



Improving soil moisture and runoff simulations over Europe using a high-resolution data-assimilation modeling framework

Bibi S. Naz^{1,2}, Wolfgang Kurtz^{1,2}, Carsten Montzka¹, Wendy Sharples^{2,3}, Klaus Goergen^{1,2}, Jessica Keune⁴, Huilin Gao⁵, Anne Springer⁶, Harrie-Jan Hendricks Franssen^{1,2}, Stefan Kollet^{1,2}

- 5 ¹Research Centre Jülich, 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
³Research Centre Jülich, Jülich Supercomputing Centre, Jülich 52425, Germany
⁴Laboratory of Hydrology and Water Management, Ghent University, Ghent 9000, Belgium
⁵Zachry Department of Civil Engineering, Texas A&M University, College Station, TX 77843, USA
10 ⁶Institute of Geodesy and Geoinformation, Bonn University, Nussallee 17, Bonn 53115, Germany

Correspondence to: Bibi S. Naz (b.naz@fz-juelich.de)

Abstract. Accurate and reliable hydrologic simulations are important for many applications such as water resources management, future water availability projections and predictions of extreme events. However, the accuracy of water balance estimates is limited by the lack of observations at large scales and the uncertainties of model simulations due to errors in model structure and inputs (e.g. hydrologic parameters and atmospheric forcings). In this study, we assimilated ESA CCI soil moisture (SM) information to improve the estimation of continental-scale soil moisture and runoff. The assimilation experiment was conducted over a time period from 2000 to 2006 with the Community Land Model, version 3.5 (CLM3.5) integrated with the Parallel Data Assimilation Framework (PDAF) at spatial resolution of 0.0275° (~3 km) over Europe. The model was forced with the high-resolution reanalysis COSMO-REA6 from the Hans-Ertel Centre for Weather Research (HERZ). Our results show that estimates of soil moisture have improved, particularly in the summer and autumn seasons when cross-validated with independent CCI-SM observations. On average, the mean bias in soil moisture was reduced from 0.1 mm³/mm³ in open-loop simulations to 0.004 mm³/mm³ with SM assimilation. The assimilation experiment also shows overall improvements in runoff, particularly during peak runoff. The results demonstrate the potential of assimilating satellite soil moisture observations to improve high-resolution soil moisture and runoff simulations at the continental scale, which is useful for water resources assessment and monitoring.

1 Introduction

Soil moisture (SM) constitutes a key variable in major processes of the hydrologic cycle related to infiltration and runoff generation, root water uptake and plant transpiration, and evaporation (Vereecken et al., 2016). Thus, soil moisture strongly influences the partitioning of incoming radiative energy into latent and sensible heat and significantly affects the land surface energy and water budgets. Consequently, accurate estimates of SM at large scale are needed for: (1) hydrologic predictions (such as soil moisture and discharge) (Western et al., 2002) and water resource management and planning (e.g. groundwater



recharge, mitigation of droughts) (Dobriyal et al., 2012; Andreassen et al., 2013; Sridhar et al., 2008), (2) identifying regions susceptible to extreme events such as droughts and floods (Seneviratne et al., 2010), (3) numerical weather predictions (Drusch, 2007), and (4) irrigation management and agriculture practices (Shock et al., 1998; Bolten et al., 2010).

At continental space and inter-annual time scales, SM typically exhibits large variability (Brocca et al., 2010), depending on rainfall distribution, topography, soil physical properties, vegetation characteristics, and human impacts, such as irrigation. However, monitoring of soil moisture remains challenging due to the scarcity of *in situ* SM observations networks. Recent advancements in satellite-based sensors offer great potential to monitor SM over large scales for continental water resources assessment, particularly in areas where ground observation networks are sparse (Mohanty et al., 2017). Conventionally, satellite observations have been used in global water balance studies to provide information on the water cycle components, such as precipitation, evapotranspiration, soil moisture, water storage and runoff (Running et al., 2004; Kiehl and Trenberth, 1997; Vinukollu et al., 2011; Trenberth et al., 2007). However, sparse data coverage in satellite observations limits their ability to provide spatially and temporally consistent time series of water balance estimates. Another approach to facilitate studies at a regional to global scale is to estimate water budget components using land surface models forced with precipitation and other atmospheric data (such as the Community Land Model (CLM) (Lawrence et al., 2011), the Variable Infiltration Capacity (VIC) model (Liang et al., 1994; 1996), or the Joint UK Land Environment Simulator (JULES) (Best et al., 2011; Clark et al., 2011). The simulated soil moisture distributions from the land surface models provide spatially and temporally continuous information, yet their accuracy is limited by model deficiencies, and uncertainties in both model parameters and atmospheric forcing variables (Draper et al., 2009; Chen et al., 2013). Estimates from observational and modeling approaches can be merged by data assimilation to provide improved estimates of hydrologic variables at large scales (Lahoz and De Lannoy, 2014). Recent studies showed that assimilation of soil moisture data into hydrologic modeling could improve water balance predictions such as evaporation and runoff (e.g. Crow et al., 2017; Mohanty et al., 2013; Brocca et al., 2012; Matgen et al., 2012; Draper et al., 2011; Pauwels et al., 2002; 2001). However, many of these studies mainly focused on improved predictions at watershed scales using *in situ* observations. Only few studies demonstrated the potential of using satellite observations to improve runoff estimates at regional and global scales (e.g. Liu and Mishra, 2017; López López et al., 2016; Renzullo et al., 2014; Crow and Ryu, 2009; Pan et al., 2008).

Several studies evaluated the assimilation of satellite SM observations into land surface models to produce optimal soil moisture estimates (e.g., Reichle and Koster, 2005; Han et al., 2014; Lievens et al., 2015; De Lannoy and Reichle, 2016). For example, Rains et al. (2017) assimilated SMOS data into CLM over Australia for drought monitoring purposes. Han et al. (2014) evaluated the joint state and parameter estimation method for the coupled CLM and Community Microwave Emission Model (CMEM) (de Rosnay et al., 2009) through assimilation of synthetic microwave brightness temperature data. Similarly, Liu and Mishra (2017) also used assimilation of satellite SM data at the global scale to evaluate the performance of the community land surface model (CLM4.5) in simulating hydrologic fluxes such as SM, ET and runoff at 0.5° spatial resolution. They found that assimilating satellite SM data into the CLM4.5 model improved the soil moisture simulations, which also lead to better representation of other hydro-meteorological variables in the model, such as ET and runoff. Despite



these important advancements, the current spatial resolution of these global scale studies is too coarse to provide locally relevant information (Wood et al., 2011, Bierkens et al., 2015). For example, predicting water cycle processes for scientific and applied assessment of the terrestrial water cycle requires a high-resolution modeling framework on the order of 10^0 km. The spatial mismatch between coarse-resolution satellite data and high-resolution hydrologic models constitutes a great challenge. To address this issue, the spatial mismatch between observations and modeling approaches needs to be taken into account either in the data assimilation algorithm (Sahoo et al., 2013; De Lannoy et al., 2012), or through pre-processing of satellite products to match the model resolution (Merlin et al., 2010; Verhoest et al., 2015). Another challenge is the availability of computational resources, since the computational burden increases (non-) linearly with increasing model resolution, the number of ensemble members in the data assimilation system as well as the complexity of simulated processes.

The objective of this study is to investigate the performance of a high-resolution, continental land surface model in simulating soil moisture and runoff for different climatic zones using data assimilation to incorporate coarser resolution satellite soil moisture data. We used CLM3.5 coupled to the parallel data assimilation framework (PDAF) library (Kurtz et al., 2016; Neger and Hiller, 2013). PDAF is computationally efficient due to its parallelization of data assimilation routines and is suitable for applications at large spatial scales and high-resolution over long time periods (Kurtz et al., 2016). The remainder of this paper is organized as follows: the model and observational data sets as well as methods, including the CLM-PDAF setup and experimental design are described in Sect. 2; the results, including model validation and analysis of simulated soil moisture and runoff are documented in Sect. 3; while the conclusions are presented in Sect. 4.

2 Methods and Data

2.1 Model Description

In this study, the Community Land Model version CLM3.5 (Oleson et al., 2004) was applied to represent land surface processes such as surface and subsurface runoff, snow, soil moisture evolution, evaporation from soil and vegetation, transpiration and interception of precipitation by vegetation canopy, throughfall and infiltration. Specifically, runoff is parameterized using a simple TOPMODEL-based scheme (SIMTOP; Niu et al., 2005). Soil water is calculated by solving the one-dimensional Richards equation (Zeng and Decker, 2009). Groundwater table depth and recharge to groundwater from the soil column is updated dynamically using the algorithm described in Niu et al. (2007). The snow model in CLM explicitly simulates multilayer snow depending on the total snow depth, and includes processes such as snow-melting, surface frost and sublimation, liquid water retention and thawing-freezing processes, (Dai et al., 2003, Dickinson et al. 2006; Stöckli et al., 2008). Total runoff is calculated as the sum of the subsurface runoff, surface runoff and runoff generated from lakes, glaciers, and wetlands (Oleson et al., 2004).

CLM3.5 offers significant improvements in estimating the subcomponents of the terrestrial water cycle compared to earlier versions (Oleson et al., 2008), including improvements in soil water availability and resistance terms to reduce the soil



evaporation which was overestimated in earlier versions (Niu et al., 2005; Oleson et al., 2008; Yang and Niu 2003). For a detailed description of CLM3.5, the readers are referred to Oleson et al. (2008). CLM3.5 was used in this study, instead of its most recent version, to keep the modeling framework consistent to Kurtz et al. (2016).

2.2 Data assimilation framework

5 The Parallel Data Assimilation Framework (PDAF) (Nerger and Hiller, 2013) was used to assimilate satellite soil moisture into CLM3.5. PDAF provides data assimilation methods such as the ensemble Kalman filter (EnKF) (Evensen, 2003; Burgers et al., 1998) and the local ensemble transform Kalman filter (LETKF) (Hunt et al., 2007). In this study, we used the EnKF, which is a relatively simple and flexible technique for assimilating satellite data into land surface models (e.g. Draper et al., 2012; Reichle et al., 2002, 2008; Kumar et al., 2008, 2009; Pipunic et al., 2008; Crow and Wood, 2003;). It used
 10 ensembles of model states to approximate the model state error covariance matrix in order to optimally merge model predictions with observations.

The EnKF calculates the ensemble of updated states variable \mathbf{x}_t^+ at each time t of the model estimated state variable \mathbf{x}_t , as

$$\mathbf{x}_t^+ = \mathbf{x}_t + \mathbf{k}_t[\mathbf{l}_t - \mathbf{H}_t\mathbf{x}_t] \quad (1)$$

15 Where \mathbf{l}_t is the perturbed observation vector and \mathbf{k}_t is the Kalman gain vector defined as:

$$\mathbf{k}_t = \mathbf{P}_t\mathbf{H}_t^T(\mathbf{R}_t + \mathbf{H}_t\mathbf{P}_t\mathbf{H}_t^T)^{-1} \quad (2)$$

where \mathbf{H}_t^T is the transpose matrix of the observation model at time t , \mathbf{R}_t is the measurement error matrix, which is defined a
 20 priori based on the expected measurement error of the ESA CCI soil moisture product and \mathbf{P}_t is the state error covariance matrix of the model predictions calculated as:

$$\mathbf{P}_t = \frac{\sum_{n=1}^N(\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^T}{N-1} \quad (2)$$

25

where $\bar{\mathbf{x}}$ is the vector which contains the ensemble average soil moisture contents for the different grid cells and N is the number of ensemble members.

Kurtz et al. (2016) recently provided a framework to couple PDAF with the land surface–subsurface part of the Terrestrial Systems Modelling Platform (TerrSysMP; Gasper et al., 2014; Shrestha et al., 2014). They showed the efficient use of
 30 parallel computational resources by TerrSysMP-PDAF, which is needed to simulate predicted states and fluxes over large spatial domains and long simulations. In this study, we used the CLM-PDAF setup, in which PDAF is coupled with the



stand-alone CLM3.5 for soil moisture assimilation. Readers are referred to Kurtz et al. (2016) for technical descriptions of coupling and model performance.

2.3 Data

2.3.1 Land surface data and atmospheric forcing

5 The land surface static input data used in this study consist of topography, soil properties, plant functional types, and physiological vegetation parameters (Fig. 1). Digital elevation model (DEM) data were acquired from the 1km Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) (Danielson et al., 2010) as shown in Fig. 1a. The land use data was based on the Moderate Resolution Imaging Spectroradiometer (MODIS) data set (Friedl et al., 2002) (Figure 1b), where the land use types are transferred to Plant Functional Types (PFT). The properties of each of the sub-grid land fractions, such as
10 the leaf area index, the stem area index, and the monthly heights of each PFT, were calculated based on the global CLM3.5 surface data set (Oleson et al., 2008). To provide soil texture data in the model (Fig. 1c and 1d), sand and clay percentages were prescribed based on pedotransfer functions from Schaap and Leij (1998) for 19 soil classes derived from the FAO/UNESCO Digital Soil Map of the World (Batjes, 1997).

[Figure 1]

15 The high-resolution reanalysis COSMO-REA6 dataset (Bollmeyer et al., 2015) from the Hans-Ertel Centre for Weather Research (HErZ; Simmer et al., 2016) for the time period of 2000-2006 was used as the atmospheric forcing for CLM3.5. The essential meteorological variables applied in this study, such as barometric pressure, precipitation, wind speed, specific humidity, near surface air temperature, downward shortwave radiation and downward longwave radiation were downloaded from the German Weather Service (DWD; <ftp://ftp-cdc.dwd.de/pub/REA/>). The COSMO-REA6 reanalysis is based on the
20 COSMO model and available at 0.055° (~6 km) covering the CORDEX EUR-11 domain (Gutowski et al., 2016). COSMO-REA6 was produced through the assimilation of observational meteorological data using the existing nudging scheme in COSMO with boundary conditions from ERA-Interim data. Bollmeyer et al., 2015 compared the COSMO-REA6 precipitation data with the precipitation data from Global Precipitation Climatology Centre and showed that COSMO-REA6 performed well compared to observations with small underestimations of precipitation in mid and southern Europe and
25 overestimations of precipitation in Scandinavia, Russia and along the Norwegian coast. Additionally, Springer et al. (2017) assessed the closure of the water budget in the 6-km COSMO-REA6 and compared to global reanalyses (Interim ECMWF Reanalysis (ERA-Interim), Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2)) for major European river basins. In their study, Springer et al. (2017) found that the COSMO-REA6 closes the water budget within the error estimates whereas the global reanalyses underestimate the precipitation minus evapotranspiration deficit in
30 most river basins. A more comprehensive assessment of the precipitation of the HErZ reanalysis can be found in Wahl et al. (2017), albeit based on the 2 km data product, only available for central Europe.



2.3.2 ESA CCI microwave soil moisture

The European Space Agency (ESA) Climate Change Initiative (CCI) program provides daily soil moisture (CCI-SM) at 0.25° spatial resolution for approximately the top few millimeters to centimeters of soil from 1978 to 2016. The daily CCI-SM product (v03.2) is produced at 0.25° spatial resolution from the microwave retrieved surface soil moisture data and is merged from multiple sensors (Dorigo et al., 2017; Liu et al., 2012; Liu et al., 2011; Wagner et al., 2012; <http://www.esa-oilmoisture-cci.org>). For the study period of 2000 to 2006, the CCI-SM data are based on passive microwave observations (i.e. DMSP SSM/I, TRMM TMI, Aqua AMSR-E and Coriolis WindSat; Owe et al., 2008), whereas the active data products are based on observations from the C-band scatterometers on board of the ERS-1 and ERS-2 (Wagner et al., 2013; Bartalis et al., 2007) satellites. In this product, the absolute soil moisture was re-scaled against the 0.25° land surface modeling soil moisture (GLDAS-NOAH, Rodell et al., 2004) using cumulative density function matching. In this study, we used the merged product of active and passive soil moisture data which showed better accuracy than either of the passive or active data alone (Liu et al., 2011). To match the spatial resolution of our CLM3.5 setup, the original SM values were re-sampled and re-gridded to 0.0275° using the first-order conservative interpolation method (Jones, 1999) which is based on the ratio of source cell area overlapped with the corresponding destination cell area. The conservative regridding scheme preserves the physical flux fields between the source and destination grid.

The CCI-SM dataset shows large gaps in data availability over the European continent during the four seasons (December - February (DJF; Winter), March-May (MAM; Spring), June-August (JJA; Summer), and September-November (SON; Autumn); Fig. 2b). According to Fig. 2b, the temporal coverage (i.e. the ratio between the number of days and the total number of days in a season) is generally low during the winter and spring seasons, ranging from less than 30% (Scandinavian regions) to about 60% in southern Europe. SM observations show the highest temporal coverage during the summer and autumn. Due to the sparseness of the SM data at daily temporal resolution, 100 grid cells were randomly selected covering the complete model domain (Fig. 2a). The satellite CCI-SM daily soil moisture data at these locations were assimilated in the data assimilation framework. However, the number of observations for each day ranged between 2 to 75 depending on the availability of the daily CCI-SM data. As shown in Fig. 2c, there is a higher level of noise in the CCI-SM data for the first two years (2000 and 2001) which might be due to the fact that data from other sensors such as AMSRE-E and Windsat become available after 2002. Moreover, availability of selected observations was lower during winter and spring, while summer soil moisture was well covered during years 2003 to 2006. This seasonal difference in data availability is related to the occurrence of soil freezing events and snow cover.

[Figure 2]

2.3.3 Observational gridded monthly runoff

In order to evaluate the potential of improving runoff estimates by assimilating soil moisture observations, observational gridded monthly runoff data from Gudmundsson and Seneviratne (2016, E-RUN version 1.1) were used. This product



provides monthly pan-European runoff estimates from 1950 to 2015 at 0.5° resolution. The monthly runoff rates were generated using a collection of streamflow observations from small catchments combined with gridded precipitation and temperature data using a machine learning approach (Gudmundsson and Seneviratne, 2016). Monthly runoff was estimated using a regression model, which was trained with a subset of observed runoff rates and E-OBS precipitation and temperature.

5 The fitted model was subsequently applied to all grid cells of the E-OBS data to derive pan-European estimates of monthly runoff (Gudmundsson and Seneviratne, 2016). Using this cross-validation method, Gudmundsson and Seneviratne (2016) reported higher accuracy in central and western Europe, while accuracy was lower in other regions due to low density of available stations. In the current study, the half degree monthly runoff rates were resampled and re-gridded to 0.0275° using the first-order conservative interpolation method for comparison with the CLM3.5 simulated total runoff.

10 2.4 CLM-PDAF experimental design

The assimilation experiments were performed for the time period of January 2000 to December 2006. A spinup of 45 years, by simulating the time period from 1997 to 2006 five times, was performed in order to obtain equilibrium initial state variables. In this study, we implemented CLM3.5 for the EURO-CORDEX domain with a spatial resolution of 0.0275° (~3km), inscribed into the official EUR-11 grid at 0.11° spatial resolution. The model was run with 1h time step and the time window for soil moisture updates was set to 1 day. In this study, we assumed a spatially uniform observational error of 0.02 mm³/mm³ for CCI-SM in the CLM-PDAF setup.

The outputs of a land surface model are sensitive to both atmospheric forcings and soil characteristics. To account for uncertainties in atmospheric forcing and soil texture, precipitation and soil texture (%sand and %clay) were perturbed in this study. Log-normally distributed, spatially homogeneous and temporally uncorrelated multiplicative perturbations were added to precipitation. The mean and standard deviation of the applied perturbation factors for precipitation were equal to one and 0.15, respectively. Sand and clay content were perturbed using a random noise with a standard deviation of 10%. In order to guarantee the physical meaning of the soil parameters, the sand and clay content were constrained to have a sum of 100%. The initial ensemble size was set to 12 for the precipitation and soil texture in the simulation/assimilation experiment to update the volumetric soil water content (SWC) of the top soil layer (~2cm).

25 Our main experiment consists of two CLM-PDAF simulations: (a) an open-loop simulation (no data assimilation, CLM-OL) and (b) an ensemble simulation with data assimilation of ESA CCI-SM data (CLM-DA) at 100 random locations (Figure 2a). We evaluated the results of both simulations by a cross-validation with ESA CCI-SM data as shown in Figure 2a. The soil moisture validation of the CLM-DA and CLM-OL simulations used all the available CCI-SM data in the time period of 2000 to 2006. This approach also allowed us to independently cross-validate the SM values over grid cells that were not used in the data assimilation. For SM comparison, the average of simulated SWC in the top two layers (i.e. at 0.007 and 0.03 m depth) was used. Additionally, the monthly runoff dataset E-RUN as described in Sect. 2.3 was used to validate runoff as simulated by CLM-OL and CLM-DA.



To assess the skill of the assimilation experiments, the root mean square error (RMSE) and the mean bias error (BIAS) were used as validation measures.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - Y_{obs,i})^2} \quad (4)$$

$$BIAS = \frac{1}{n} \sum_{i=1}^n (Y_i - Y_{obs,i}) \quad (5)$$

- 5 where n is the total number of time steps; Y_i and $Y_{obs,i}$ represent the simulated ensemble mean and observation values at time step i , respectively.

3 Results and discussion

In this section, the impact of assimilating the ESA CCI-SM data into CLM3.5 on the terrestrial hydrologic cycle is analyzed focusing on soil moisture and runoff. The results are presented for the complete CORDEX EUR-11 domain and for 8 pre-
10 defined analysis regions from the “Prediction of Regional scenarios and Uncertainties for Defining European Climate change risks and Effects” (PRUDENCE) project (Christensen et al., 2007) as shown in Fig. 1a. We refer to these regions as the “PRUDENCE” regions.

3.1 Regional evaluation of soil moisture

3.1.1 Seasonal mean comparison

15 Figure 3 shows a comparison of the seasonal mean volumetric SWC (mm^3/mm^3) from the CLM3.5 experiments (CLM-OL, CLM-DA) with the satellite seasonal mean CCI-SM data. The CLM-OL simulation exhibits higher SWC in all seasons over most part of Europe compared to the CLM-DA simulations. Seasonally, the spatial distribution of SWC in summer and autumn is better reproduced in the CLM-DA simulations than in the CLM-OL when compared with the CCI-SM (Fig. 3c and 3d). Furthermore, Fig. 4 shows the comparison of 2000 to 2006 temporally averaged SM estimated by CLM-OL and CLM-
20 DA with the CCI-SM dataset over PRUDENCE regions. Generally, CLM-OL overestimated the SWC values for all sub-regions and in all seasons. This overestimation of soil moisture in CLM was also reported by Cai et al. 2014 when compared with other land surface models and observations over Continental US. However, using data assimilation, this overestimation was reduced consistently in all sub-regions, as shown with CLM-DA. Noticeably, assimilation also helped to reduce the spatial variability, as indicated by the narrow spread of quartiles of CLM-DA estimated SWC compared to CLM-OL in Fig.
25 4. Similarly, validating the simulations with CCI-SM data, the improvements of the CLM-DA vary within PRUDENCE regions and seasons. Improvements were more prominent for the UK, France, and Central Europe (for all seasons), while for other regions SWC was slightly overestimated in spring (Fig. 4b) and underestimated in summer and autumn (Fig. 4c and Fig. 4d). The underestimation of SWC was particularly pronounced over the Iberian Peninsula and the Mediterranean regions in summer (Fig. 4c).



[Figure 3]

[Figure 4]

In order to validate the skill of CLM-DA relative to CLM-OL, a cross-validation with CCI-SM observations was performed and RMSE and BIAS for soil moisture were calculated using daily values for each PRUDENCE region and each season, as shown in Fig. 5. Note, that for calculating these statistics, model data were only used for the days when satellite data were available. Over Scandinavia, the Alpine, the Mediterranean and Eastern Europe PRUDENCE regions, CLM-DA showed a consistently lower RMSE than CLM-OL for all seasons, except for winter, where improvements in SWC were comparatively small (Fig. 5a and Fig. 5b). The BIAS for CLM-OL indicates a clear overestimation of soil moisture relative to satellite CCI-SM observations (Fig. 5c), whereas the BIAS for soil moisture from CLM-DA is significantly reduced (Fig. 5d). The mean BIAS dropped from $0.1 \text{ mm}^3/\text{mm}^3$ (CLM-OL) to $0.004 \text{ mm}^3/\text{mm}^3$ (CLM-DA) for all regions. However, data assimilation introduced a dry BIAS in summer (up to $-0.03 \text{ mm}^3/\text{mm}^3$), as indicated by CLM-DA in Fig. 5d.

[Figure 5]

3.1.2 Daily validation

The long-term (January 2000 to December 2006) daily SM averaged over PRUDENCE regions (Fig. 1) in Europe, as simulated by CLM-OL and CLM-DA, and observed by CCI-SM are shown in Fig. 6. The assimilated CCI-SM data improved the simulations of surface soil moisture in CLM-DA. The daily soil moisture patterns simulated by CLM-DA show a strong agreement with the CCI-SM observations, with peaks and troughs generally coinciding for all regions and over the European domain except for the years 2000 and 2001. The CCI-SM observations show increased variability and drier soil moisture values for the years 2000 and 2001 compared to the full period. This can be explained by the strong contribution of the X-band passive microwave data of SSM/I and TRMM to the final CCI-SM product. Wang (1987) showed that X-band data have a very shallow soil penetration depth of a few millimeters and sensitive to vegetation cover. After implementing the C-band radiometer data of AMSR-E in 2002 and Windsat in 2003 into CCI-SM, noise level and bias was reduced (Dorigo et al., 2017). Overall, the daily soil moisture values estimated by the CLM-DA show a slightly better agreement with the CCI-SM data for the summer and autumn seasons than for the spring and winter seasons. The winter season bias is pronounced particularly over Scandinavia, the Alpine region and Eastern Europe. The poorer performance of CLM-DA in these regions might be due to the limited amount of CCI-SM data in the winter season (Fig. 2b), dense vegetation, frozen soil (e.g. in the Scandinavian regions) and/or CLM3.5 model errors related to simulating soil moisture in colder regions (Oleson et al. 2008, Decker and Zeng, 2009). Additionally, the magnitudes of the bias and variance of the CCI-SM observational error could be important. As indicated by Dorigo et al. (2017), the CCI-SM error variance is low where the satellite track density increases and the error variance is high in areas with more data gaps. Note that the setup of CLM-DA in this study assumed a spatially uniform observational error for CCI-SM.

[Figure 6]



3.2 Regional evaluation of seasonal runoff

Figure 7 shows the runoff estimates of the two experiments, i.e. CLM-OL and CLM-DA, compared to the E-RUN observational product. CLM-DA reduces the runoff bias compared to CLM-OL. CLM-OL simulates higher magnitudes of runoff (on average 1.42 mm/day) over most parts of Europe compared to CLM-DA (on average 0.25 mm/day) in all seasons. The overestimation in the CLM-OL simulations was more pronounced in the summer and spring seasons (Fig. 7b and Fig. 7c). Compared to CLM-OL, regional runoff patterns simulated by CLM-DA agree better with runoff observations. At the seasonal scale, CLM-DA shows similar spatial runoff distributions in winter and spring as the E-RUN runoff data set (Fig. 7a and Fig. 7b). However, CLM-DA underestimates runoff in summer and autumn particularly in Mid- and Southern Europe (Fig. 7c and Fig. 7d). It is obvious that the assimilation of soil moisture led to an overall improvement in the simulated total runoff for all regions as shown in Fig. 8. However, improvements in runoff are more prominent over the alpine, Scandinavia and Eastern Europe regions, where CLM-DA minimized the difference to E-RUN from 1.41 mm/day, 1.22 mm/day and 0.85 mm/day, to -0.5 mm/day, -0.06 mm/day and -0.03 mm/day, respectively (Table 1). The improvements over other regions, such as the UK, Iberian Peninsula, France and Mediterranean, were comparatively small. These findings indicate the potential of satellite soil moisture assimilation in CLM3.5 to improve other terrestrial components of the water cycle as a basis for more accurate water balance analyses.

[Figure 7]

[Figure 8]

[Table 1]

Figure 9 shows that simulations based on CLM-OL have higher RMSE and BIAS values than CLM-DA simulations (with respect to E-RUN data). The RMSE and BIAS for CLM-OL for mean monthly runoff over all regions and in all seasons vary between 0.4 and 1.5 mm/day and -0.11 to 2.5 mm/day, respectively. CLM-DA reduces the range of both RMSE and BIAS of mean monthly runoff to 0.2 to 1 mm/day and -1.5 to 0.27 mm/day, respectively, across all regions (Fig. 9b and Fig. 9d). However, the data assimilation in CLM-DA also introduces a negative BIAS for most regions, which indicates underestimation of monthly runoff relative to E-RUN. This underestimation might be related to model limitations to correctly represent saturation excess runoff processes at the large spatial scales, particularly in the arid to semi-arid regions. In the dry regions, assimilation of soil moisture data may result in reduction of soil moisture values close to the residual water content values which may lead to small runoff generation. Currently, the CLM-PDAF setup only performs state updates. In future, joint update of states and hydraulic parameters related to soil texture and runoff generation may further improve runoff estimates (Huang et al., 2013).

[Figure 9]

The time series of monthly runoff, as illustrated in Figure 10, show that CLM-OL highly overestimates the magnitude of runoff. The CLM-DA reduces these biases over all regions, but at the same time induces a dry bias for some regions. At the regional scale, the CLM-DA performs better when compared to E-RUN in Mid-Europe, Scandinavia, Alpine and Eastern



Europe regions in capturing peaks and low runoff. In the UK, Iberian Peninsula, France and the Mediterranean regions, however, peak runoff in winter is underestimated whereas low runoff in summer is in correspondence with observed monthly runoff data. The relatively poor performance of CLM-OL may be related to several limitations in the CLM model (Li et al., 2011). For example, Li et al. (2011) showed unrealistic behavior of subsurface runoff and high runoff peaks in CLM4.0, which they attributed to the exponential form of the surface runoff parameterization. Additionally, the assumption of topographically controlled surface runoff generation in the CLM model (Oleson et al. 2008) is problematic in areas with flat topography, thick soils, or deep groundwater (Li et al., 2011). Another reason may be uncertainties of E-RUN runoff data used in this study, which are derived from gridded atmospheric variables at coarser resolution ($0.5^\circ \times 0.5^\circ$ grid resolution) and flow observations. In future, additional observational data need to be explored in assimilation experiments and assessment of the results.

[Figure 10]

3.3 Uncertainties and limitations

This study demonstrates that the assimilation of coarse-scale satellite CCI soil moisture data is beneficial and improves the high-resolution CLM model simulations of soil moisture and runoff over a large spatial domain. However, we note a number of limitations in this study. The spatial mismatch between the coarser resolution CCI-SM and our high-resolution land surface model was tackled by rescaling the CCI-SM data to the model resolution ($\sim 3\text{km}$) without any bias correction. An adequate bias correction of CCI-SM data for high resolutions may require more elaborated methods and techniques to account for the increased variability over time. A further possibility is the multiscale assimilation of the CCI-SM data, which would allow to update various model grid cells covered by a satellite observation (Montzka et al., 2012). In multiscale assimilation, the average soil moisture content for the group of grid cells covered by the satellite measurement is compared with the satellite-based soil moisture content which may result in slightly improved CLM simulation results, but was beyond the scope of this study.

In addition to discrepancies at the spatial scale, uncertainties in soil moisture estimations may result from data gaps in satellite soil moisture retrievals, which are limited in regions of pronounced topography, standing water, areas of dense vegetation and snow covered areas and frozen soil. Additionally, CCI-SM is a merged product from a variety of sensors leading to inconsistencies due to differences in viewing angle, sensor characteristics and soil moisture retrieval algorithms (Dorigo et al., 2017). In future, more observations are needed to independently validate model and assimilation experiments. In this work, the CCI-SM dataset was also used for verification over grid cells that were not used in the data assimilation. However, it would be preferable to validate with another independent dataset at the continental scale. The problem is that at the model grid scale only very limited independent (in situ) soil moisture data are available.

Furthermore, in order to reduce model errors from parameter uncertainties, this study only allows to account for uncertainties in the soil texture parameters. In data assimilation, it is preferable to account for additional model parameter uncertainties towards runoff that shows a high sensitivity. Alternatively, prior model calibration can be considered to constrain model



parameters better and reduce systematic biases and uncertainties in CLM3.5 before applying the assimilation framework. For improvement of hydrologic predictions, joint assimilation of additional datasets such as river discharge and snow data may also be considered in future research.

4 Summary and conclusions

5 A soil moisture data assimilation framework at the continental scale was applied to generate long-term daily soil moisture and runoff estimates as part of a terrestrial systems monitoring framework for Europe at approximately 3 km resolution for the years 2000 to 2006. An ensemble was generated by perturbing precipitation and soil texture properties. This ensemble was used as input in the CLM-PDAF data assimilation framework (Kurtz et al., 2016) and used to assimilate CCI-SM soil moisture data. The impact of satellite soil moisture assimilation on daily soil moisture and runoff was evaluated and cross-validated with CCI-SM data and gridded runoff from E-RUN observations at regional and seasonal scales. Using this high-resolution CLM-PDAF setup, the conclusions of this study are:

1. This study showed that assimilation of satellite SM improved the soil moisture simulations over most parts of Europe relative to open-loop simulations. Open loop simulations overestimated SM in most parts of Europe and in all seasons. For the study domain, on average, the mean bias in soil moisture was reduced from 0.1 mm³/mm³ in open-loop simulations to 0.004 mm³/mm³ with SM assimilation.
2. Regionally, significant improvements were achieved for soil moisture across most regions, except over Scandinavia and Eastern Europe. The low performance of CLM-OL and CLM-DA in these regions might be due to lack of data in space and time, as caused by track changes, radio-frequency interference, dense vegetation, and frozen soil limiting the assimilation of soil moisture data in land surface processes simulations. Similarly, both CLM-OL and CLM-DA performed poorly for years 2000 and 2001, which appears to be related to large data gaps and higher noise levels in CCI-SM satellite data in these years. This indicates adequate suitability of ESA CCI soil moisture for data assimilation studies from 2002 onwards, whereas the accuracy and noise levels of earlier periods data is not appropriate for this purpose.
3. At the seasonal time scale, the CLM-DA simulations performed better in the summer and autumn seasons than in the winter and spring seasons. This might be again related to large data gaps in the winter season or model limitations to correctly represent complex cold region processes such as frozen soil.
4. The assimilation of CCI-SM data into CLM3.5 also improved the overall performance of the CLM3.5 model in simulating total runoff (i.e. surface and subsurface runoff). The largest improvements were achieved in the simulation of peak runoff as result of soil moisture assimilations. The improvement in peak runoff could be particular importance in the management of extreme events such as flooding.

The results from this study are not only useful as a standalone, high-resolution product for evaluating trends in soil moisture patterns across Europe, but can also be used as an independent dataset for validation of other land surface models.



Additionally, by selecting the ESA CCI soil moisture product for assimilation, the potential impact of a long term soil moisture observations on hydrologic simulations can be assessed for climate change studies. Moreover, CLM3.5 is also part of the fully coupled Terrestrial Systems Model Platform (TerrSysMP; Shrestha et al., 2014, Gasper et al., 2014, Keune et al., 2016) that simulates the full terrestrial hydrologic cycle including feedbacks between atmosphere, land-surface and subsurface compartments of the water cycle. The impact of satellite soil moisture assimilation on other water cycle variables across the soil-vegetation-atmosphere system using TerrSysMP and its effects on the accuracy of model simulations at the continental scale remains to be explored.

Acknowledgements

The authors gratefully acknowledge the computing time granted by the JARA-HPC Vergabegremium on the supercomputer JURECA at Forschungszentrum Jülich. This work was supported by the Energy oriented Centre of Excellence (EoCoE), grant agreement number 676629, funded within the Horizon2020 framework of the European Union. C. Montzka was funded by BELSPO (Belgian Science Policy Office) in the framework of the STEREO III programme – HYDRAS+ (SR/00/302). W. Kurtz was supported by SFB-TR32 “Patterns in soil-vegetation-atmosphere systems: monitoring, modelling and data assimilation” funded by the German Science Foundation (DFG).

15

References

- Andreasen, M., Andreasen, L. A., Jensen, K. H., Sonnenborg, T. O., and Bircher, S.: Estimation of regional groundwater recharge using data from a distributed soil moisture network, *Vadose Zone J.*, 12, doi: 10.2136/vzj2013.01.0035, 2013.
- Bartalis, Z., Wagner, W., Naeimi, V., Hasenauer, S., Scipal, K., Bonekamp, H., Figa, J., and Anderson, C.: Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT), *Geophys. Res. Lett.*, 34, doi: 10.1029/2007gl031088, 2007.
- Batjes, N.: A world dataset of derived soil properties by FAO–UNESCO soil unit for global modelling, *Soil use and management*, 13, 9-16, doi:10.1111/j.1475-2743.1997.tb00550.x, 1997.
- Best, M., Pryor, M., Clark, D., Rooney, G., Essery, R., Ménard, C., Edwards, J., Hendry, M., Porson, A., and Gedney, N.: The Joint UK Land Environment Simulator (JULES), model description–Part 1: energy and water fluxes, *Geosci. Model Dev.*, 4, 677-699, doi:10.5194/gmdd-4-595-2011, 2011.
- Bierkens, M. F., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., de Roo, A., Döll, P., Drost, N., and Famiglietti, J. S.: Hyper-resolution global hydrological modelling: What is next? Everywhere and locally relevant, *Hydrol. Process*, 29, 310-320, doi: 10.1002/hyp.10391, 2015.



- Bollmeyer, C., Keller, J., Ohlwein, C., Wahl, S., Crewell, S., Friederichs, P., Hense, A., Keune, J., Kneifel, S., and Pscheidt, I.: Towards a high-resolution regional reanalysis for the European CORDEX domain, *Q. J. Royal Meteorol. Soc.*, 141, 1-15, doi: 10.1002/qj.2486, 2015.
- Bolten, J. D., Crow, W. T., Zhan, X., Jackson, T. J., and Reynolds, C. A.: Evaluating the utility of remotely sensed soil moisture retrievals for operational agricultural drought monitoring, *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, 3, 57-66, doi: 10.1109/jstars.2009.2037163, 2010.
- Bonan, G. B., Lawrence, P. J., Oleson, K. W., Levis, S., Jung, M., Reichstein, M., Lawrence, D. M., and Swenson, S. C.: Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data, *J. Geophys. Res-Biogeophys.*, 116, doi: 10.1029/2010jg001593, 2011.
- 10 Brocca, L., Melone, F., Moramarco, T., and Morbidelli, R.: Spatial-temporal variability of soil moisture and its estimation across scales, *Water Resour. Res.*, 46, doi: 2010.
- Brocca, L., Moramarco, T., Melone, F., Wagner, W., Hasenauer, S., and Hahn, S.: Assimilation of surface-and root-zone ASCAT soil moisture products into rainfall-runoff modeling, *IEEE T. Geosci. Remote*, 50, 2542-2555, doi: 10.1109/tgrs.2011.2177468, 2012.
- 15 Burgers, G., Jan van Leeuwen, P., and Evensen, G.: Analysis scheme in the ensemble Kalman filter, *Mon. weather rev.*, 126, 1719-1724, doi: 10.1175/1520-0493(1998)126<1719:asitek>2.0.co;2, 1998.
- Cai, X., Yang, Z. L., Xia, Y., Huang, M., Wei, H., Leung, L. R., and Ek, M. B.: Assessment of simulated water balance from Noah, Noah-MP, CLM, and VIC over CONUS using the NLDAS test bed, *J. Geophys. Res-Atmos.*, 119, doi: 10.1002/2014jd022113, 2014.
- 20 Chen, Y., Yang, K., Qin, J., Zhao, L., Tang, W., and Han, M.: 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, doi: 10.1002/jgrd.50301, 2013.
- Christensen, J. H., and 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, doi: 10.1007/s10584-006-9210-7, 2007.
- 25 Clark, D., Mercado, L., Sitch, S., Jones, C., Gedney, N., Best, M., Pryor, M., Rooney, G., Essery, R., and Blyth, E.: The Joint UK Land Environment Simulator (JULES), model description-Part 2: carbon fluxes and vegetation dynamics, *Geosci. Model Dev.*, 4, 701, doi: 10.5194/gmd-4-701-2011, 2011.
- Crow, W., and Ryu, D.: A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals, *Hydrol. Earth Sys. Sc.*, 13, 1-16, doi: 10.5194/hess-13-1-2009, 2009.
- 30 Crow, W., Chen, F., Reichle, R., and Liu, Q.: L-band microwave remote sensing and land data assimilation improve the representation of pre-storm soil moisture conditions for hydrologic forecasting, *Geophys. Res. Lett.*, doi: 2017.



- Crow, W. T., and Wood, E. F.: The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using ensemble Kalman filtering: A case study based on ESTAR measurements during SGP97, *Adv. Water Resour.*, 26, 137-149, doi: 10.1016/s0309-1708(02)00088-x, 2003.
- Dai, Y., Zeng, X., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G., Denning, A. S., Dirmeyer, P. A., Houser, P. R., and Niu, G.: The common land model, *B. Am. Meteorol. Soc.*, 84, 1013-1023, doi: 10.1175/BAMS-84-8-1013, 2003.
- Danielson, J. J., and Gesch, D. B.: Global multi-resolution terrain elevation data 2010 (GMTED2010), US Geological Survey 2331-1258, 2011.
- De Lannoy, G. J., Reichle, R. H., Houser, P. R., Arsenault, K. R., Verhoest, N. E., and Pauwels, V. R.: Satellite-scale snow water equivalent assimilation into a high-resolution land surface model, *J. Hydrometeorol.*, 11, 352-369, doi: 10.1175/2009jhm1192.1, 2010.
- De Lannoy, G. J., Reichle, R. H., Arsenault, K. R., Houser, P. R., Kumar, S., Verhoest, N. E., and Pauwels, V.: Multiscale assimilation of Advanced Microwave Scanning Radiometer–EOS snow water equivalent and Moderate Resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado, *Water Resour. Res.*, 48, doi: 10.1029/2011wr010588, 2012.
- De Lannoy, G. J., and 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, doi: 10.1175/jhm-d-15-0037.1, 2016.
- De Rosnay, P., Drusch, M., Boone, A., Balsamo, G., Decharme, B., Harris, P., Kerr, Y., Pellarin, T., Polcher, J., and Wigneron, J. P.: AMMA Land Surface Model Intercomparison Experiment coupled to the Community Microwave Emission Model: ALMIP-MEM, *J. Geophys. Res-Atmos.*, 114, doi: 10.1029/2008jd010724, 2009.
- Dickinson, R. E., Oleson, K. W., Bonan, G., Hoffman, F., Thornton, P., Vertenstein, M., Yang, Z.-L., and Zeng, X.: The Community Land Model and its climate statistics as a component of the Community Climate System Model, *J. Climate*, 19, 2302-2324, doi: 10.1175/jcli3742.1, 2006.
- Dobriyal, P., Qureshi, A., Badola, R., and Hussain, S. A.: A review of the methods available for estimating soil moisture and its implications for water resource management, *J. Hydrol.*, 458, 110-117, doi: 10.1016/j.jhydrol.2012.06.021, 2012.
- Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., and Gruber, A.: ESA CCI Soil Moisture for improved Earth system understanding: state-of-the art and future directions, *Remote Sens. Environ.*, doi: 10.1016/j.rse.2017.07.001, 2017.
- Draper, C., Mahfouf, J. F., and Walker, J.: An EKF assimilation of AMSR-E soil moisture into the ISBA land surface scheme, *J. Geophys. Res-Atmos.*, 114, doi: 10.1029/2008jd011650, 2009.
- Draper, C., Mahfouf, J.-F., Calvet, J.-C., Martin, E., and Wagner, W.: Assimilation of ASCAT near-surface soil moisture into the SIM hydrological model over France, *Hydrol. Earth Syst. Sc.*, 15, 3829, doi: 10.5194/hess-15-3829-2011, 2011.



- Draper, C., Reichle, R., De Lannoy, G., and Liu, Q.: Assimilation of passive and active microwave soil moisture retrievals, *Geophys. Res. Lett.*, 39, doi: 10.1029/2011gl050655, 2012.
- Drusch, M.: Initializing numerical weather prediction models with satellite-derived surface soil moisture: Data assimilation experiments with ECMWF's Integrated Forecast System and the TMI soil moisture data set, *J. Geophys. Res-Atmos.*, 5 112, doi: 10.1029/2006JD007478, 2007.
- Evensen, G.: The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean dynam.*, 53, 343-367, doi: 10.1007/s10236-003-0036-9, 2003.
- Friedl, M. A., McIver, D. K., Hodges, J. C., Zhang, X., Muchoney, D., Strahler, A. H., Woodcock, C. E., Gopal, S., Schneider, A., and Cooper, A.: Global land cover mapping from MODIS: algorithms and early results, *Remote Sens. Environ.*, 83, 287-302, doi: 10.1109/igarss.2002.1027129, 2002.
- 10 Gasper, F., Goergen, K., Shrestha, P., Sulis, M., Rihani, J., Geimer, M., and Kollet, S.: Implementation and scaling of the fully coupled Terrestrial Systems Modeling Platform (TerrSysMP v1. 0) in a massively parallel supercomputing environment—a case study on JUQUEEN (IBM Blue Gene/Q), *Geosci. Model Dev.*, 7, 2531-2543, doi: 10.5194/gmdd-7-3545-2014, 2014.
- 15 Gudmundsson, L., and Seneviratne, S. I.: Observational gridded runoff estimates for Europe (E-RUN version 1.0), *Earth Sys. Sci. Data*, 2016, 1-1, doi: 10.5194/essd-2015-38, 2016.
- Gutowski Jr, W. J., Giorgi, F., Timbal, B., Frigon, A., Jacob, D., Kang, H.-S., Raghavan, K., Lee, B., Lennard, C., and Nikulin, G.: WCRP coordinated regional downscaling experiment (CORDEX): a diagnostic MIP for CMIP6, *Geosci. Model Dev.*, 9, 4087, doi: 10.5194/gmd-2016-120, 2016.
- 20 Han, X., Franssen, H. J. H., Montzka, C., and Vereecken, H.: Soil moisture and soil properties estimation in the Community Land Model with synthetic brightness temperature observations, *Water Resour. Res.*, 50, 6081-6105, doi: 10.1002/2013wr014586, 2014.
- Huang, M., Hou, Z., Leung, L. R., Ke, Y., Liu, Y., Fang, Z., and Sun, Y.: Uncertainty analysis of runoff simulations and parameter identifiability in the community land model: evidence from MOPEX basins, *J. Hydrometeorol.*, 14, 1754-25 1772, doi: 10.1175/jhm-d-12-0138.1, 2013.
- Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, *Physica D: Nonlinear Phenomena*, 230, 112-126, doi: 10.1016/j.physd.2006.11.008 2007.
- Keune, J., Gasper, F., Goergen, K., Hense, A., Shrestha, P., Sulis, M., and Kollet, S.: Studying the influence of groundwater representations on land surface-atmosphere feedbacks during the European heat wave in 2003, *J. Geophys. Res-Atmos.*, 30 121, doi: 10.1002/2016jd025426, 2016.
- Kiehl, J. T., and Trenberth, K. E.: Earth's annual global mean energy budget, *B. Am. Meteorol. Soc.*, 78, 197-208, doi: 10.1175/1520-0477(1997)078<0197:eagmeb>2.0.co;2, 1997.



- Kumar, S. V., Reichle, R. H., Peters-Lidard, C. D., Koster, R. D., Zhan, X., Crow, W. T., Eylander, J. B., and Houser, P. R.: A land surface data assimilation framework using the land information system: Description and applications, *Adv. Water Resour.*, 31, 1419-1432, doi: 10.1016/j.advwatres.2008.01.013, 2008.
- Kumar, S. V., Reichle, R. H., Koster, R. D., Crow, W. T., and Peters-Lidard, C. D.: Role of subsurface physics in the
5 assimilation of surface soil moisture observations, *J. Hydrometeorol.*, 10, 1534-1547, doi: 10.1175/2009jhm1134.1, 2009.
- Kurtz, W., He, G., Kollet, S. J., Maxwell, R. M., Vereecken, H., and Hendricks Franssen, H.-J.: 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, <https://doi.org/10.5194/gmd-9-1341-2016>, 2016.
- 10 Lahoz, W. A., and De Lannoy, G. J.: Closing the gaps in our knowledge of the hydrological cycle over land: conceptual problems, *Surv. Geophys.*, 35, 623-660, doi: 10.1007/978-94-017-8789-5_8, 2014.
- Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., Zeng, X., Yang, Z. L., Levis, S., and Sakaguchi, K.: Parameterization improvements and functional and structural advances in version 4 of the Community Land Model, *J. Adv. Model Earth Sy.*, 3, doi: 10.1029/2011ms00045, 2011.
- 15 Li, H., Huang, M., Wigmosta, M. S., Ke, Y., Coleman, A. M., Leung, L. R., Wang, A., and Ricciuto, D. M.: Evaluating runoff simulations from the Community Land Model 4.0 using observations from flux towers and a mountainous watershed, *J. Geophys. Res-Atmos.*, 116, doi: 10.1029/2011jd016276, 2011.
- Li, M., and Ma, Z.: Soil moisture drought detection and multi-temporal variability across China, *Science China. Earth Sci.*, 58, 1798, doi: 10.1007/s11430-015-5076-8, 2015.
- 20 Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, *J. Geophys. Res-Atmos.*, 99, 14415-14428, doi: 10.1029/94jd00483, 1994.
- Liang, X., Wood, E. F., and Lettenmaier, D. P.: Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification, *Global Planet. Change*, 13, 195-206, doi: 1996.
- Lievens, H., Tomer, S. K., Al Bitar, A., De Lannoy, G., Drusch, M., Dumedah, G., Franssen, H.-J. H., Kerr, Y., Martens, B.,
25 and Pan, M.: SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia, *Remote Sens. Environ.*, 168, 146-162, doi: 10.1016/j.rse.2015.06.025, 2015.
- Liu, D., and Mishra, A. K.: Performance of AMSR_E soil moisture data assimilation in CLM4. 5 model for monitoring hydrologic fluxes at global scale, *J. Hydrol.*, 547, 67-79, doi: 10.1016/j.jhydrol.2017.01.036, 2017.
- Liu, Y., Dorigo, W. A., Parinussa, R., de Jeu, R. A., Wagner, W., McCabe, M. F., Evans, J., and Van Dijk, A.: Trend-preserving blending of passive and active microwave soil moisture retrievals, *Remote Sens. Environ.*, 123, 280-297, doi:
30 10.1016/j.rse.2012.03.014, 2012.
- Liu, Y. Y., Parinussa, R., Dorigo, W. A., De Jeu, R. A., Wagner, W., Van Dijk, A., McCabe, M. F., and Evans, J.: Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals, *Hydrol. Earth Syst. Sci.*, 15, 425, doi: 10.5194/hessd-7-6699-2010, 2011.



- Matgen, P., Fenicia, F., Heitz, S., Plaza, D., de Keyser, R., Pauwels, V. R., Wagner, W., and Savenije, H.: Can ASCAT-derived soil wetness indices reduce predictive uncertainty in well-gauged areas? A comparison with in situ observed soil moisture in an assimilation application, *Adv. Water Resour.*, 44, 49-65, doi: 10.1016/j.advwatres.2012.03.022, 2012.
- Merlin, O., Al Bitar, A., Walker, J. P., and Kerr, Y.: An improved algorithm for disaggregating microwave-derived soil moisture based on red, near-infrared and thermal-infrared data, *Remote Sens. Environ.*, 114, 2305-2316, doi: 5 10.1016/j.rse.2010.05.007, 2010.
- Mohanty, B. P., Cosh, M., Lakshmi, V., and Montzka, C.: Remote sensing for vadose zone hydrology—a synthesis from the vantage point, *Vadose Zone J.*, 12, doi: 10.2136/vzj2013.07.0128, 2013.
- Mohanty, B. P., Cosh, M. H., Lakshmi, V., and Montzka, C.: Soil moisture remote sensing: State-of-the-science, *Vadose Zone J.*, 16, doi: 0.2136/vzj2016.10.0105, 2017.
- Montzka, C., Pauwels, V. R., Han, X., Franssen, H. J. H. and Vereecken, H.: Multiscale and multivariate data assimilation in terrestrial systems: A review, *Sensors*, 12, 16291-16333, doi:10.3390/s121216291, 2012.
- Nerger, L., and Hiller, W.: Software for ensemble-based data assimilation systems—Implementation strategies and scalability, *Comput. Geosci.*, 55, 110-118, doi: 10.1016/j.cageo.2012.03.026, 2013.
- 15 Niu, G. Y., Yang, Z. L., Dickinson, R. E., and Gulden, L. E.: A simple TOPMODEL-based runoff parameterization (SIMTOP) for use in global climate models, *J. Geophys. Res-Atmos.*, 110, doi: 10.1029/2005jd006111, 2005.
- Niu, G. Y., Yang, Z. L., Dickinson, R. E., Gulden, L. E., and Su, H.: Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data, *J. Geophys. Res-Atmos.*, 112, doi: 10.1029/2006jd007522, 2007.
- 20 Oleson, K., Dai, Y., Bonan, G., Bosilovich, M., Dickinson, R., Dirmeyer, P., Hoffman, F., Houser, P., Levis, S., and Niu, G.: Technical description of the community land model (CLM), NCAR Technical Note NCAR/TN-461+ STR, National Center for Atmospheric Research, Boulder, CO, 2004.
- Oleson, K., Niu, G. Y., Yang, Z. L., Lawrence, D., Thornton, P., Lawrence, P., Stöckli, R., Dickinson, R., Bonan, G., and Levis, S.: Improvements to the Community Land Model and their impact on the hydrological cycle, *J. Geophys. Res-*
- 25 *Biogeo.*, 113, doi: 10.1029/2007jg000563, 2008.
- Owe, M., de Jeu, R., and Holmes, T.: Multisensor historical climatology of satellite-derived global land surface moisture, *J. Geophys. Res-Earth*, 113, doi: 10.1029/2007jf000769, 2008.
- Pan, M., Wood, E. F., Wójcik, R., and McCabe, M. F.: Estimation of regional terrestrial water cycle using multi-sensor remote sensing observations and data assimilation, *Remote Sens. Environ.*, 112, 1282-1294, doi: 30 10.1016/j.rse.2007.02.039, 2008.
- Pauwels, V., Hoeben, R., Verhoest, N. E., De Troch, F. P., and Troch, P. A.: Improvement of TOPLATS-based discharge predictions through assimilation of ERS-based remotely sensed soil moisture values, *Hydrol. Process.*, 16, 995-1013, doi: 10.1002/hyp.315, 2002.



- Pauwels, V. R., Hoeben, R., Verhoest, N. E., and 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, doi: 10.1016/s0022-1694(01)00440-1, 2001.
- Pipunic, R., Walker, J., and Western, A.: Assimilation of remotely sensed data for improved latent and sensible heat flux prediction: A comparative synthetic study, *Remote Sens. Environ.*, 112, 1295-1305, doi: 10.1016/j.rse.2007.02.038, 2008.
- Reichle, R., Crow, W., Koster, R., Sharif, H., and Mahanama, S.: Contribution of soil moisture retrievals to land data assimilation products, *Geophys. Res. Lett.*, 35, doi: 10.1029/2007gl031986, 2008.
- Reichle, R. H., McLaughlin, D. B., and Entekhabi, D.: Hydrologic data assimilation with the ensemble Kalman filter, *Mon. Weather Rev.*, 130, 103-114, doi: 10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2, 2002.
- Reichle, R. H., and Koster, R. D.: Global assimilation of satellite surface soil moisture retrievals into the NASA Catchment land surface model, *Geophys. Res. Lett.*, 32, doi: 10.1029/2004gl021700, 2005.
- Renzullo, L. J., Van Dijk, A., Perraud, J.-M., Collins, D., Henderson, B., Jin, H., Smith, A., and McJannet, D.: Continental satellite soil moisture data assimilation improves root-zone moisture analysis for water resources assessment, *J. Hydrol.*, 15 519, 2747-2762, doi: 10.1016/j.jhydrol.2014.08.008, 2014.
- Rodell, M., Houser, P., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C., Arsenault, K., Cosgrove, B., Radakovich, J., and Bosilovich, M.: The global land data assimilation system, *B. Am. Meteorol. Soc.*, 85, 381-394, doi: 10.1175/BAMS-85-3-381, 2004.
- Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., and Hashimoto, H.: A continuous satellite-derived measure of global terrestrial primary production, *BioScience*, 54, 547-560, doi: 10.1641/0006-3568(2004)054[0547:ACSMOG]2.0.CO;2, 2004.
- Sahoo, A. K., De Lannoy, G. J., Reichle, R. H., and Houser, P. R.: Assimilation and downscaling of satellite observed soil moisture over the Little River Experimental Watershed in Georgia, USA, *Adv. Water Resour.*, 52, 19-33, doi: 10.1016/j.advwatres.2012.08.007, 2013.
- Schaap, M. G., and Leij, F. J.: Database-related accuracy and uncertainty of pedotransfer functions, *Soil Sci.*, 163, 765-779, doi: 10.1097/00010694-199810000-00001, 1998.
- Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., and Teuling, A. J.: Investigating soil moisture-climate interactions in a changing climate: A review, *Earth-Sci Rev.*, 99, 125-161, doi: 10.1016/j.earscirev.2010.02.004, 2010.
- Shock, C. C., Barnum, J. M., and Seddigh, M.: Calibration of watermark soil moisture sensors for irrigation management. *Proc. Int. Irrig. Show, San Diego California USA*, 139-146, doi: 10.1007/bf00007875, 998.
- Shrestha, P., Sulis, M., Masbou, M., Kollet, S., and Simmer, C.: A scale-consistent terrestrial systems modeling platform based on COSMO, CLM, and ParFlow, *Mon. weather rev.*, 142, 3466-3483, doi: 10.1175/mwr-d-14-00029.1, 2014.



- Simmer, C., Adrian, G., Jones, S., Wirth, V., Göber, M., Hohenegger, C., Janjic, T., Keller, J., Ohlwein, C., and Seifert, A.:
HErZ: The German Hans-Ertel Centre for Weather Research, *B. Am. Meteorol. Soc.*, 97, 1057-1068, doi:
10.1175/bams-d-13-00227.1, 2016.
- Springer, A., Eicker, A., Bettge, A., Kusche, J., and Hense, A.: Evaluation of the Water Cycle in the European COSMO-
5 REA6 Reanalysis Using GRACE, *Water*, 9, 289, doi:10.3390/w9040289, 2017.
- Sridhar, V., Hubbard, K. G., You, J., and Hunt, E. D.: Development of the soil moisture index to quantify agricultural
drought and its “user friendliness” in severity-area-duration assessment, *J. Hydrometeorol.*, 9, 660-676, doi:
10.1175/2007jhm892.1, 2008.
- Stöckli, R., Lawrence, D., Niu, G. Y., Oleson, K., Thornton, P. E., Yang, Z. L., Bonan, G., Denning, A., and Running, S. W.:
10 Use of FLUXNET in the Community Land Model development, *J. Geophys. Res-Bioge.*, 113, doi:
10.1029/2007jg000562, 2008.
- Trenberth, K. E., Smith, L., Qian, T., Dai, A., and Fasullo, J.: Estimates of the global water budget and its annual cycle using
observational and model data, *J. Hydrometeorol.*, 8, 758-769, doi: 10.1175/jhm600.1, 2007.
- Vereecken, H., Schnepf, A., Hopmans, J. W., Javaux, M., Or, D., Roose, T., Vanderborght, J., Young, M. H., Amelung, W.,
15 Aitkenhead, M., Allison, S. D., Assouline, S., Baveye, P., Berli, M., Brüggemann, N., Finke, P., Flury, M., Gaiser, T.,
Govers, G., Ghezzehei, T., Hallett, P., Hendricks Franssen, H. J., Heppell, J., Horn, R., Huisman, J. A., Jacques, D.,
Jonard, F., Kollet, S., Lafolie, F., Lamorski, K., Leitner, D., McBratney, A., Minasny, B., Montzka, C., Nowak, W.,
Pachepsky, Y., Padarian, J., Romano, N., Roth, K., Rothfuss, Y., Rowe, E. C., Schwen, A., Šimůnek, J., Tiktak, A., Van
Dam, J., van der Zee, S. E. A. T. M., Vogel, H. J., Vrugt, J. A., Wöhling, T., and Young, I. M.: Modeling Soil
20 Processes: Review, Key Challenges, and New Perspectives, *Vadose Zone J.*, 15, 10.2136/vzj2015.09.0131, 2016.
- Verhoest, N. E., van den Berg, M. J., Martens, B., Lievens, H., Wood, E. F., Pan, M., Kerr, Y. H., Al Bitar, A., Tomer, S. K.,
and Drusch, M.: Copula-based downscaling of coarse-scale soil moisture observations with implicit bias correction,
IEEE T. Geosci. Remote, 53, 3507-3521, doi: 10.1109/tgrs.2014.2378913, 2015.
- Vinukollu, R. K., Wood, E. F., Ferguson, C. R., and Fisher, J. B.: Global estimates of evapotranspiration for climate studies
25 using multi-sensor remote sensing data: Evaluation of three process-based approaches, *Remote Sens. Environ.*, 115,
801-823, doi: 10.1016/j.rse.2010.11.006, 2011.
- Wagner, W., Dorigo, W., de Jeu, R., Fernandez, D., Benveniste, J., Haas, E., and Ertl, M.: Fusion of active and passive
microwave observations to create an essential climate variable data record on soil moisture, *ISPRS Annals of the
Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Annals)*, 7, 315-321, 2012.
- 30 Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., Figa-Saldaña, J., de Rosnay, P., Jann, A., and
Schneider, S.: The ASCAT soil moisture product: A review of its specifications, validation results, and emerging
applications, *Meteorol. Z.*, 22, 5-33, doi: 10.1127/0941-2948/2013/0399, 2013.
- Wang, J. R.: Microwave emission from smooth bare fields and soil moisture sampling depth, *IEEE transactions on
geoscience and remote sensing*, 616-622, 1987, doi: 10.1109/TGRS.1987.289840.



- Western, A. W., Zhou, S.-L., Grayson, R. B., McMahon, T. A., Blöschl, G., and Wilson, D. J.: Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes, *J. Hydrol.*, 286, 113-134, doi: 10.1016/j.jhydrol.2003.09.014, 2004.
- Wood, E. F., Roundy, J. K., Troy, T. J., Van Beek, L., Bierkens, M. F., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaffè, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A., and Whitehead, P.: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, *Water Resour. Res.*, 47, 10.1029/2010WR010090, 2011.
- 5 J., Gochis, D., van de Giesen, N., Houser, P., Jaffè, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A., and Whitehead, P.: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, *Water Resour. Res.*, 47, 10.1029/2010WR010090, 2011.
- Wu, D. D., Anagnostou, E. N., Wang, G., Moges, S., and Zampieri, M.: Improving the surface-ground water interactions in the Community Land Model: Case study in the Blue Nile Basin, *Water Resour. Res.*, 50, 8015-8033, doi: 10.1002/2013wr014501, 2014.
- 10 10.1002/2013wr014501, 2014.
- Yang, Z.-L., and Niu, G.-Y.: The versatile integrator of surface and atmosphere processes: Part 1. Model description, *Global Planet. Change*, 38, 175-189, doi: 10.1016/S0921-8181(03)00028-6, 2003.
- Zeng, X., and Decker, M.: Improving the numerical solution of soil moisture-based Richards equation for land models with a deep or shallow water table, *J. Hydrometeorol.*, 10, 308-319, doi: 10.1175/2008jhm1011.1, 2009.

15

20

25

30



Tables and Figures

5

Table 1: Monthly mean bias (CLM minus E-RUN) in mean seasonal runoff (mm/day) for CLM-OL and CLM-DA for all PRUDENCE regions and all seasons, i.e. winter (DJF), spring (MAM), summer (JJA) and autumn (SON).

Regions	Winter		Spring		Summer		Autumn		Annual	
	CLM-OL (mm/d)	CLM-DA (mm/d)	CLM-OL (mm/d)	CLM-DA (mm/d)	CLM-OL (mm/d)	CLM-DA (mm/d)	CLM-OL (mm/d)	CLM-DA (mm/d)	CLM-OL (mm/d)	CLM-DA (mm/d)
BI	1.56	-1.07	1.50	-0.52	0.93	-0.27	0.41	-0.95	1.10	-0.70
IP	0.24	-0.49	0.53	-0.34	0.32	-0.03	-0.11	-0.20	0.24	-0.27
FR	1.26	-0.44	1.42	-0.33	0.71	-0.06	0.26	-0.23	0.91	-0.26
ME	0.57	-0.33	1.27	-0.15	0.53	-0.10	0.35	-0.23	0.68	-0.20
SC	0.79	-0.14	0.53	-0.10	2.48	0.29	1.09	-0.29	1.22	-0.06
AL	1.26	-0.49	1.30	-0.76	2.05	-0.13	1.02	-0.68	1.41	-0.52
MD	0.53	-0.49	1.30	-0.23	0.67	0.04	0.18	-0.16	0.67	-0.21
EA	0.54	-0.10	1.34	0.10	0.91	-0.02	0.62	-0.09	0.85	-0.03
EU	0.84	-0.44	1.15	-0.29	1.08	-0.04	0.48	-0.35	0.89	-0.28

10

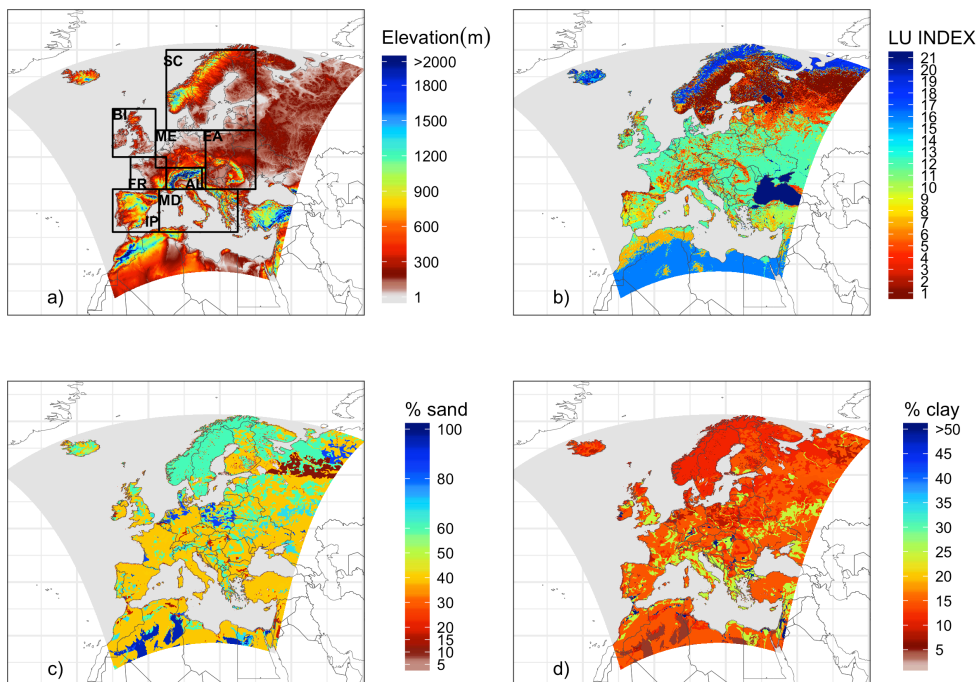


Figure 1: Model surface input data: a) USGS GMTED2010 DEM, b) dominant land use type based on MODIS data, c) percent sand content, and d) percent clay content based on global FAO soil database. The land use indices in (b) are defined as: 1:Evergreen Needleleaf Forest, 2:Evergreen Broadleaf Forest, 3:Deciduous Needleleaf Forest, 4:Deciduous Broadleaf Forest, 5:Mixed Forests, 6:Closed Shrublands, 7:Open Shrublands, 8: Woody Savannas, 9:Savannas, 10:Grasslands, 11:Permanent Wetlands, 12: Croplands, 13: Urban and Built-Up, 14: Cropland/Natural Vegetation Mosaic, 15: Snow and Ice, 16: Barren or Sparsely Vegetated, 17: Water, 18:Wooded Tundra, 19:Mixed Tundra, 20:Barren Tundra, 21: Lakes. The inner boxes in (a) show the boundaries of the PRUDENCE regions (FR: France, ME: mid-Europe, SC: Scandinavia, EA: Eastern Europe, MD: Mediterranean, IP: Iberian Peninsula, BI: British Islands, AL: Alpine region; Christensen et al., 2007).

5

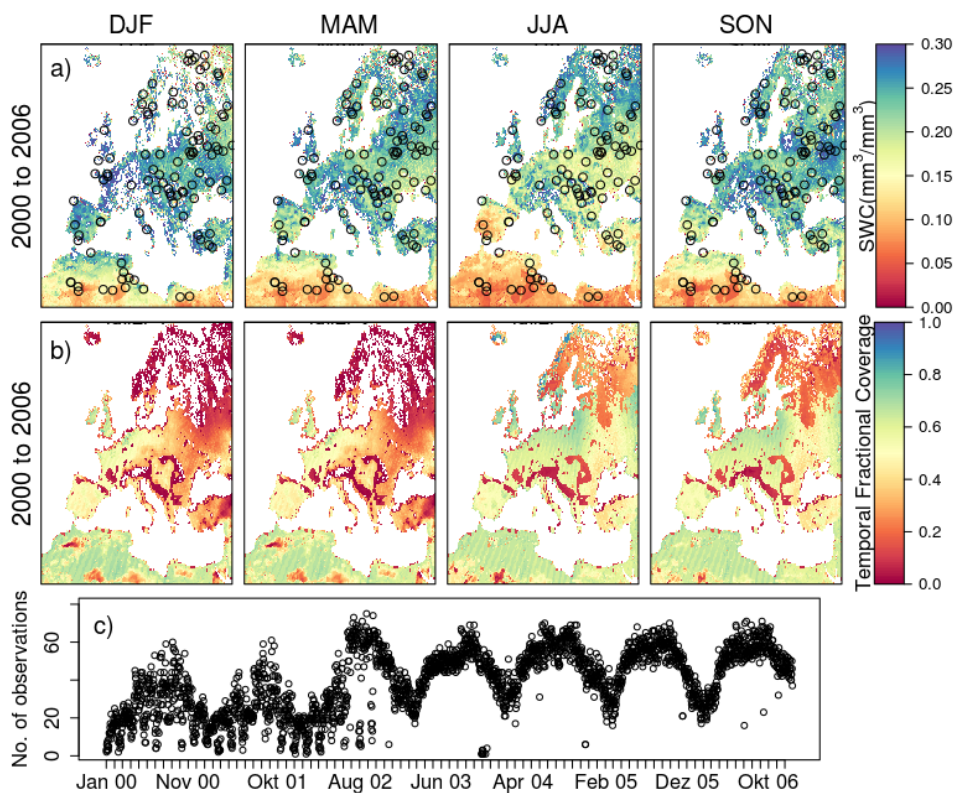


Figure 2: Satellite ESA-CCI soil moisture data resampled to 0.0275° resolution for the time period of 2000 to 2006 over EU-CORDEX. (a) Temporally averaged soil moisture content for different seasons, (b) fraction of days that soil moisture observations were reported during different seasons, and (c) number of selected observations with valid data for the respective day over the 2000-2006 period used for assimilating SM in data assimilation experiment. Black circles in (a) indicate the location of grid cells selected for data assimilation.

5

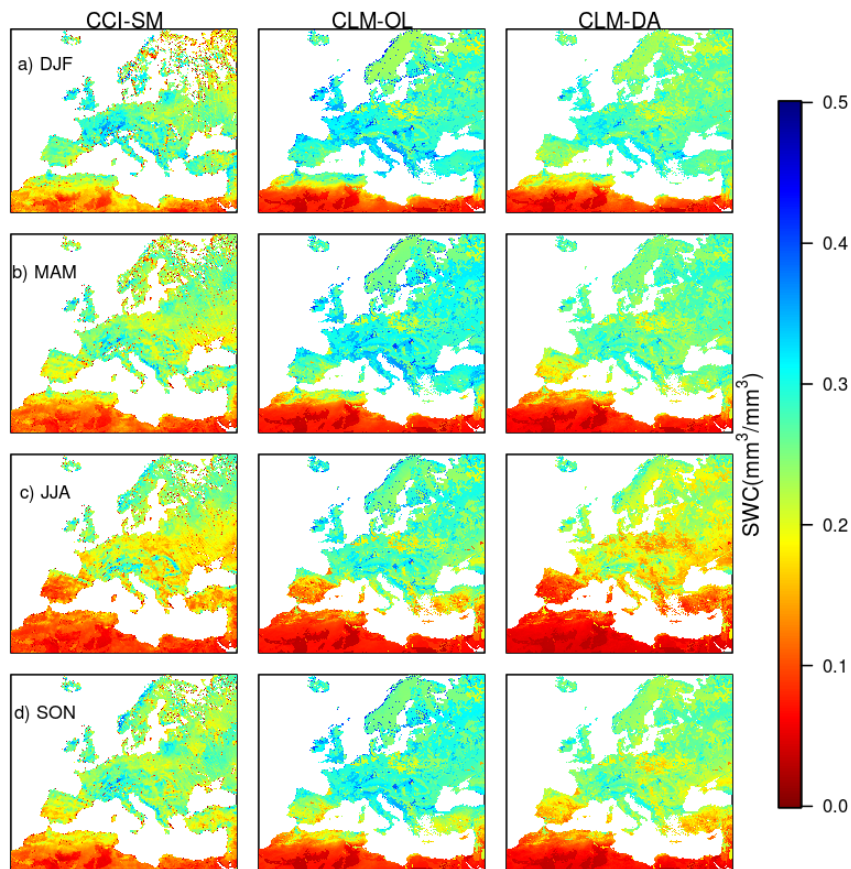


Figure 3: Temporally averaged soil moisture (mm^3/mm^3) content over different seasons simulated by CLM-OL and CLM-DA for 2000-2006 for a) DJF (December, January and February), (b) MAM (March, April, May), c) JJA (June, July and August) and d) SON (September, October, November) seasons. Temporally averaged seasonal soil moisture data from CCI-SM are also shown for comparison.

5

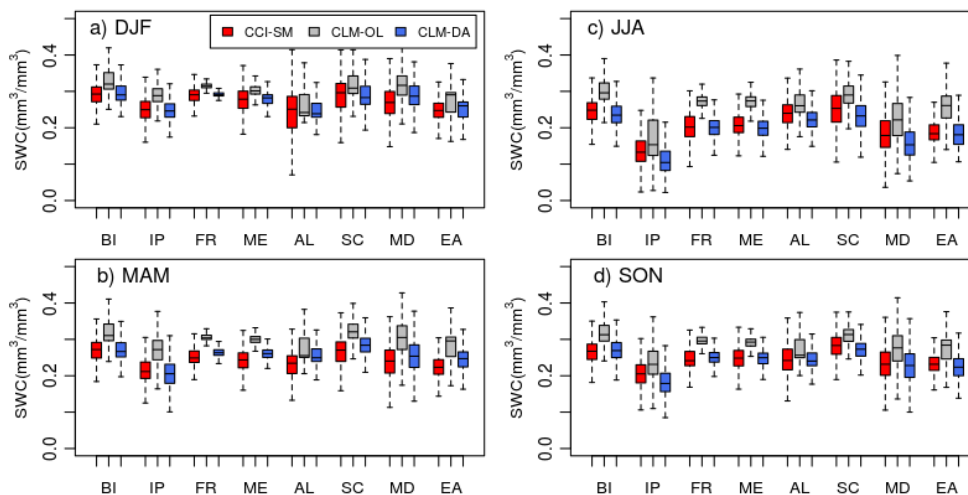


Figure 4: Box plots showing the spread of seasonally averaged soil water content (mm^3/mm^3) over the 2000 – 2006 time period and in the PRUDENCE regions for (a) DJF, (b) MAM, (c) JJA and (d) SON seasons. The boxplots illustrate the spatial distribution of SWC with quartiles, median and extreme values marked by solid lines.

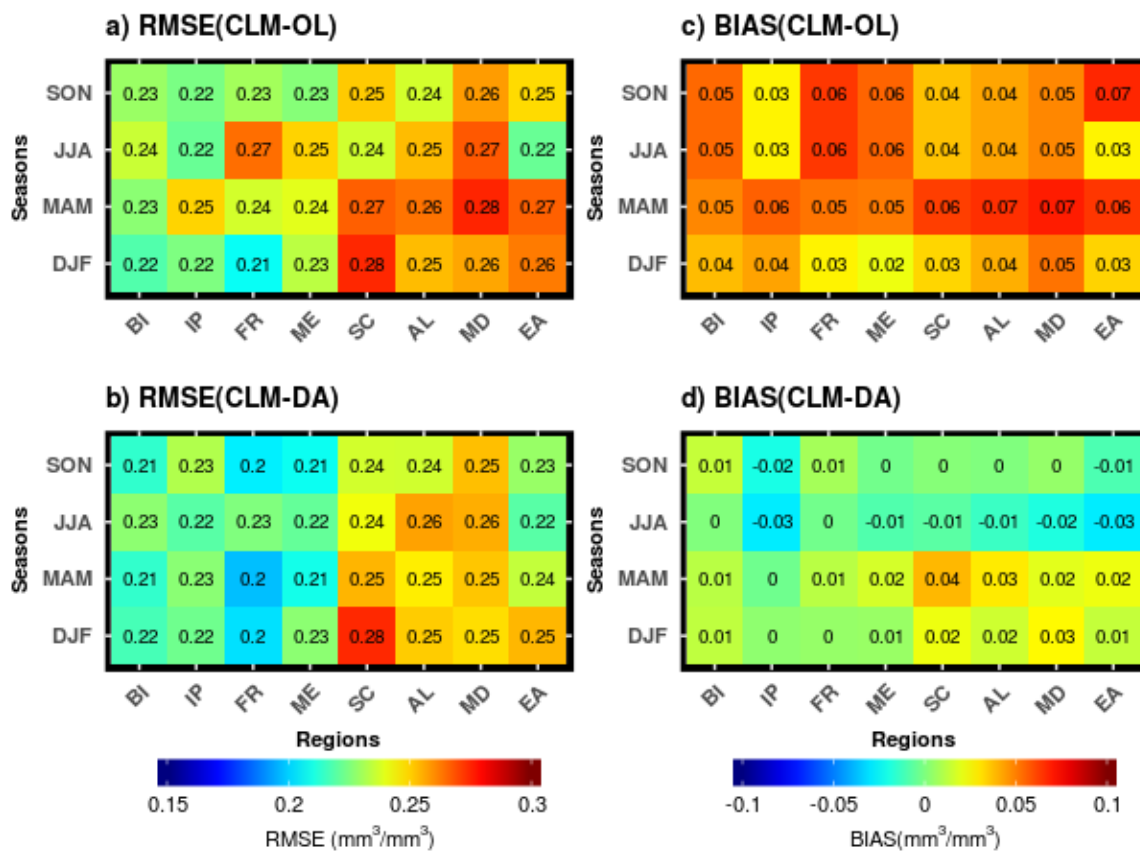


Figure 5: RMSE and BIAS for daily soil water content over different seasons and PRUDENCE regions for (a,c) CLM-OL and (b,d) CLM-DA simulations over the years 2000-2006.

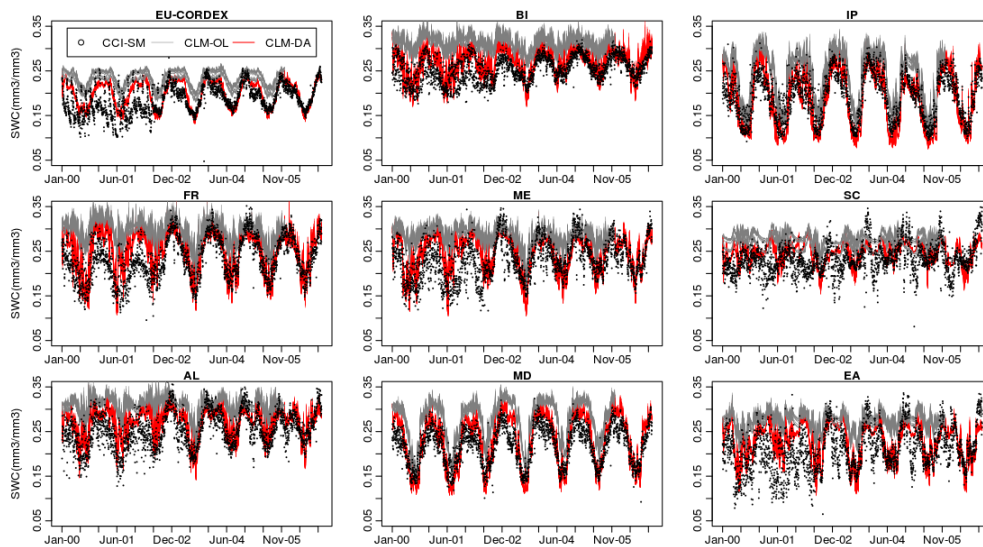


Figure 6: Spatially averaged daily soil water content (SWC) simulated with CLM-DA and CLM-OL and compared with CCI-SM data for the years 2000 – 2006 over the PRUDENCE regions.

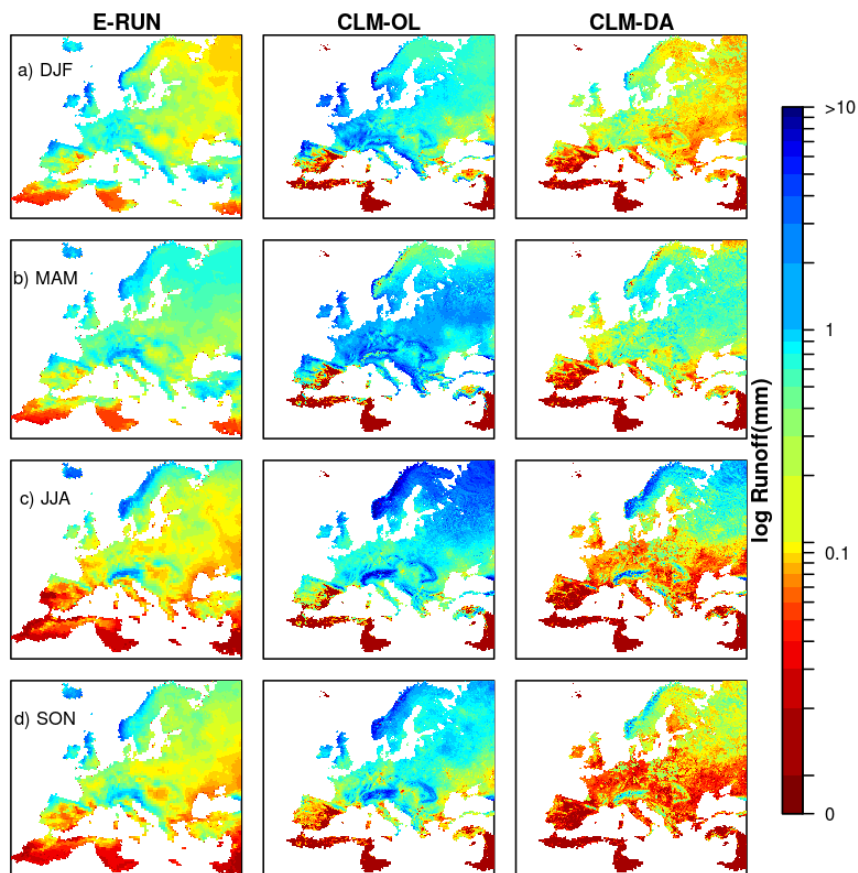


Figure 7: Temporally averaged monthly runoff (mm/day) over different seasons simulated by CLM-OL and CLM-DA for the years 2000 – 2006 for a) DJF, (b) MAM, c) JJA and d) SON seasons. Temporally averaged monthly runoff from E-RUN is shown for comparison.

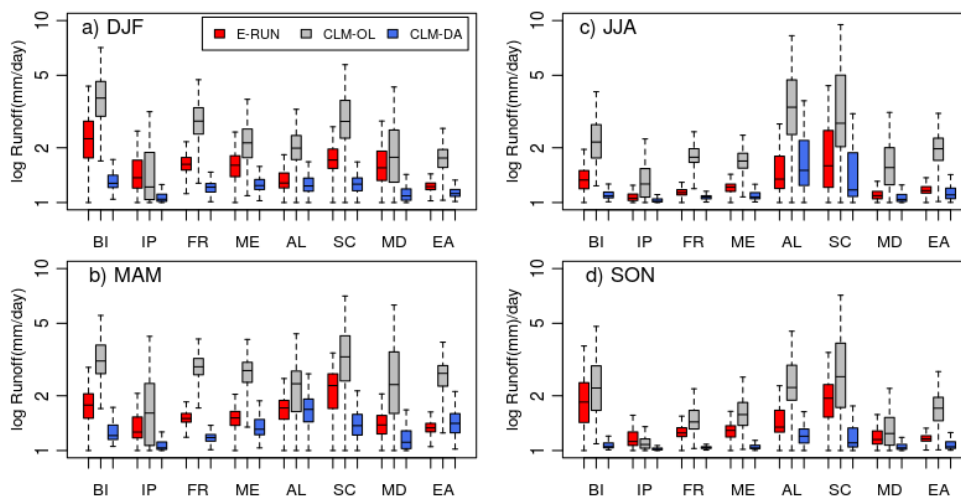


Figure 8: Boxplots of temporally averaged runoff (mm/day) over the years 2000 – 2006 for all PRUDENCE region and seasons, i.e. (a) DJF, (b) MAM, (c) JJA and (d) SON. The boxplots indicate the spatial distribution of monthly averaged runoff over each region.

5

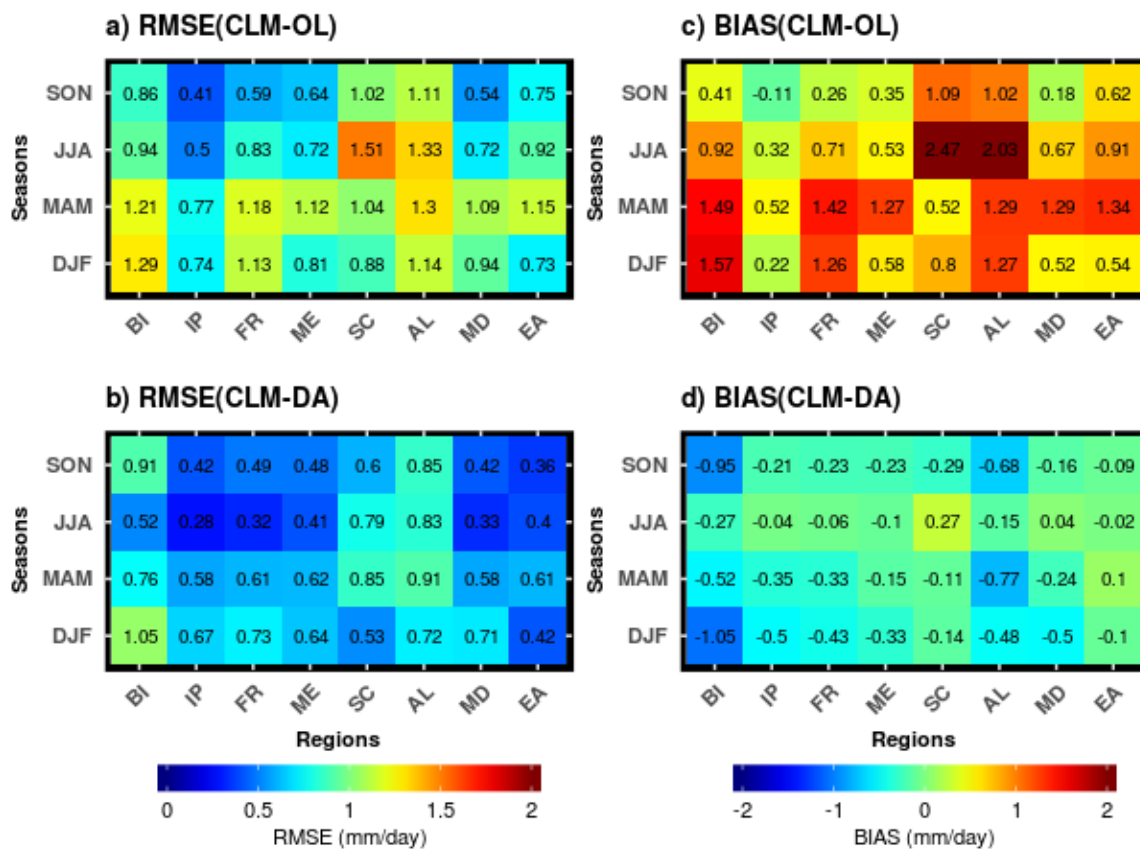


Figure 9: Root mean square error and mean bias error for runoff (mm/day) over different seasons in the PRUDENCE regions for (a,c) CLM-OL and (c,d) CLM-DA simulations over the years 2000-2006.

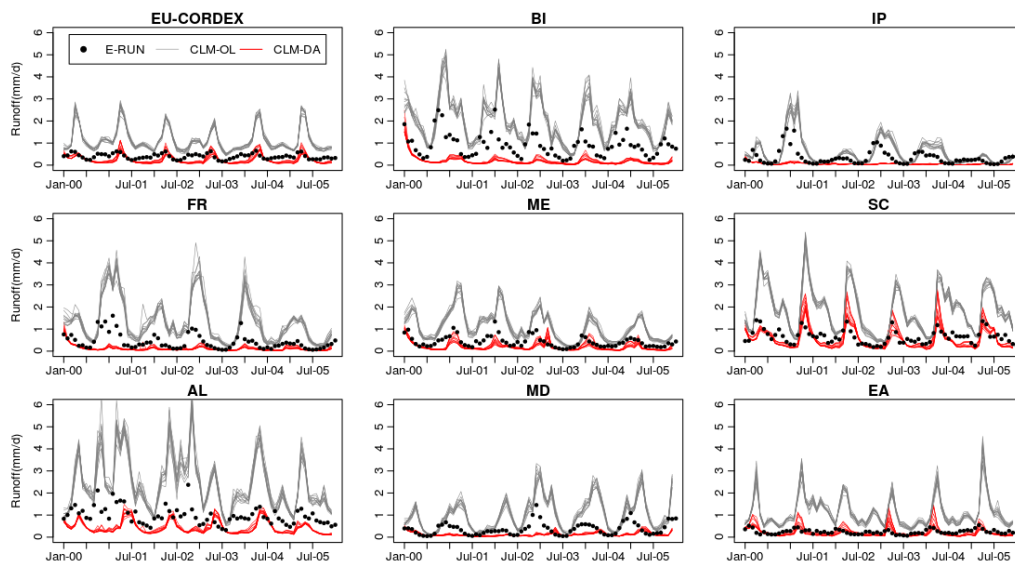


Figure 10: Monthly time series of runoff from CLM-DA and CLM-OL simulation and compared with E-RUN runoff observation data for the years 2000 – 2006 over the PRUDENCE regions.