (LAI) into a global land data assimilation system (LDAS-Monde). De Santis et al. (2021) found relatively minor improvements in predicting river flow by assimilating the ESA CCI SM product. Finally, Naz et al. (2020) assimilated the ESA CCI SM product into the community land model (CLM) over Europe, producing a high-resolution (3 km) reanalysis dataset for soil moisture. They found an improvement over in situ sites compared to the model-only simulation.

The studies above thus report mixed results, partly due to the use of different model and DA configurations. In this study, we assess the impact of such configurations on the performance of the DA system. Specifically, we consider three choices regarding the design of the DA system with its LSM and error parameterization. First, we focus on the chosen observation error (perturbation magnitude) and how this parameter affects the skill of the DA system and its internal diagnostics. Second, we consider two approaches for treating observation-minus-forecast bias: cumulative density function (CDF) matching through either the commonly used climatological rescaling (Reichle and Koster 2004) or seasonal monthly rescaling (Barbu et al. 2014). Third, we address the impact of the forcing input on the LSM by comparing two atmospheric reanalyses: the fifth-generation European Center for Medium-Range Weather Forecasts (ECMWF) Reanalysis (ERA5; Hersbach et al. 2020) and the NASA Modern-Era Retrospective Analysis for Research and Applications version 2 (MERRA-2; Gelaro et al. 2017). While comparisons between these datasets have been made previously for a variety of different applications (Olauson 2018; Seethala et al. 2021; Song and Wei 2021; Taszarek et al. 2021), this study presents, to our knowledge, the first assessment of the impact of ERA5 versus MERRA-2 forcing on Noah-MP soil moisture simulations with and without soil moisture DA. Additionally, we look into how external factors affect the in situ performance of the DA system. As in previous studies (e.g., Draper et al. 2012), we assess the impact of retrieval quality and land cover on the added value of DA using the ESA CCI SM product.

We assimilate the combined active–passive ESA CCI SM product into the Noah LSM with multiparameterization options (Noah-MP; Niu et al. 2011), which dynamically simulates both soil moisture and vegetation, over the European continent. We evaluate the model-only and DA performances for soil moisture against in situ measurements. However, the limited number of in situ sites does not represent the entire domain regarding soil

texture, land cover, and climate. Therefore, we also evaluate the DA diagnostics (Reichle et al. 2017), which are available over the entire simulation domain.

Section 2 introduces the experiment setup and evaluation methods. Section 3 presents the results in view of the three design choices outlined above. The conclusions are summarized in section 4.

2. Methodology

a. Model configuration

Model simulations are performed using the Noah-MP model (Niu et al. 2011; Yang et al. 2011) version 4.01 with dynamic vegetation, implemented within the NASA Land Information System (LIS; Kumar et al. 2006). Noah-MP simulates soil moisture at depths of 0–10 cm (SM_1), 10–40 cm (SM_2), 40–100 cm (SM_3), and 100–200 cm (SM_4). We will refer to soil moisture in the top 10 cm as the surface soil moisture (sfsm, equivalent to SM_1) and the top 100 cm as the root-zone soil moisture (rzsm). We derive the model rzsm as the weighted average of the top three layers, with the thickness of the layers serving as weights. We use a regular grid over the European continent (11.50°W–40.25°E, 29.75°–71.75°N) at a spatial resolution of 0.25°. Starting from a uniform initial condition, we spin up the LSM by looping three times from 2000 to 2010 to obtain a reliable climatology to initialize the year 2000. A 2-yr period from 1 January 2000 through 31 December 2001 then serves to spin up the 24 perturbed ensemble members (see section 2d). All experiments (see section 2e) are conducted from 1 January 2002 through 31 December 2019 using a model integration time step of 15 min and daily averaged model output at 0000 UTC.

The used land cover data are obtained from NASA's Moderate Resolution Imaging Spectroradiometer (MODIS; Friedl et al. 2002) while the Harmonized World Soil Database (HWSD; Nachtergaele et al. 2010) is used for soil texture information and NASA's Shuttle Radar Topography Mission (SRTM; Farr et al. 2007) provides the elevation map. Land cover and soil texture are spatially transformed to the model grid using a tiling approach. Land cover is assumed to be constant over time as in most studies (unlike Maertens et al. 2021, for example).

b. Meteorological forcing

The Noah-MP LSM is forced by the latest two major climate reanalysis products in separate experiments (see section 2e): ERA5 (Hersbach et al. 2020) lowest model level forecasts with a resolution of 0.25° × 0.25°, and MERRA-2 (Gelaro et al. 2017) lowest model level analyses with a resolution of 0.5° × 0.625°. Both datasets are matched to the model grid using a bilinear interpolation approach. The MERRA-2 forcing includes a gauge-based precipitation correction similar to the one developed for MERRA-Land (Reichle et al. 2011; Reichle and Liu 2014). For both forcings, elevation-based atmospheric temperature corrections are applied in LIS.

c. Satellite retrievals

We use satellite retrievals from the combined active–passive ESA CCI SM product version 7.1 (Dorigo et al. 2017;

Gruber et al. 2019) for assimilation into the Noah-MP LSM. This product merges soil moisture retrievals of multiple satellites by combining both active and passive microwave products after rescaling them to the Global Land Data Assimilation System Noah LSM (GLDAS Noah; Rodell et al. 2004), thereby obtaining a consistent long-term record. Passive products are retrieved using the Land Parameter Retrieval Model (LPRM; Owe et al. 2008) algorithm and include, for our study period, the Special Sensor Microwave Imager (SSM/I), the Tropical Rainfall Measuring Mission's Microwave Imager (TMI), the Advanced Microwave Scanning Radiometer (AMSR)-E and AMSR-2, WindSat, the Soil Moisture Ocean Salinity mission (SMOS), and the Soil Moisture Active Passive mission (SMAP). Active products are retrieved using the TU Wien Water Retrieval Package (WARP; Wagner et al. 1999; Naeimi et al. 2009) algorithm and include the Active Microwave Instrument Wind Scatterometer (AMI-WS), and the Advanced Scatterometer (ASCAT)-A and ASCAT-B.

The ESA CCI SM product has a temporal resolution of one day on a 0.25° regular grid (same as the model grid). It is quality controlled, with retrievals masked out over rain forests, during frozen or snow-covered periods, or if flags of the retrieval algorithm are raised. Furthermore, we do not assimilate observations over grid cells for which the land cover is predominantly "urban and built-up land" or "snow or ice," or where the topographic complexity (TC) exceeds 10%. TC is defined as the standard deviation of the elevation within a grid cell, normalized between 0% and 100%, and is provided with the ESA CCI SM product. We do assimilate retrievals over forested areas if they do not have a warning flag in the ESA CCI SM product.

For experiments using monthly CDF matching (see section 2e), we ensure a reliable rescaling for each location by omitting months with less than 100 observations across the 18 simulated years. For experiments using climatological CDF matching, we omit locations with less than 100 observations in at least three different years from the assimilation.

d. Data assimilation method

A one-dimensional ensemble Kalman filter (EnKF; Reichle et al. 2002; Evensen 2003) with 24 ensemble members is used for assimilation of the soil moisture retrievals into the Noah-MP LSM at 0000 UTC; the daily ESA CCI SM retrievals are centered around this time. CDF matching is applied to remove the differences in climatology between the model and the retrievals. Differences between the forecasted and observed sfsm, also called innovations, are mapped to updates in prognostic volumetric soil moisture content in the four LSM soil layers. The magnitude of the updates, also called increments, depends on the assumed error in the forecasts and observations.

Perturbations to meteorological fields and the model's four soil moisture prognostic variables approximate the forecast errors. The settings for these perturbations are calibrated for the open loop through ensemble verification measures (De Lannoy et al. 2006). Settings for cross correlations of the errors are taken from Kumar et al. (2014). Note that the effective correlation of the forecast errors used in the Kalman filter is a combination of the correlations imposed via the perturbations and correlations from the dynamical coupling which is present in the model because of the physical interaction between the layers. As listed in Table 1, we perturb precipitation and downward shortwave radiation with multiplicative perturbations drawn from a lognormal distribution, whereas the downward longwave radiation and four soil moisture variables are perturbed with additive perturbations drawn from a normal distribution. This means that samples are drawn from a standard normal distribution (and truncated if their value exceeds 2.5) and multiplied by the standard deviation in Table 1. Forcing perturbations have a temporal correlation of 24 h, whereas the perturbations in the four prognostic variables are not temporally correlated. This temporal correlation relates to the time scale of the AR(1) process. No spatial correlations for the perturbations are enabled, and perturbations to the prognostic variables are not cross correlated to those of the forcings. We applied the correction described by Ryu et al. (2009) to remove unintended perturbation bias in the forecasts. This means that the ensemble average of an open loop will be the same as a deterministic forecast. The characterization of the random and systematic observation errors is treated as a design choice and is discussed below.

TABLE 1. Overview of ensemble perturbations to forcing fields—downward shortwave radiation (SW), downward longwave radiation (LW), and precipitation (P)—and prognostic variables of soil moisture in the four soil layers (SM$_i$, $i = 1, \ldots, 4$). Additive ($+$) perturbations have a mean of 0 and are drawn from a normal distribution, while multiplicative ($\times$) perturbations have a mean of 1 and are drawn from a lognormal distribution. We applied temporal correlations of the time series through a first-order autoregressive model. Perturbations to prognostic variables are uncorrelated to the forcing perturbations.

| | Type | Mean | Standard deviation | Temporal correlation | Cross correlations with other perturbations | | | | | | |
					SW	LW	P	SM$_1$	SM$_2$	SM$_3$	SM$_4$
SW	$\times$	1	0.3	24 h		-0.5	-0.8				
LW	$+$	0	50 W m^{-2}	24 h	-0.5		0.5				
P	$\times$	1	0.5	24 h	-0.8	0.5					
SM$_1$	$+$	0	0.004 00 m^3 m^{-3}	0					0.6	0.4	0.2
SM$_2$	$+$	0	0.000 07 m^3 m^{-3}	0				0.6		0.6	0.4
SM$_3$	$+$	0	0.000 04 m^3 m^{-3}	0				0.4	0.6		0.6
SM$_4$	$+$	0	0.000 02 m^3 m^{-3}	0				0.2	0.4	0.6	

Table 2. Overview of the performed experiments.

	CDF matching	Forced by ERA5	Forced by MERRA-2
Open loop	—	OL_E	OL_M
Data assimilation	Monthly	DA_E	DA_M
	Climatological	DA_E^c	—

e. Performed experiments and impact assessments

Table 2 gives an overview of the experiments discussed in this paper. Two open loop (model-only) experiments OL_x are performed, with x referring to the reanalysis dataset used to force the LSM, i.e., ERA5 for $x = E$ and MERRA-2 for $x = M$. Similarly, the reference data assimilation experiments are denoted by DA_x. These experiments use monthly CDF matching (see below) and an observation error (see below) of $S = 0.025 \text{ m}^3 \text{ m}^{-3}$ whereas other DA experiments use climatological CDF matching, denoted by DA_x^c, or a different value for S, denoted by $DA_x(S)$. All OL and DA experiments have the same forecast perturbation settings listed in Table 1.

1) Impact of the observation error

The magnitude of the observation error S refers to the assumed observation error standard deviation. Several experiments $DA_E(S)$ are performed with $S \in \{0.00625, 0.0125, 0.025, 0.05, 0.1, 0.2\}$ ($\text{m}^3 \text{ m}^{-3}$), leaving all other settings unchanged. The reference experiments DA_E and DA_M use a value of $S = 0.025 \text{ m}^3 \text{ m}^{-3}$, which was determined based on an optimization of innovation diagnostics (Reichle et al. 2002). This value is much smaller than the total error of the ESA CCI retrievals because anomaly observations are assimilated (after CDF matching) and the perturbations determined by S only reflect the error on the short-term and interannual signal in the observations.

2) Impact of the rescaling approach

Because the Kalman filter assumes unbiased observations and forecasts, we use a pixel-based CDF-matching approach to rescale the retrievals to the model climatology (Reichle and Koster 2004). In the reference experiments DA_E and DA_M, the CDF matching is done separately for each month of the year (Barbu et al. 2014) based on the 18 years of retrievals and model-only simulations. This approach was also used by Kumar et al. (2015), de Rosnay et al. (2020), and Aires et al. (2021), for example. In most other studies (e.g., Reichle and Koster 2005; Draper et al. 2012; Kumar et al. 2014; Lievens et al. 2015), however, the CDF matching is based on the entire time series of observations, here referred to as climatological CDF matching. We use this more common climatological rescaling approach in the DA_E^c experiment, keeping all other configurations unchanged.

3) Impact of the meteorological forcing

Most land surface DA systems correct for short-term and interannual errors introduced by meteorological forcings. The only difference between experiments DA_E and DA_M is the

reanalysis dataset used to force the Noah-MP LSM, as shown in Table 2.

f. Evaluation sites and skill metrics

Evaluation sites are acquired from two sources. First, the International Soil Moisture Network (ISMN; Dorigo et al. 2011, 2013, 2021) provides in situ measurements of soil moisture at various depths at individual points in the study domain. The data have been quality controlled, and we only use data points flagged as "good" for evaluation. Second, soil moisture observations from the SMAP/In Situ Core Validation Site Land Surface Parameters Matchup Data, Version 1, are available at 9-km grid-scale core validation sites as part of the SMAP Calibration/Validation Program (Cal/Val; Colliander et al. 2017). Figure 1a marks all 94 grid cells for which in situ sfsm is available for evaluation, after removing sites for locations where no DA takes place (e.g., mountainous areas). We aggregate the in situ data to daily resolution and require at least 200 daily observations over at least three different years per site before including it in the evaluation. When multiple sites are available within a 0.25° grid cell, skill metrics are computed for each site separately, after which the metric values are averaged to obtain a single value for the grid cell. This decreases the weight of areas with multiple sensors when computing the overall skill improvements over the domain. Similarly, there are 63 locations with in situ rzsm data (not shown). In situ sites are only used to assess rzsm if at least one sensor is present at a depth below 10 cm.

Figure 1b shows a considerable variation in the period in which in situ sfsm is available for evaluation. For example, the GTK network in Finland and the ORACLE network in central France cover periods in the first part of the simulations, whereas Cal/Val sites and data from the FMI network enable evaluation of later parts of the simulations. Some networks cover relatively long periods (e.g., SMOSMANIA), whereas others do not (e.g., GROW). The temporal coverage of the reference data matters. Indeed, we expect larger DA skill improvements at sites with data in later periods when the assimilated ESA CCI SM product has higher-quality retrievals. The supplemental material provides a detailed overview of the ISMN networks used and their references in Table S1.

Similar to other studies (e.g., Kumar et al. 2019), we mainly rely on the anomaly correlation (R_{anom}) as skill metric, because it removes climatological differences between in situ and model (or retrieval) estimates of soil moisture (Koster et al. 2009). We compute the anomaly values by subtracting the multiyear mean seasonal cycle from the soil moisture data, after which we calculate the Pearson correlation coefficient between two anomaly time series. We also use the unbiased root-mean-square difference (ubRMSD; Entekhabi et al. 2010), defined as $\text{ubRMSD} = \sqrt{\text{RMSD}^2 - \text{bias}^2}$, with

$$\text{bias} = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i) \text{ and } \text{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2}. \quad (1)$$

Here x_i represents the soil moisture estimated by the OL or DA experiment (or satellite retrievals), and y_i represents the
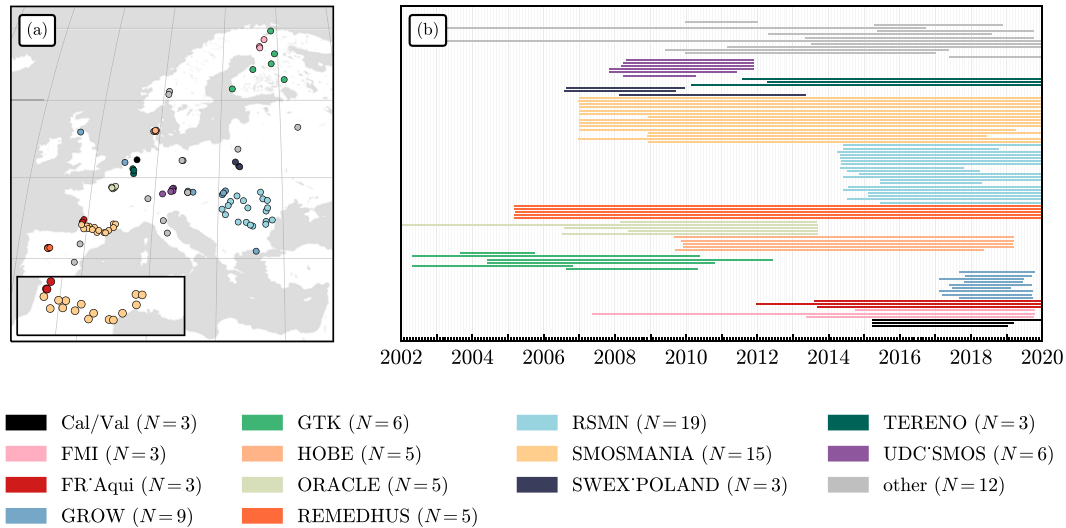
FIG. 1. (a) Location of the ISMN in situ evaluation sites and the Cal/Val reference sites for sfsm, colored per network. The inset shows a close-up of the SMOSMANIA network in southern France. (b) The lines indicate the start and end dates for which sfsm information is available at individual sites, colored per network. Gaps in the data may be present between the start and end dates (not shown).

in situ observed soil moisture at day i (N days in total). In situ observations when the soil is frozen or snow covered are removed before computing the skill metrics.

g. Regional data assimilation diagnostics

Apart from evaluating over stations with in situ data, we also assess the DA system through internal diagnostics (Gelb et al. 1974). These diagnostics are based on statistics of the innovations, or observation-minus-forecast residuals, and are thus available at all times and locations for which soil moisture retrievals are assimilated (Reichle et al. 2017). Following previous studies (e.g., Reichle et al. 2002; De Lannoy and Reichle 2016a,b; Reichle et al. 2017), the DA diagnostics are used to explore whether the assumed errors of the system are consistent with the actual errors. If the DA system extracts the information from the observations optimally, the innovations ν should form a zero-mean white noise sequence in time for each location (Gelb et al. 1974; Reichle et al. 2017).

The normalized innovations $\tilde{\nu}$ are obtained by dividing the innovations ν by the square root of the sum of the model and observation ensemble variances. If we choose the assumed model and observation errors optimally, the standard deviation of $\tilde{\nu}$, $s_{\tilde{\nu}}$, should be close to unity. Values greater than one indicate underestimated forecast and observation error variances, whereas values smaller than one indicate an overestimation of the error variances (Reichle et al. 2002; Desroziers et al. 2005; Crow and Bolten 2007; Crow and Van den Berg 2010).

Increments are the state space equivalent of the innovations: we derive them for each of the four soil layers of the model. Like the innovations, they should follow a zero-mean white noise sequence in time. We will primarily focus on the relations between soil moisture increments of different soil layers and the impact of the three considered design choices on these relations.

3. Results and discussion

a. Model-only experiments

In this section, we compare the results of the two ensemble model-only experiments OL_E and OL_M. We do not assimilate retrievals in these runs but use the same perturbations to forcing fields and prognostic variables as the subsequent DA experiments (Table 1).

The map of time-average rzsm for the OL_E experiment in Fig. 2a shows the expected patterns, with drier soils over southern Europe and northern Africa. Sandy soils, in Norway and Finland for example, are also substantially drier. The results for sfsm (not shown) are very similar, but with drier topsoil in northern Africa. Figure 2b shows the difference in the average rzsm between OL_E and OL_M (an artificial pattern is visible in this map due to the difference in resolution between both reanalyses). MERRA-2 yields drier soils in most regions of Europe, most pronounced over mountainous areas such as the Pyrenees and the Carpathians. Over northern Africa, the LSM forced by ERA5 generally yields drier soils, especially over the Atlas Mountain ranges. The bias over Scandinavia is small, with soils in OL_E slightly drier. We observe that the soil moisture differences strongly relate to precipitation differences. Indeed, we observe higher average precipitation over most of Europe for the ERA5 reanalysis compared to MERRA-2 (Fig. S1). The magnitude of this difference is also linked to topography. An explanation for this may be that the ERA5 reanalysis is not rain gauge corrected over our study domain, and several studies (e.g., Bandhauer et al. 2022 over Europe) have found that the mean precipitation is overestimated. Over Scandinavia, the precipitation in both datasets is
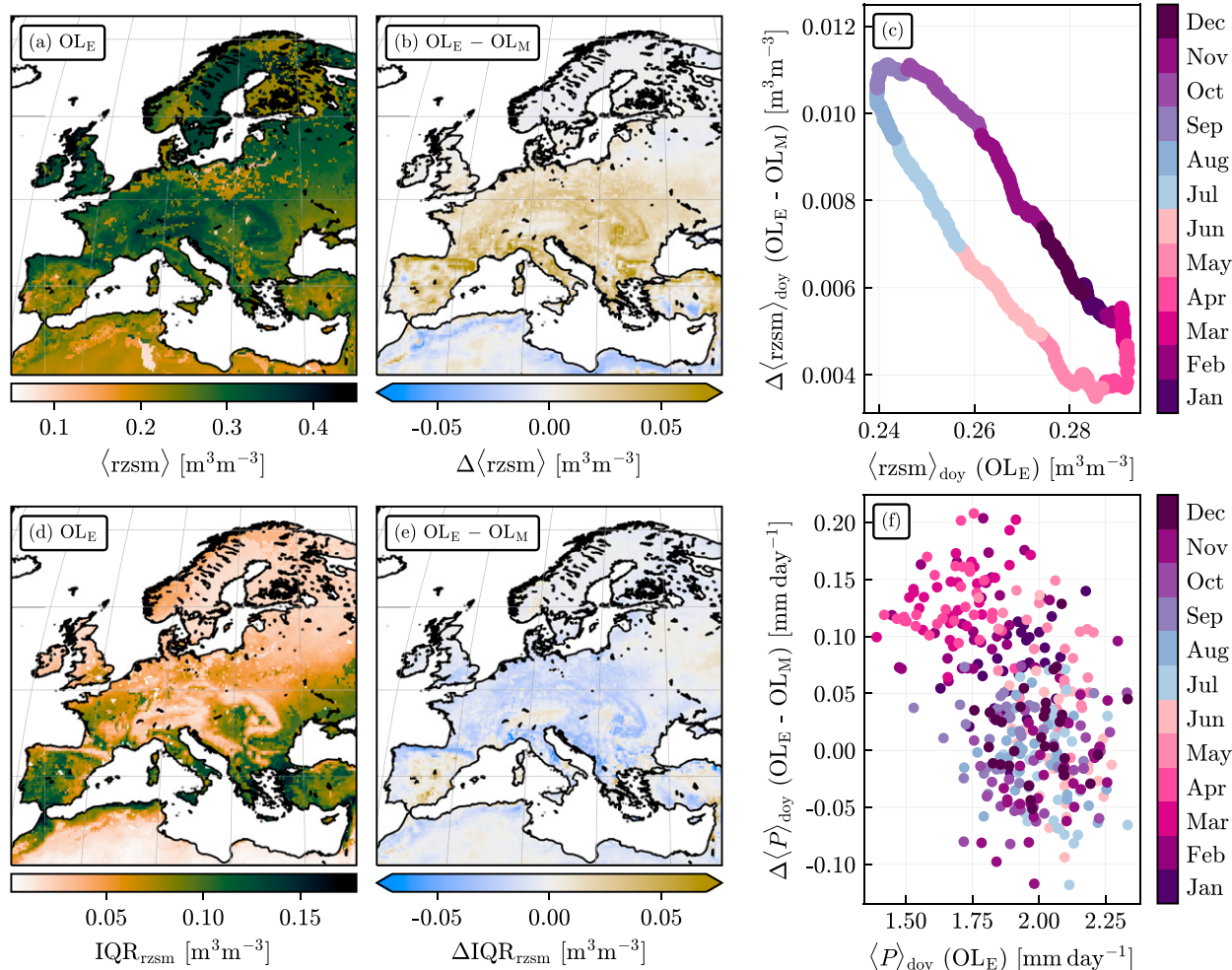
FIG. 2. (a) Temporally averaged (2002–19) rzsm, $\langle \mathrm{rzsm} \rangle$, in the OL_E experiment. (b) Difference in $\langle \mathrm{rzsm} \rangle$ between OL_E and OL_M: brown indicates higher values for OL_E, and blue indicates higher values for OL_M. (c) Difference in the spatial average for each day of the year, $\Delta \langle \mathrm{rzsm} \rangle_{\mathrm{doy}}$, between OL_E and OL_M as a function of $\langle \mathrm{rzsm} \rangle_{\mathrm{doy}}$ for OL_E. (d),(e) As in (a) and (b), but for the interquartile range $\mathrm{IQR}_{\mathrm{rzsm}}$. (f) As in (c), but for the precipitation P.

comparable, while MERRA-2 has more precipitation over northern Africa, in line with rzsm in Fig. 2b.

Figure 2d depicts the temporal variability of the rzsm through the interquartile range (IQR). There is a clear latitudinal effect, with southern Europe and Turkey showing the most variability in soil moisture over time and northern Europe the least. Mountainous areas such as the Pyrenees, the Alps, and the Carpathian Mountains show little variation in simulated soil moisture. Again, the maps for sfsm are very similar, but sfsm exhibits more variability than rzsm over northern Africa (not shown). The difference in rzsm variability between OL_E and OL_M is visualized in the map of Fig. 2e. Contrary to the map of the temporal average (Fig. 2b), most regions in Europe have a higher variability over time for OL_M. This behavior can be understood by looking at the soil moisture dynamics in Fig. 2c, which compares the time series of spatially averaged rzsm for each day of the year across all simulation years, $\langle \mathrm{rzsm} \rangle_{\mathrm{doy}}$. The difference $\Delta \langle \mathrm{rzsm} \rangle_{\mathrm{doy}}$ between OL_E and

OL_M is shown as a function of $\langle \mathrm{rzsm} \rangle_{\mathrm{doy}}$ in the OL_E experiment. This difference increases during summer, to reach its largest value during the driest period toward the end of summer. The difference between both runs then decreases again to reach its minimum during the wettest period at the end of winter. This means that soils dry out more during summer in the simulations forced by MERRA-2 than in simulations forced by ERA5, explaining the larger variability for the former observed in Fig. 2e. This effect is larger for rzsm than for sfsm (not shown). One important driver for the dynamics depicted in Fig. 2c is the precipitation P. Figure 2f portrays the difference $\Delta \langle P \rangle_{\mathrm{doy}}$ between the ERA5 and MERRA-2 reanalyses for each day of the year, as a function of $\langle P \rangle_{\mathrm{doy}}$. The differences in precipitation between the two reanalyses can indirectly explain the differences in rzsm shown in Fig. 2c: $\langle P \rangle_{\mathrm{doy}}$ is typically larger for ERA5 during light rainfall events, while $\langle P \rangle_{\mathrm{doy}}$ is larger for MERRA-2 during heavy rainfall events. Note that the difference in color between Fig. 2c and Fig. 2f
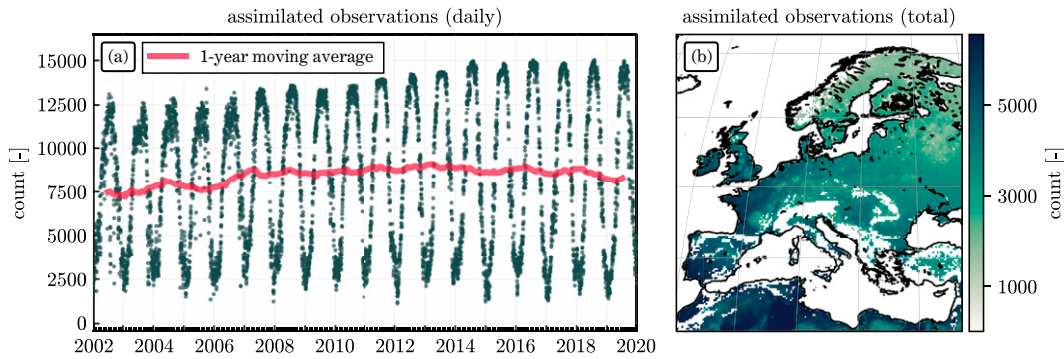
FIG. 3. (a) Daily count of assimilated ESA CCI soil moisture retrievals over the study domain. (b) Total number of retrievals assimilated for each grid cell over the complete simulation period. No assimilation took place over white grid cells. Both figures correspond to the DA_E experiment.

indicates a time delay between $\langle P \rangle_{\mathrm{doy}}$ and $\langle \mathrm{rzsm} \rangle_{\mathrm{doy}}$ of about 6 months.

b. Number of assimilated observations

The number of assimilated soil moisture retrievals varies both in space and time for the DA_E experiment. Figure 3a shows daily counts of assimilated observations over the study domain. Apart from a clear seasonal pattern, the number of assimilated retrievals increases over time as additional sensors are included in the combined active–passive ESA CCI SM product. This is most pronounced in the peaks of the summer months, when the number of retrievals increases from 12 500 in 2002 to 15 000 in 2019. The impact of adding *MetOp-A* ASCAT (January 2007) or AMSR2 (July 2012) is visible, for example (Dorigo et al. 2017). While the inclusion of some products (e.g., SMAP in April 2015) does not have a direct impact on the number of assimilated retrievals, it allows for a higher-quality daily averaging and should decrease the product uncertainty.

The map for DA_E in Fig. 3b shows that retrievals are never assimilated in mountainous areas such as the Alps, the Pyrenees, and the Carpathian Mountains as satellite retrievals in such areas are masked out (section 2c). Similarly, the map also shows fewer assimilated retrievals in areas with cold winters such as Northern Europe due to the masking of retrievals during frozen conditions. Drawing the map for the DA_E^c experiment yields a similar pattern, with slightly more assimilated observations throughout the domain, and in the northern latitudes specifically. The pattern for DA_M follows that of DA_E closely.

c. Evaluation of the DA skill improvements

Figure 4a shows boxplots of R_{anom} for the model-only (OL_E) and data assimilation (DA_E) sfsm forced by ERA5, at the ISMN and SMAP Cal/Val sites. The mean difference between both R_{anom} distributions is significantly different from zero at the 5% level (paired two-sample t test, $p = 0.0015$), but the difference is small with mean $R_{\mathrm{anom}} = 0.52$ for DA_E and mean $R_{\mathrm{anom}} = 0.51$ for OL_E. Figure 4b shows the spatial pattern of the DA impact, with a skill improvement at sites in southern France (the SMOSMANIA network) and

eastern Europe (the RSMN network) and a deterioration in Finland (GTK and FMI networks). As mentioned earlier, the SMOSMANIA and RSMN networks provide evaluation data toward the end of the simulation, when retrievals are more numerous and of higher quality, whereas the GTK network mostly consists of in situ information that can be used to evaluate the first half of the simulation (Fig. 1). We are thus not only comparing different regions but also different periods, which may be an additional explanation of the differences in skill improvements. Maps of the sfsm ubRMSD (not shown) display a similar spatial pattern, with mean values of 0.052 and 0.051 $m^3\,m^{-3}$ for the OL_E and DA_E experiments, respectively. The evaluation of the rzsm (Figs. 4c,d) yields similar results as for the sfsm, with mean R_{anom} values of 0.51 and 0.49 for the OL_E and DA_E experiments, respectively (ubRMSD: 0.042 and 0.043 $m^3\,m^{-3}$). Overall, the significant but small improvements found in our study align with Reichle et al. (2021) who show that the impact of soil moisture DA over Europe can be limited compared to other (drier) regions, or regions where the forcing data are presumably of lower quality.

In Fig. 5a, we visualize (through R_{anom} with the in situ data) the impact of model-only (OL_E) F5 and observation skill on the skill improvement by the DA_E system. For evaluation sites located in the top left of this figure, OL_E outperforms the retrievals, whereas the retrievals outperform the model in the bottom right of the figure. No DA skill degradation is observed when the observations outperform the model, but this is the case only for a minority of the evaluation sites (18 out of 94). However, it is clear that improvements can still occur in other locations, where the retrievals perform worse than the model.

Linking the skill of the model and/or the retrievals, and thus the expected DA skill improvement, to environmental factors such as land cover would allow making a priori assumptions regarding locations where the DA is expected to have a substantial positive or negative impact. Figure 5b shows the distribution of forest fraction (based on the model land cover input) in the grid cells for sites showing skill improvement (blue boxplot) and sites with skill deterioration (red boxplot). A two-sample t test indicates that the means of the two distributions are different at the 5%
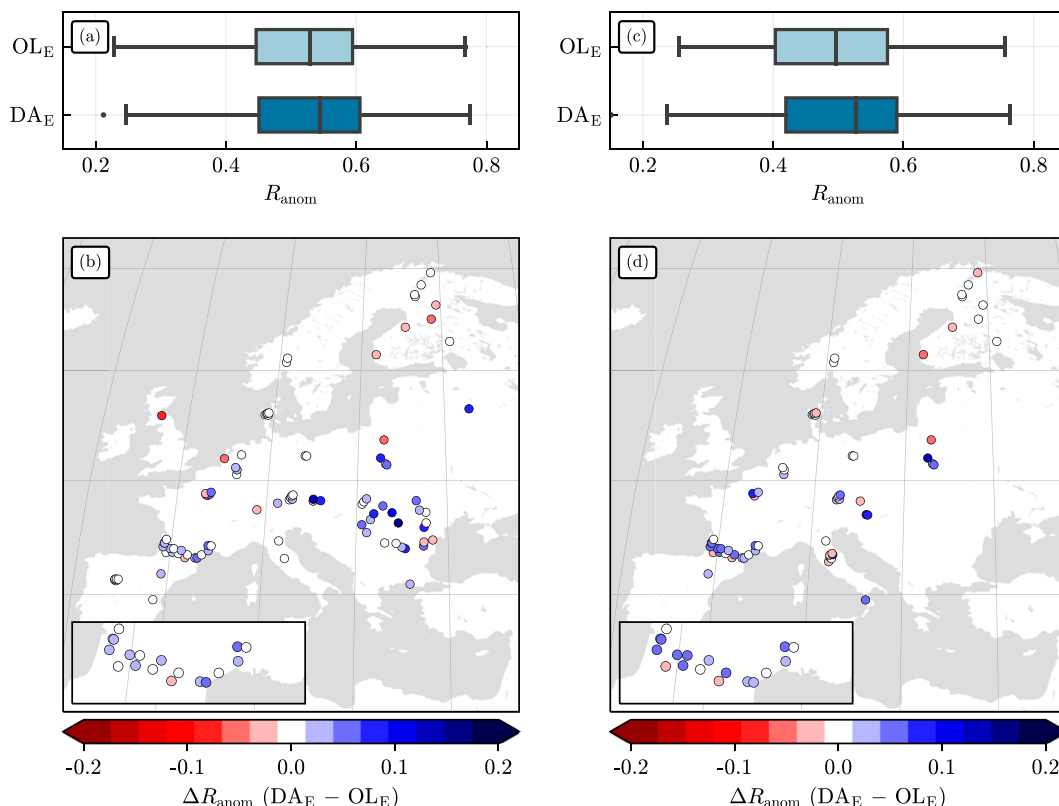
FIG. 4. (a) Comparing spatial distributions of sfsm R_{anom} for the OL_E and DA_E experiments relative to ISMN and Cal/Val evaluation data. (b) Change in anomaly correlation (ΔR_{anom}) of sfsm: blue sites show an improvement in skill by the DA, whereas red sites show a degradation. The inset shows a close-up of the SMOSMANIA network in southern France. (c),(d) As in (a) and (b), but for the rzsm.

significance level ($p = 0.014$). Locations with skill degradation are more likely located in (densely) forested areas, such as the ISMN sites in Finland. Figure 5c demonstrates that locations where retrievals outperform the model (OL_E) indeed have a relatively low forest fraction, although exceptions occur.

Similarly, Figs. 5d and 5e reveal statistically significant ($p = 0.009$) skill improvement by DA at grid cells with larger cropland fractions. Model and satellite retrieval performance differences at such sites are generally smaller than over forested sites. This is consistent with the findings of Draper et al. (2012) and is linked to croplands being more suited for microwave remote sensing than forests, and the Noah-MP model not accounting for cropping practices such as irrigation. A map showing the spatial distributions of forests and croplands over our study domain can be found in the supplemental material (Fig. S2).

Previous studies over the contiguous United States (De Lannoy and Reichle 2016b; Kumar et al. 2019) show larger skill improvements over the east (wetter soils) than the west (drier soils) of the country. We therefore also examined the link between mean (modeled) rzsm and skill improvement (Figs. 5f,g) but found no statistically significant effect ($p = 0.104$).

1) Impact of the observation error

Figure 6 shows the distribution of sfsm and rzsm skill improvements ΔR_{anom} (compared to OL_E) at ISMN and SMAP Cal/Val sites for $\mathrm{DA}_E(S)$ with different values of the observation error S. Superimposed on the boxplots are the individual evaluation sites, with blue colors indicating higher retrieval quality and red colors indicating lower retrieval quality (relative to the model skill). As discussed previously, higher satellite retrieval qualities result in larger skill improvements, but observations with lower skill than the model may still improve the DA results. Figure 6 displays a trade-off for both sfsm and rzsm when choosing the "optimal" value of the observation error S. If S is too small, the DA puts too much weight on the observations: while the best-performing sites show very strong improvements in this scenario, many sites also show strong deteriorations. As S increases, the DA remains closer to the OL and ΔR_{anom} values converge toward zero. A larger value for S is thus a "safe" choice if the objective is to assimilate a product of low(er) quality (compared to the model) without harming the model simulations. The value $S = 0.025$ m^3 m^{-3} results in the highest median skill improvement for both the sfsm ($\Delta R_{\mathrm{anom}} = 0.007$) and the rzsm ($\Delta R_{\mathrm{anom}} = 0.008$).
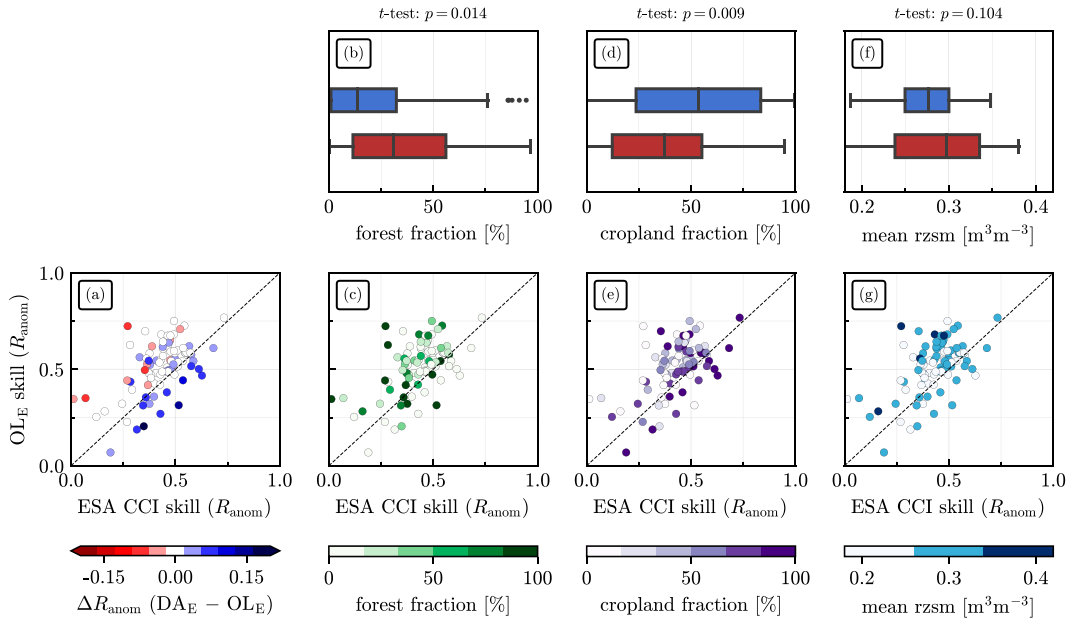
FIG. 5. (a) Skill change (ΔR_{anom}; $DA_E - OL_E$) for sfsm (color), as a function of OL_E skill (vertical axis) and observation skill (horizontal axis). Blue indicates improvements and red indicates deteriorations. (b) Fraction of the evaluated grid cells occupied by forest, for all sites with improvements (blue boxplot) and deteriorations (red boxplot). The p value indicating the significance of the difference between the means of both distributions is shown on top (two-sample t test). (c) As in (a), but with the color replaced by the forest fraction. (d),(e) As in (b) and (c), but for the cropland fraction. (f),(g) As in (b) and (c), but for the mean rzsm.

2) IMPACT OF THE RESCALING APPROACH

Skill improvements ΔR_{anom} for DA with the monthly (DA_E) and climatologically (DA_E^c) rescaled observations are shown in Figs. 7a and 7b for the sfsm and rzsm respectively. For both rescaling methods, the DA skill gain is virtually identical when averaged over all stations and also exhibits low interstation variability (sfsm: average ΔR_{anom} of 0.014 and 0.012 for DA_E and DA_E^c, respectively; rzsm: average ΔR_{anom} of 0.015 and 0.016 for DA_E and DA_E^c, respectively). The lower panels of Fig. 7 are discussed in the following subsection.

3) IMPACT OF THE METEOROLOGICAL FORCING

Figure 7c (sfsm) and Fig. 7d (rzsm) compare the skill improvements ΔR_{anom} between DA_E and DA_M experiments. The DA skill changes, both improvements and degradations, are more moderate for DA_M than for DA_E, as indicated by the slope of the regression that is significantly smaller than 1, while the intercept is not significantly different from 0 (sfsm: average ΔR_{anom} of 0.014 and -0.003 for DA_E and DA_M, respectively; rzsm: average ΔR_{anom} of 0.015 and 0.012 for DA_E and DA_M, respectively). The OL simulations of sfsm have a
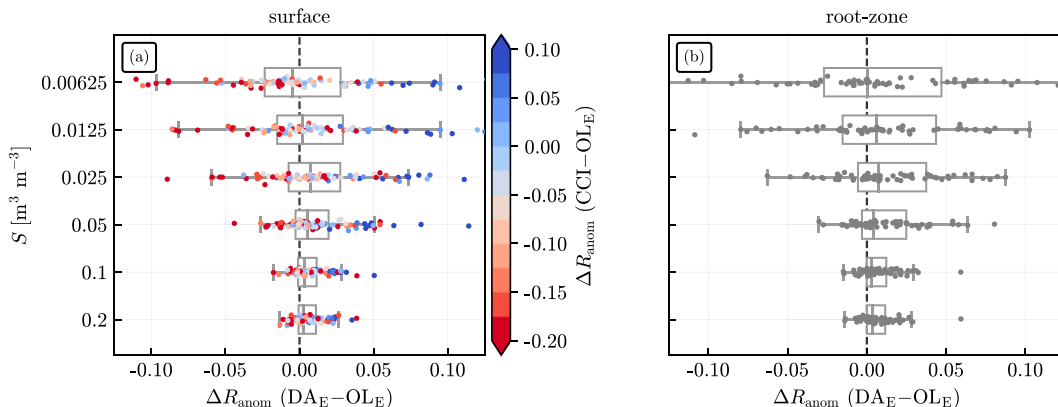

FIG. 6. (a) Distribution of skill improvements in sfsm (horizontal axis) as a function of the observation error S (vertical axis). Individual evaluation sites are superimposed with an indication of the retrieval skills relative to OL_E. (b) As in (a), but for the rzsm. Since the CCI product only represents surface soil moisture, no comparison between satellite retrievals and in situ rzsm measurements is shown.
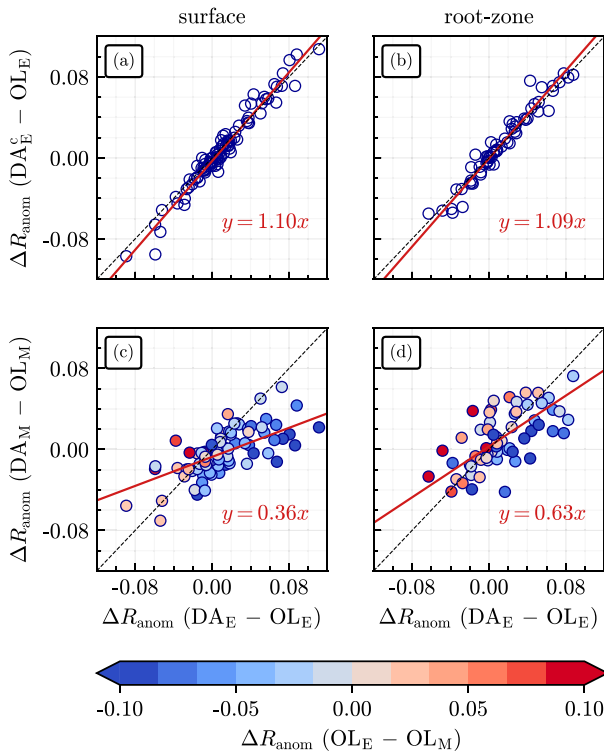
FIG. 7. Relation between skill improvements of DA_E and skill improvements of (a),(b) DA_E^c and (c),(d) DA_M relating to (left) sfsm and (right) rzsm. The color bar represents the skill difference between OL_E and OL_M in (c) and (d).

larger R_{anom} when forced with MERRA-2 than when forced with ERA5 for about 79% of the stations. This may partly explain why the skill improvements are generally larger for DA_E than for DA_M, as they are compared to their respective OL experiments. Despite the smaller skill improvements, the DA_M experiment still slightly outperforms DA_E in terms of R_{anom} (not shown).

d. Statistics of the data assimilation innovations

By design, the temporal average of the innovations is small throughout the domain, with values ranging from -0.001 to 0.003 m^3 m^{-3} (5th and 95th spatial percentiles, respectively) for DA_E (Fig. S3), and similar values for DA_E^c and DA_M. The impact of the rescaling approach and meteorological forcing on the temporal innovation mean is not significant. However, the innovation time series show a seasonal pattern in DA_E^c, with innovations steadily increasing during one-half of the year (May–November) and decreasing the other half of the year. This is reduced in DA_E (this is shown for the corresponding DA increments in Fig. S5). Unsurprisingly, rescaling the observation via monthly CDF matching thus results in a time series resembling a white noise sequence more closely. This is our main motivation to use monthly rescaling as the reference configuration in the DA_E and DA_M experiments. In the remainder of this section, we will discuss the temporal standard deviation of the normalized innovations $s_{\tilde{\nu}}$, which should be close to unity. Figure

S4 shows this quantity's temporal and spatial distribution for the DA_E experiment. It reveals smaller values over desert regions (Sahara), and larger values over agricultural regions (e.g., Ukraine), consistent with the findings of De Lannoy and Reichle (2016b) and Reichle et al. (2017).

1) IMPACT OF THE OBSERVATION ERROR

Figure 8a shows the spatial distributions of the temporal standard deviation of the normalized innovations $s_{\tilde{\nu}}$ for different values of the observation error S. The reference value $S = 0.025$ m^3 m^{-3} is depicted in light green and is centered around a value of $s_{\tilde{\nu}}$ between 1/2 and 1. The fact that $s_{\tilde{\nu}} < 1$ for most grid cells indicates that the total assumed forecast plus observation error of the system is slightly overestimated. Increasing S shifts the distribution of $s_{\tilde{\nu}}$ to smaller values (stronger overestimation of the errors), while excessively decreasing S shifts the distribution to larger values than unity (underestimation of the errors).

While our experiments with $S = 0.025$ m^3 m^{-3} happen to simultaneously result in the best value for $\langle s_{\tilde{\nu}} \rangle$ (Fig. 8a) and ΔR_{anom} (Fig. 6), there are several reasons why this would not always be the case. First, the innovation diagnostics cover the complete spatial and temporal domain of the experiments, while the in situ sites only cover the sparse locations and times shown in Fig. 1. Second, we optimize the observation error while keeping the assumed forecast perturbation parameters constant. Ideally, both should be tuned simultaneously as both their sum and ratio impact innovation diagnostics (Crow and Bolten 2007). Furthermore, our tuning assumes time invariance of forecast and observation error parameters.

2) IMPACT OF THE OBSERVATION RESCALING APPROACH

Figure 8b shows the joint spatial $s_{\tilde{\nu}}$ distributions of the DA_E experiment (monthly CDF matching), and the DA_E^c experiment (climatological CDF matching). Most bins are located above the diagonal, indicating that $s_{\tilde{\nu}}$ is systematically larger (and hence closer to unity) for the DA_E^c than for the DA_E experiment. The actual remaining total error (after bias removal) of the DA_E experiment is larger than that of the DA_E^c experiment. Therefore, as expected, we find that the assumed total error of the DA system is overestimated more if the observations are rescaled monthly instead of climatologically when all other settings remain unchanged.

3) IMPACT OF THE METEOROLOGICAL FORCING

The choice of meteorological forcing dataset does not have a significant impact on the $s_{\tilde{\nu}}$, as can be seen from Fig. 8c. The panel shows the relation of the spatial $s_{\tilde{\nu}}$ distributions in the DA_E and the DA_M experiment. The bins are scattered more or less evenly around the diagonal, indicating that the choice of ERA5 or MERRA-2 forcing does not have a major impact on how the total error of the system is being estimated.

e. Statistics of the data assimilation increments

The results for the increments in each of the four soil layers are in line with those of the innovations, i.e., the DA_E^c
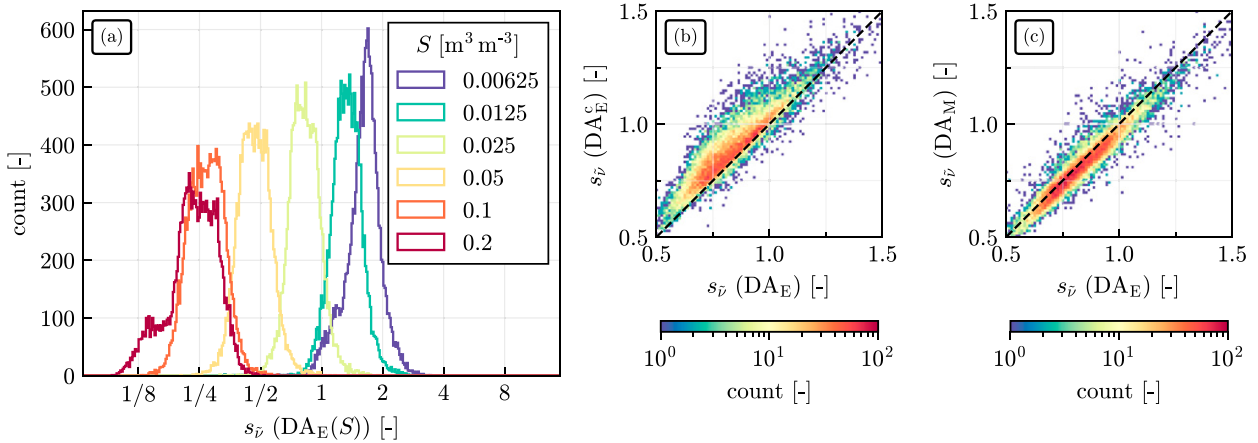
FIG. 8. (a) Spatial histograms of the normalized innovations $s_{\tilde{\nu}}$ for the DA$_E(S)$ experiments, each with different observation error S. The horizontal axis has a logarithmic scale. (b) Bivariate histogram of $s_{\tilde{\nu}}$ for DA$_E$ on the horizontal axis (monthly CDF matching) vs DA$_E^c$ on the vertical axis (climatological CDF matching). (c) Bivariate histogram of $s_{\tilde{\nu}}$ for DA$_E$ on the horizontal axis (ERA5 as forcing) vs DA$_M$ on the vertical axis (MERRA-2 as forcing).

experiment yields zero-mean time series that have a strong seasonal pattern. The increments of the DA$_E$ experiment resemble a white noise sequence more closely (Fig. S5). Increments in deeper soil layers are smaller than in the top layers: the difference in increment size between layers 1 and 4 is more than an order of magnitude.

In this section, we will focus on the relation between increments of different layers. Figure 9a shows the bivariate distribution of individual increments, i.e., not aggregated over space or time, for layers 1 and 2 (relations between deeper layers look similar and are provided in Fig. S7). The increments between both layers are positively correlated, and two regimes can be discerned from the histogram: one of strong increment coupling between soil layers (close to the vertical zero line) and one of weak increment coupling (close to the horizontal zero line). For the latter, increments of layer 1 are relatively large but close to zero for layer 2. There are even some occasions where the

increments of two different layers have an opposite sign (i.e., they contribute negatively to the correlation). The relatively low spatiotemporal increment correlation between the layers is in line with the assumed ensemble error correlation when perturbing the soil moisture forecasts (Table 1).

Before moving on to the other panels of Fig. 9, Fig. 10 links the different regimes in the bivariate increment distributions to other variables. The same two-dimensional bins between layers 1 and 2 as in Fig. 9a are shown, but the color now indicates the average value of some explanatory variables in that bin, rather than the count of increments. Figure 10a displays the vertical coupling strength (VCS), which is defined here as the anomaly correlation between the time series of sfsm (top layer) and rzsm (composite of the top three layers) in the model-only experiment OL$_E$ (Kumar et al. 2009). A map of the VCS over the study domain (Fig. S8) displays strong geographical differences, with the smallest values over Spain,
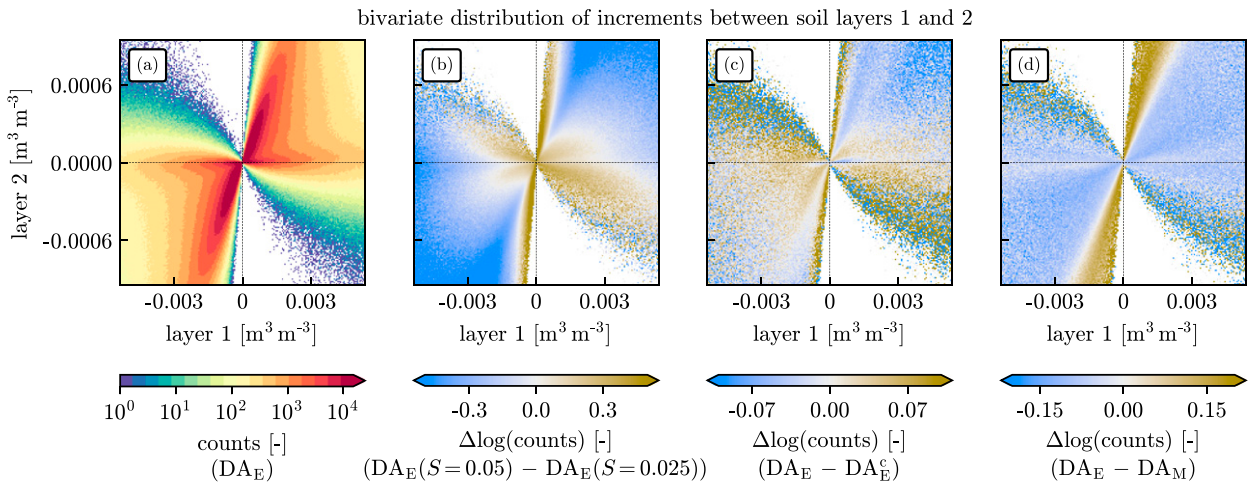

FIG. 9. (a) Bivariate histogram of increments between layers 1 and 2 of the DA$_E$ experiment. The other panels show the difference between increment histograms of the DA$_E$ experiment and (b) DA$_E(S = 0.025)$, (c) DA$_E^c$, and (d) DA$_M$ experiments.

bivariate distribution of increments between soil layers 1 and 2: average value of variable per bin
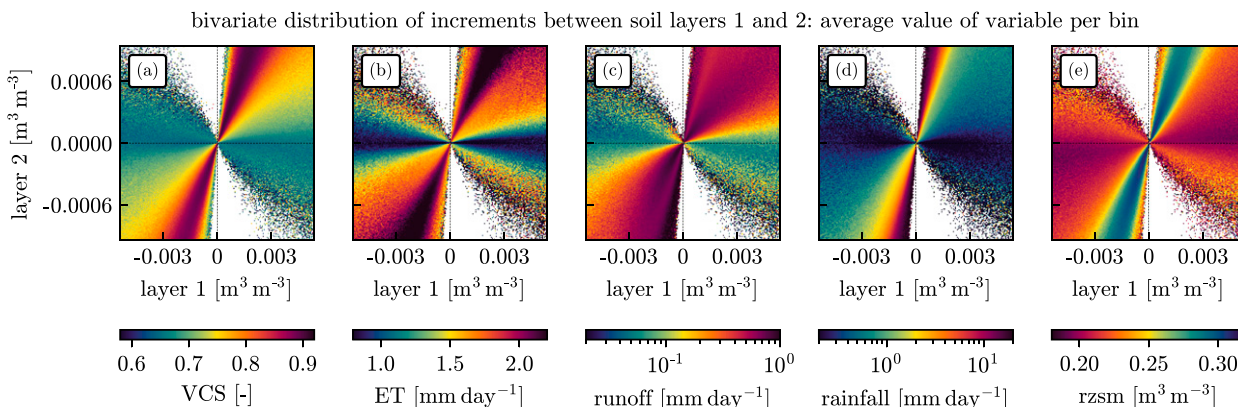

FIG. 10. As in Fig. 9a, but the colors now indicate the average value of different variables at the times and locations of the increments: (a) VCS, (b) ET, (c) total runoff, (d) rainfall rate, and (e) rzsm.

Turkey, and Africa and a clear influence of the soil texture over the rest of Europe (e.g., a lower VCS over sandy soils). From Fig. 10a, we find that lower values of the VCS typically coincide with the weak increment coupling regime (large increments for the surface layer and near-zero increments for deeper layers), whereas higher values of the VCS typically coincide with the strong increment coupling regime (large increments for both surface and deeper layers). This is consistent with the findings of Kumar et al. (2009). The effect of other land surface variables, i.e., the evapotranspiration (ET), total runoff, rainfall, and rzsm, is visualized in Figs. 10b–e. The panels show that wetter soil and larger hydrological fluxes typically lead to stronger correlations between the increments of different layers. Note that all these variables have an important latitudinal dependence over our study domain. Indeed, the regions with little ET, runoff, rainfall, and rzsm are mostly located in northern Africa, southern Europe, and Turkey. These regions thus dominate the weak increment coupling regime.

1) IMPACT OF THE OBSERVATION ERROR

Figure 9b shows the difference between histograms of increments for the $DA_E(S)$ experiments with $S = 0.025$ m^3 m^{-3} (as shown in Fig. 9a) and $S = 0.05$ m^3 m^{-3}. For the larger observation error S, the distribution shows more increments close to zero while the number of counts at the tails decreases. This is unsurprising, as we expect smaller DA updates for larger observation uncertainties.

2) IMPACT OF THE RESCALING APPROACH

The impact of rescaling the observations via monthly or climatological CDF matching is shown in Fig. 9c, which compares experiments DA_E and DA_E^c. Monthly rescaling of the observations results in smaller increments, as the distributions show positive differences centered around zero and negative differences toward the tails. This aligns with the smaller remaining innovation values when the observations are rescaled through monthly CDF matching.

3) IMPACT OF THE METEOROLOGICAL FORCING

Figure 9d shows that the DA_M experiment has more increments in the weak increment coupling regime, whereas DA_E has more increments in the strong coupling regime. This means that the correlations between layer 1 and deeper increments are larger with ERA5 meteorological forcings. As a result, we find more near-zero increments in the deeper layers for DA_M and in the top layer for DA_E. A possible explanation may be the higher number of heavy rainfall events in the ERA5 forcing compared to the MERRA-2 forcing (not shown), and the associated wetter rzsm for ERA5 (Fig. 2). Indeed, Figs. 10d and 10e show that heavy rainfall events and wetter rzsm are strongly linked to the strong coupling regime in the bivariate histograms.

Figure 11 offers a closer look at the impact of rainfall on the increment distribution. The top panel again shows a bivariate histogram, this time between the top soil layer increments for the DA_E and the DA_M experiments, respectively (Fig. S6 shows the histograms for all soil layers). Again, the distribution is centered around zero and decreases quickly for larger increments values. Most increments are located around the diagonal, resulting in a relatively high Pearson correlation of 0.72 between increments of both experiments. This correlation decreases for deeper layers, as shown in Fig. S6, indicating that the impact of choosing another forcing dataset is larger for the rzsm than for the sfsm. This can be linked to the fact that rzsm has a longer memory of distinct updates. A limited number of increments have opposite signs in the DA_E and DA_M experiments (top-left and bottom-right quadrants of Fig. 11a). An explanation for this may be rainfall events that are not be captured well in one or the other forcing dataset. Figure 11b shows that for increments close to the diagonal (the large majority), the average difference in rainfall rate ΔP between both forcing datasets is nearly zero. However, the small number of increments that have opposite signs in both experiments show large absolute values in ΔP, suggesting that the DA is correcting for a missed precipitation event in one of both forcing datasets.
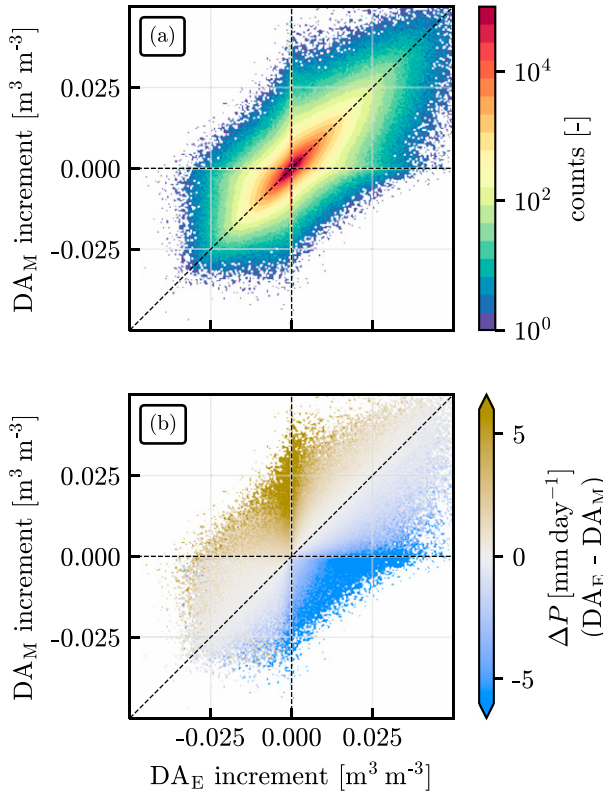
FIG. 11. (a) Bivariate distribution between increments of the DA_E and DA_M experiments for soil layer 1. (b) Same binning as in (a), but the color represents the average difference in rainfall ΔP between both meteorological forcings: brown indicates higher rainfall rates in DA_E and blue indicates higher rainfall rates in DA_M.

4. Conclusions

The combined active–passive ESA CCI SM product v7.1 was assimilated into the Noah-MP land surface model over Europe and adjacent regions to produce consistent estimates of surface and root-zone soil moisture as well as other land surface variables (incl. dynamic vegetation). The data assimilation (DA) system was first evaluated using in situ soil moisture data. We found that the distributions of anomaly correlations R_{anom} across various in situ sites are similar for open loop (OL) and DA experiments. However, in some regions, the DA improves soil moisture estimates (e.g., eastern Europe), whereas in other regions it deteriorates the skill (Scandinavia). We found that DA often leads to degradation of skill over locations that are dominated by forests, whereas the largest improvements can be found over croplands. The DA typically introduces large improvements at sites where the R_{anom} between the ESA CCI SM retrievals and the in situ soil moisture is high compared to the R_{anom} between the model-only simulations and in situ soil moisture, but at most sites, the model outperforms the ESA CCI SM. This may be attributed to the fact that the reanalysis forcings over Europe have a high quality due to an abundance of observations from

land stations, radiosondes, aircraft, etc., which help the model simulations.

A second assessment of the DA system used regional innovation and increment diagnostics. The bivariate distributions of increments in pairs of soil layers show regimes with strong and weak increment coupling between soil layers. A strong increment coupling is found in regions with high evapotranspiration, rainfall, and runoff values, i.e., outside of northern Africa, southern Europe, and Turkey.

In this study, we examined how the maximal amount of information from remotely sensed observations can be used by assessing the design choices of the DA system and tuning them for optimal performance. As such, we systematically examined the impact of three design choices for both the in situ and regional assessments outlined above:

1) The magnitude of the observation error S. This parameter has an important impact on the in situ soil moisture skill, with a trade-off between fostering large improvements at sites with high-quality retrievals (requiring small S) and limiting the deteriorations at sites with low-quality retrievals (requiring large S). A reference value of $S = 0.025$ m^3 m^{-3} was found to have a temporal standard deviation of the normalized innovations close to the ideal value of $\langle s_{\tilde{\nu}} \rangle = 1$. This value for S also yielded the largest median improvement in terms of ΔR_{anom} at the in situ sites. Additionally, we showed the impact of S on the distribution of the increments, which are narrower for larger S. A suggestion for future work is the exploration of temporally varying errors, as these may reflect the dataset properties more accurately and thus lead to better DA performance.

2) The observation rescaling via monthly or climatological CDF matching. Although this choice does not significantly impact the in situ evaluation results, we found that innovations and increments follow the desired temporal white noise sequence much more closely if the observations are rescaled on a monthly basis. The increments, and thus possible water budget imbalances, are smaller in the DA experiment with monthly rescaling.

3) The choice of meteorological forcing dataset. Separate OL and DA experiments were performed using ERA5 and MERRA-2 atmospheric reanalyses as forcings. We found that the skill improvements in the experiments forced by MERRA-2 are smaller, at least in part because in this scenario the R_{anom} values of the model-only simulations are better at the validation sites to begin with. No clear effect of the forcing was visible in the statistics of the innovations. The increments showed a stronger propagation to deeper soil layers in the DA experiment forced by ERA5 than for that forced with MERRA-2. One possible explanation may be that the ERA5 reanalysis simulates more heavy rainfall events that result in wetter soils over mainland Europe (not for northern Africa), which thus result in increments belonging to the regime of strong vertical increment coupling.

It should be noted that each of the three design choices has been explored independently from one another, whereas in

reality, we can expect to see interactions between them. For example, observation (and model) errors could be improved further by tuning them specifically for the meteorological forcing being used. The perturbations applied to the precipitation in particular (Table 1) would be chosen differently for land surface models forced by ERA5 and MERRA-2 in an optimal system.

In conclusion, our study demonstrates the potential benefits of assimilating the combined active–passive ESA CCI SM product into the Noah-MP LSM over Europe, with the optimization of key design choices leading to better estimates of soil moisture than can be achieved from satellite retrievals or models alone. This can in turn help in improving our understanding of the water, energy, and carbon cycles as these DA updates in soil moisture are linked to changes in other geophysical variables such as leaf area index and evapotranspiration.

Data availability statement. The datasets presented in this paper are too large to be retained or publicly archived with available resources. Documentation and methods used to support this study are available from zdenko.heyvaert@kuleuven.be at KU Leuven.

REFERENCES

Aires, F., P. Weston, P. de Rosnay, and D. Fairbairn, 2021: Statistical approaches to assimilate ASCAT soil moisture information—I. Methodologies and first assessment. *Quart. J. Roy. Meteor. Soc.*, **147**, 1823–1852, https://doi.org/10.1002/qj.3997.

Albergel, C., and Coauthors, 2017: Sequential assimilation of satellite-derived vegetation and soil moisture products using SURFEX_v8.0: LDAS-Monde assessment over the Euro-Mediterranean area. *Geosci. Model Dev.*, **10**, 3889–3912, https://doi.org/10.5194/gmd-10-3889-2017.

Bandhauer, M., and Coauthors, 2022: Evaluation of daily precipitation analyses in E-OBS (v19.0e) and ERA5 by comparison to regional high-resolution datasets in European regions. *Int. J. Climatol.*, **42**, 727–747, https://doi.org/10.1002/joc.7269.

Barbu, A.-L., J.-C. Calvet, J.-F. Mahfouf, and S. Lafont, 2014: Integrating ASCAT surface soil moisture and GEOV1 leaf area index into the SURFEX modelling platform: A land data assimilation application over France. *Hydrol. Earth Syst. Sci.*, **18**, 173–192, https://doi.org/10.5194/hess-18-173-2014.

Blyverket, J., P. D. Hamer, L. Bertino, C. Albergel, D. Fairbairn, and W. A. Lahoz, 2019: An evaluation of the EnKF vs. EnOI and the assimilation of SMAP, SMOS and ESA CCI soil moisture data over the contiguous US. *Remote Sens.*, **11**, 478, https://doi.org/10.3390/rs11050478.

Colliander, A., and Coauthors, 2017: SMAP/in situ core validation site land surface parameters match-up data version 1. NASA National Snow and Ice Data Center Distributed Active Archive Center, accessed 10 December 2021, https://doi.org/10.5067/DXAVIXLY18KM.

Crow, W. T., and J. D. Bolten, 2007: Estimating precipitation errors using spaceborne surface soil moisture retrievals. *Geophys. Res. Lett.*, **34**, L08403, https://doi.org/10.1029/2007GL029450.

——, and M. J. van den Berg, 2010: An improved approach for estimating observation and model error parameters in soil moisture data assimilation. *Water Resour. Res.*, **46**, W12519, https://doi.org/10.1029/2010WR009402.

De Lannoy, G. J. M., and R. H. Reichle, 2016a: Assimilation of SMOS brightness temperatures or soil moisture retrievals into a land surface model. *Hydrol. Earth Syst. Sci.*, **20**, 4895–4911, https://doi.org/10.5194/hess-20-4895-2016.

——, and ——, 2016b: Global assimilation of multiangle and multipolarization SMOS brightness temperature observations into the GEOS-5 catchment land surface model for soil moisture estimation. *J. Hydrometeor.*, **17**, 669–691, https://doi.org/10.1175/JHM-D-15-0037.1.

——, P. R. Houser, V. R. Pauwels, and N. E. C. Verhoest, 2006: Assessment of model uncertainty for soil moisture through ensemble verification. *J. Geophys. Res.*, **111**, D10101, https://doi.org/10.1029/2005JD006367.

de Rosnay, P., J. Muñoz-Sabater, C. Albergel, L. Isaksen, S. English, M. Drusch, and J.-P. Wigneron, 2020: SMOS brightness temperature forward modelling and long term monitoring at ECMWF. *Remote Sens. Environ.*, **237**, 111424, https://doi.org/10.1016/j.rse.2019.111424.

De Santis, D., D. Biondi, W. T. Crow, S. Camici, S. Modanesi, L. Brocca, and C. Massari, 2021: Assimilation of satellite soil moisture products for river flow prediction: An extensive experiment in over 700 catchments throughout Europe. *Water Resour. Res.*, **57**, e2021WR029643, https://doi.org/10.1029/2021WR029643.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 3385–3396, https://doi.org/10.1256/qj.05.108.

Dorigo, W., and Coauthors, 2017: 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.

——, and Coauthors, 2021: 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.

Dorigo, W. A., and Coauthors, 2011: The international soil moisture network: A data hosting facility for global in situ soil moisture measurements. *Hydrol. Earth Syst. Sci.*, **15**, 1675–1698, https://doi.org/10.5194/hess-15-1675-2011.

——, and Coauthors, 2013: Global automated quality control of in situ soil moisture data from the international soil moisture network. *Vadose Zone J.*, **12**, 1–21, https://doi.org/10.2136/vzj2012.0097.

Draper, C. S., R. H. Reichle, G. J. M. De Lannoy, and Q. Liu, 2012: Assimilation of passive and active microwave soil moisture retrievals. *Geophys. Res. Lett.*, **39**, L04401, https://doi.org/10.1029/2011GL050655.

Entekhabi, D., R. H. Reichle, R. D. Koster, and W. T. Crow, 2010: Performance metrics for soil moisture retrievals and

application requirements. *J. Hydrometeor.*, **11**, 832–840, https://doi.org/10.1175/2010JHM1223.1.

——, and Coauthors, 2014: SMAP handbook. NASA, 180 pp., https://smap.jpl.nasa.gov/system/internal_resources/details/original/178_SMAP_Handbook_FINAL_1_JULY_2014_Web.pdf.

Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, https://doi.org/10.1007/s10236-003-0036-9.

Farr, T. G., and Coauthors, 2007: The Shuttle Radar Topography Mission. *Rev. Geophys.*, **45**, RG2004, https://doi.org/10.1029/2005RG000183.

Friedl, M. A., and Coauthors, 2002: 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.

Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *J. Climate*, **30**, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1.

Gelb, A., and Coauthors, 1974: *Applied Optimal Estimation*. MIT Press, 384 pp.

Gruber, A., T. Scanlon, R. van der Schalie, W. Wagner, and W. Dorigo, 2019: Evolution of the ESA CCI soil moisture climate data records and their underlying merging methodology. *Earth Syst. Sci. Data*, **11**, 717–739, https://doi.org/10.5194/essd-11-717-2019.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Kerr, Y. H., and Coauthors, 2010: The SMOS mission: New tool for monitoring key elements of the global water cycle. *Proc. IEEE*, **98**, 666–687, https://doi.org/10.1109/JPROC.2010.2043032.

Koster, R. D., Z. Guo, R. Yang, P. A. Dirmeyer, K. Mitchell, and M. J. Puma, 2009: On the nature of soil moisture in land surface models. *J. Climate*, **22**, 4322–4335, https://doi.org/10.1175/2009JCLI2832.1.

Kumar, S. V., and Coauthors, 2006: Land Information System: An interoperable framework for high resolution land surface modeling. *Environ. Modell. Software*, **21**, 1402–1415, https://doi.org/10.1016/j.envsoft.2005.07.004.

——, R. H. Reichle, R. D. Koster, W. T. Crow, and C. D. Peters-Lidard, 2009: Role of subsurface physics in the assimilation of surface soil moisture observations. *J. Hydrometeor.*, **10**, 1534–1547, https://doi.org/10.1175/2009JHM1134.1.

——, and Coauthors, 2014: Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation. *J. Hydrometeor.*, **15**, 2446–2469, https://doi.org/10.1175/JHM-D-13-0132.1.

——, C. D. Peters-Lidard, J. A. Santanello, R. H. Reichle, C. S. Draper, R. D. Koster, G. Nearing, and M. F. Jasinski, 2015: Evaluating the utility of satellite soil moisture retrievals over irrigated areas and the ability of land data assimilation methods to correct for unmodeled processes. *Hydrol. Earth Syst. Sci.*, **19**, 4463–4478, https://doi.org/10.5194/hess-19-4463-2015.

——, M. Jasinski, D. M. Mocko, M. Rodell, J. Borak, B. Li, H. K. Beaudoing, and C. D. Peters-Lidard, 2019: NCA-LDAS land analysis: Development and performance of a multisensor, multivariate land data assimilation system for the national climate assessment. *J. Hydrometeor.*, **20**, 1571–1593, https://doi.org/10.1175/JHM-D-17-0125.1.

Lievens, H., and Coauthors, 2015: SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia. *Remote Sens. Environ.*, **168**, 146–162, https://doi.org/10.1016/j.rse.2015.06.025.

——, and Coauthors, 2017: Joint Sentinel-1 and SMAP data assimilation to improve soil moisture estimates. *Geophys. Res. Lett.*, **44**, 6145–6153, https://doi.org/10.1002/2017GL073904.

Liu, Y., W. Wang, and Y. Liu, 2018: ESA CCI soil moisture assimilation in SWAT for improved hydrological simulation in Upper Huai River Basin. *Adv. Meteor.*, **2018**, 7301314, https://doi.org/10.1155/2018/7301314.

Loew, A., M. Schwank, and F. Schlenz, 2009: Assimilation of an L-band microwave soil moisture proxy to compensate for uncertainties in precipitation data. *IEEE Trans. Geosci. Remote Sens.*, **47**, 2606–2616, https://doi.org/10.1109/TGRS.2009.2014846.

Maertens, M., G. J. M. De Lannoy, S. Apers, S. V. Kumar, and S. P. P. Mahanama, 2021: Land surface modeling over the Dry Chaco: The impact of model structures, and soil, vegetation and land cover parameters. *Hydrol. Earth Syst. Sci.*, **25**, 4099–4125, https://doi.org/10.5194/hess-25-4099-2021.

Mahfouf, J.-F., 2010: Assimilation of satellite-derived soil moisture from ASCAT in a limited-area NWP model. *Quart. J. Roy. Meteor. Soc.*, **136**, 784–798, https://doi.org/10.1002/qj.602.

Nachtergaele, F., and Coauthors, 2010: The harmonized world soil database. *Proc. 19th World Congress of Soil Science, Soil Solutions for a Changing World*, Brisbane, Australia, International Union of Soil Sciences, 34–37, https://edepot.wur.nl/154132.

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

Naz, B. S., S. Kollet, H.-J. H. Franssen, C. Montzka, and W. Kurtz, 2020: A 3 km spatially and temporally consistent European daily soil moisture reanalysis from 2000 to 2015. *Sci. Data*, **7**, 111, https://doi.org/10.1038/s41597-020-0450-6.

Niu, G.-Y., and Coauthors, 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *J. Geophys. Res.*, **116**, D12109, https://doi.org/10.1029/2010JD015139.

Olauson, J., 2018: ERA5: The new champion of wind power modelling? *Renewable Energy*, **126**, 322–331, https://doi.org/10.1016/j.renene.2018.03.056.

Owe, M., R. de Jeu, and T. Holmes, 2008: Multisensor historical climatology of satellite-derived global land surface moisture. *J. Geophys. Res. Earth Surf.*, **113**, F01002, https://doi.org/10.1029/2007JF000769.

Pal, M., and R. Maity, 2021: Assimilation of remote sensing based surface soil moisture to develop a spatially varying vertical soil moisture profile database for entire Indian mainland. *J. Hydrol.*, **601**, 126807, https://doi.org/10.1016/j.jhydrol.2021.126807.

Paloscia, S., S. Pettinato, E. Santi, C. Notarnicola, L. Pasolli, and A. Reppucci, 2013: Soil moisture mapping using Sentinel-1 images: Algorithm and preliminary validation. *Remote Sens. Environ.*, **134**, 234–248, https://doi.org/10.1016/j.rse.2013.02.027.

Reichle, R. H., and R. D. Koster, 2004: Bias reduction in short records of satellite soil moisture. *Geophys. Res. Lett.*, **31**, L19501, https://doi.org/10.1029/2004GL020938.

——, and ——, 2005: Global assimilation of satellite surface soil moisture retrievals into the NASA catchment land surface model. *Geophys. Res. Lett.*, **32**, L02404, https://doi.org/10.1029/2004GL021700.

——, and Q. Liu, 2014: Observation-corrected precipitation estimates in GEOS-5. NASA/TM-2014-104606, Vol. 35, 18 pp., https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20150000725.pdf.

——, D. B. McLaughlin, and D. Entekhabi, 2002: Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Wea. Rev.*, **130**, 103–114, https://doi.org/10.1175/1520-0493(2002)130<0103: HDAWTE>2.0.CO;2.

——, R. D. Koster, G. J. M. De Lannoy, B. A. Forman, Q. Liu, S. P. P. Mahanama, and A. Touré, 2011: Assessment and enhancement of MERRA land surface hydrology estimates. *J. Climate*, **24**, 6322–6338, https://doi.org/10.1175/JCLI-D-10-05033.1.

——, and Coauthors, 2017: Global assessment of the SMAP Level-4 surface and root-zone soil moisture product using assimilation diagnostics. *J. Hydrometeor.*, **18**, 3217–3237, https://doi.org/10.1175/JHM-D-17-0130.1.

——, and Coauthors, 2019: Version 4 of the SMAP Level-4 soil moisture algorithm and data product. *J. Adv. Model. Earth Syst.*, **11**, 3106–3130, https://doi.org/10.1029/2019MS001729.

——, Q. Liu, J. V. Ardizzone, W. T. Crow, G. J. M. De Lannoy, J. Dong, J. S. Kimball, and R. D. Koster, 2021: The contributions of gauge-based precipitation and SMAP brightness temperature observations to the skill of the SMAP Level-4 soil moisture product. *J. Hydrometeor.*, **22**, 405–424, https://doi.org/10.1175/JHM-D-20-0217.1.

Rodell, M., and Coauthors, 2004: The Global Land Data Assimilation System. *Bull. Amer. Meteor. Soc.*, **85**, 381–394, https://doi.org/10.1175/BAMS-85-3-381.

Ryu, D., W. T. Crow, X. Zhan, and T. J. Jackson, 2009: Correcting unintended perturbation biases in hydrologic data assimilation. *J. Hydrometeor.*, **10**, 734–750, https://doi.org/10.1175/2008JHM1038.1.

Seethala, C., and Coauthors, 2021: On assessing ERA5 and MERRA2 representations of cold-air outbreaks across the Gulf Stream. *Geophys. Res. Lett.*, **48**, e2021GL094364, https://doi.org/10.1029/2021GL094364.

Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling, 2010: 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.

Song, Y., and J. Wei, 2021: Diurnal cycle of summer precipitation over the North China Plain and associated land–atmosphere interactions: Evaluation of ERA5 and MERRA-2. *Int. J. Climatol.*, **41**, 6031–6046, https://doi.org/10.1002/joc.7166.

Srinivasan, G., A. Robock, J. K. Entin, L. Luo, K. Y. Vinnikov, and P. Viterbo, 2000: Soil moisture simulations in revised AMIP models. *J. Geophys. Res.*, **105**, 26 635–26 644, https://doi.org/10.1029/2000JD900443.

Taszarek, M., N. Pilguj, J. T. Allen, V. Gensini, H. E. Brooks, and P. Szuster, 2021: Comparison of convective parameters derived from ERA5 and MERRA-2 with rawinsonde data over Europe and North America. *J. Climate*, **34**, 3211–3237, https://doi.org/10.1175/JCLI-D-20-0484.1.

Wagner, W., G. Lemoine, and H. Rott, 1999: A method for estimating soil moisture from ERS scatterometer and soil data. *Remote Sens. Environ.*, **70**, 191–207, https://doi.org/10.1016/S0034-4257(99)00036-X.

Yang, Z.-L., and Coauthors, 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins. *J. Geophys. Res.*, **116**, D12110, https://doi.org/10.1029/2010JD015140.