Using data assimilation to optimize pedotransfer functions using large-scale in-situ soil moisture observations.

Elizabeth Cooper[1], Eleanor Blyth[1], Hollie Cooper[1], Rich Ellis[1], Ewan Pinnington[3], and Simon J. Dadson[1,2]

[1]UK Centre for Ecology and Hydrology, Wallingford, UK.
[2]School of Geography and the Environment, South Parks Road, Oxford OX1 3QY
[3]National Center for Earth Observation, Department of Meteorology, University of Reading, Reading, UK

Correspondence: Elizabeth Cooper (elicoo@ceh.ac.uk)

Abstract. Soil moisture predictions from land surface models are important in hydrological, ecological and meteorological applications. In recent years the availability of wide-area soil-moisture measurements has increased, but few studies have combined model-based soil moisture predictions with in-situ observations beyond the point scale. Here we show that we can markedly improve soil moisture estimates from the JULES land surface model using field scale observations and data assimilation techniques. Rather than directly updating soil moisture estimates towards observed values, we optimise constants in the underlying pedotransfer functions, which relate soil texture to JULES soil physics parameters. In this way we generate a single set of newly calibrated pedotransfer functions based on observations from a number of UK sites with different soil textures. We demonstrate that calibrating a pedotransfer function in this way can improve the performance of land surface models, leading to the potential for better flood, drought and climate projections.

1 Introduction

Soil moisture is an important physical variable, significant in agriculture (Pinnington et al., 2018), flood events (Koster et al., 2010; Berghuijs et al., 2019), and processes related to weather and climate (Seneviratne et al., 2010). Land surface models such as the Joint UK Land Environment Simulator (JULES) can be used to make predictions of soil moisture, and generally rely on empirical pedotransfer functions (PTFs) to relate readily available or easy-to-measure soil characteristics such as soil texture to the soil hydraulics parameters required by the model (see e.g.Van Looy et al. (2017)).

There are a number of different types of pedotransfer function, as noted in Van Looy et al. (2017) and Hodnett and Tomasella (2002), with different inputs and outputs depending partly on the requirements of the chosen land surface model. In 'class' approaches, soil types are clustered into groups, and hydraulic model parameters are then obtained from a look-up table (Wösten et al., 1999); this results in discrete soil hydraulics parameter sets. Alternatively, continuous pedotransfer functions take soil characteristic information from each sample of interest and apply the function to produce continuous soil hydraulics parameter sets (e.g.Cosby et al. (1984), Hodnett and Tomasella (2002), Schaap et al. (2001)). To date, pedotransfer functions have been derived by fitting to results from field or laboratory experiments on point or small-scale soil samples (cm to m), despite the fact that land surface models are generally applied at larger (field to km) scales. The recent development of novel in-situ techniques for measuring soil moisture over field, rather than point scale presents an opportunity to test whether land surface models, in conjunction with commonly used pedotransfer functions, are able to reproduce field-scale soil moisture observations. In this paper, we have compared JULES soil moisture predictions with soil moisture observations from the COSMOS-UK dataset (Stanley et al., 2019); these observations are measured by cosmic ray neutron sensors (CRNS) over a footprint of up to 12 ha. We have then used the LaVEnDAR four dimensional ensemble variational data assimilation framework (Pinnington et al., 2020) to combine COSMOS-UK soil moisture observations at 16 sites with equivalent JULES soil moisture estimates. We have thereby optimized constants in the Cosby pedotransfer function (Cosby et al., 1984). This results in a newly calibrated set of pedotransfer functions based on field-scale soil moisture observations across 16 sites with a range of soil types. This approach allows us to test whether we can improve the performance of the model by optimising the pedotransfer functions for larger scales using field-scale soil moisture observations. Our approach also allows comparison of the soil hydraulics parameters generated using large-scale ($\sim$ hundred metre) soil moisture measurements with those generated by the original pedotransfer functions, which are based on small-scale ($\sim$ cm) measurements. We chose to optimize the pedotransfer functions rather than directly optimizing soil physics parameters to better ensure physically consistent parameter sets; this approach also has the advantage that we can assimilate observations from all sites simultaneously to produce one set of pedotransfer functions applicable at all sites and beyond.

Larger scale soil moisture measurements are also increasingly available from satellite products and these have been used to good effect in data assimilation frameworks with land surface models in e.g. Pinnington et al. (2018), Liu et al. (2011), De Lannoy and Reichle (2016) and Yang et al. (2016). The advantage of the CRNS measurements used here is that they provide a more direct soil moisture measurement than those from satellites. CRNS soil measurements are also representative of depths of approximately 10 to 30cm, compared to the top 5 to 10 cm for satellite retrievals.

An alternative approach to ~~assimilation of~~ CRNS soil moisture measurements into land surface models is taken in Brunetti et al. (2019) and Han et al. (2015). These studies both use neutron counts from CRNS instruments as observations, combined with the method presented in Shuttleworth et al. (2013) to map modelled soil moisture estimates into equivalent neutron counts. In this study we instead directly compare modelled and CRNS derived soil moisture.

The rest of the paper is organised as follows: in section 2 we outline the JULES land surface model and the COSMOS-UK data used in this study; we also describe the data assimilation experiment we have performed and introduce the metric by which we ~~have~~ measured how well the model fits the observations. In section 3 we present results, showing that we can use COSMOS-UK observations from 2017 to improve the fit between the JULES model output and observations over two years at all the sites we included. We discuss our results in the context of changes ~~to~~ the JULES soil physics parameters in section 4. In section 5 we conclude that it is possible to optimise pedotransfer functions for field scale soil moisture measurements, and that this markedly improves the fit of JULES soil moisture estimates to COSMOS-UK observations.

2 Methods

2.1 JULES land surface model

JULES uses the Darcy-Richards equation to model soil hydraulic processes (Best et al., 2011), so that the downward water flux, W, between adjacent soil layers is given by

$$W = K\left(\frac{\partial \Psi}{\partial z} + 1\right) \tag{1}$$

where Ψ is the soil matric suction, K is the soil hydraulic conductivity and z is distance from the soil surface in the vertical direction.

JULES provides two options for representing the relation between soil water content, θ, matric suction, and hydraulic conductivity; in this paper we use the Brooks and Corey soil physics option (Best et al. (2011),Brooks and Corey (1964)), where we assume

$$\frac{\theta}{\theta_s} = \left(\frac{\Psi}{\Psi_s}\right)^{-\frac{1}{b}} \tag{2}$$

and

$$\frac{K}{K_s} = \left(\frac{\theta}{\theta_s}\right)^{2b+3}. \tag{3}$$

In equations (2) and (3), θ_s, K_s and Ψ_s are values of soil moisture, hydraulic conductivity and soil matric suction at saturation; b is a soil-dependent constant with a value usually determined through a pedotransfer function. The soil physics parameters used in the implementation of Brooks and Corey soil physics in JULES are briefly described in table 1; more details are available in Best et al. (2011) or Jules user guide. (2020).

The values of the eight soil physics parameters outlined in table 1 are generally calculated via a set of pedotransfer functions. Here we use the Cosby pedotransfer functions, which have the following mathematical formulation, (Cosby et al., 1984; Marthews et al., 2014)

$$b = \kappa_1 + \kappa_2 f_{clay} - \kappa_3 f_{sand} \tag{4}$$

$$\theta_s = \kappa_4 - \kappa_5 f_{clay} - \kappa_6 f_{sand} \tag{5}$$

$$\Psi_s = 0.01 \times 10^{\kappa_7 - \kappa_8 f_{clay} - \kappa_9 f_{sand}} \tag{6}$$

Parameter name and symbol	**Description**
satcon, K_s	Hydraulic conductivity at saturation (kgm^{-2})
sathh, Ψ_s	Absolute value of the soil matric suction at saturation (m)
vsat, θ_s	Volumetric water content at saturation (m^{3}m^{-3})
vcrit, θ_{crit}	Volumetric soil moisture content at -33kPa (critical point) m^{3}m^{-3}
vwilt, θ_{wilt}	Volumetric soil moisture content at -1500kPa (wilting point) m^{3}m^{-3}
b	Exponent in soil hydraulic characteristic
$hcap$	Dry soil heat capacity (J m^{-3}K^{-1}).
$hcon$	Dry soil thermal conductivity (W m^{-1}K^{-1}).

Table 1. Soil physics parameters

$$K_s = 10^{-\kappa_{10}-\kappa_{11}f_{clay}+\kappa_{12}f_{sand}} \times \frac{25.4}{3600} \tag{7}$$

$$\theta_{crit} = \theta_s \left(\frac{\Psi_s}{3.364}\right)^{1/b}, \tag{8}$$

$$\theta_{wilt} = \theta_s \left(\frac{\Psi_s}{152.9}\right)^{1/b} \tag{9}$$

$$h_{cap} = (1-\theta_s)(2.376\times10^6 f_{clay} + 2.133\times10^6 f_{silt} + 2.133\times10^6 f_{sand}) \tag{10}$$

$$h_{con} = 0.025^{\theta_s} \times 1.16^{f_{clay}(1-\theta_s)} \times 1.57^{f_{sand}(1-\theta_s)} \times 1.57^{f_{silt}(1-\theta_s)} \tag{11}$$

where f_{clay}, f_{sand} and f_{silt} are fractions of clay and silt in the soil. The values of the constants κ_1 to κ_{12} usually used in equations (4) to (11) are those given in Cosby et al. (1984); we present them in table 2. These values are empirically determined from 1448 small soil samples (cm dimensions) taken from 23 states in the United States (for further details of the soil samples and sampling methods see Rawls (1976) and Holtan (1968)).

Constant	Value from Cosby et al. (1984)
κ_1	3.10
κ_2	15.70
κ_3	0.3
κ_4	0.505
κ_5	0.037
κ_6	0.142
κ_7	2.17
κ_8	0.63
κ_9	1.58
κ_{10}	0.6
κ_{11}	0.64
κ_{12}	1.26

Table 2. Values of the constants commonly used in the Cosby pedotransfer functions

JULES requires meteorological driving data to produce soil moisture estimates. In this paper we have used half-hourly meteorological observations measured at COSMOS-UK sites as driving data; in this way we can use JULES to give soil moisture predictions at any COSMOS-UK sites with sufficiently complete meteorological data.

JULES provides estimates of soil moisture at various depths; in the standard configuration used here these correspond to four layers, with depths [0,10cm], [10cm to 35cm], [35cm to 100cm] and [100cm to 300cm]. The JULES layers are often referred to by their thicknesses, which are 10cm, 25cm, 65cm and 200cm respectively. Here, we refer to the soil moisture estimates for the four layers as SM_{10},SM_{25},SM_{65} and SM_{200}.

2.2 COSMOS-UK soil moisture data

The COSMOS-UK project comprises a network of soil moisture monitoring stations across the United Kingdom, providing long-term soil moisture measurements at around 50 sites. Data for 2013 to 2017 are available in the EIDC archive (Stanley et al. (2019)). Soil moisture observations are made using an innovative Cosmic Ray Neutron Sensor (CRNS) instrument at each site; these provide a measurement of soil moisture over an area of up to 12 ha (30 acres) (Antoniou et al., 2019; Evans et al., 2016). We have used daily-averaged soil moisture data from 16 COSMOS-UK sites as observations in this paper. The sites were selected based on completeness of soil moisture and meteorological data over a three year period from 2016-2018 and are listed in Table 3, with details of land cover and broad soil descriptions taken from Antoniou et al. (2019). Locations of the sites are shown in figure 1. For more details of each of the sites, see Antoniou et al. (2019). The Cosby pedotransfer function was designed to work for mineral soils and we therefore consider that the first seven sites listed in table 3 are those at which the JULES model can be expected to provide a good match to observations via our chosen PTF; soil types at the remaining sites mean that JULES may not be able to represent the observed soil moisture time series as accurately.

Sitename and abbreviation	**Land cover**	**Soil description**
Cardington (CARDT)	Grassland	Typical mineral soil
Bickley Hall (BICKL)	Grassland	Typical mineral soil
Crichton (CRICH)	Grassland	Typical mineral soil
Waddesdon (WADDN)	Grassland	Typical mineral soil
Hollin Hill (HOLLN)	Grassland	Typical mineral soil
Easter Bush (EASTB)	Grassland	Typical mineral soil
Rothamstead (ROTHD)	Grassland	Typical mineral soil
Chimney Meadows (CHIMN)	Grassland	Calcereous mineral soil
Sheepdrove (SHEEP)	Grassland	Mineral soil; fairly high organic carbon content
Porton Down (PORTN)	Grassland	Highly calcereous mineral soil
Hartwood Home (HARTW)	Grassland/woodland	Typical mineral soil
Gisburn Forest (GISBN)	Coniferous forest	Mineral soil; high organic carbon content
Chobham Common (CHOBH)	Heath	Highly variable soil
Lullington Heath (LULLN)	Grassland/heath	Highly calcareous mineral soil
Moorhouse (MOORH)	Grassland/heath	Mineral soil with very high organic content
Sourhope (SOURH)	Grassland	Mineral with very high organic content

Table 3. COSMOS-UK sites selected for this study

Both the depth and the footprint over which the CRNS measure soil moisture change with soil moisture (Evans et al. (2016); Köhli et al. (2015) Antoniou et al. (2019)), with the footprint and depth of the measurement both becoming smaller as soil moisture increases. The COSMOS-UK dataset includes estimates of the depth over which each daily soil moisture value is valid, known as a $D86$ value. Measurements of several other environmental variables are made at COSMOS-UK sites, using a suite of instrumentation. These include point soil moisture and temperature measurements at various depths in the soil, and meteorological variables. We have used half-hourly in-situ meteorological data from the COSMOS-UK dataset as driving data for the JULES model.

2.3 Data assimilation

Data assimilation is a group of methods in which information from models and observations is combined in order to give the best estimate of the state of a physical system and/or model parameter values. In this paper we have used the four-dimensional ensemble variational data assimilation technique, LaVEnDAR, which is introduced in Pinnington et al. (2020) and is based on Liu et al. (2008). We use LaVEnDAR here to optimise 12 constants, κ_1 to κ_{12}, in the Cosby pedotransfer functions (eqns

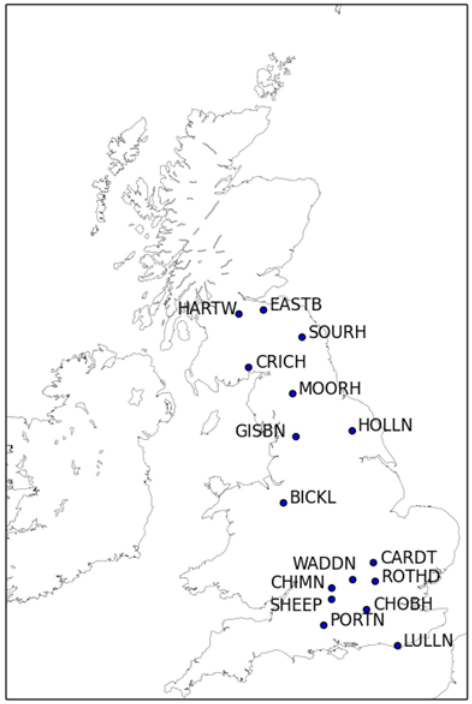

Figure 1. Locations of COSMOS-UK sites used in this study

(4) to (11)) based on estimates of soil moisture from JULES and corresponding large-scale observations of soil moisture from COSMOS-UK. The values of κ_1 to κ_{12} are assumed to be constant in time and space; the same values are used across all sites to generate soil JULES moisture estimates via the pedotransfer functions.

2.4 Experimental details

In order to use COSMOS-UK data with JULES outputs in the LaVEnDAR scheme, we require both sets of soil moisture values to correspond to the same soil depth. We have therefore devised a weighted depth approach, in which we extract from each JULES prediction an average soil moisture corresponding to the UK-COSMOS $D86$ observed depth. For each day, we calculate a depth-adjusted JULES soil moisture estimate, SM_{depth}, depending on the 75m $D86$ value provided for that day, such that

$$SM_{depth} = \begin{cases} SM_{10}, & \text{if } D86 \leq 10\text{cm}, \\ \frac{10}{D86} SM_{10} + \frac{(D86-10)}{D86} SM_{25} & \text{if } 10\text{cm} < D86 \leq 35\text{cm}, \\ \frac{10}{D86} SM_{10} + \frac{25}{D86} SM_{25} + \frac{(D86-35)}{D86} SM_{65}, & \text{if } 35\text{cm} < D86 \leq 65\text{cm}, \end{cases} \tag{12}$$

where SM_{10}, SM_{25} and SM_{65} are the JULES predicted soil moisture values from the [0,10cm], [10cm to 35cm] and [35cm to 100cm] layers respectively, and the D86 value is given in cm. In this way, thickness-weighted contributions to the soil moisture are taken from every JULES layer which would be wholly or partly contained within the D86 depth. We have not taken the COSMOS-UK variable footprint into account in this study.

In this paper we have used an ensemble size of 50, as in related experiments in Pinnington et al. (2020) and Liu et al. (2008). In order to implement the LaVEnDAR scheme we

1. Generated a 50-member ensemble of each of the 12 PTF constants κ_1 to κ_{12}. These were obtained by sampling from a Gaussian distribution centred on the value given in table 3, with standard deviation equal to 10% of the mean. This standard deviation value was chosen fairly arbitrarily; future work could assess the sensitivity of the results to the values chosen for each PTF constant.

2. Assembled 50 unique sets of 12 constants κ_1 to κ_{12}.

3. Used each unique set of constants in equations (4) to (11) to generate 50 sets of soil physics parameters for each site. Soil texture information for each site was taken from the Harmonised World Soil Database (HWSD) (Fischer et al., 2008)

4. Used the soil parameter sets to run 50 realisations of JULES at each of our selected sites over a 2 year time window to create a prior ensemble of 50 soil moisture time series per site.

5. Used the LaVEnDAR scheme to generate a new, posterior ensemble of values for each of the 12 PTF constants, taking into account COSMOS-UK soil moisture observations from 2017. Here, we assumed uncorrelated observation errors of 50 $\%$ of the mean soil moisture value at each site.

6. Used the new posterior ensemble of PTF constants to generate 50 posterior sets of soil physics variables at each site.

7. Ran 50 posterior realisations of JULES at each site to create posterior soil moisture time series.

These steps are also shown in schematic form in figure 2.

We have assumed a high observation error value in this experiment. The daily soil moisture measurements we use are averaged from 30 minute soil moisture measurements; analysis of these data shows that uncertainty in the daily values is approximately 20%. We have inflated this here to 50% observation error as we found that smaller observation error values impacted negatively on the posterior soil moisture results. Inflation of the observation error here compensates for the assumption of uncorrelated observation error; in fact there will likely be intra-site correlations between observation errors due to site-specific instrument calibration. We note that similar observation error covariance inflation techniques have been used in e.g. assimilation of satellite observations in numerical weather prediction (Fowler et al. (2018), Hilton et al. (2009)). We have used COSMOS-UK measurements from 2017 only in our data assimilation experiments, but compared the prior and posterior JULES runs from 2017 and 2018 with observations.

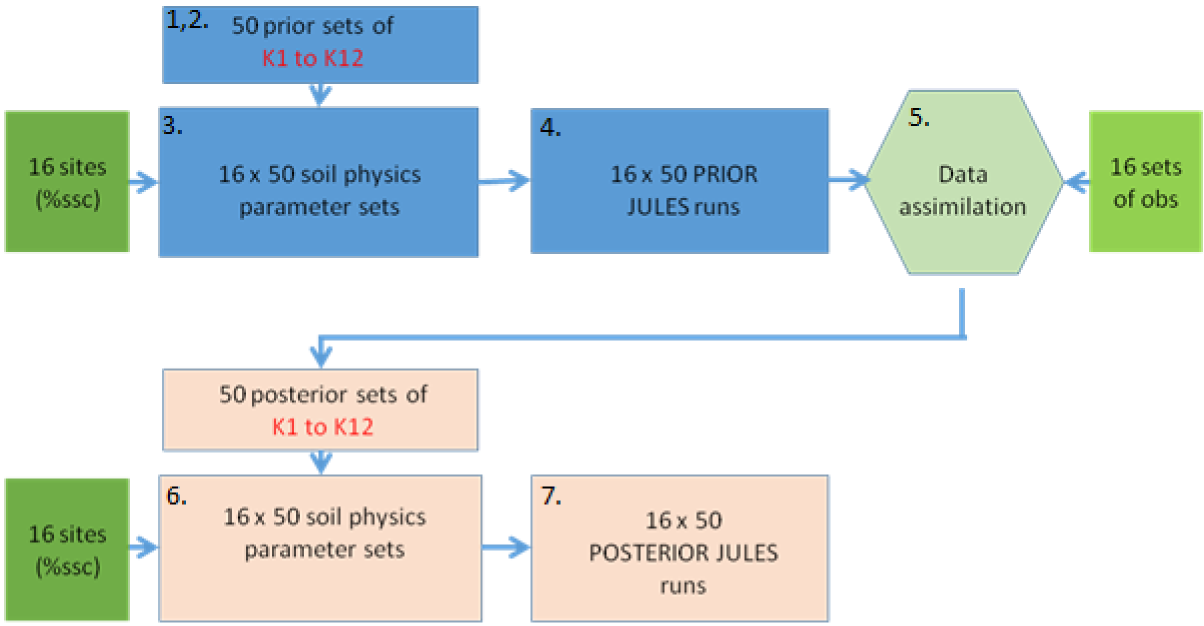

Figure 2. Schematic showing data assimilation experimental design; %ssc refers to site specific fractions of sand, silt and clay in the soil

2.5 Metrics

In order to assess how well our prior and posterior JULES runs match COSMOS-UK observations we require a metric. Here we have used the Kling-Gupta efficiency metric, as described in Gupta et al. (2009); Knoben et al. (2019) to compare the goodness of fit between observed and modelled (ensemble mean) soil moisture times series. The Kling Gupta efficiency (KGE) is given by

$$KGE = 1 - \sqrt{(1-r)^2 + (1-\alpha)^2 + (1-\beta)^2}, \tag{13}$$

where

$$\alpha = \frac{\sigma_{model}}{\sigma_{obs}} \tag{14}$$

and

$$\beta = \frac{\mu_{model}}{\mu_{obs}}. \tag{15}$$

In equations (14) and (15), μ_{model} and μ_{obs} are mean values of the modelled and measured soil moisture time series respectively; σ_{model} and σ_{obs} are the standard deviations in the modelled and observed soil moisture time series. The value of r is the Pearson correlation coefficient between the model and the observation time series data, and can vary between -1 (anti-correlation) and 1 (perfect correlation), with score of 0 indicating no correlation. The value of α reflects how well the spread in the modelled soil moisture values matches that of the observations, with a value of 1 corresponding to perfect matching. Equation (15) shows that the value of β represents bias between the model and observations, with a value of 1 indicating zero bias. Since α and β can be larger or smaller than 1, the value of the KGE can range between 1 (perfect model fit to data) to very large negative values. In Knoben et al. (2019) the authors argue that while in some studies a threshold of KGE $\geq$0 has been used to denote 'good' model performance, a lower threshold of KGE $\geq$-0.41 is required for the model to perform better than a mean persistence forecast.

3 Results

3.1 Effect of data assimilation on JULES soil moisture predictions

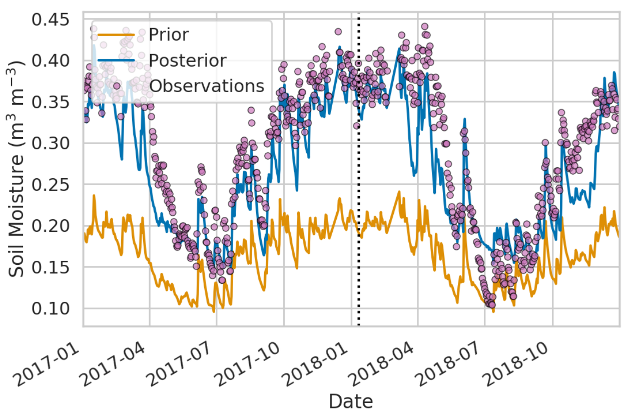

Figure 3. Observed and modelled (ensemble mean) soil moisture time series at Bickley Hall (BICKL). The dotted line separates the period over which observations used for assimilation (2017) from the period in which no observations have been assimilated (2018).

Figures 3 to 6 show measured and modelled soil moisture time series for 2017 and 2018 at two representative COSMOS-UK sites. In all cases the modelled soil moisture series is the ensemble mean. These figures show that the JULES runs using posterior PTF constants produce soil moisture estimates which are a better match to the observations than the JULES runs using the prior PTF constants. Figures 3 and 4 show results from Bickley Hall (BICKL), which is a site at which we expect soil moisture to be well represented by JULES via the Cosby PTF (this site has a typical mineral soil). Figures 5 and 6 represent results from a site at which the high organic content of the soil means that we do not expect our JULES setup to match the observations so successfully.

Figure 7a shows the KGE values for prior and posterior JULES runs at all 16 sites included in our study. These metrics show how closely the prior and posterior JULES runs match the observations over the period of 2017 and 2018 before and after assimilation of observations from 2017. Figure 7a shows that data assimilation makes fits to observations significantly better at all sites according to the Kling Gupta metric; all the analysis Kling-Gupta efficiency scores are closer to the ideal value

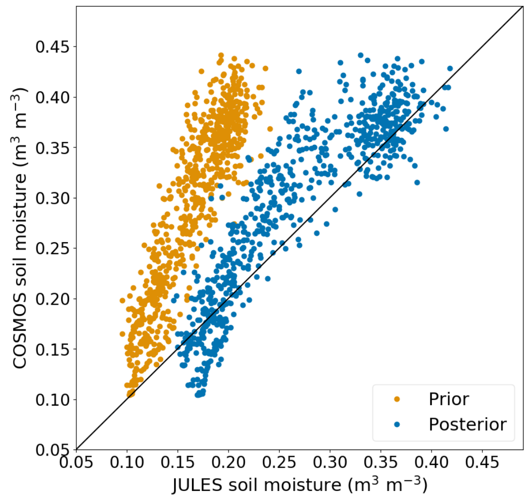

Figure 4. Observed vs modelled (ensemble mean) soil moisture at Bickley Hall (BICKL) for prior and posterior JULES runs. Diagonal line shows 1:1 perfect correspondence line.

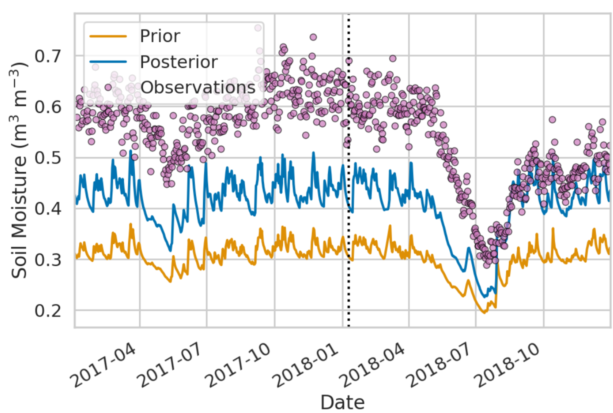

Figure 5. Observed and modelled (ensemble mean) soil moisture time series at Gisburn Forest (GISBN). The dotted line separates the period over which observations used for assimilation (2017) from the period in which no observations have been assimilated (2018).

of 1 than the prior values. We note that for all sites, the match between model and measurements is better in 2017 and 2018 even though only observations from 2017 were used in the optimization process. This indicates that the new values for the PTF constants allow JULES to simulate large-scale soil moisture measurements better than the original (prior) PTF constants. Figure 7b shows that the prior and posterior correlation coefficients, r, are very similar at most sites, although there are a number of sites at which the r value gets slightly worse following data assimilation. Despite this, the prior and posterior r values are all greater than 0.8 at all sites with a typical mineral soil. The r value stays low at Moorhouse (MOORH), perhaps because the soil at this site is too highly organic for the Cosby parameters to really be applicable, and for the COSMOS-UK

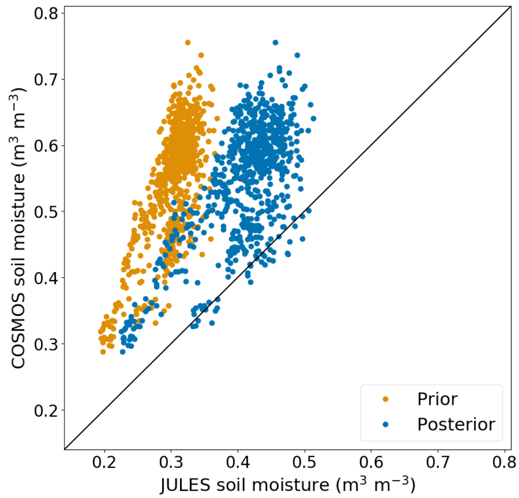

Figure 6. Observed vs modelled (ensemble mean) soil moisture at Gisburn Forest (GISBN) for prior and posterior JULES runs. Diagonal line shows 1:1 perfect correspondence line.

measurements to be reliable. The r value also stays low at Gisburn Forest (GISBN), which is likely due to the fact that there are a large number of trees at this site and interception processes are therefore likely to be more important here than at other sites; the COSMOS-UK 'soil moisture' measurement may actually include water held on the canopy of trees, which is not included in the JULES modelled soil moisture value here. Figure 7c shows that a significant contribution to improved KGE at all sites comes from improvement in the alpha component, which is much closer to the ideal value of 1 for all of the posterior JULES runs than the prior JULES runs. The alpha component represents how well the spread in the model matches the spread in the observations. We saw in time series plots such as figures 3 and 5 that the spread in JULES soil moisture was too small at all sites; our results show that the data assimilation has acted to correct this by updating the value of the PTF constants. Figure 7d shows that the beta parameter is closer to the ideal value of 1 after data assimilation than before at all sites except for Cardington, i.e. data assimilation is correcting a bias in the JULES outputs at all but one site. The prior bias at Cardington is in the opposite direction to bias at all of the other sites.

3.2 Effect of data assimilation on soil physics parameters

Figures 8 and 9 show that the mean value of K_s gets smaller (4 to 5 times smaller) at each site after data assimilation, and that the posterior distribution of the K_s parameter is narrower than the prior distribution. The results in figures 8 and 9 also show that the site-to-site variability of the b parameter reduces following data assimilation; the largest mean prior values of b are reduced, and the distributions with the smallest mean values are shifted to slightly larger values. Figures 8 and 9 show that the mean value θ_s has increased at all the sites following data assimilation, and the distribution of θ_s at each site has become much narrower. The mean values of the θ_{crit} and θ_{wilt} distributions have stayed broadly similar or increased slightly after data assimilation. We also see that at all sites Ψ_s becomes very small ($\sim$ 30 times smaller) after data assimilation.

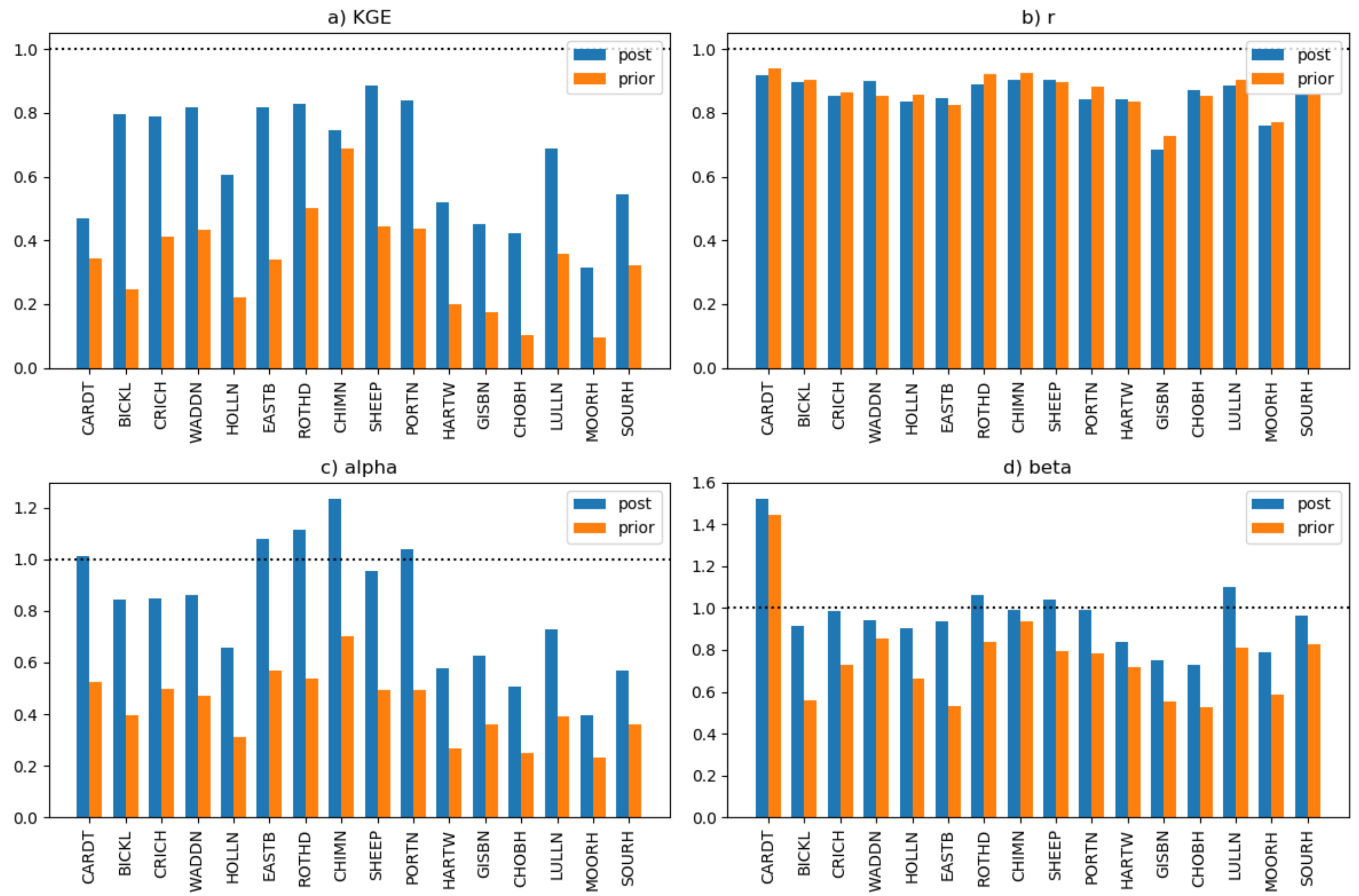

Figure 7. Kling Gupta efficiency scores for JULES runs using prior and posterior PTF variable values. Dotted horizontal lines show value of metric for perfect match between model and observation

Figures 8 and 9 show that h_{cap} and h_{con} change through data assimilation. However, this translates into minimal differences between the prior and post soil temperatures; both prior and post data assimilation temperature estimates are close to the in-situ COSMOS-UK measurements (not shown).

3.3 Effect of data assimilation on pedotransfer function constants

Figure 10 shows prior (orange) and posterior (blue) distributions of the 12 PFT constants, κ_1 to κ_{12}. These plots demonstrate how the dependence of the soil physics parameters on texture is changed in equations (4) to (11) via data assimilation. The values of κ_1, κ_2 and κ_3 control the magnitude of the soil physics parameter b through equation (4). The decreases of κ_2 and κ_3 after data assimilation translate to a decreased dependence of b on clay and sand fractions through equation (4). Changes to κ_4, κ_5 and κ_6 contribute to changes to θ_s through equation (5). The large increase in κ_4 values allows larger values of θ_s

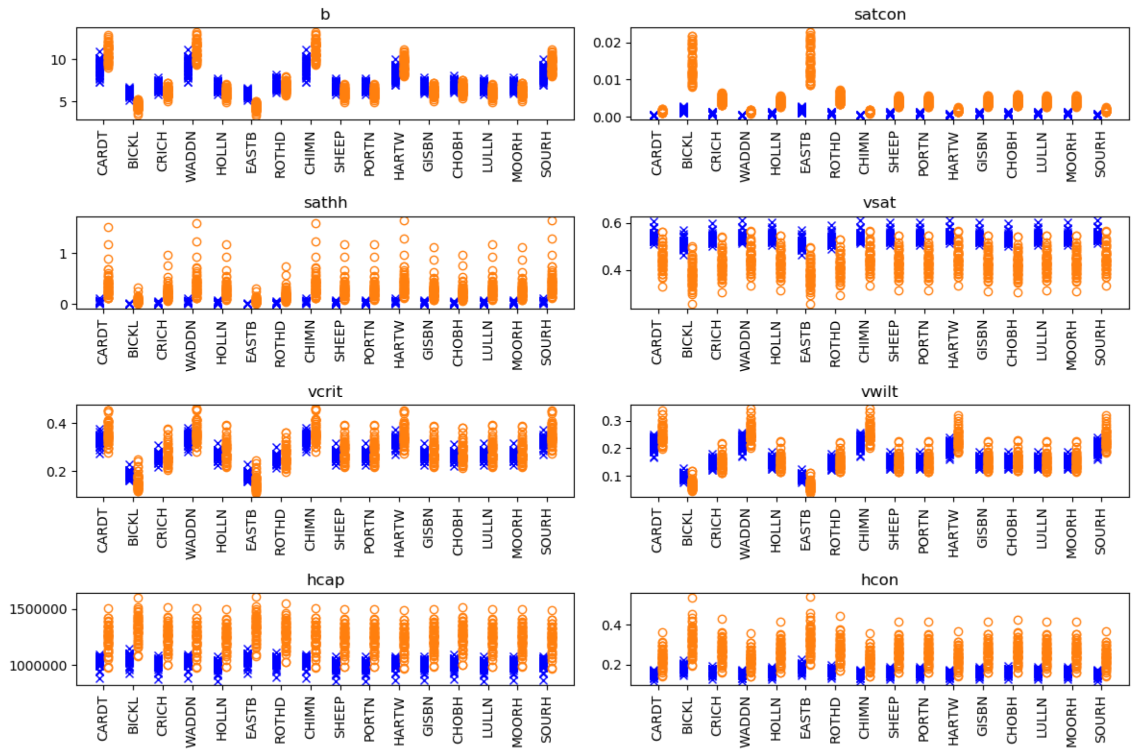

Figure 8. Ensemble prior (orange) and posterior (blue) parameter values at each site. These are ‘topsoil’ results, which we have assumed to correspond to the top two soil layers in JULES.

to be realised after data assimilation. The parameter Ψ_s controlled by κ_7, κ_8 and κ_9. The mean value of κ_7 is greatly reduced following data assimilation, and this leads to the much smaller posterior values of Ψ_s seen in figures 8 and 9. The constants κ_{10}, κ_{11} and κ_{12} determine the values of K_s through equation (6). The shift in the κ_{10} distribution to larger values leads to the reduction in values of K_s seen in figures 8 and 9.

4 Discussion

The results in section 3.1 show that we have been able to successfully update the constants in a Cosby-like pedotransfer function based on field scale in-situ soil moisture measurements. The new set of constants obtained in this way generate soil physics parameters at each studied COSMOS-UK site such that there is a large improvement in the match between modelled and observed field-scale soil moisture ~~values~~ at all sites.

Our results suggest that it is primarily a combination of the changes to θ_s, Ψ_s and K_s distributions which result in a better match to the observations after data assimilation. The new distributions allow the model to access higher soil moisture values, potentially correcting for a deficiency in supporting datasets, parameter values or process representation in JULES. We

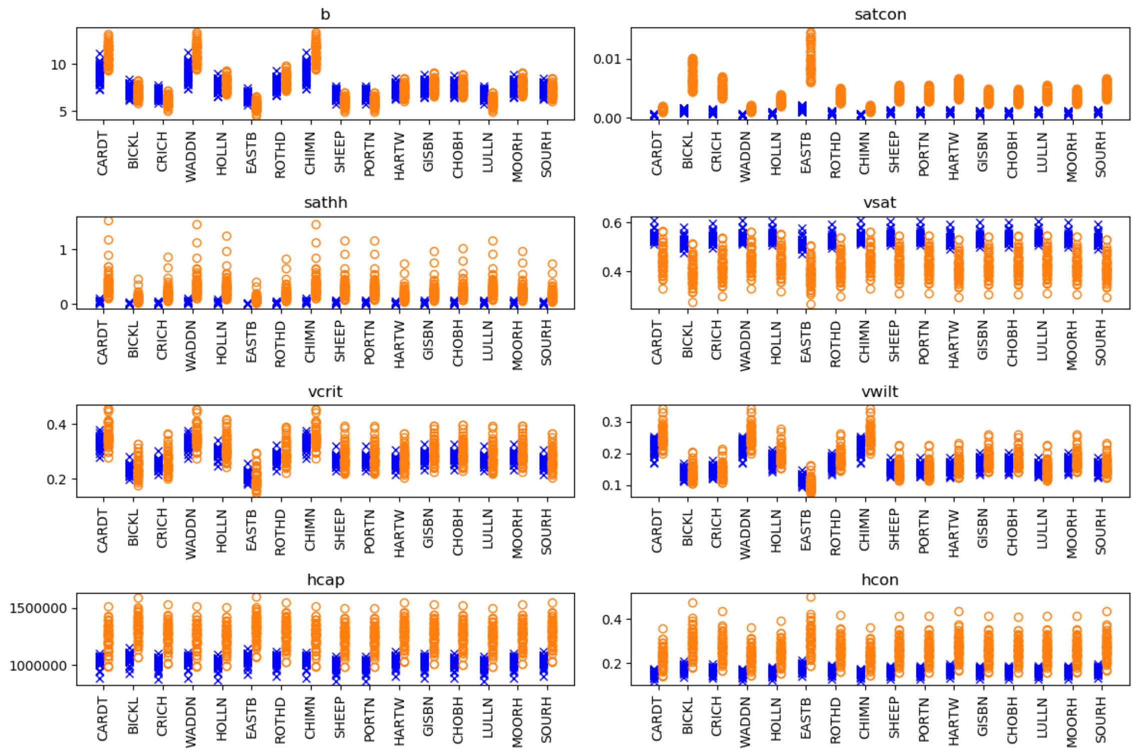

Figure 9. Ensemble prior (orange) and posterior (blue) parameter values at each site. These are ‘subsoil’ results, which we have assumed to correspond to the deeper two soil layers in JULES.

suggest that the data assimilation is effectively acting to slow the drainage of water in JULES, especially close to saturation, by increasing θ_s and decreasing K_s.

The improvements seen here were obtained by assimilating all the soil moisture values across 16 sites simultaneously rather than on a per site basis. This strengthens our implicit assumption that the same physical processes can be modelled (through JULES and the Cosby pedotransfer function) for a range of different UK sites and soil types.

We note that the soil physics parameter values calculated here may not exactly match physically expected values for a number of reasons. Firstly, we have fitted to COSMOS large-scale measurements; differences in parameter values from the prior values may therefore reflect the different scales on which they were calculated. Additionally, the COSMOS-UK soil moisture observations likely include contributions from processes which are important to soil moisture but we have not taken account of here with JULES, such as ponding of water on the soil surface, interception of water on vegetation, groundwater processes and local soil compaction. Therefore, we may be effectively parameterising for these processes (and others not included in JULES) through our new soil physics parameters. In this experiment we have mainly used grass sites, so interception is not likely to play a large role in daily averaged moisture values (JULES outputs show the amount of water intercepted to be, at most, of the order 100 times smaller than the amount of water in the top soil layer).

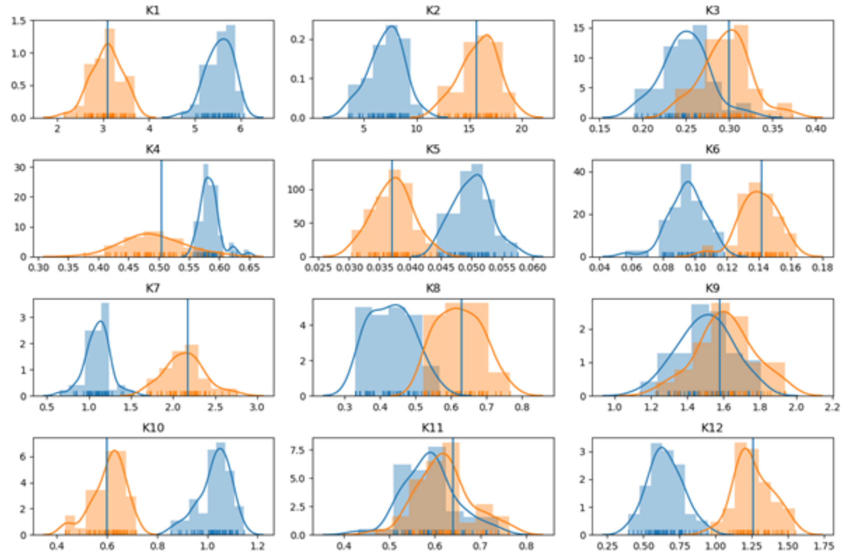

Figure 10. Prior and post PTF variable value distributions. Orange shows prior and blue posterior

5 Conclusions

We have shown that it is possible to use the LaVEnDAR data assimilation framework to improve JULES estimates of soil moisture based on one year's worth of large-scale COSMOS-UK soil moisture measurements across 16 sites. We have demonstrated improved fit to observations over a two year period at all 16 sites by adjusting the values of constants in the underlying pedotransfer function. Averaging across all the sites we see an improvement in the average KGE metric from 0.33 (range 0.10 to 0.69) before data assimilation to an average of 0.66 after data assimilation (range 0.31 to 0.89).

The method we propose here could be used for any different choice of land surface model and/or pedotransfer function; our choice of PTF here was motivated by the fact that it is widely used and has a relatively simple mathematical formulation. Calibrating PTFs for the soils on which they are to be used and at the scales at which they are applied, rather than on small-scale field or lab soil samples, will ultimately improve the performance of land surface models. This will allow better estimates from flood forecasting models, earth system models and numerical weather prediction.

Code availability. TEXT

Data availability. TEXT

Code and data availability. The code used in these experiments is available from the MetOffice JULES repository (https://code.metoffice.gov.uk/trac/jules) under Rose suite number u-bq016.

The LAVENDAR data assimilation first release is available here: https://github.com/pyearthsci/lavendar.

COSMOS-UK data are deposited annually in the Environmental Information Data Centre (EIDC) (eidc.ac.uk); additional data not included in the online repository are available on request (cosmos.ceh.ac.uk). https://doi.org/10.5285/a6012796-291c-4fd6-a7ef-6f6ed0a6cfa5)

Author contributions. EC, EP and RE devised the experiments, with input from EB and SD. EP created the LaVEnDAR data assimilation framework. EC and EP designed the rose-suite used here and ran the experiments. EC, RE, EP, EB and SD all contributed to analysis of results. HC provided access to COSMOS-UK data and site-specific information for model setup. EC prepared the manuscript with inputs from all co-authors.

Competing interests. The authors have no competing interests.

Disclaimer. TEXT

Acknowledgements. This work was supported by the Natural Environment Research Council grant number NE/S017380/1 as part of the Hydro-JULES programme. The authors gratefully acknowledge the provision by UKCEH of hydrometeorological and soil data collected by the COSMOS-UK project. COSMOS-UK is funded by the Natural Environment Research Council award number NE/R016429/1 as part of the UK-SCAPE programme.

References

Antoniou, V., Askquith-Ellis, A., Bagnoli, S., Ball, L., Bennett, E., Blake, J., Boorman, D., Brooks, M., Clarke, M., Cooper, H., Cowan, N., Cumming, A., Doughty, L., Evans, J., Farrand, P., Fry, M., Hewitt, N., Hitt, O., Jenkins, A., Kral, F., Libre, J., Lord, W., Roberts, C., Morrison, R., Parkes, M., Nash, G., Newcomb, J., Rylett, D., Scarlett, P., Singer, A., Stanley, S., Swain, O., Thornton, J., Trill, E., Vincent, H., Ward, H., Warwick, A., Winterbourn, B., and Wright, G.: COSMOS-UK user guide: users' guide to sites, instruments and available data (version 2.10), Tech. rep., Wallingford, http://nora.nerc.ac.uk/id/eprint/524801/, [Accessed 5th August 2020], 2019.

Berghuijs, W. R., Harrigan, S., Molnar, P., Slater, L. J., and Kirchner, J. W.: The Relative Importance of Different Flood-Generating Mechanisms Across Europe, Water Resources Research, 55, 4582–4593, https://doi.org/10.1029/2019WR024841, 2019.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Menard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., and Harding, R. J.: The Joint UK Land Environment Simulator (JULES), model description Part 1: Energy and water fluxes, Geoscientific Model Development, 4, 677–699, https://doi.org/10.5194/gmd-4-677-2011, 2011.

Brooks, R. H. and Corey, A. T.: Hydraulic properties of porous media, Hydrological Papers 3, 3, 1964.

Brunetti, G., Šimůnek, J., Bogena, H., Baatz, R., Huisman, J. A., Dahlke, H., and Vereecken, H.: On the Information Content of Cosmic-Ray Neutron Data in the Inverse Estimation of Soil Hydraulic Properties, Vadose Zone Journal, 18, 1–24, https://doi.org/10.2136/vzj2018.06.0123, 2019.

Cosby, B. J., Hornberger, G. M., Clapp, R. B., and Ginn, T. R.: A Statistical Exploration of the Relationships of Soil Moisture Characteristics to the Physical Properties of Soils, Water Resources Research, 20, 682–690, https://doi.org/10.1029/WR020i006p00682, 1984.

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

Evans, J. G., Ward, H. C., Blake, J. R., Hewitt, E. J., Morrison, R., Fry, M., Ball, L. A., Doughty, L. C., Libre, J. W., Hitt, O. E., Rylett, D., Ellis, R. J., Warwick, A. C., Brooks, M., Parkes, M. A., Wright, G. M. H., Singer, A. C., Boorman, D. B., and Jenkins, A.: Soil water content in southern England derived from a cosmic-ray soil moisture observing system - COSMOS-UK, Hydrological Processes, 30, 4987–4999, https://doi.org/10.1002/hyp.10929, 2016.

Fischer, G., Nachtergaele, F., Prieler, S., Van Velthuizen, H., Verelst, L., and Wiberg, D.: Global agro-ecological zones assessment for agriculture (GAEZ 2008), IIASA, Laxenburg, Austria and FAO, Rome, Italy, 10, http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/, [Accessed 5th August 2020], 2008.

Fowler, A. M., Dance, S. L., and Waller, J. A.: On the interaction of observation and prior error correlations in data assimilation, Quarterly Journal of the Royal Meteorological Society, 144, 48–62, https://doi.org/10.1002/qj.3183, 2018.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Han, X., Franssen, H.-J. H., Rosolem, R., Jin, R., Li, X., and Vereecken, H.: Correction of systematic model forcing bias of CLM using assimilation of Cosmic-Ray neutrons and land surface temperature: a study in the Heihe Hydrologic impacts of deforestation View project Correction of systematic model forcing bias of CLM using assimilation of cosmic-ray Neutrons and land surface temperature: a study in the Heihe Catchment, China, Hydrol. Earth Syst. Sci, 19, 615–629, https://doi.org/10.5194/hess-19-615-2015, 2015.

Hilton, F., Collard, A., Guidard, V., Randriamampianina, R., and Schwaerz, M.: ECMWF/EUMETSAT NWP-SAF Workshop on the assimilation of IASI in NWP, Assimilation of IASI Radiances at European NWP Centres, Tech. rep., https://www.ecmwf.int/sites/default/files/elibrary/2009/9896-assimilation-iasi-radiances-european-nwp-centres.pdf, [Accessed 5th August 2020], 2009.

Hodnett, M. G. and Tomasella, J.: Marked differences between van Genuchten soil water-retention parameters for temperate and tropical soils: a new water-retention pedo-transfer functions developed for tropical soils, Geoderma, 108, 155–180, https://doi.org/https://doi.org/10.1016/S0016-7061(02)00105-2, 2002.

Holtan, H.: Moisture-tension data for selected soils on experimental watersheds, Agricultural Research Service, U.S. Dept. of Agriculture, 1968.

Jules user guide., A.: JULES documentation, http://jules-lsm.github.io/latest/index.html, [Accessed 5th August 2020], 2020.

Knoben, W. J. M., Freer, J. E., and Woods, R. A.: Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, Hydrology and Earth System Sciences, 23, 4323–4331, https://doi.org/10.5194/hess-23-4323-2019, 2019.

Köhli, M., Schrön, M., Zreda, M., Schmidt, U., Dietrich, P., and Zacharias, S.: Footprint characteristics revised for field-scale soil moisture monitoring with cosmic-ray neutrons, Water Resources Research, 51, 5772–5790, https://doi.org/10.1002/2015WR017169, 2015.

Koster, R., Mahanama, S., Livneh, B., Lettenmaier, D., and Reichle, R.: Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow., Nature Geoscience, 3, 613–616, https://doi.org/10.1038/ngeo944, 2010.

Liu, C., Xiao, Q., and Wang, B.: An Ensemble-Based Four-Dimensional Variational Data Assimilation Scheme. Part I: Technical Formulation and Preliminary Test, Monthly Weather Review, 136, 3363–3373, https://doi.org/10.1175/2008MWR2312.1, 2008.

Liu, Q., Reichle, R. H., Bindlish, R., Cosh, M. H., Crow, W. T., De Jeu, R., De Lannoy, G. J., Huffman, G. J., and Jackson, T. J.: The contributions of precipitation and soil moisture observations to the skill of soil moisture estimates in a land data assimilation system, Journal of Hydrometeorology, 12, 750–765, https://doi.org/10.1175/JHM-D-10-05000.1, 2011.

Marthews, T. R., Quesada, C. A., Galbraith, D. R., Malhi, Y., Mullins, C. E., Hodnett, M. G., and Dharssi, I.: High-resolution hydraulic parameter maps for surface soils in tropical South America, Geoscientific Model Development, 7, 711–723, https://doi.org/10.5194/gmd-7-711-2014, 2014.

Pinnington, E., Quaife, T., and Black, E.: Impact of remotely sensed soil moisture and precipitation on soil moisture prediction in a data assimilation system with the JULES land surface model, Hydrology and Earth System Sciences, 22, 2575–2588, https://doi.org/10.5194/hess-22-2575-2018, 2018.

Pinnington, E., Quaife, T., Lawless, A., Williams, K., Arkebauer, T., and Scoby, D.: The Land Variational Ensemble Data Assimilation Framework: LAVENDAR v1.0.0, Geoscientific Model Development, 13, 55–69, https://doi.org/10.5194/gmd-13-55-2020, 2020.

Rawls, W.: Calibration of selected infiltration equations for the Georgia Coastal Plain, 1976.

Schaap, M. G., Leij, F. J., and van Genuchten, M. T.: ROSETTA: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions, Journal of Hydrology, 251, 163–176, https://doi.org/https://doi.org/10.1016/S0022-1694(01)00466-8, 2001.

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-Science Reviews, 99, 125–161, https://doi.org//10.1016/j.earscirev.2010.02.004, 2010.

Shuttleworth, J., Rosolem, R., Zreda, M., and Franz, T.: The COsmic-ray Soil Moisture Interaction Code (COSMIC) for use in data assimilation, Hydrology and Earth System Sciences, 17, 3205–3217, https://doi.org/10.5194/hess-17-3205-2013, 2013.

Stanley, S., Antoniou, V., Ball, L., Bennett, E., Blake, J., Boorman, D., Brooks, M., Clarke, M., Cooper, H., Cowan, N., Evans, J., Farrand, P., Fry, M., Hitt, O., Jenkins, A., Kral, F., Lord, W., Morrison, R., Nash, G., Rylett, D., Scarlett, P., Swain, O., Thornton, J., Trill, E., Warwick, A., and Winterbourn, J.: Daily and sub-daily hydrometeorological and soil data (2013-2017) [COSMOS-UK], https://doi.org/10.5285/a6012796-291c-4fd6-a7ef-6f6ed0a6cfa5, https://doi.org/10.5285/a6012796-291c-4fd6-a7ef-6f6ed0a6cfa5, 2019.

Van Looy, K., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., Montzka, C., Nemes, A., Pachepsky, Y. A., Padarian, J., Schaap, M. G., Tóth, B., Verhoef, A., Vanderborght, J., van der Ploeg, M. J., Weihermüller, L., Zacharias, S., Zhang, Y., and Vereecken, H.: Pedotransfer Functions in Earth System Science: Challenges and Perspectives, Reviews of Geophysics, 55, 1199–1256, https://doi.org/10.1002/2017RG000581, 2017.

Wösten, J. H. M., Lilly, A., Nemes, A., and Bas, C. L.: Development and use of a database of hydraulic properties of European soils, Geoderma, 90, 169–185, https://doi.org/https://doi.org/10.1016/S0016-7061(98)00132-3, 1999.

Yang, K., Zhu, L., Chen, Y., Zhao, L., Qin, J., Lu, H., Tang, W., Han, M., Ding, B., and Fang, N.: Land surface model calibration through microwave data assimilation for improving soil moisture simulations, Journal of Hydrology, 533, 266–276, https://doi.org/10.1016/j.jhydrol.2015.12.018, 2016.