



OPEN

DATA DESCRIPTOR

A 3 km spatially and temporally consistent European daily soil moisture reanalysis from 2000 to 2015

Bibi S. Naz^{1,2}✉, Stefan Kollet^{1,2}, Harrie-Jan Hendricks Franssen^{1,2}, Carsten Montzka^{1,2} & Wolfgang Kurtz³

High-resolution soil moisture (SM) information is essential to many regional applications in hydrological and climate sciences. Many global estimates of surface SM are provided by satellite sensors, but at coarse spatial resolutions (lower than 25 km), which are not suitable for regional hydrologic and agriculture applications. Here we present a 16 years (2000–2015) high-resolution spatially and temporally consistent surface soil moisture reanalysis (ESSMRA) dataset (3 km, daily) over Europe from a land surface data assimilation system. Coarse-resolution satellite derived soil moisture data were assimilated into the community land model (CLM3.5) using an ensemble Kalman filter scheme, producing a 3 km daily soil moisture reanalysis dataset. Validation against 112 *in-situ* soil moisture observations over Europe shows that ESSMRA captures the daily, inter-annual, intra-seasonal patterns well with RMSE varying from 0.04 to 0.06 m³m⁻³ and correlation values above 0.5 over 70% of stations. The dataset presented here provides long-term daily surface soil moisture at a high spatiotemporal resolution and will be beneficial for many hydrological applications over regional and continental scales.

Background & Summary

Soil moisture (SM) is characterized by complex dynamics across a wide range of spatial and temporal scales¹ that can impact hydrological processes such as runoff, evaporation and transpiration from vegetation through changing soil moisture². As a result, accurate characterization of spatial distribution and temporal variations of SM is critical for many regional-scale applications, including weather predictions, subsurface hydrology, flood forecasting, drought monitoring, agriculture and climate change impact studies^{1–3}. However, SM remains a difficult variable to obtain over large scale with reasonable temporal and spatial resolution, because there are no high-resolution soil moisture observations available at the continental scale and observations of SM from measurements are very sparse. While remote sensing (RS) products give reliable estimates and cover large areas^{4–6}, most long-term soil moisture data from spaceborne remote sensors have relatively low spatial resolution (in the order of 25 to 50 km) and they are spatially and temporally discontinuous^{7,8}. An alternative source of high-resolution SM estimates is from land surface models (LSM). However, predictions are often poor due to inadequate model physics, poor parameter estimates and erroneous atmospheric forcings⁹. Soil moisture reanalysis products are therefore needed which can provide downscaled estimates of SM with complete spatiotemporal coverage by merging coarse-resolution SM observations with a high resolution LSM using data assimilation (DA) techniques^{3,10–13}. These products overcome the shortcomings of sparse spatial and temporal distributions in observations and provide a better estimate of SM than obtained only by modeling or by satellite observations alone. Soil moisture reanalysis provides unique and consistent datasets for studying complex spatial patterns of SM from regional to global scales and temporal variability from daily to annual scales^{1,14}. Moreover, the relationship to other essential climate variables, such as runoff or evapotranspiration, can be investigated in more detail. It can also be used as initial input for climate change analysis and numerical simulations and for cross-validation of SM outputs in modeling studies.

¹Jülich Research Center GmbH, Institute of Bio- and Geosciences: Agrosphere (IBG 3), Jülich, 52425, Germany. ²Centre for High-Performance Scientific Computing in Terrestrial Systems, Geoverbund ABC/J, Jülich, 52425, Germany. ³Leibniz Supercomputing Centre, Boltzmannstr. 1, 85748, Garching, Germany. ✉e-mail: b.naz@fz-juelich.de

Product	ERA5	MERRA2-land	GLEAM-3.2a	GLDAS-2.1	NCEP-CFSR
Time period	1979–present	1980–present	1980–2018	2000–2019	1979–present
Spatial resolution (km ²)	~31	~50	25	25	50
Spatial coverage	Global	Global	Global	Global	Global
Temporal resolution	Hourly	Hourly	Daily	3 hourly	3 hourly
Assimilation scheme	4D-VAR	3D-VAR	Newtonian nudging	EnKF	3D-VAR

Table 1. Main characteristics of commonly used long-term soil moisture global datasets from reanalysis products.

Several commonly used long-term soil moisture global datasets exist from land surface DA systems^{15–17}, as listed in Table 1. The focus of these reanalysis systems has been on the assimilation of meteorological observations, except for the Global Land Evaporation Amsterdam Model (GLEAM, v3.2a)¹⁸ product which assimilate surface SM. The overall goal of these products is to provide estimates of atmospheric, land and oceanic climate variables. At the European regional scale, there have been few studies which provide soil moisture reanalysis through DA techniques by assimilating surface soil moisture information from satellite into land surface models^{10,19–23}. Though these global and regional reanalysis products are an attractive data source, they have a relatively coarse resolution (typically at 25–50 km grid spacing) and may not provide locally representative information of soil moisture which is important for regional hydrologic and agriculture applications.

The land surface DA system CLM-PDAF consisting of the Community Land Model (CLM)²⁴ and the Parallel Data Assimilation Framework (PDAF)^{25,26} was used to utilize the coarse resolution satellite soil moisture data to update the soil moisture estimates in the land surface model. The DA structure in CLM-PDAF allows to directly ingest remotely sensed observations of land surface conditions to produce accurate, spatially and temporally consistent fields of land surface states, with reduced uncertainties through an ensemble based DA method. Recently, the European Space Agency Climate Change Initiative (ESACCI)²⁷ provides a homogeneous and the longest time series of SM data to date, covering the period 1979–2018, and has been widely used for various Earth system research^{28–30} and in DA studies^{19,31,32}. We selected ESACCI SM data for DA because of its availability at longer timescales, which also makes it possible to construct a long-term high-resolution SM reanalysis at continental scale.

Using CLM-PDAF, the daily SM data at 0.25° resolution from ESACCI were assimilated into CLM using an ensemble Kalman filter (EnKF) DA method^{33,34} producing the first 3 km European SSM reanalysis (called ESSMRA hereafter) dataset. Figure 1 shows the schematic flow of the CLM-PDAF setup to develop the ESSMRA dataset. The purpose of ESSMRA is to provide a long-term (2000–2015) spatially and temporally consistent data source with high spatiotemporal resolution (3 km, daily) and high quality for the research community to use in hydrological and climate applications as well as to study the spatial/temporal variability of SM over Europe. The relatively longer time scale and fine spatial resolution of this new European gridded ESSMRA dataset could provide a valuable data source for many hydrological applications over larger regimes and to regional and continental scale studies.

Methods

The 3 km ESSMRA was generated using three main steps: (1) the regional land surface model setup over Europe, (2) implementation of a DA framework, and (3) validation of ESSMRA based on observations and other reanalysis products.

Land surface model, parameters and data. The community land model (CLM), available through the National Centre of Atmospheric Research (NCAR) as part of the Community Climate System Model (CCSM) is used in this soil moisture reanalysis system. The coupled land surface data assimilation system (CLM-PDAF) uses the CLM version 3.5 (CLM3.5) which offers significant improvements in estimating the components of the terrestrial water cycle compared to earlier versions (i.e. CLM2.0 and 3.0)²⁴. Later versions of CLM (4.0 and 4.5) may be used in the future, however, a previous study showed that the differences between CLM3.5 and later versions of CLM (4.0 and 4.5) with respect to soil moisture variability remained small when compared to observations³⁵. The CLM3.5 simulates the hydrological cycle over land by taking into account interception of water by plants, throughfall, infiltration, runoff, soil water and accumulation and melting of snow. In CLM3.5 the soil profile is divided into 10 soil layers (0–3.8 m). The input soil texture information (sand fraction, clay fraction) is available for the surface layer only. For simplicity, sand fraction and clay fraction information for 19 soil classes in the first layer were also used for the deeper layers. The movement of moisture between these layers is calculated using Richard's equation. The bottom soil layer is also coupled with an unconfined aquifer to account for groundwater recharge and discharge processes²⁴. The current setup of CLM3.5 does not consider different geological conditions of bedrock.

To account for land surface variability within a grid cell, a CLM grid cell consists of one or more columns to capture surface heterogeneity through land unit (e.g. glacier, wetland, lake, and vegetation). The vegetated fraction can be further divided into 17 different plant functional types (PFTs). The water and energy balance equations are solved for each land cover type and aggregated back to the grid cell level. CLM requires several static surface input parameters related to vegetation, soils and topography³⁶. Table 2 lists the source for each input parameter used in this study. The land cover information for each PFT in our model setup was estimated based on the Moderate Resolution Imaging Spectro radiometer (MODIS) MCD12Q1 (version5) land cover product³⁷. The

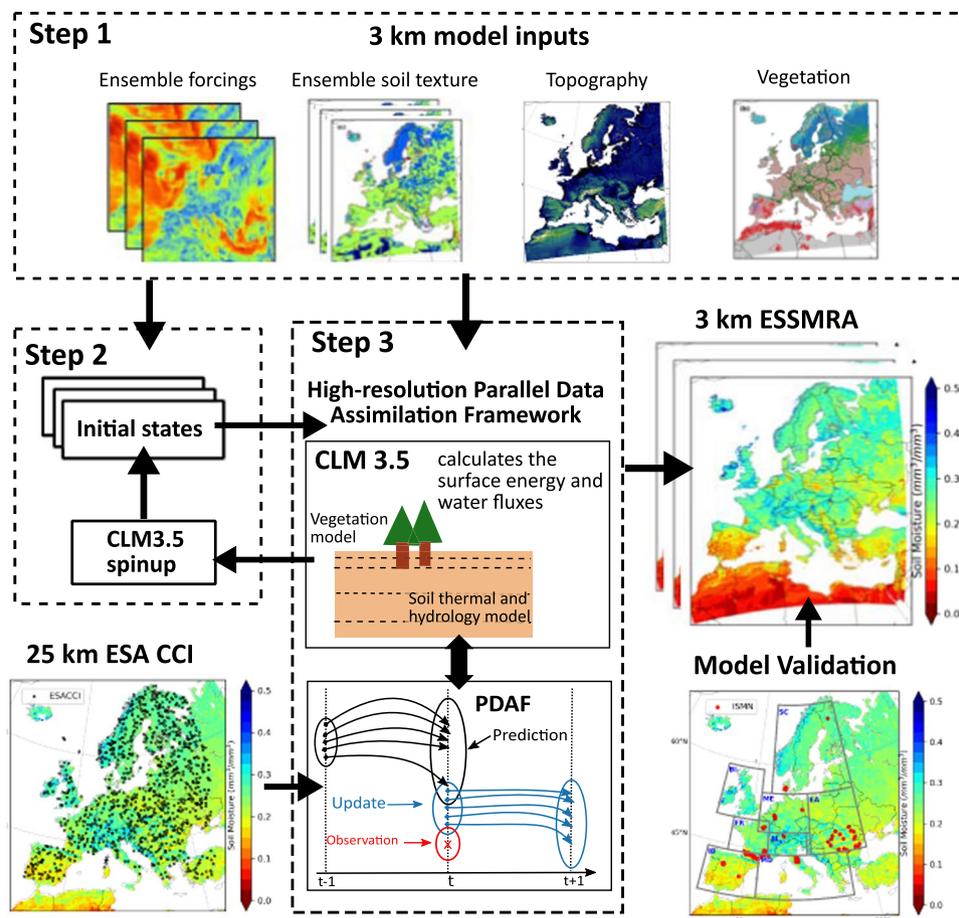


Fig. 1 Schematic of CLM-PDAF workflow adopted to generate high resolution ESSMRA product. The first step was to prepare an ensemble of 3 km input data for the CLM model for the EU-CORDEX domain. Second, CLM was initialized for each ensemble member with the equilibrium initial state variables. In the third step, selected ESA CCI soil moisture observations (resolution 25 km) were assimilated into CLM using the Ensemble Kalman filter to generate the 3 km ESSMREA dataset.

Dataset	Spatial resolution (km)	Temporal Resolution	Source	model input
COSMO-REA6	6	Hourly	ftp://opendata.dwd.de	Meteorological Inputs
MODIS (MCD12Q1)	0.5	Yearly		Land Cover
GLASS LAI	1	Weekly	http://glcf.umd.edu	LAI
FAO Soil	10	NA	http://www.fao.org	Soil Texture
GMTED2010	1	NA	https://www.usgs.gov/	DEM

Table 2. Model input datasets used to generate ESSMRA.

1 km Global Land Surface Satellite (GLASS) LAI (leaf area index) product³⁸ was used to estimate 12 monthly LAI values for every 3 km grid cell which allow spatially distributed monthly LAI values for each PFT. Additionally, yearly model runs were performed where the LAI information was updated at the start of each year to account for annual variability in LAI. Additional properties such as the stem area index and the monthly heights of each PFT were calculated based on the global CLM3.5 surface dataset²⁴. Soil texture data such as sand and clay percentages were determined for 19 soil classes derived from the Food and Agricultural Organization (FAO) database³⁹ and soil characteristics dataset developed by Miller *et al.*⁴⁰. Topography data were acquired from the 1 km Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010)⁴¹.

The atmospheric forcing data such as solar radiation, temperature, pressure, near-surface wind speed, specific humidity and precipitation rate were prepared using the regional atmospheric reanalysis COSMO-REA6 dataset⁴² from the Hans Ertel Centre for Weather Research (HErZ)⁴³. It is based on the numerical weather prediction COSMO (Consortium for Small Scale Modelling) model⁴⁴ and spans the period 1995–2017 with hourly data of atmospheric variables at 0.055° (~6 km) over Europe. The COSMO-REA6 was corrected through the

assimilation of observational meteorological data using the existing nudging scheme in COSMO model with boundary conditions from ERA-Interim data. A more comprehensive description of the dataset is available in previous studies^{42,45}.

The SM satellite observations from the combined ESACCI dataset at 0.25° resolution were used for the DA experiment, as described in the next section. In the past decade, a number of different satellite missions have been launched to provide SM retrievals with high temporal resolution over large regions. Examples are Soil Moisture and Ocean Salinity^{46,47} (SMOS; launched in 2009) and the Soil Moisture Active Passive (SMAP, launched in 2015) missions⁴⁸. The complete list of operational remotely sensed surface SM products is also given in Babaeian *et al.*⁸. These recent data products are only available for the last decade and cannot be used to apply soil moisture information in a land surface reanalysis for extended time periods. The combined ESACCI product which combines both active and passive microwave sensors provides large spatiotemporal coverage and offers a good opportunity to improve LSM estimates with DA techniques. In this merged product, the absolute soil moisture values are rescaled to common climatology using soil moisture estimates from GLDAS-Noah model through CDF-matching method. The quality of the ESACCI SM product has been evaluated on a global scale by several studies²⁸. By comparing the ESACCI SM dataset with *in-situ* measurements, a previous study found that the product was able to capture the annual cycle of SM and its short-term variability²⁹. The quality of the product has been shown to increase with time due to the addition of new satellites and methods used to merge them²⁸.

Generation of ESSMRA dataset using CLM-PDAF. For the generation of the ESSMRA dataset, we used the CLM-PDAF framework, in which PDAF is coupled with CLM for soil moisture assimilation. The ESSMRA product was generated by performing three main steps as shown in Fig. 1. First, CLM3.5 was implemented for the EURO-CORDEX domain with a spatial resolution of 0.0275° (3 km), inscribed into the official EUR11 grid. The model was driven with COSMO-REA6 reanalysis dataset for the time period 2000–2015. To match the spatial resolution of CLM3.5 setup, the COSMO-REA6 dataset at 6 km resolution was re-gridded to 0.0275° (3 km) using the first-order conservative interpolation method⁴⁹. In the second step, a 30 years spin-up run (simulating time period of 2000–2006 five times) of CLM forced by atmospheric fields was carried out to obtain equilibrium initial state variables which were used to initialize the model. In the third step, the model was run for 2000–2015 at hourly time step with the assimilation of ESACCI soil moisture data into the model once a day using the EnKF algorithm. The EnKF algorithm uses ensembles of model simulations to approximate the model state and parameter error covariance matrix in order to optimally merge model predictions with observations^{3,50–52}. The PDAF is designed for high-performance computing infrastructures and can efficiently cope with the high computational burden of ensemble-based DA^{25,26}. Because of this feature, it was possible to produce a pan-European longer-term and high spatial resolution land surface DA product. To generate ensembles of forecast states, we perturbed the precipitation and the soil parameters (sand and clay percentage) by applying log normally distributed multiplicative perturbations (with a mean of 1 and standard deviation of 0.15) to the precipitation field and random noise drawn from spatially uniform distribution ($\pm 10\%$) to the sand and clay content, respectively. In the present study, we only updated the soil moisture state variable and kept the soil texture constant for individual ensemble members throughout our simulations instead of joint state and parameter updates of soil moisture and soil texture in the DA approach as used in our previous study³¹. The ensemble size was set to 20 in our assimilation experiment using similar methodology as used in Naz *et al.*³¹. Our initial study found slightly improved SM estimates when ensemble size was increased from 12 to 20³¹. However, increasing ensemble size is quite challenging for such a large-scale high-resolution model because of the memory and storage requirements. In the DA approach, another challenge is the spatial mismatch between coarse-resolution satellite data and high-resolution hydrologic models. To address the resolution mismatch between the ESACCI data and model (i.e. 0.25° and 0.0275°, respectively), the ESACCI grid cell nearest to the model grid cell was identified and considered as an independent data point for DA. While this approach avoids the additional step of downscaling the satellite data to model resolution, it may smooth out the high-resolution features of the LSM. In future multiscale assimilation (i.e. to update various model grid cells covered by a satellite observation) of the ESACCI SM data into CLM could be explored⁵³. Another limitation associated with our method is that due to the large number of grid cells (1544 × 1590) and required computational resources, it was not possible to assimilate all of the data available from the ESACCI satellite data into CLM. In the current framework, we randomly selected 1000 grid cells (5% of total grid cells over land). The SM observations at the selected grid cells were then assimilated into the CLM model. The non-assimilated data were used in the model validation step. This approach allowed us to evaluate the impact of DA at other locations where the data were not assimilated. However, it should be noted that this approach might negatively affect the SM estimates that are further apart from the assimilated grid cells because of the use of global EnKF in our approach. In the future, the local ensemble transform Kalman filter (LETKF) could be used to avoid this limitation. The strength and limitations associated with our methods are also discussed in detail in Naz *et al.*³¹.

Using the above setup, DA experiments were conducted using CLM-PDAF over Europe (Fig. 2). We selected 2000–2015 as our period of analysis because of the availability of most model input data in our experiment for this time period. This experiment allowed us to generate a 16-year high resolution ESSMRA dataset at daily time scale. A second experiment was also performed to evaluate the impact of DA using the same model setup, but without assimilating the ESACCI observations into the model. We referred to these experiments as “CLM-DA” (data assimilation) and “CLM-OL” (open loop simulation), respectively.

Observation and reanalysis products. The *in-situ* soil moisture data from 11 networks across Europe were acquired from the International Soil Moisture Network (ISMN)⁵⁴, which provides globally available soil moisture measurements. The surface soil moisture data from 112 stations for the top 5 to 10 cm surface layer were collected to evaluate the ESSMRA product in the top two CLM soil layers (about 3 cm). *In-situ* data were collected for 2000–2015, but their availability does not necessarily cover the whole period. For comparison with ESSMRA

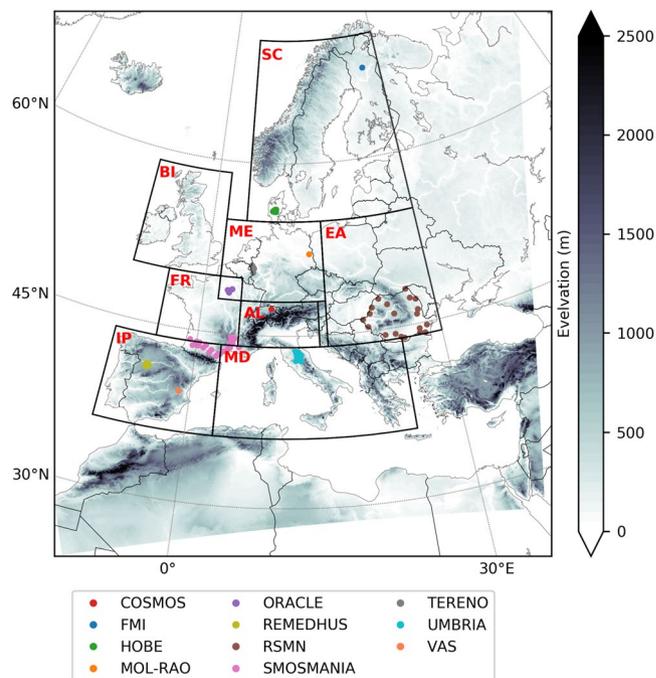


Fig. 2 Map of EURO-CORDEX domain (1544×1592 grid cells) showing surface elevation, boundaries of PRUDENCE regions and locations of International Soil Moisture Network (ISMN) stations used for data validation. Black boxes correspond to PRUDENCE regions with abbreviated letters in red colour indicating names of the regions (FR: France, ME: Mid-Europe, SC: Scandinavia, EA: Eastern Europe, MD: Mediterranean, IP: Iberian Peninsula, BL: the British Isles, AL: Alpine region). The solid dots represent locations of the *in-situ* ISMN soil moisture stations.

Network	Country	# of Stations	Soil depth (m)	Start date	End date
COSOMO	Austria	1	0.0–0.1	2010-12-13	2015-12-31
FMI	Finland	1	0.0–0.05	2011-09-18	2015-12-31
HOBE	Denmark	23	0.0–0.05	2009-09-08	2015-01-01
MOL-RAO	Germany	1	0.0–0.08	2003-01-01	2014-01-01
ORACLE	France	3	0.0–0.05	2000-01-07	2013-09-09
REMEDHUS	Spain	24	0.0–0.05	2005-03-15	2015-12-31
RSMN	Romania	20	0.0–0.05	2014-04-09	2015-12-31
SMOSMANIA	France	21	0.0–0.05	2007-01-01	2015-12-31
TERENO	Germany	5	0.0–0.05	2009-12-31	2015-12-31
UMBRIA	Italy	13	0.0–0.05	2002-10-09	2015-12-31
VAS	Spain	2	0.0–0.05	2010-01-10	2012-01-01

Table 3. Information of selected *in-situ* soil moisture stations used in the study.

daily estimates, the measurements with hourly time scale were aggregated to daily time scale. If more stations are located within one 3 km grid cell, the average of those stations was used for comparison. The characteristics of the selected *in-situ* networks are presented in Table 3.

The ESSMRA data were also compared with other global soil moisture reanalysis products from the European Centre for Medium-Range Weather Forecasts Reanalysis 5 (ERA5)¹⁷, Global Land Data Assimilation System (GLDAS)¹⁵ and Global Land Evaporation Amsterdam Model (GLEAMv3.2a)¹⁸ available at 0.25° resolution and hourly temporal resolution. The aim is to understand the spatiotemporal patterns of ESSMRA relative to other existing reanalysis products and how ESSMRA differs from other SM reanalysis products at regional scale. The SM from these reanalysis datasets have been widely used by many studies^{18,55–62}. For instance, a comparison of GLDAS and ERA5 SM with *in-situ* data in Europe showed good agreement in terms of temporal dynamics of SM data⁶². Similarly, the SM estimates from GLEAM (v3) compared well with *in-situ* surface SM and was further improved than its previous versions with assimilation of surface soil moisture observations into the GLEAM¹⁸.

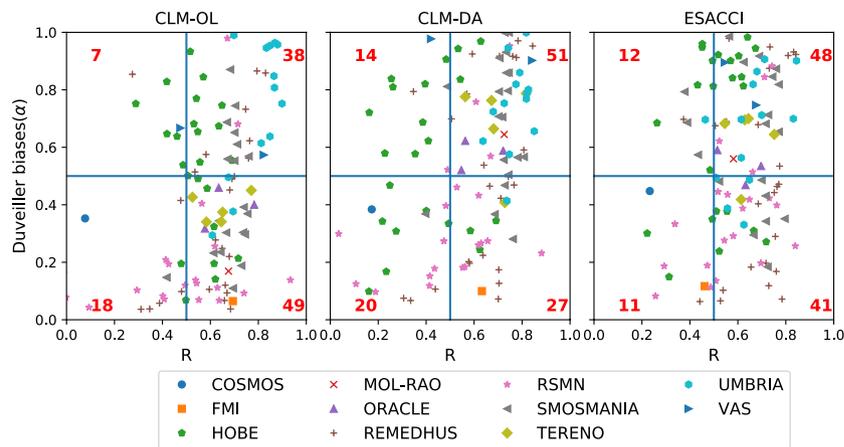


Fig. 3 Scatter plot of R and α of surface SM from CLM-DA, CLM-OL and ESACCI against *in-situ* observations used in this study. Horizontal and vertical lines divide the graph into four quadrants based on thresholds of R and α . The total number of stations falling into each category is labelled in each quadrant of the graph. The stations are color-coded by soil moisture network.

Data Records

The ESSMRA dataset in NetCDF format is freely available for download from PANGAEA data repository⁶³ as well as at the Jülich Supercomputing Centre data repository⁶⁴. The dataset consists of ensemble mean of daily surface soil moisture for the period of 2000–2015 and is available at a monthly temporal frequency using the following naming convention as:

EU_ESSMRA_daily_ensmean_CLM-PDAF_3Km_v1.yyyymm.nc.

The netcdf files contain the variable “H2OSOIL” which is the volumetric soil moisture at the 0–3 cm layer [m^3m^{-3}]. For example the file, EU_ESSMRA_daily_ensmean_CLM-PDAF_3Km_v1.200101.nc contains the daily soil moisture values for the month of January 2001. Each file also contains the definition of the geographical coordinate system of the grid (latitudes, longitudes and rotated pole).

Technical Validation

The newly developed ESSMRA dataset was validated at different spatiotemporal scales in four steps. First, the ESSMRA dataset was validated using independent stations data. Second, the performance of ESSMRA was evaluated at regional scale with respect to ESACCI data to assess the impact of DA at other locations where the data were not assimilated. Third, the ability of ESSMRA to capture the monthly and yearly climatologies for different regions in Europe was evaluated against existing global reanalysis products (ERA5, GLDAS and GLEAM). Finally, ESSMRA was compared with ERA5, GLDAS and GLEAM reanalysis products to understand the spatial variability of SM at the European scale.

For evaluation against *in-situ* station measurements and other commonly used SM products we used the Pearson correlation coefficient (R), root mean square error (RMSE), unbiased root-mean-square error (ubRMSE) and a metric α proposed by Duveiller *et al.*⁶⁵ which represents additive/multiplicative bias between two datasets. For α , 0 represents full bias and 1 indicates no bias. For this validation, we extracted the ESSMRA data to the nearest location of the station. However, if more stations are located within one 3 km grid cell, we used the average of those stations. For regional analysis, the results were presented for eight predefined analysis regions from the “Prediction of Regional scenarios and Uncertainties for Defining European Climate change risks and Effects” (PRUDENCE) project⁶⁶ (FR: France, ME: Mid-Europe, SC: Scandinavia, EA: eastern Europe, MD: Mediterranean, IP: Iberian Peninsula, BI: British Isles, AL: Alpine region) as shown in Fig. 2. We referred to these regions as the “PRUDENCE” regions.

For ESSMRA validation, the average of simulated SM in the top two layers (i.e., at 0.007 and 0.03 m depth) of CLM was used. Because of high computational cost and storage requirement associated with implementing the continental scale 3 km integrated hydrologic and DA framework, currently, we only analysed the top 3 cm soil moisture, which is the limitation of this study. However, prior study⁶⁷ showed no major differences for latent heat flux estimation using information content of surface SM or enhanced with soil moisture in deeper layers. With advancing capabilities in computing and storage, ESSMRA dataset can be extended to root zone data analysis using CLM-PDAF in the future.

Validation using independent *in-situ* station observations. The surface SM from CLM-DA and CLM-OL experiments were validated against *in-situ* measurements and also compared with the ESACCI satellite merged product (shown in Fig. 3). The surface SM observations were obtained from the ISMN database using all the data available over Europe between 2000 and 2015 (Table 3) and were used for the independent validation. The Pearson correlation coefficient (R) and α for each product were calculated at the *in-situ* station locations.

Networks	R			alpha			RMSE			ubRMSE		
	min	average	max	min	average	max	min	average	max	min	average	max
REMEDHUS (24)	0.29	0.63	0.84	0.07	0.50	0.97	0.05	0.12	0.19	0.04	0.06	0.15
HOBE(23)	-0.01	0.31	0.69	0.11	0.56	0.93	0.05	0.09	0.23	0.04	0.06	0.12
SMOSMANIA (21)	0.40	0.70	0.79	0.28	0.67	1.00	0.04	0.09	0.17	0.04	0.06	0.09
RSMN(20)	0.01	0.49	0.88	0.09	0.34	0.95	0.06	0.12	0.18	0.04	0.06	0.08
UMBRIA(13)	0.63	0.75	0.82	0.40	0.74	1.00	0.04	0.09	0.16	0.04	0.07	0.12
TERENO(5)	0.57	0.67	0.79	0.42	0.68	0.77	0.06	0.07	0.09	0.05	0.06	0.07
ORACLE(3)	0.55	0.61	0.71	0.51	0.55	0.62	0.10	0.11	0.12	0.06	0.08	0.09
VAS(2)	0.41	0.63	0.84	0.90	0.94	0.98	0.05	0.06	0.07	0.04	0.05	0.07
COSMOS (1)	NA	0.09	NA	NA	0.35	NA	NA	0.12	NA	NA	0.08	NA
FMI (1)	NA	0.01	NA	NA	0.08	NA	NA	0.16	NA	NA	0.05	NA
MOL-RAO(1)	NA	0.70	NA	NA	0.64	NA	NA	0.08	NA	NA	0.05	NA

Table 4. Statistical scores for the comparison between ESSMRA and *in-situ* SM for all 112 stations available during 2000–2015.

These statistics were computed taking all station measurements into account for the period 2000–2015. Figure 3 shows the scatter plot of these scores against *in-situ* data for CLM-DA, CLM-OL and ESACCI.

Using a threshold of 0.5 for both R and α , stations fell into four categories, (1) stations with higher agreement (i.e. both R and α are greater than 0.5), (2) stations with higher R value but lower α values (i.e. there is higher difference in magnitude but temporal dynamics match well), (3) stations with higher values of α and low R values, and (4) stations with both low values of R and α . The thresholds for both R and α are based on the general rule of thumb that $R > 0.5$ represent moderate to high correlation. The performance of CLM-DA experiments in comparison to CLM-OL experiment and ESACCI data was evaluated in each of these four categories for bias and correlation. Based on these thresholds, Fig. 3 shows that SM from CLM-DA is in good agreement with observations over half of the stations (i.e. 51 out of 112 fell into category 1) for both R and α having values greater than 0.5 whereas CLM-OL and ESACCI shows higher agreement with observation for 38 and 48 stations, respectively. Overall CLM-DA performs well in matching the magnitude (i.e. α values > 0.5) over 58% of stations and higher correlation values for 70% of the stations, while ESACCI shows α values greater than 0.5 for 53% of the stations but exhibits higher correlation values (i.e. > 0.5) for 80% of the stations. CLM-DA did a poor job for 17% of the stations where both magnitude and temporal dynamics are not matching well with observations (i.e. α and R values < 0.5). These locations were mostly for the *in-situ* stations of RSMN and HOBE networks, where ESSMRA SM is overestimated with respect to the stations data. This overestimation is also reported in other studies^{68,69} as well based on the comparison with satellite-based SM products.

Table 4 shows statistical scores for ESSMRA at all *in-situ* locations across all networks. On average, correlations ranged from 0.01 to 0.70 while α values ranged from 0.08 to 0.94. The average ubRMSE and RMSE values ranged from 0.04 to 0.06 m^3m^{-3} and 0.05 to 0.12 m^3m^{-3} , respectively. A potential cause for lower correlation values of ESSMRA for some of the networks such as FMI, HOBE (located in Scandinavian region) and COSMOS (located in Alpine region) might be related to the limitations and uncertainties of ESACCI retrieval algorithm which is sensitive to vegetation, frozen soil and complex topography²⁷. The SM in these areas is also influenced by soil freezing and thawing processes, dense forest, soil organic matter and the presence of numerous water bodies and bogs^{69,70}. These processes are not well represented in the land surface models.

There are also some caveats regarding the use of *in-situ* observations for validation of model estimates. First, because of the differences in the spatial representativeness between different products, it is complicated to evaluate the coarser resolution product against point measurements i.e. the local measurement may not properly represent the large-scale average. For example at the point scale point, measurements cover $\sim 1 \text{ dm}^3$, while the model has a grid resolution of approximately 3 km. Second, CLM near-surface soil moisture variable represents an average over the top 5 cm of soil, whereas the *in-situ* measurements do not represent such a depth average, the surface soil moisture measurements instead represent conditions at a depth of about 5 cm. The spatial representativeness error and the vertical mismatch between the *in-situ* measurements and the modeled soil moisture variable will influence the skill metrics we computed for the validation.

Despite these issues, ESSMRA product shows overall good agreement with *in-situ* observations at daily time scale as shown in Fig. 4. Because of the issues stated above with spatial representativeness, the average of the *in-situ* observations of all stations was compared with the averaged soil moisture of all grids within each ISMN listed network. This comparison shows that at daily scale the model is able to reproduce the daily variations in soil moisture fairly well, except for the RSMN network. The overestimation of ESSMRA SM data over RSMN is in line with findings by previous studies^{68,69}.

Regional validation using ESACCI SM data. The ESACCI SM data, which were excluded from the DA procedure, were used to evaluate the impact of assimilating soil water content (m^3m^{-3} ; SWC) on model estimated SM at regional scale. We calculated the skill score (SS) of RMSE, ubRMSE, R and α for SWC using following equations:

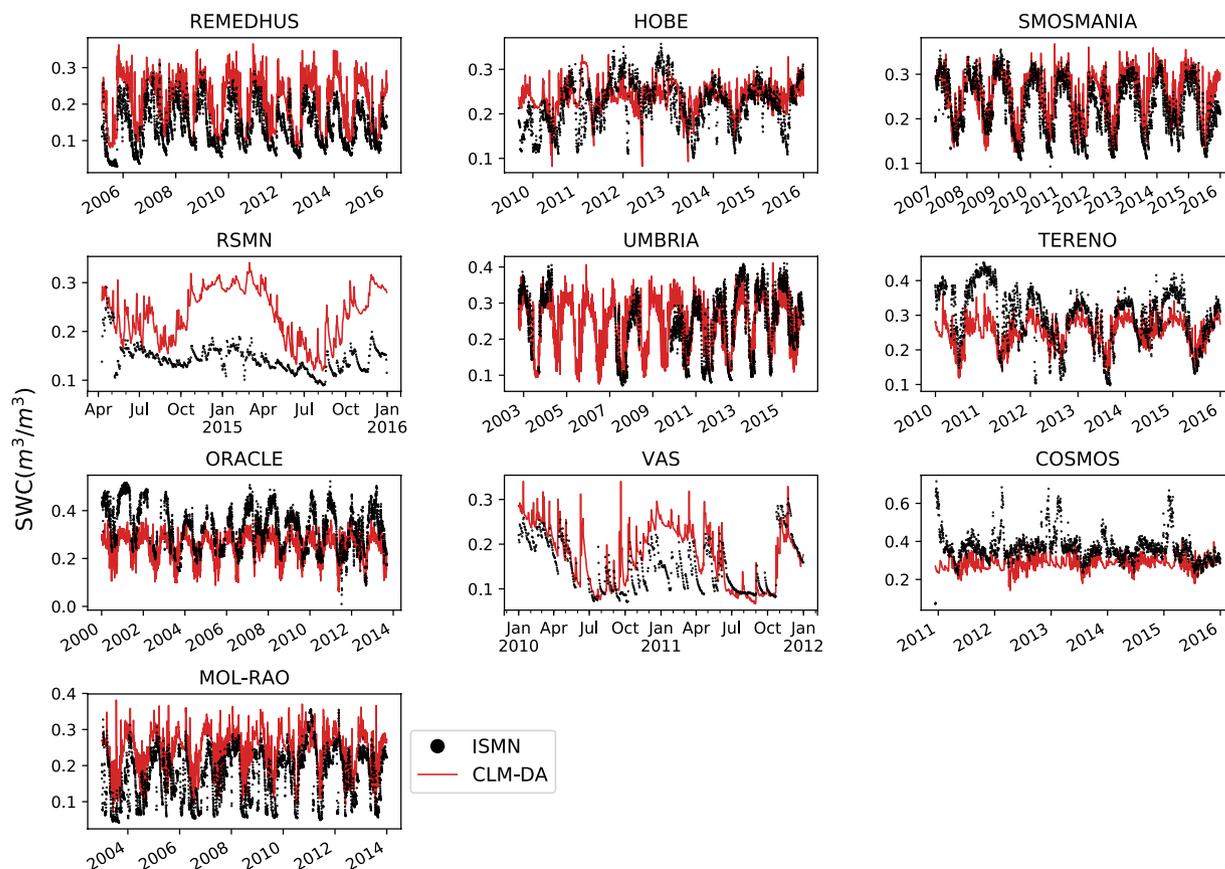


Fig. 4 Comparison of daily time series of soil water content (m^3/m^3) from CLM-DA and *in-situ* observations from the ISMN networks. The average of the *in-situ* observations of all stations within the ISMN network was first calculated and then compared with the averaged soil moisture of all grids within the same ISMN network.

$$SS_{RMSE} = 1 - \frac{DA_{RMSE}}{OL_{RMSE}} \quad (1)$$

$$SS_{ubRMSE} = 1 - \frac{DA_{ubRMSE}}{OL_{ubRMSE}} \quad (2)$$

$$SS_R = R_{DA} - R_{OL} \quad (3)$$

$$SS_\alpha = \alpha_{DA} - \alpha_{OL} \quad (4)$$

Positive SS values in Eq. (1) to Eq. (4) indicate improvement as a result of DA relative to the open loop, while $SS < 0$ indicates a degradation in assimilation results.

The impact of data assimilation on CLM-DA model performance is illustrated in Fig. 5 on the basis of skill scores for RMSE, ubRMSE, R and α for spatially averaged SWC over the PRUDENCE regions. For CLM-DA, assimilation of ESACCI observation shows positive SS_{ubRMSE} (i.e. reduced RMSE against CLM-OL) over all PRUDENCE regions with significant improvement over FR region and least improvement over SC region (Fig. 5a). However, assimilation of ESACCI had little impact on the model performance to capture the observed temporal variations in SWC as indicated by small positive values of SS_{ubRMSE} (ranges from 0.1 to 0.2) over regions FR, ME, AL and EA and negative values over BI, IP and MD regions (i.e. -0.02 , -0.05 and -0.08 , respectively; Fig. 5b). Figure 5c shows that assimilating ESACCI overall give little improvement in terms of correlation ($SS_R > 0.02$) with slightly degraded correlation over the BI region (-0.01). However, most regions show significant improvements in terms of reducing biases in SWC as indicated by positive SS_α values (> 0.12) (Fig. 5d). The smaller improvements in R are likely due to the fact that temporal dynamics of CLM estimated SM are captured well in the CLM-OL, in which case there is little benefit from assimilation.

Regional validation using reanalysis products. To further explore the quality of ESSMRA, we evaluated the skill of ESSMRA over PRUDENCE regions in comparison with commonly used SM reanalysis products as shown in Fig. 6. For this analysis, R and α were computed using spatially-averaged soil moisture over

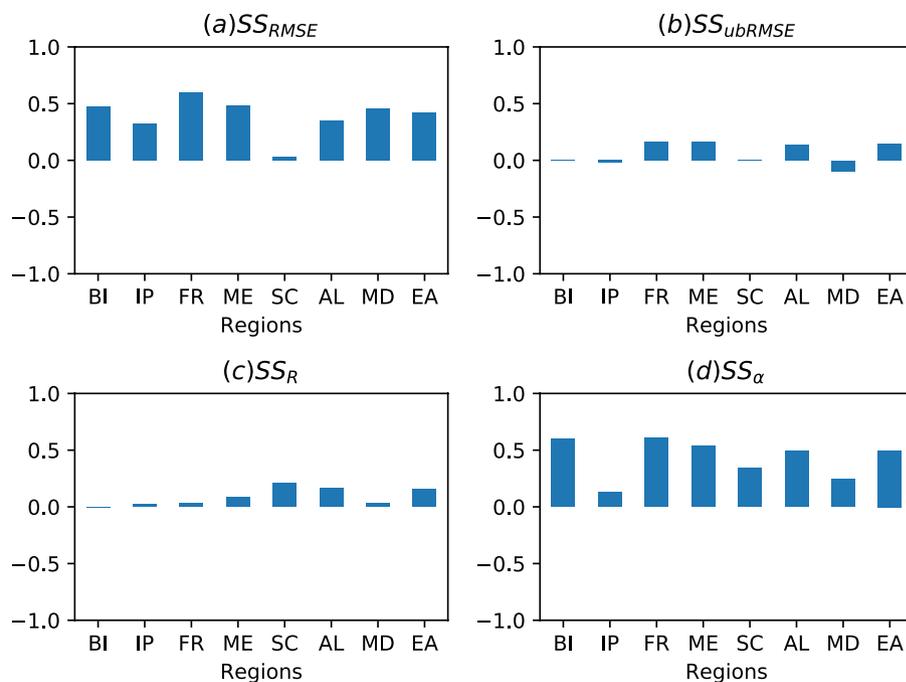


Fig. 5 Performance evaluation of CLM-OL and CLM-DA against ESACCI SM. The skill score of (a) SS_{RMSE} , (b) SS_{ubRMSE} , (c) SS_R and (d) SS_{α} were calculated using spatially averaged SWC over each PRUDENCE region for years 2000–2015 from CLM-OL and CLM-DA simulations.

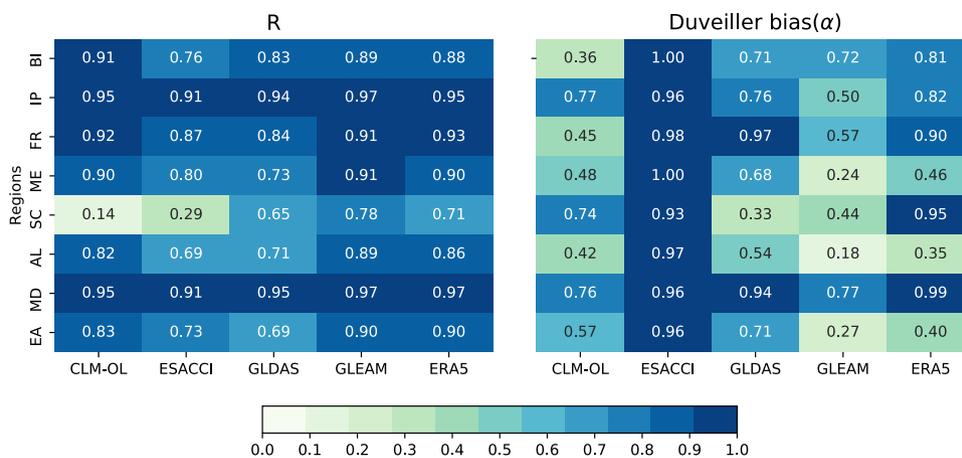


Fig. 6 Performance evaluation of CLM-DA (ESSMRA) against GLDAS, GLEAM, ERA5, ESACCI and CLM-OL over PRUDENCE regions. Skill scores of (a) R and (b) Duveiller bias(α) were calculated using spatially averaged soil water content for years 2000–2015 over each PRUDENCE regions.

PRUDENCE regions between ESSMRA (CLM-DA) and other products (GLDAS, GLEAM and ERA5, ESACCI). This analysis shows that overall ESSMRA has higher correlation (i.e. $R > 0.7$) with other reanalysis products, which indicates higher agreement of ESSMRA with other products in terms of timing and relative magnitude in time series. For the bias results, overall ESSMRA has higher agreement with ERA5 ($\alpha > 0.8$), followed by GLDAS ($\alpha > 0.71$), while it has lower agreement with GLEAM ($\alpha < 0.5$) over most regions. Higher agreement with the ESACCI ($\alpha > 0.90$) than the CLM-OL, again, shows the positive impact of DA through correcting of errors in the timing and magnitude of the soil moisture.

In addition, ESSMRA's ability to capture the monthly and yearly climatologies over PRUDENCE regions is also evaluated for the period 2000–2015 against existing reanalysis products as shown in Supplementary Figs. S1 and S2. These results suggest that ESSMRA follows the seasonal variations fairly well, indicating that the timing and magnitude of SM at monthly and annual scales is reasonably accurate. As shown in Supplementary Fig. S1, however, in the dryer regions such as IP and MD, the soil moisture estimates by ESSMRA are lower than the other products particularly in summer. This might be due to the fact that satellite soil moisture tends to underestimate the true soil moisture content in dry conditions due to systematic retrieval errors⁷¹ which may also affect the

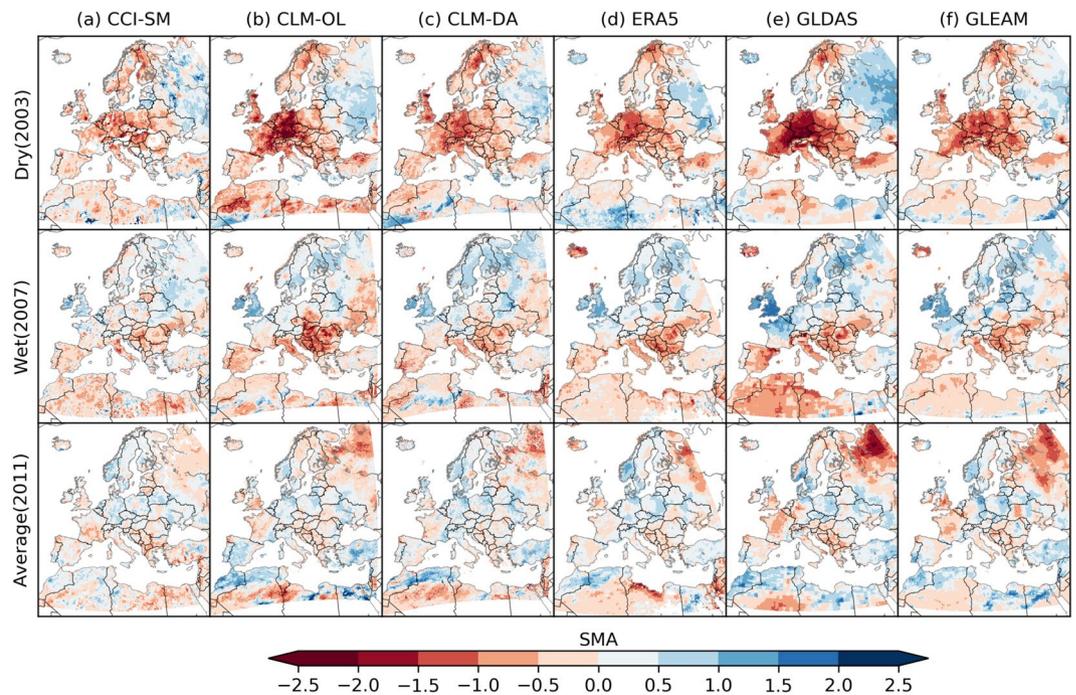


Fig. 7 Spatial distribution of the standardized summer (JJA) soil moisture anomaly for dry, wet and average years. The soil moisture anomalies were calculated for year 2003 (dry year), 2007 (wet year) and 2011 (average year) and compared across existing data and reanalysis products of (a) satellite (ESACCI), (b) CLM-OL (c) CLM-DA, (d) ERA5, (e) GLDAS and (f) GLEAM.

accuracy of assimilated SM estimation. On the other hand, the dry bias could be related to the different spatial scales and soil layers depths between the products. For example, in GLDAS the near-surface soil moisture variable represents an average over the top 10 cm of soil, whereas ERA5 and GLEAM estimate surface soil moisture over the 5 cm of soil depth. Moreover, small differences between SM estimates from these datasets might also be related to mismatch between input datasets, model parameterizations and assumptions used by different LSMs for generation of these products. Nevertheless, above results suggest overall a good agreement at daily, monthly and annual scale with SM reanalysis products, which increases confidence in ESSMRA SM estimates.

Soil moisture variability. To assess the ability of ESSMRA to capture short term soil moisture variability in comparison to other reanalysis products, seasonal and annual standardized anomalies (SMA) were calculated as follows:

$$SMA = \frac{SM_t - \overline{SM}}{SM_\sigma} \quad (5)$$

where SM_t is the average soil moisture value for a current year, \overline{SM} is long-term average and SM_σ is the standard deviation, which are both calculated for the same period of 2000–2015.

Figure 7 illustrates the spatial distribution of SSM anomalies from the ESACCI satellite merged product, GLDAS, GLEAM, ERA5 and ESSMRA developed in this study along with CLM-OL. Summer anomalies of SM for a dry year (2003), wet year (2007) and average year (2011) were calculated using average SM values over June, July and August (JJA) relative to the mean JJA SM for the 2000–2015 period. The dry, wet and average years were selected by comparing yearly precipitation amounts to the long term average precipitation of 2000–2015 over Europe. The spatial distribution shows similar patterns of positive and negative SM anomalies over Europe across all datasets for dry, wet and average years. For the dry year 2003 (a record heat wave over Europe), CLM-DA shows a similar area extent of negative anomalies as the ESACCI dataset and ERA5, whereas CLM-OL, GLDAS and GLEAM exhibit much stronger negative anomalies over central Europe. The SM anomaly from CLM-DA for the wet and average years (2007, 2011) shows a good match with ESACCI and other reanalysis datasets except GLDAS which shows much stronger wet and dry anomalies than others.

At the annual scale, the time series of soil moisture anomalies over the PRUDENCE regions for 2000–2015 (Fig. 8) shows that SM anomalies from CLM-DA could capture the temporal SM variability and agrees well with SM anomalies from ESACCI, GLDAS, GLEAM and ERA5 across all regions. However, the agreement between different products was higher for the average years than for the dry and wet years.

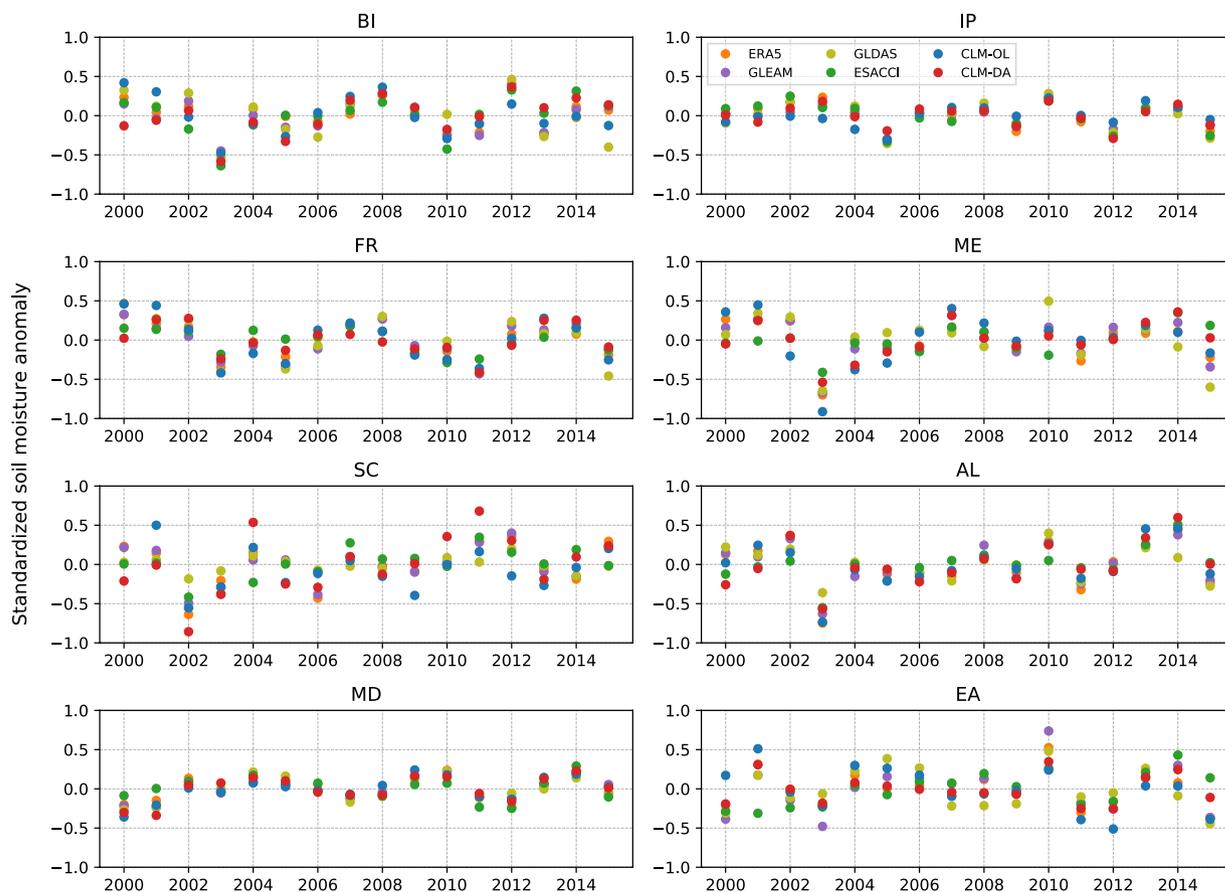


Fig. 8 The standardized annual soil moisture anomalies from ESACCI, CLM-OL, CLM-DA, ERA5, GLDAS and GLEAM over the PRUDENCE regions. The standardized annual soil moisture anomalies were calculated using spatially averaged soil water content for years 2000–2015 over each PRUDENCE region.

Usage Notes

The soil moisture dataset at high spatiotemporal resolution could be used for many practical applications. For example, it can be used as an initial input data for climate change analysis and for numerical weather prediction models to improve the model forecast in terms of location and amount of extreme precipitation events. Because of the scarcity of the *in-situ* soil moisture observations over large areas, this dataset can also be used for validation of SM outputs in modelling studies. This dataset will also be useful to understand the development and persistence of extreme weather events such as droughts, floods and heatwaves.

However, the ESSMRA dataset may still include some uncertainties. For example, uncertainties in ESSMRA may exist in regions where satellite soil moisture retrievals are sparse due to topography, standing water, dense vegetation, frozen soil and/or snow-covered areas. Apart from data gaps, ESACCI is a merged product and may contain inconsistencies because of differences in sensor characteristics and soil moisture retrieval algorithms²⁹. These inconsistencies may also induce uncertainties in the ESSMRA data, particularly in some regions such as over Northern Europe or in Alpine regions and need to be improved in the future.

Code availability

The CLM-PDAF setup is available through the Terrestrial System Modelling Platform (TSMP). TSMP is provided through a git repository available at the model's website (<https://www.terrsysmp.org/>). The users are required to register to the git repository to get access to the code, pre- and post-processing tools and documentations for installing the code with examples setups. TSMP is released without the component models. For the coupled CLM-PDAF configuration, the code for PDAF library is available through website (pdaf.awi.de) which also provide links to the documentation and the source code. The CLM (version 3.5), as used in this study, is available as an open source model through the official CLM website (<http://www.cgd.ucar.edu/tss/clm/distribution/clm3.5/index.html>) which offers all links to documentation, source code, and input data for the stand-alone version release of CLM.

Received: 14 October 2019; Accepted: 12 March 2020;

Published online: 03 April 2020

References

- Brocca, L., Melone, F., Moramarco, T. & Morbidelli, R. Spatial-temporal variability of soil moisture and its estimation across scales: Soil Moisture Spatiotemporal Variability. *Water Resour. Res.* **46**, W02516 (2010).
- Meza, F. J., Montes, C., Bravo-Martínez, F., Serrano-Ortiz, P. & Kowalski, A. S. Soil water content effects on net ecosystem CO₂ exchange and actual evapotranspiration in a Mediterranean semiarid savanna of Central Chile. *Sci. Rep.* **8**, 8570 (2018).
- Brocca, L. *et al.* Assimilation of surface- and root-zone ASCAT soil moisture products into rainfall-runoff modeling. *IEEE Trans. Geosci. Remote Sens.* **50**, 2542–2555 (2012).
- Mohanty, B. P., Cosh, M. H., Lakshmi, V. & Montzka, C. Soil moisture remote sensing: State-of-the-science. *Vadose Zone J.* **16**, 1 (2017).
- Escorihuela, M. J. & Quintana-Seguí, P. Comparison of remote sensing and simulated soil moisture datasets in Mediterranean landscapes. *Remote Sens. Environ.* **180**, 99–114 (2016).
- Vinukollu, R. K., Wood, E. F., Ferguson, C. R. & Fisher, J. B. Global estimates of evapotranspiration for climate studies using multi-sensor remote sensing data: Evaluation of three process-based approaches. *Remote Sens. Environ.* **115**, 801–823 (2011).
- Peng, J., Loew, A., Merlin, O. & Verhoest, N. E. C. A review of spatial downscaling of satellite remotely sensed soil moisture: Downscale Satellite-Based Soil Moisture. *Rev. Geophys.* **55**, 341–366 (2017).
- Babaeian, E. *et al.* Ground, Proximal and Satellite Remote Sensing of Soil Moisture. *Rev. Geophys.* **57**, 530–616 (2019).
- Dungan, J. L., Wang, W., Michaelis, A., Votava, P. & Nemani, R. *Sources of uncertainty in predicting land surface fluxes using diverse data and models*. Report No. ARC-E-DAA-TN1640 (NASA Ames Research Center, 2010).
- Brocca, L. *et al.* Improving runoff prediction through the assimilation of the ASCAT soil moisture product. *Hydrol. Earth Syst. Sci.* **14**, 1881–1893 (2010).
- Lievens, H. *et al.* SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia. *Remote Sens. Environ.* **168**, 146–162 (2015).
- Lievens, H. *et al.* Assimilation of SMOS soil moisture and brightness temperature products into a land surface model. *Remote Sens. Environ.* **180**, 292–304 (2016).
- De Lannoy, G. J. & Reichle, R. H. Global assimilation of multiangle and multipolarization SMOS brightness temperature observations into the GEOS-5 catchment land surface model for soil moisture estimation. *J. Hydrometeorol.* **17**, 669–691 (2016).
- Berg, A. *et al.* Impact of Soil Moisture-Atmosphere Interactions on Surface Temperature Distribution. *J. Climate* **27**, 7976–7993 (2014).
- Rodell, M. *et al.* The global land data assimilation system. *Bull. Amer. Meteor. Soc.* **85**, 381–394 (2004).
- Reichle, R. H. *et al.* Assessment and Enhancement of MERRA Land Surface Hydrology Estimates. *J. Climate* **24**, 6322–6338 (2011).
- Copernicus Climate Change Service (C3S) ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. *Copernicus Climate Change Service Climate Data Store (CDS)*, <https://cds.climate.copernicus.eu/#/search?text=ERA5&type=dataset> (2017).
- Martens, B. *et al.* GLEAM v3: satellite-based land evaporation and root-zone soil moisture. *Geosci. Model Dev.* **10**, 1903–1925 (2017).
- Albergel, C. *et al.* 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 (2017).
- De Rosnay, P. *et al.* A simplified Extended Kalman Filter for the global operational soil moisture analysis at ECMWF. *Q. J. Roy. Meteor. Soc.* **139**, 1199–1213 (2013).
- Draper, C., Mahfouf, J.-F., Calvet, J.-C., Martin, E. & Wagner, W. Assimilation of ASCAT near-surface soil moisture into the SIM hydrological model over France. *Hydrol. Earth Syst. Sci.* **15**, 3829–3841 (2011).
- Ni-Meister, W., Houser, P. R. & Walker, J. P. Soil moisture initialization for climate prediction: Assimilation of scanning multifrequency microwave radiometer soil moisture data into a land surface model. *J. Geophys. Res.-Atmos.* **111**, D20102 (2006).
- Saha, S. *et al.* The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.* **91**, 1015–1058 (2010).
- Oleson, K. W. *et al.* Improvements to the Community Land Model and their impact on the hydrological cycle. *J. Geophys. Res.-Biogeo.* **113**, G01021 (2008).
- Nerger, L. & Hiller, W. Software for ensemble-based data assimilation systems—Implementation strategies and scalability. *Comput. Geosci.* **55**, 110–118 (2013).
- Kurtz, W. *et al.* TerrSysMP-PDAF (version 1.0): a modular high-performance data assimilation framework for an integrated land surface-subsurface model. *Geosci. Model Dev.* **9**, 1341–1360 (2016).
- Wagner, W. *et al.* Fusion of active and passive microwave observations to create an essential climate variable data record on soil moisture. *ISPRS Annal. Photogramm. Remote Sens. Spat. Inf. Sci.* **7**, 315–321 (2012).
- Gruber, A., Scanlon, T., van der Schalie, R., Wagner, W. & Dorigo, W. Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology. *Earth Syst. Sci. Data* **11**, 717–739 (2019).
- Dorigo, W. *et al.* ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sens. Environ.* **203**, 185–215 (2017).
- McNally, A. *et al.* Evaluating ESA CCI soil moisture in East Africa. *Int. J. Appl. Earth Obs.* **48**, 96–109 (2016).
- Naz, B. S. *et al.* Improving soil moisture and runoff simulations at 3 km over Europe using land surface data assimilation. *Hydrol. Earth Syst. Sci.* **23**, 277–301 (2019).
- Liu, Y., Wang, W. & Liu, Y. ESA CCI Soil Moisture Assimilation in SWAT for Improved Hydrological Simulation in Upper Huai River Basin. *Adv. Meteorol.* **2018**, 1–13 (2018).
- Burgers, G., Jan van Leeuwen, P. & Evensen, G. Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev.* **126**, 1719–1724 (1998).
- Evensen, G. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean dyn.* **53**, 343–367 (2003).
- Lawrence, D. M. *et al.* Parameterization improvements and functional and structural advances in version 4 of the Community Land Model. *J. Adv. Model. Earth Syst.* **3**, M03001 (2011).
- Lawrence, P. J. & Chase, T. N. Representing a new MODIS consistent land surface in the Community Land Model (CLM 3.0). *J. Geophys. Res.-Biogeo.* **112**, G01023 (2007).
- Friedl, M. A. *et al.* Global land cover mapping from MODIS: algorithms and early results. *Remote Sens. Environ.* **83**, 287–302 (2002).
- Xiao, Z. *et al.* Use of General Regression Neural Networks for Generating the GLASS Leaf Area Index Product From Time-Series MODIS Surface Reflectance. *IEEE Trans. Geosci. Remote Sens.* **52**, 209–223 (2014).
- Batjes, N. H. A world dataset of derived soil properties by FAO-UNESCO soil unit for global modelling. *Soil use and manage.* **13**, 9–16 (1997).
- Miller, D. A. & White, R. A. A Conterminous United States Multilayer Soil Characteristics Dataset for Regional Climate and Hydrology Modeling. *Earth Interact.* **2**, 1–26 (1998).
- Danielson, J. J. & Gesch, D. B. Global multi-resolution terrain elevation data 2010 (GMTED2010). U.S. Geological Survey Open-File Report 2011–1073, 26 (2011).
- Bollmeyer, C. *et al.* Towards a high-resolution regional reanalysis for the European CORDEX domain. *Q. J. Roy. Meteor. Soc.* **141**, 1–15 (2015).
- Simmer, C. *et al.* Herz: The German Hans-Ertel Centre for Weather Research. *Bull. Amer. Meteor. Soc.* **97**, 1057–1068 (2016).
- Baldauf, M. *et al.* Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities. *Mon. Wea. Rev.* **139**, 3887–3905 (2011).

45. Wahl, S. *et al.* A novel convective-scale regional reanalyses COSMO-REA2: Improving the representation of precipitation. *Meteorol. Z.* **26**, 345–361 (2017).
46. Kerr, Y. H. *et al.* Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE Trans. Geosci. Remote Sens.* **39**, 1729–1735 (2001).
47. Mecklenburg, S. *et al.* ESA's Soil Moisture and Ocean Salinity mission: From science to operational applications. *Remote Sens. Environ.* **180**, 3–18 (2016).
48. Entekhabi, D. *et al.* The soil moisture active passive (SMAP) mission. *Proc. IEEE* **98**, 704–716 (2010).
49. Jones, P. W. First-and second-order conservative remapping schemes for grids in spherical coordinates. *Mon. Wea. Rev.* **127**, 2204–2210 (1999).
50. Mohanty, B. P., Cosh, M., Lakshmi, V. & Montzka, C. Remote sensing for vadose zone hydrology—a synthesis from the vantage point. *Vadose Zone J.* **12**, 1 (2013).
51. Pauwels, V. R., Hoeben, R., Verhoest, N. E. & De Troch, F. P. The importance of the spatial patterns of remotely sensed soil moisture in the improvement of discharge predictions for small-scale basins through data assimilation. *J. Hydrol.* **251**, 88–102 (2001).
52. Pauwels, V. R., Hoeben, R., Verhoest, N. E., De Troch, F. P. & Troch, P. A. Improvement of TOPLATS-based discharge predictions through assimilation of ERS-based remotely sensed soil moisture values. *Hydrol. Process.* **16**, 995–1013 (2002).
53. Montzka, C., Pauwels, V., Franssen, H.-J. H., Han, X. & Vereecken, H. Multivariate and multiscale data assimilation in terrestrial systems: A review. *Sensors* **12**, 16291–16333 (2012).
54. Dorigo, W. A. *et al.* The International Soil Moisture Network: a data hosting facility for global *in situ* soil moisture measurements. *Hydrol. Earth Syst. Sci.* **15**, 1675–1698 (2011).
55. Albergel, C. *et al.* ERA-5 and ERA-Interim driven ISBA land surface model simulations: which one performs better? *Hydrol. Earth Syst. Sci.* **22**, 3515–3532 (2018).
56. Betts, A. K., Chan, D. Z. & Desjardins, R. L. Near-Surface Biases in ERA5 Over the Canadian Prairies. *Front. Environ. Sci.* **7**, 129 (2019).
57. Albergel, C. *et al.* LDAS-Monde Sequential Assimilation of Satellite Derived Observations Applied to the Contiguous US: An ERA-5 Driven Reanalysis of the Land Surface Variables. *Remote Sens.* **10**, 1627 (2018).
58. Chen, Y. *et al.* Evaluation of AMSR-E retrievals and GLDAS simulations against observations of a soil moisture network on the central Tibetan Plateau. *J. Geophys. Res-Atmos.* **118**, 4466–4475 (2013).
59. Bi, H., Ma, J., Zheng, W. & Zeng, J. Comparison of soil moisture in GLDAS model simulations and *in situ* observations over the Tibetan Plateau: Evaluate GLDAS Soil Moisture Over TP. *J. Geophys. Res-Atmos.* **121**, 2658–2678 (2016).
60. Spennemann, P. C., Rivera, J. A., Saulo, A. C. & Penalba, O. C. A Comparison of GLDAS Soil Moisture Anomalies against Standardized Precipitation Index and Multisatellite Estimations over South America. *J. Hydrometeorol.* **16**, 158–171 (2015).
61. Zawadzki, J. & Kędzior, M. Statistical analysis of soil moisture content changes in Central Europe using GLDAS database over three past decades. *Open Geosci.* **6**, 344–353 (2014).
62. Piles, M., Ballabrera-Poy, J. & Muñoz-Sabater, J. Dominant Features of Global Surface Soil Moisture Variability Observed by the SMOS Satellite. *Remote Sens.* **11**, 95 (2019).
63. Naz, B. S., Kollet, S., Hendricks-Franssen, H.-J., Montzka, C. & Kurtz, W. ESSMRA V1.1: 3 km surface soil moisture reanalysis over Europe (2000–2015). *PANGAEA*, <https://doi.org/10.1594/PANGAEA.907036> (2019).
64. Naz, B. S., Kollet, S., Hendricks-Franssen, H.-J., Montzka, C. & Kurtz, W. ESSMRA: 3 km surface soil moisture reanalysis over Europe (2000–2015). *Data Publication Server Forschungszentrum Jülich*, <https://datapub.fz-juelich.de/slots/essmra/index.html> (2019).
65. Duveiller, G., Fasbender, D. & Meroni, M. Revisiting the concept of a symmetric index of agreement for continuous datasets. *Sci. Rep.* **6**, 19401 (2016).
66. Christensen, J. H. & Christensen, O. B. A summary of the PRUDENCE model projections of changes in European climate by the end of this century. *Climatic change* **81**, 7–30 (2007).
67. Qiu, J., Crow, W. T. & Nearing, G. S. The Impact of Vertical Measurement Depth on the Information Content of Soil Moisture for Latent Heat Flux Estimation. *J. Hydrometeorol.* **17**, 2419–2430 (2016).
68. Al-Yaari, A. *et al.* Evaluating soil moisture retrievals from ESA's SMOS and NASA's SMAP brightness temperature datasets. *Remote Sens. Environ.* **193**, 257–273 (2017).
69. Zeng, J., Chen, K.-S., Bi, H. & Chen, Q. A Preliminary Evaluation of the SMAP Radiometer Soil Moisture Product Over United States and Europe Using Ground-Based Measurements. *IEEE Trans. Geosci. Remote Sens.* **54**, 4929–4940 (2016).
70. Rautiainen, K. *et al.* L-Band Radiometer Observations of Soil Processes in Boreal and Subarctic Environments. *IEEE Trans. Geosci. Remote Sens.* **50**, 1483–1497 (2012).
71. Cheng, M. *et al.* A Study on the Assessment of Multi-Source Satellite Soil Moisture Products and Reanalysis Data for the Tibetan Plateau. *Remote Sens.* **11**, 1196 (2019).

Acknowledgements

The authors gratefully acknowledge funding from the European Commission Horizon 2020 research and innovation program under Grant Agreement No. 824158 (EoCoE-II). The authors also gratefully acknowledge the computing time granted through JARA-HPC and the VSR commission on the supercomputer JURECA at Forschungszentrum Jülich through compute time projects cjb31.

Author contributions

B.S.N. and S.K. designed the study. B.S.N. performed the data assimilation experiments. W.K. and H.-J.H.F. contributed to the data assimilation experiments setup. C.M. helped with data validation. B.S.N. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41597-020-0450-6>.

Correspondence and requests for materials should be addressed to B.S.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2020