Land surface model performance using cosmic-ray and point scale soil moisture measurements for calibration

Joost Iwema[1], Rafael Rosolem[1,2], Mostaquimur Rahman[1], Eleanor Blyth[3], Thorsten Wagener[1,2]

[1]Department of Civil Engineering, University of Bristol, Bristol, UK
5 [2]Cabot Institute, University of Bristol, Bristol, UK
[3]Centre for Ecology and Hydrology, Wallingford, OX10 8BB, UK

Correspondence to: Joost Iwema (joost.iwema@bristol.ac.uk)

Abstract. At the so-called hyper-resolution scale (i.e. grid cells of 1 km^2) land surface model parameters can be calibrated with Eddy-Covariance flux data and point-scale soil moisture data. However, measurement scales of Eddy-Covariance and 10 point-scale data differ substantially. In our study, we investigated the impact of reducing the scale mismatch between surface energy flux and soil moisture observations by replacing point-scale soil moisture data with observations derived from Cosmic-Ray Neutron Sensors made at larger spatial scales. Five soil and evapotranspiration parameters of the Joint UK Land Environment Simulator (JULES) were calibrated against point-scale and Cosmic-Ray Neutron Sensor soil moisture data separately. We calibrated the model for twelve sites in the USA representing a range of climatic, soil, and vegetation conditions. 15 The improvement in latent heat flux estimation for the two calibration solutions was assessed by comparing to Eddy-Covariance flux data and to JULES simulations with default parameter values. Calibrations against the two soil moisture products alone did show an advantage for the Cosmic-Ray technique. However, further analyses of two-objective calibrations with soil moisture and latent heat flux showed no substantial differences between both calibration strategies. This was mainly caused by the limited effect of calibrating soil parameters on soil moisture dynamics and surface energy fluxes. Other factors 20 that played a role were limited spatial variability in surface fluxes implied by soil moisture spatio-temporal stability, and data quality issues.

1 Introduction

The land surface water and energy balances are coupled through the process of evapotranspiration. Soil moisture is one of the main water reservoirs near the land surface and can hence importantly control the surface water and energy balances. Soil 25 moisture provides a first order (i.e. direct) control on evapotranspiration when there is insufficient water to meet the evaporative demand (Manabe et al., 1969; Budyko, 1956; Seneviratne et al., 2010). An indirect effect of soil moisture on surface energy flux partitioning is for instance the damping effect on soil and air temperature, which in its turn affects humidity, evapotranspiration, boundary-layer stability, and in some cases precipitation (Seneviratne et al., 2010). The control of soil moisture on temperature at seasonal scales is especially strong in transitional climate regions (Koster et al., 2004).

Land Surface Models (LSMs) solve the surface mass (including water), energy, and momentum balances to provide the weather and climate prediction models with lower boundary conditions. The land surface has been shown to play an important role in global atmospheric circulation (Koster et al., 2004). Because the soil moisture state and surface fluxes are so closely connected, it is important to accurately simulate these simultaneously (Henderson-Sellers et al., 1996; Richter et al., 2004; Seneviratne et al., 2010; Dirmeyer, 2011; Dirmeyer et al., 2013).

The increasing complexity of land surface models over the past decades (Sellers, 1997; Seneviratne et al., 2010) has brought with it an increasing number of parameters, with values not easily defined with in-situ measurements because of scale mismatch. For instance, stomatal resistance measured at leaf level is not the same as canopy stomatal resistance needed for LSMs (Blyth et al., 1993). Soil hydraulic parameter values (e.g. soil hydraulic conductivity) are often obtained from laboratory experiments on soil cores for cubic centimetres to cubic decimetres. The soil properties and processes at this scale can however differ from those at the LSM grid cell sizes, which are often as large as hundreds to tens of thousands of square kilometres (Pitman 2003). Due to the different governing processes upscaling the soil hydraulic properties from soil core scale to field scale is non-trivial (Vinnikov 1996; Crow et al., 2012).

In an effort to develop global hydrometeorological monitoring and prediction capabilities (Wood et al., 2011), hydrological models and LSMs are now increasingly being applied at the finer `hyper-resolution scale' with grid cells of about 1 km^2. Typically, parameters are calibrated and validated at this hyper-resolution against in-situ measurements from sources such as eddy-covariance flux towers and point-scale (PS) soil moisture sensors (e.g. Time Domain Transmissivity; and Time Domain Reflectometry) (Stockli et al., 2008; Richter et al., 2004; Blyth et al., 2010; Blyth et al., 2011; Rosolem et al., 2013). Such calibration/validation data has become more widely available at hyperresolution scale (Baldocchi, 2001; Smith, 2012). However, the horizontal footprints of different measurement techniques' vary from each other: Eddy-covariance surface energy flux data represent a downwind footprint of 100 m^2 to 1 km^2, while in-situ soil moisture probes link to much smaller surface areas by representing a support volume (Blöschl and Sivapalan, 1995) of ~4 dm^3 only (Running et al., 1999; Kurc and Small, 2007; Vivoni et al., 2008). Soil moisture is spatially non-uniform within the eddy-covariance footprint due to heterogeneity in soil properties, vegetation, and topography. Therefore, soil moisture measurements best (i.e. most effectively) representing the soil below the eddy-covariance tower's footprint should be used when the performance of a land surface model is evaluated. If soil moisture is measured at only a single or few locations with limited support volume, like with PS sensors, the measured soil moisture content might be different from the effective soil moisture state that controls the surface exchange processes. It is therefore often assumed that soil moisture measured at a scale closer to the footprint of an eddy-covariance tower is more informative than a single or a couple of PS sensor profiles for studying land-surface processes and constraining model parameters at scales of >~100 m^2 (Robinson et al., 2008). This poses a potential scale mismatch issue when a single or a few PS sensors are used. On the other hand, past research has shown soil moisture measured at only one or a few points within an area of similar size as an eddy-covariance footprint, can have similar value to surface energy flux simulation as soil moisture measured at a larger scale (e.g. Vachaud et al., 1985; Teuling et al., 2006; Mittelbach and Seneviratne, 2012). These studies showed that points within a soil moisture observation network keep their rank with respect to the mean soil moisture (anomaly),

i.e. they either under –or overestimate the mean (anomaly), so-called spatio-temporal stability. The physical principle behind the spatio-temporal stability theory is that different time variant and hydrological processes either create or destroy spatial soil moisture variability whereas time invariant land surface characteristics induce an effective offset in the spatial variability (Albertson and Montaldo, 2003; Teuling and Troch, 2005; Vanderlinden et al., 2012). When soil moisture reaches values below the critical point (i.e. transpiration becomes moisture limited), spatial variability in soil moisture and fluxes is reduced (Teuling and Troch, 2005). Soil moisture dynamics was found to be a small portion of total soil moisture variability (Mittelbach and Seneviratne, 2012), while having a greater effect on surface energy fluxes than absolute soil moisture in land surface models (Dirmeyer et al., 1999; Teuling et al., 2009). The implication of the spatio-temporal stability theory therefore is that the spatial scale mismatch issue might have limited implications to surface energy flux simulation. The question which then rises is,

d*oes reduced observation scale mismatch improve LSM energy flux estimates?*

Based on the spatio-temporal stability theory, we phrased the following hypothesis for our research question:

reduced observation scale mismatch does not lead to LSM energy fluxes closer to eddy covariance observations.

In recent years, new soil moisture measurement techniques have been developed that have, compared to point-scale soil moisture sensing techniques, a reduced scale mismatch with eddy-covariance surface energy flux measurements. Improvement in wireless technology and remote data collection technology have made the development of PS soil moisture sensor networks more feasible (Cardell-Oliver, 2005; Ritsema et al., 2009; Trubilowicz, 2009; Bogena et al., 2010; Robinson et al., 2008). Newer soil moisture sensor techniques, for instance one which makes use of the Global Positioning System (GPS; Larson et al 2008, 2010), and the Cosmic-Ray Neutron Sensor (CRNS; Zreda et al., 2008) have the advantage that their installation requires less time and work effort because only a single above ground sensor is needed. The CRNS (Zreda et al., 2008) is an above-ground passive sensor which utilises natural cosmic-ray neutron radiation to estimate soil moisture content in the top 10-70 cm. The sensor's footprint area has a radius of about 100 to 300 m surrounding the above-ground sensor (Desilets and Zreda, 2013; Köhli et al., 2015). Franz et al. (2012) showed soil moisture estimated from CRNS neutron counts differed less than 20% from the average of a co-located point scale soil moisture sensor network at a site in Arizona. Networks of CRNS have been established in various countries (e.g. Cosmic-ray Soil Moisture Observation System; COSMOS (Zreda et al., 2012), COSMOS-UK (Evans et al., 2016), the Australian National Cosmic-Ray Soil Moisture Monitoring Facility CosmOZ (Hawdon et al., 2004), and TERrestrial ENvironmental Observatoria; TERENO (Baatz et al., 2015)).

Unlike wireless point scale sensor networks, both the GPS and CRNS technology provide an integrated soil moisture measurement over the entire support volume (Larson et al., 2008; Zreda et al., 2008). We chose to answer our research question using the CRNS technology because the COSMOS network provides publicly available data for multiple years at a range of sites co-located with AmeriFlux/FLUXNET eddy-covariance towers (ORNL-DAAC, 2015). Twelve of these sites provided sufficient LSM forcing data, PS soil moisture data, CRNS data, and eddy-covariance LE and sensible heat flux (H) data.

Before our modelling exercise we first compared the PS and CRNS data. The outcomes of this data analysis were mainly used to see if the results from the calibration and validation yielded larger differences in surface energy flux estimation at sites where the two soil moisture observation products showed higher deviation from each other. To investigate our research question we made the LSM simulated soil moisture content match the observed PS or CRNS data as closely as possible. We did this by calibrating parameters of the Joint UK Land Environment Simulator (JULES; Best et al., 2011) against point-scale and Cosmic-Ray Neutron data separately. We subsequently validated the results against eddy-covariance observed data over the same periods. To assess the change in soil moisture and surface energy fluxes after calibration we compared the calibrated runs against a default run with parameter values computed from a widely used soil properties database. We emphasise that we compared the two different soil moisture measurement techniques' scales and not the techniques as such.

2 Data and Methods

2.1 Calibration and Validation data: PS, CRNS, and eddy-covariance data

Point scale (PS) soil moisture and CRNS neutron count data from twelve AmeriFlux/COSMOS sites were used (Figure 1; full COSMOS site names are shown in this figure). These twelve sites covered eight of twenty Ecoclimatic domains of the US National Ecological Observatory Network (NEON; www.neonscience.org) (Figure 1). These twelve sites hence represent a variety of climates and land cover types, but also different soil types (Table 1).

Hourly PS data for nine sites were obtained from the publicly available AmeriFlux Level 2 data source (ORNL). Data for the three California Climate Gradients sites (DC, CS, and SO) were obtained at http://www.ess.uci.edu/~california/ (data version 3.4; Goulden et al., 2015). The number of PS profiles, the installation depths, and sensor types differed between the twelve study sites (Table A1.1 in Appendix 1). We used point scale soil moisture data from the soil layers up to 30 cm depth only for consistency among all sites. There were only two sites reporting soil moisture data at greater depths: WR at 50 cm, and MO at 100 cm). Our main objective was to investigate the difference in information content due to two soil moisture measurement techniques' different horizontal scales in relation to the eddy-covariance footprint, rather than to compare the measurement techniques themselves. Quality control was applied to filter out spurious and unrealistic data points due to sensor errors. The PS data was then interpolated to the JULES soil layer on which the model was calibrated.

Hourly CRNS neutron count data were obtained from the COSMOS network website (www.cosmos.hwr.arizona.edu). Corrections were applied as by Zreda et al. (2012). Water vapour corrections (Rosolem et al., 2013) were applied with respect to dry air (Bogena et al., 2013). The quality control approach used for the PS analysis was also applied to CRNS neutron count data series to remove unrealistic points. Snow cover periods were also removed for the analysis. A 5-hour moving average window was applied to the observed CRNS neutron counts (following Shuttleworth et al., 2013; Rosolem et al., 2014).

We used the exact same data for the soil moisture data comparison as for the calibration (providing that both PS and CRNS were available in the same period). As validation data, latent heat (LE) flux hourly data from AmeriFlux Level 2 data source

was used for nine sites. We used data version 3.4 (Goulden et al., 2015) for the three California Climate Gradients sites. Quality control was applied to the LE and H flux data to remove outliers and unrealistic data points.

2.2 Soil moisture data comparison methodology

To compare PS and CRNS soil moisture data we computed the Mean Squared Deviation (MSD) and its decomposition (Gupta

5   et al., 2009) into:

(1) The squared difference between the means (structural bias)

(2) The squared difference between the standard deviations (indicates different seasonality)

(3) A term relating to the coefficient of linear correlation (indicates different dynamics)

Which yields the following equation:

10    $$MSD = (\mu_{PS} - \mu_{CRNS})^2 + (\sigma_{PS} - \sigma_{CRNS})^2 + 2 \cdot \sigma_{PS} - \sigma_{CRNS} \cdot (1 - r)$$

Where μ (m^3m^{-3}) is the observed mean, σ (m^3m^{-3}) is standard deviation, and r (-) is the coefficient of linear correlation. We however scaled the relative contributions of the three MSD components to the Root Mean Squared Deviation (RMSD) instead of MSD, to keep units of m^3m^{-3}. We then ranked the sites from lowest to highest RMSD. According to our hypothesis, we would expect to see a larger difference in simulated surface energy fluxes after calibration when the two soil moisture time

15   series differ most. We compared PS soil moisture with CRNS soil moisture values computed from vertically homogeneous soil moisture values obtained from the observed neutron counts using the COsmic-ray SoiI Moisture Interaction Code (COSMIC; Shuttleworth et al., 2013).

2.3 JULES forcing data and initial conditions

JULES requires precipitation, air temperature, atmospheric pressure, wind speed, specific humidity, downward shortwave

20   radiation, and downward longwave radiation as meteorological forcing data. Quality controlled hourly data was obtained from AmeriFlux Level 2 and the three California Climate Gradients sites. At some of the sites however, certain specific forcing data was not available from AmeriFlux and hence data from different sources were used (Table A1.2 of Appendix 1).

Model input data was gap-filled following Rosolem et al. (2010) because JULES requires continuous time series except for precipitation where gaps were set to zero. For all sites gaps smaller than 3 hours were filled by linear interpolation. Larger

25   gaps of up to 30 days were gap-filled using the average diurnal pattern of the preceding and following 15 days. In addition, some remaining gaps in Downward Shortwave Radiation and Downward Longwave Radiation at Wind River (WR) were filled using linear least squares relationships with NLDAS-2 data (http://disc.sci.gsfc.nasa.gov/uui/datasets?keywords=NLDAS). At Soaproot (SO) and Coastal Sage (CS) data gaps in the atmospheric pressure time series were filled with NLDAS data. At CS NLDAS data were also used to gap fill air temperature. Gap filling at Santa Rita Creosote was done with data from the nearby

30   Sahuarita site followed by the gap-filling procedure described above.

2.4 Calibration and validation methodology

2.4.1 Joint UK Land Environment Simulator (JULES)

We used the Joint UK Land Environment Simulator (JULES; Best et al., 2011; Clark et al., 2011) in this study. JULES can be coupled as lower boundary condition to the UK Met Office Unified Model (Cullen, 1993). Within JULES choices can be made (e.g. canopy radiation model type) and certain modules (e.g. vegetation dynamics) can be switched on or off to operate at different levels of complexity. In addition, we chose the UK Variable resolution configuration (UKV) because it is the standard setting when JULES is run coupled with the UK Met Office Unified Model. The UKV land grid cell size is 1 km by 1 km. However, our study focused on JULES standalone simulations at the 12 grid points located at the sites investigated. The UKV setting employs the multi-layer canopy radiation module with surface heat capacity and snow beneath the canopy, the single canopy layer `big leaf' approach for leaf-level photosynthesis (which computes radiation absorption with Beer's law). Soil heat conductivity was calculated using the approach of Dharssi et al. (2009). We used the default JULES-UKV soil layering (Supplement 1, Figure S1.1). The hydraulic bottom boundary condition in JULES is free drainage.

JULES computes the transport of water through the soil using a finite difference representation of the Richards' equation. The vertical fluxes are computed with the Buckingham-Darcy equation. JULES-UKV uses the Mualem-Van genuchten (Van Genuchten, 1980; Mualem, 1983) soil water retention equations. The Van Genuchten equation calculates soil water content θ ($m^3 m^{-3}$) from soil hydraulic pressure head ψ (m):

$$\frac{\theta - \theta_{res}}{\theta_{sat} - \theta_{res}} = \frac{1}{[1 + (\alpha\Psi)^n]^{1-1/n}},$$ (1)

with shape parameter n (-), α (m^{-1}) representing the inverse of the water entry pressure, θ_{res} ($m^3 m^{-3}$) is the empirical residual soil moisture content (without physical meaning), and θ_{sat} (or smsat; $m^3 m^{-3}$) is the saturated soil moisture content. In JULES parameter n is defined as b = 1/(n-1) and sathh=α^{-1} (m).

The Mualem equation computes the unsaturated hydraulic conductivity K:

$$K = K_{sat} \frac{\theta - \theta_{res}}{\theta_{sat} - \theta_{res}} \left[1 - (1 - \frac{\theta - \theta_{res}}{\theta_{sat} - \theta_{res}}^{1/(1-\frac{1}{n})})^{1-\frac{1}{n}} \right]^2,$$ (2)

where K_{sat} (or satcon; mm·s^{-1}) is the saturated hydraulic water conductivity.

The values of the Mualem-Van Genuchten parameters need to be defined by the user for each grid cell/point based on soil characteristics.

In JULES soil moisture directly interacts with transpiration (through root water uptake) and bare soil evaporation as described hereafter (see also Figure S1.2 of Supplement 1). JULES first computes the potential photosynthesis, which is a function of three limiting factors: Rubisco limitation, radiation limitation, and photosynthetic product transport limitation. The potential photosynthesis is multiplied with a soil moisture reduction factor to obtain the actual photosynthesis. To obtain this soil moisture reduction factor the model first computes a limiting factor for each layer:

$$\beta_i = \begin{cases} 1, & \theta_i \geq \theta_{crit} \\ \dfrac{(\theta_i - \theta_{wilt})}{(\theta_{crit} - \theta_{wilt})}, & \theta_{wilt} < \theta_i < \theta_{crit}, \\ 0, & x \leq \theta_{wilt} \end{cases} \tag{3}$$

where θ_i is the unfrozen soil moisture content in layer i, θ_{crit} is the critical point soil moisture content below which soil moisture is limiting the root water uptake (matrix water potential -330 cm in JULES), and θ_{wilt} is the wilting point soil moisture content below which no root water uptake occurs (-15000 cm in JULES). These reduction factors are multiplied with the root density in the layer. These weighted reduction factors are then summed to obtain the root zone soil moisture reduction factor. From the actual photosynthesis the plant stomatal conductance is computed. Separately the bare soil surface conductance, which is a function of the soil moisture content in the upper soil layer and the critical soil moisture, is computed. The surface conductance is then computed as a function of the stomatal conductance and the bare soil surface conductance.

The potential evapotranspiration is also calculated separately. This variable is multiplied with the saturated land fraction to compute the free water evaporation (e.g. lake and canopy evaporation). The rest of the potential evapotranspiration is multiplied with the surface conductance to obtain the bare soil evaporation + plant transpiration. Together these fluxes are the actual evapotranspiration (water) or latent heat flux (energy). The amount of water drawn from the top soil layer through bare soil evaporation depends on the bare soil surface conductance. The distribution of the root water uptake between the layers depends on the weighted soil moisture limitation factor for each layer. The water extraction from the soil in its turn directly affects the soil moisture content in the different layers at the start of the next time step. These soil moisture contents then affect the soil moisture redistribution, surface runoff, and deep drainage.

JULES-UKV also requires a number of initial conditions: the amount of unfrozen water and snow stored on the surface (on canopy and on soil surface; set to zero in this study), snow properties (set to JULES-UKV defaults), the surface temperature (set to the air temperature of the hour before the first simulation time step), the soil temperature of each layer (set to the soil temperature from AmeriFlux data the hour before initial time step) the soil water content in each layer (set to the soil water content from the PS observed moisture content of the hour before the first simulation time step and applied homogeneously throughout the profile). Soil moisture was spun up by running a maximum of five cycles and stopped when soil moisture convergence was lower than or equal to 10% compared to the previous cycle.

2.4.2 Calibration and validation approaches

At each site we calibrated JULES against PS observed soil moisture and against CRNS observed neutron counts respectively. We chose to calibrate simulated neutron counts against CRNS observed neutron counts using COSMIC (Shuttleworth et al., 2013) to translate simulated soil moisture profiles into equivalent neutron counts. We computed the Root Mean Squared Error (RMSE) between simulated and observed hourly time series. To better match the observed soil moisture/neutron count time series, we calibrated five JULES parameters that influence the model soil moisture state (Figure S1.2 of Supplement 2 These included Mualem-Van Genuchten shape parameter (b), the water entry pressure parameter (sathh), and the saturation hydraulic conductivity (K_{sat}). The critical point (θ_{crit}) and wilting point (θ_{wilt}) soil moisture content parameters from the evapotranspiration

limitation factor were also calibrated. We chose these parameters because they are, in theory, directly linked to the movement of moisture in the soil and to the effects of soil moisture on transpiration in JULES.

To assess the effects of calibration on soil moisture and surface energy flux simulation we compared the calibrated solutions against a default run at each site. The parameter values for the default case were derived from soil properties (percentages clay, loam, and organic matter) reported by the Harmonised World Soil Database (HWSD; FAO, 2009) for each of the twelve sites. These properties were used in the Wösten Pedotransfer Function (Wösten et al., 1997) to obtain values for b, sathh, and K_{sat}. Parameter values for θ_{crit} and θ_{wilt} were subsequently obtained with the Van Genuchten formula.

The parameter calibration ranges were the same for all sites (Table 2). They were constructed by computing the minimum and maximum parameter values for the entire soil texture triangle (based on Wösten Pedotranser Function). Three organic matter contents were taken into consideration, yielding three triangles. Clay percentages above 70% were excluded to avoid extreme values for parameter b especially. The range for θ_{crit} was set to 10-90 % of the saturated soil moisture content. The saturated soil moisture content was computed as a function of the dry soil bulk density ($\rho_{bd,dry}$): $\theta_{sat} = 1 - \rho_{bd,dry} / 2.65$ (Brady and Weil, 1996). Soil bulk density values obtained from the COSMOS network were used. This yielded different θ_{crit} ranges in terms of soil moisture content (m^3m^{-3}) for different sites. The residual soil moisture content (defined implicitly in JULES) was set to zero because the Wösten Pedotransfer Function does not consider it.

JULES' remaining two ancillary parameters; the soil heat conductivity and soil heat capacity, were for each site computed as a function of soil properties (HWSD) with De Vries' (1963) method. The bare soil albedo was set constant at 0.38 (-) for all sites. Plant Functional Type parameters were set to JULES defaults except for the e-folding rooting depth (depth above which 86 % of plant roots are present) and the canopy height, at sites where more specific information was available from AmeriFlux/COSMOS site information or from site specific literature.

We calibrated JULES using the BORG Multi-Objective Algorithm (BORG-MOEA or BORG; Hadka and Reed, 2013). This calibration tool was designed for multi-objective problems but also works for single-objective calibration. BORG employs multiple optimisation algorithms simultaneously to obtain convergence while also keeping the searched parameter space wide. The algorithm measures progress with the epsilon-progress technique, which uses the objective function space divided in boxes with sides of size epsilon. Epsilon is a user defined value for each objective function (we used epsilon values of 0.001 m^3m^{-3} for PS and 1 cph for CRNS, Kollat et al., 2012). Only if a new solution resides inside a box with a better objective function value, BORG considers it is progress. If no progress was obtained after 200 runs, the algorithm had stagnated. In this case the BORG algorithm triggers a restart, which consists of (among other techniques) changing population size to maintain a diverse population and to escape local optima. We used a maximum number of 3,000 runs and an initial population size of 100 runs.

As validation metric we chose RMSE between the observed and simulated latent heat flux (LE). Because data quality issues with eddy-covariance data are often observed during night time (Goulden et al., 1996; 2006; 2012; Aubinet et al., 2010), we computed these metrics over day time hours only. We defined day time as downward shortwave radiation $> 20 \, Wm^{-2}$. To avoid extreme RMSE-EF values, we used hours with both observed and simulated LE values $\geq 1 \, Wm^{-2}$ only. Otherwise a few hours

with small LE or H values would have dominated the RMSE values, while this would probably have been due to forcing or eddy-covariance data inconsistencies and would not relate to soil moisture temporal variability.

It is known that single-objective calibration is often insufficient to constrain parameters to simulate different states and fluxes well (Gupta et al., 1999). In addition, Vereecken et al. (2008) argued that calibrating soil hydraulic parameters against soil moisture only does not guarantee better surface energy fluxes. To find out whether it was actually feasible to expect better surface energy flux simulation when soil moisture was improved we performed calibrations where we optimised the model for two objectives simultaneously. We employed the BORG algorithm to simultaneously optimise the RMSE of (day time) latent leat flux and the RMSE of all day soil moisture (using PS soil moisture and CRNS neutron counts separately). We analysed the trade-off between the two objective functions. We computed the compromise solution for each two-objective-calibration. Plotted within the normalised two-objective solution space, the compromise solution is the model run which has the smallest distance to the origin. This means no other solution can be obtained that yields a better approximation for one objective function without deteriorating the other.

3 Results and Discussion

3.1 Soil moisture data analyses

In Figure 2 comparison between PS and CRNS soil moisture time series shows that the seasonal trends of the two soil moisture observation products were similar. The two soil moisture products however also differed from each other, in distinct ways at different sites. PS soil moisture observations were systematically higher than CRNS soil moisture observations at eight of twelve sites. At three sites (DC, SO, CS) PS soil moisture dried down quicker than CRNS soil moisture, while at ME the opposite behaviour was observed. At KE, MM, TR, and MO PS showed a higher seasonality signal (up to 50% higher) than CRNS. Peaks in PS soil moisture were at three sites (UM, KE, TR) up to twice as high as in CRNS soil moisture. In addition, the CRNS data appears noisier than the PS data, which is an effect of inherent randomness in neutron radiation reaching the CRNS sensor element (Zreda et al., 2012). This effect was more pronounced for lower neutron intensity.

The differences seen in Figure 2 are also summarised in Figure 3, which shows a gradual site to site increase in RMSD between PS and CRNS soil moisture (RMSD-SM$_{obs}$). Overall, bias contributed to 50% or more of the total error at seven out of twelve sites.

Additional analyses indicated that differences between the two soil moisture estimates could not be clearly related to any differences in site physical characteristics other than the mean soil wetness. Dominant vegetation type seemed not to have an effect on the similarity between the two soil moisture data products: both forested and grass sites included those with relatively small RMSD-SM$_{obs}$ and those with relatively high RMSD-SM$_{obs}$. Only bare/shrub covered sites (DC, SR, CS) were all below the sites' average RMSD-SM$_{obs}$ of 0.05 m^3m^{-3}. These four sites were however also relatively dry. Soil type and soil bulk density were also investigated for correlations with RMSD-SM$_{obs}$, but no trends were discovered. Larger differences between

PS and CRNS soil moisture could be expected at sites with more heterogeneous soil or vegetation. Static satellite photos of the sites from the COSMOS website did not indicate systematically more heterogeneous conditions at the sites where PS and CRNS soil moisture differed more. Site info (e.g. topography, presence of rocks) from COSMOS and AmeriFlux did also not clearly show more horizontally heterogeneous soil properties for those sites. The fact that the soil moisture time series of PS and CRNS differed from each other in various ways could be related to a number of issues. First, PS sensor types and numbers of sensors differed between the sites (Table A1 of Appendix A). Secondly, the exact installation locations of the PS sensors may in certain cases have been for instance next to a macropore, or near roots, while at other sites they were coincidently installed in a homogeneous soil patch.

The presence of neutron mitigating factors other than soil moisture (e.g. atmospheric pressure, sensor type, biomass, intercepted water, and water in litter layer) also affect the observed CRNS neutron count. Because different hydrogen pools are more present at certain sites than at others, the uncertainty on neutron count observations varies between sites. The results did however not show effects of land cover and soil properties on the similarity between the two soil moisture products. Another factor is that, the quality of the calibration of COSMIC could possibly be different for different sites. Finally, at multiple sites, the PS and CRNS soil moisture time series were similar as shown with RMSD values. This could be expected at rather homogeneous sites. Moreover, Köhli et al. (2015) suggested the CRNS footprint to be around 150-200 m instead of 300 m as reported by Desilets and Zreda (2013). In that case the differences between the two soil moisture observation techniques could be smaller than initially thought.

We derived vertically constant CRNS soil moisture values from observed neutron counts with COSMIC. This method contains inherent uncertainty because in reality soil moisture is often not vertically homogeneous. The outcomes of the calibration (against PS and CRNS) and validation provide insight in the effects of the differences between the two soil moisture products on JULES' surface energy flux partitioning and latent heat flux simulation.

3.2 Single objective calibration against soil moisture observations

The degree to which the objective function (RMSE-SM; Figure 4) values decreased differed between sites and the two calibration strategies (PS or CRNS), with decreases of 21% (AR-PS) to 93% (UM-CRNS). While the errors of the default runs existed mostly of systematic bias, after calibration the difference in dynamics was the largest source of uncertainty and in 16 out of 24 cases this contribution actually increased in absolute terms. The increase in difference in dynamics was due to the selected objective function (RMSE), which reduces the mean error between modelled and observed data. Previous research (e.g. Teuling et al., 2009) has shown calibrating soil parameters has a large effect on simulated absolute soil moisture values , but substantially less on soil moisture seasonality and dynamics. Our finding supports this.

The RMSE values reduced relatively more for the CRNS calibration (70% on average over the twelve sites) than for the PS calibration (55% on average). The calibration method could possibly explain this. CRNS calibration was against observed neutron counts, while PS calibration was against observed soil moisture contents. Because neutron counts have an inverse relationship with soil moisture content, the PS calibration was possibly governed by avoiding larger errors occurring during a

few brief soil moisture peaks. While focussing on getting the fitting for those peaks right, the PS calibration neglected the smaller errors during dry periods. This would then result in relatively smaller decrease in RMSE values than for the CRNS calibrations because those were fitted with heavier weights to the drier periods.

Figure 4 also shows that the relative improvement was not systematically lower or higher for sites with higher similarity between the two observed soil moisture time series. Actually the largest improvement was for the CRNS calibration at UM. Therefore, it appears that the quality of the default runs was hence predominated by the quality of the chosen default parameter values.

Figure 5 shows soil moisture time series for four selected sites: UM was chosen because PS and CRNS soil moisture were most similar there, SR was a site with moderate difference, and at WR and MO PS and CRNS soil moisture were most different. Simulated soil moisture improved especially at UM and WR, where the default runs overestimated soil moisture contents. Simulated soil moisture dynamics became more similar to both PS and CRNS observed soil moisture dynamics, even though observed soil moisture dynamics differed from each other substantially at sites WR and MO. The relatively high noise in the CRNS soil moisture time series is due to the relatively low neutron count at this site, and possibly due to temporal variations in other hydrogen pools like water intercepted in the forest canopy and in the litter layer.

3.3 Validation of the single-objective calibrations against eddy-covariance observations

While calibration errors decreased for soil moisture, latent heat flux estimation improved for fourteen out of twenty-four calibrations (Figure 6). This means that an improvement in simulated soil moisture did not necessarily lead to better estimation of surface energy fluxes. Calibration against PS soil moisture improved RMSE-LE at six sites (UM, KE, DC, SR, SO, and ME), while calibration against CRNS neutron counts improved RMSE-LE at eight sites (KE, DC, SR, ME, SO, AR, MM, and WR). At five sites RMSE-LE improved after calibration against both PS soil moisture and CRNS neutron counts. Calibration yielded lower RMSE-LE after calibration against CRNS neutron counts than after calibration against PS soil moisture at all but three sites (UM, KE, and WR). Figure 6 also shows that RMSE-SM decreased substantially less (i.e. >20% difference) after calibration against PS soil moisture than after calibration against CRNS neutron counts (i.e. >20% difference between both strategies on the horizontal axis of Figure 6, which occurred at six sites). At five of these six sites the relative change in surface energy flux performance was also smaller (sites MO, AR, TR, ME, and SR). This indicates that further improvement in soil moisture simulation after calibration against PS data could have yielded better surface energy fluxes. The differences between the two calibration strategies can also be seen from the different locations of the centres-of-mass of the two point clouds. The cloud of the CRNS strategy is clearly located more to the lower left corner in Figure 6.

In Figure 6 10% change in latent heat flux estimation was chosen to distinguish substantial change from non-substantial change, derived from the approximate error in eddy-covariance data (Sellers and Hall, 1992; Finkelstein and Sims, 2001). Improvement in latent heat flux was actually substantial in four cases for PS calibration and five cases for CRNS calibration. Using this threshold also revealed that RMSE-LE did not change substantially in six cases for calibrations with a more than 60% change in RMSE-SM. This again shows that a change in simulated soil moisture did not necessarily mean a substantial change in

surface energy flux simulation. Analysis of the RMSE of evaporative fraction (EF=LE/(LE+H)), which shows the ability to simulate surface energy partitioning, yielded similar overall results as our analysis of the RMSE of latent heat flux (Supplement 2, Figure S2.1).

One factor causing some of these limited improvements was that, when mean simulated root zone soil moisture (weighted with root density) increased after calibration, the values of the wilting point and critical point soil moisture parameters moved along (data not shown), yielding similar soil moisture stress. This happened for both calibration strategies at site KE, and for the calibration against CRNS neutron counts at sites MO and TR. This could relate to the limited value of simulated absolute soil moisture for surface energy flux estimation in land surface models (Dirmeyeret al., 2000; Koster et al., 2009). However, we also found the distance between wilting point values and critical point values to decrease after calibration. This occurred with a simultaneous decrease in standard deviation of the simulated soil moisture. The self-adjusting behaviour of the wilting point and critical point parameters was also indicated by parameter sensitivity analysis (Appendix 2), which showed soil moisture was substantially more sensitive to a change in critical point value than latent heat flux.

Another issue, which occurred for instance for the PS calibration at SO and for the CRNS calibrations at site AR, was that while surface energy flux estimation improved for a certain period, it deteriorated for another period (data not shown). A third cause of limited improvement in surface energy flux estimation occurred for example at site ME (data not shown). Soil moisture stress was relatively limited (beta mostly above 0.6) and during periods of soil moisture stress, the latent heat flux was dominated by different stress factors.

To see if validation results would be different if only those periods during which soil moisture stress occurred were evaluated, we also analysed the performance over these periods (data not shown). During these periods plant water uptake limitation factor β was below one (data not shown). Only at site SO did we see somewhat better performance during these periods compared to the original validation period, after both PS and CRNS calibration.

We explored trends between relative improvement in surface energy flux estimation and soil wetness, precipitation, vegetation type, vegetation height, and soil characteristics. No clear trends were discovered. The only feature that could be distinguished was that for the two sites with a bare soil tile (DC and SR) PS and CRNS calibration improved EF and LE. Rooting depth did also not explain relative improvement in surface energy flux estimation. Finally, larger differences between the two observed soil moisture time series did generally yield more different simulated surface energy flux time series, with a clear exception for site UM, where the two calibration approaches yielded more than 20% difference in change in RMSE-LE.

In Figure 7 the monthly mean diurnal latent heat cycles of four sites (UM, SR, WR, and MO) are shown (night time data included but not used for calibration). Both calibration strategies yielded overestimation in March and April and underestimation from June to August at site UM. A month long gap in longwave and shortwave downward model forcing data occurred in April 2012. This might have affected the results for this site. At SR, both PS and CRNS calibration improved latent heat flux during periods of low evapotranspiration, while during the other periods only the CRNS calibrations yielded better results. At WR calibration against PS soil moisture yielded overestimation of latent heat flux, while CRNS calibration yielded

too low latent heat flux. The PS calibration at MO yielded latent heat flux underestimation, whereas CRNS calibration did not change LE substantially.

In summary, the single-objective calibrations against CRNS neutron counts yielded larger improvements in simulated surface energy fluxes than single-objective calibrations against PS soil moisture (Figure 6). Improvements in surface energy flux estimation were however substantial for four calibrations against PS soil moisture and five calibrations against CRNS neutron counts. Limited improvements in surface energy flux estimation after calibration could partly be attributed to the limited value of absolute soil moisture to estimate surface energy fluxes with land surface models. This seems reasonable because calibration mostly affected absolute soil moisture (Figure 4). This result corresponds with earlier research that showed model soil moisture dynamics and seasonality have a larger effect on surface energy flux simulation than absolute soil moisture (e.g. Teuling et al., 2009; Dirmeyer et al., 1999). Previous research has also indicated that soil moisture alone is insufficient to estimate soil hydraulic parameters (Vereecken et al., 2008; 2015). Our findings support this for some sites. To better understand these implications, the two-objective simulations against soil moisture and latent heat flux simultaneously (discussed in the next section) provides further insight in these results.

3.4 Two-objective calibration against soil moisture and latent heat flux

The results from the two-objective calibrations suggest that differences in latent heat flux estimation improvement between compromise solutions for both two-objective calibration strategies were limited (Figure 8). Only at three sites were these differences more than 5% (at sites SR, DC, and UM). Improvements were substantial (according to our 10% threshold for improvement in RMSE of latent heat flux) for six calibrations with PS soil moisture and for eight calibrations with CRNS soil moisture.

Scatterplots of normalised RMSE of latent heat flux (LE) and normalised RMSE of soil moisture for all sites are shown in Figure 9 and Figure 10 for the PS and CRNS two-objective calibration strategies respectively. The RMSE values were normalised with respect to the default solutions, which therefore have normalised RMSEs of 1 (-). The black dots represent individual model runs and the default model run for each site is represented by a red cross. The single-objective calibration solution of each single-objective calibration solution is shown as a blue triangle and the compromise solution of each two-objective calibration is shown as a green triangle. These two figures showed that the differences between the compromise solutions of the two calibration strategies observed for SR, DC, and UM in Figure 8, were less meaningful than analysis of Figure 8 only showed.

These findings are based on the shapes of the black point clouds in Figure 9 and Figure 10. We first look at the left edges of the black point clouds. When these edges are close to vertical, a small deterioration in RMSE-SM or RMSE-N (e.g. less than 0.05 (-)) would yield a large deterioration in normalised RMSE-LE (e.g. greater than 0.2 (-)). We observed this for three calibrations with PS soil moisture (SO, AR, and WR; indicated with pink line for WR) and in all these three cases, a less than 0.05 (-) change in normalised RMSE-SM would have yielded worse simulated latent heat flux than with the default parameter set. Seven calibrations with CRNS neutron counts (UM, DC, SO, KE, SR, CS, and WR; indicated with pink line for WR)

showed a close to vertical edge. In four cases (UM, DC, KE, and CS) this would have yielded worse simulated latent heat flux than with the default parameter set. A negative slope for this side of the point cloud means an improvement in soil moisture estimation would mean a deterioration in latent heat flux estimation. We observed such negative slope for four calibrations with PS soil moisture (CS, MM, TR, and MO; indicated with continuous black line for MO) and for two calibrations with

5   CRNS neutron counts.

Next, we look at the lower edges of the point clouds in Figure 9 and Figure 10. When this edge is horizontal, this means good latent heat flux estimation can be obtained for a wide range of soil moisture estimation performances (for instance site ME, indicated with pink lines in both figures). When this edge has a negative slope, this means the best latent heat flux estimation is obtained for worse soil moisture (for instance site TR, indicated with pink lines in both figures). We observed these two

10   features for all two-objective calibrations. Single-objective calibration against latent heat flux would however only have necessarily yielded worse soil moisture than the default parameter set for six calibrations with respect to PS soil moisture (sites UM, KE, SR, TR, AR and WR) and two calibrations with respect to CRNS neutron counts (sites TR and AR). The implication of these results is that the quality of latent heat flux simulation did not depend strongly on the quality of soil moisture simulation.

15   In summary, the two-objective calibrations against soil moisture (or neutron counts) and latent heat flux showed fewer substantial differences between calibration with PS soil moisture and calibration with CRNS neutron counts. These results indicate that the differences between both *single*-objective calibration strategies, which showed an advantage for CRNS observations, were possibly not as substantial as it seemed at first. The spatio-temporal stability theory, which implies limited spatial variability in surface energy fluxes, could be one explanation for this (Vachaud et al., 1985; Teuling et al., 2006;

20   Mittelbach and Seneviratne, 2012; Albertson and Montaldo, 2003). Another factor that has possibly played a role is that the spatial scale advantage of the CRNS was masked out by the possibly lower quality of its measurements (see also Section 3.1). Different hydrogen pools than soil moisture affect the neutron measurements at various temporal resolutions (e.g. rainfall interception; Baroni and Oswald, 2015). However, PS sensors also have their limitations. Different (electromagnetic) PS sensors have different designs and properties, which affect the data quality (Robinson et al., 2008; Blonquist et al., 2005). Soil

25   type and soil specific calibration also affect the accuracy and precision of the PS data. For instance, the relationship between electrical permittivity and soil moisture content is strong for quartz rich soils, but less accurate for clay soils (e.g. Ishida et al., 2000; Robinson et al., 2008). An issue that could have had effect on our results is the occurrence of gaps in the model forcing data. Gaps in observed meteorological data are often inevitable and must be filled for a land surface model to be run. Percentages of missing hours differed between 0% and 15% at our sites. Gaps larger than fifteen days filled with the moving

30   window gap-filling procedure occurred mainly for downward shortwave and/or downward longwave radiation; at sites UM, KE, SR, DC, TR, and CS. At site WR a data gap of 29 days in precipitation and air pressure occurred. Especially the gap in precipitation data can have negatively affected the model results at this site. In this study we used PS sensor data in the upper 30 cm of the soil only, even though data from deeper sensors was publicly available at two sites (WR and MO). This choice can have provided a disadvantage to the PS data in our comparison, especially at sites with deeper roots and during soil

moisture limiting conditions in the upper soil. Investigating the role of deeper roots was however beyond the scope of this study. Our goal was to compare the effects of the difference in horizontal footprint on latent heat flux simulation after using the two different measurement techniques to calibrate model parameters.

4 Conclusions

5   We investigated whether the spatial scale mismatch between the surface energy flux data and soil moisture data could be reduced through the use of Cosmic-Ray Neutron Sensors (CRNS). Five soil and evapotranspiration parameters of LSM JULES were calibrated against Point Scale (PS) and CRNS soil moisture data separately, for twelve COSMOS/AmeriFlux sites with different climate, land cover, and soil properties. Next, at each site, the improvement in latent heat flux for the two calibration solutions was assessed by comparing the fit with eddy-covariance data and a version of LSM JULES runs with default
10   parameter values based on a widely used soil database. Before the calibrations we compared the observed soil moisture data from the two sensor types. These analyses showed the differences between PS and CRNS soil moisture varied in nature at the investigated sites. While at eight sites there were mainly systematic biases, at three other sites the seasonality was most different, and at one other sites different time series dynamics were the main cause of differences between the two soil moisture observations. The single-objective calibration of JULES parameters against Cosmic-Ray Neutron Sensor neutron counts
15   yielded better simulated latent heat flux than single-objective calibration against point-scale soil moisture. The analysis of multi-objective calibrations (against (1) PS soil moisture and latent heat flux and (2) CRNS neutron counts and latent heat flux) however revealed that differences between calibrations with these two soil moisture observation methods did overall not yield substantially different surface energy flux estimations.

These outcomes did not provide sufficient evidence to reject our null-hypothesis "*reduced scale mismatch does not lead to*
20   *LSM flux estimates closer to eddy covariance observations*". The spatio-temporal stability theory in soil moisture, on which our hypothesis was based, can possibly explain the limited differences in surface energy flux estimation. This theory implies that spatial variability in surface energy fluxes is relatively limited within an eddy-covariance tower footprint. Therefore simulated surface energy fluxes after calibration with point scale soil moisture data would not necessarily be better than simulated surface energy fluxes after calibrations with CRNS neutron count data. Calibrating soil parameters had mostly an
25   effect on absolute soil moisture values rather than soil moisture dynamics. Soil moisture dynamics have a greater effect on surface energy flux simulation in land surface models than absolute soil moisture values do. Related to this we observed that after calibration the wilting point and critical point soil moisture parameter values adjusted themselves in a similar way as the root zone soil moisture did, yielding similar soil moisture control on transpiration despite changes in soil moisture values. In other cases we found calibration against soil moisture to improve surface energy fluxes during certain periods, but to deteriorate
30   surface energy fluxes during other periods, yielding similar overall performance. Yet in other cases evapotranspiration was not limited by soil moisture stress. The potential scale advantage of the CRNS was possibly masked out by the possibly lower measurement quality of this sensor because other hydrogen pools than soil moisture affect the neutron count observations.

Future use of CRNS soil moisture data could however benefit from improved knowledge on the effects of additional hydrogen pools (e.g. Baroni and Oswald, 2015) and of the sensor footprint (Köhli et al., 2015). In this study, our results are conditioned to a single land surface model (JULES). For additional understanding of the importance of both PS and CRNS measurements for surface energy flux simulation, this study can be extended to other models in the future.

5   Acknowledgements

This research was supported by the Queen's School of Engineering (University of Bristol) Ph.D. scholarship and the by Engineering and Physical Sciences Research Council (EP/L504919/1). Additional support for this work was also provided by the Natural Environment Research Council (A MUlti-scale Soil moisture-Evapotranspiration Dynamics study (AMUSED); grant number NE/M003086/1). Funding for AmeriFlux data resources was provided by the US. Department of Energy's Office of Science. We would like to thank the investigators of the sites used for providing us with site information through personal communication. The authors would like to thank Jos van Dam, referee Ryan Teuling, referee Roland Baatz, and one anonymous referee for providing constructive comments. Finally, the authors would like to thank the Editor, Harrie-Jan Hendricks Franssen, for guiding the revision process.

References

Albertson, J. D., and Montaldo, N.: Temporal dynamics of soil moisture variability: 1. Theoretical basis, Water Resour. Res., 39, 1274, doi:10.1029/2002WR001616, 2003.

Anderson, R.G., Goulden, M.L.: Relationship between climate, vegetation, and energy exchange across a montane gradient. J. Geophys. Res., 116, G01026, doi:10.1029/2010JG001476, 2011.

Aubinet, M., C. Feigenwinter, B. Heinesch, C. Bernhofer, E. Canepa, A. Lindroth, L. Montagnani, C. Rebmann, P. Sedlak, and Van Gorsel, E.: Direct advection measurements do not help to solve the night-time CO2 closure problem: Evidence from three different forests, Agric. For. Meteorol., 150(5), 655–664, doi:10.1016/j.agrformet.2010.01.016, 2010.

Baatz, R., Bogena, H. R., Hendricks Franssen, H.-J, Huisman, J.A., Montzka, C., and Vereecken, H.: An empirical vegetation correction for soil water content quantification using cosmic ray probes, Water Resour. Res., 51, 2030-2046, doi:10.1002/2014WR016443, 2015.

Baldocchi, D., Eva Falge, Lianhong Gu, Richard Olson, David Hollinger, Steve Running, Peter Anthoni, Ch Bernhofer, Kenneth Davis, Robert Evans, Jose Fuentes, Allen Goldstein, Gabriel Katul, Beverly Law, Xuhui Lee, Yadvinder Malhi, Tilden Meyers, William Munger, Walt Oechel, K. T. Paw, Kim Pilegaard, H. P. Schmid, Riccardo Valentini, Shashi Verma, Timo Vesala, Kell Wilson, and Steve Wofsy, 2001: FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem–Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities. Bull. Amer. Meteor. Soc., 82, 2415–2434, doi: 10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2, 2001

Baroni, G., and Oswald, S.E.: A scaling approach for the assessment of biomass changes and rainfall interception using cosmic-ray neutron sensing. J. Hydrol., 525, 264-276, doi:10.1016/j.hydrol.2015.03.053, 2015.

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., Harding, R. J.: The Joint UK Land Environment Simulator (JULES), model description - part 1: Energy and water fluxes. Geoscientific Model Development 4(3), 677-699, 2011.

Beven, K., and Germann, P.: Macropores and water flow in soils revisited, Water Resour. Res., 49, 3071–3092, doi:10.1002/wrcr.20156, 2013

Bloeschl, G., Sivapalan, M.: Scale issues in hydrological modelling: A review. Hydrological Processes 9 (3-4), 251{290. URL http://dx.doi.org/10.1002/hyp.3360090305, 1995.

Blonquist, J. M., Jones, S. B., and Robinson, D. A.: Standardizing characterization of electromagnetic water content sensors: Part 2. Evaluation of seven sensing systems. Vadose Zone J., 4, 1059-1069.

Blyth, E. M., Dolman, A. J. and Wood, N.: Effective resistance to sensible- and latent-heat flux in heterogeneous terrain. Q.J.R. Meteorol. Soc., 119: 423–442. doi:10.1002/qj.49711951104, 1993.

Blyth, E., Gash, J., and Lloyd, A.: Evaluating the JULES Land Surface Model Energy Fluxes Using FLUXNET Data. Journal of Hydrometeorology, 11, 509-519, 2010.

Blyth, E., Clark, D.B., Ellis, R., Huntingford, C., Los, S., Pryor, M., Best, M, and Sitch, S.: A comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and carbon at both the global and seasonal scale. Geosci. Model Dev., 4, 255-269, doi: 10.5194/gmd-4-255-2011, 2011.

Bogena, H.R., J.A. Huisman, H. Meier, U. Rosenbaum, and A. Weuthen: Hybrid wireless underground sensor networks: Quantification of signal attenuation in soil. Vadose Zone J. 8. 3. 755-761, 2009.

Bogena, H. R., Huisman, J. A., Baatz, R., Hendricks Franssen, H. J., Vereecken, H.: Accuracy of the cosmic-ray soil water content probe in humid forest ecosystems: The worst case scenario. Water Resources Research 49, 5778-5791, doi:10.1002/wrcr.20463, 2013.

Brady, N.C., and Weil, R.R.: The nature and properties of soils (11th ed.). Prentice Hall, New York, 1996.

Budyko, M.I.: Тепловой баланс земной поверхности (Heat Balance of the Earth's Surface). Gidrometeoizdat, Leningrad. 255 pp, 1956.

Cardell-Oliver, R., Kranz, M., Smettem, K. and Mayer, K.: A reactive soil moisture sensor network: Design and field evaluation. International journal of distributed sensor networks, 1(2), 149-162, 2005.

Cavanaugh, M.L., Kurc, S.A., Scott, R.L.: Evapotranspiration partitioning in semiarid shrubland ecosystems: a two-site evaluation of soil moisture control on transpiration. Ecohydrology 4 (5), 671e681, 2011.

Chen, X., Rubin, Y., Ma, S., Baldocchi, D.: Observations and stochastic modelling of soil moisture control on evapotranspiration in a Californian oak savanna. Water Resources Research, 44, W08409, doi:10.1029/2007WR006646, 2008.

Clark, D. B., Mercado, L. M. , Sitch, S., Jones, C.D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Menard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Blyth, E., Boucher, O. , Harding, R. J., Huntingford, C., and Cox, P. M.: The Joint UK Land Environment Simulator (JULES), model description - part 2: Carbon fluxes and vegetation dynamics. Geoscientific Model Development 4(3), 677-699, 2011.

5   Cullen, M. J. P.: The unified forecast/climate model, Meteorol. Mag., 122, 81–94, 1993.

Crow, W.T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., Rosnay, P., Ryu, D., Walker, J. P.: Upscaling sparse groundbased soil moisture observations for the validation of coarse-resolution satellite soil moisture products. Rev. Geophys. 50 (RG2002), 1-20. URL http://dx.doi.org/10.1029/2011RG000372, 2012.

De Vries, D.A.: Thermal properties of soils. In D.A. De Vries, and Afgan, N.H. (eds.): Heatand mass transfer in the biosphere.

10   I. Transfer processes in plant environment, 5-28, Washington, DC: Scripts, 1963.

Desilets, D., and Zreda, M.: Footprint diameter for a cosmic-ray soil moisture probe: Theory and Monte Carlo simulations. Water Resour. Res., 49, 3566-3575, doi:10.1002/wrcr.20187, 2013.

Dharssi, I., Vidale, P. L., Verhoef, A., Macpherson, B., Jones, C., and Best, M.: New soil physical properties implemented in the Unified Model at PS18, Met Office Technical Report 528, Met Office, Exeter, UK, 2009.

15   Dirmeyer, P.A., Zeng, F.J., Ducharne, A., Morrill, J.C., and Koster, R.D.: The Sensitivity of Surface Fluxes to Soil Water Content in Three Land Surface Schemes. J. Hydrometeor., 1, 121–134, doi: 10.1175/1525-7541(2000)001<0121:TSOSFT>2.0.CO;2, 2000.

Dirmeyer, P. A., Dolman, A. J., and Sato, N.: The pilot phase of the Global Soil Wetness Project. Bull. Amer. Meteor. Soc., 80, 851-878, 1999.

20   Dirmeyer, P. A., Zeng, F. J., Ducharne, A., Morrill, J. C., and Koster, R. D.: The Sensitivity of Surface Fluxes to Soil Water Content in Three Land Surface Schemes. J. Hydromteor., 1, 121-134, 2000.

Dirmeyer, P.A.: The terrestrial segment of soil moisture-climate coupling. Geophysical Research Letters, 38 (16), 2011.

Dirmeyer, P.A., Jin, Y., Singh, B., Y, X.: Trends in Land-Atmosphere Interactions from CMIP5 Simulations. Journal of Hydrometeorology, 14, 829-849, doi: 10.1175/JHM-D-12-0107.1, 2013.

25   Downer, C. W., and Ogden, F. L.: Appropriate vertical discretization of Richards' equation for two-dimensional watershed-scale modelling. Hydrol. Process., 18, 1-22, 2004.

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-

30   UK. Hydrol. Process., doi: 10.1002/hyp.10929, 2016

FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009. Harmonized World Soil Database (version 1.1). FAO, Rome, Italy and IIASA, Laxenburg, Austria.

Fan, Y., Li. H., and Miguez-Macho, G.: Global Patterns of Groundwater Table Depth. 339(6122, 940-943, doi:10.1126/science.1229881, 2013.

Finch, J.W. and Haria, A: The representation of chalk soils in the JULES/MOSES soil hydrology model. Rept., NERC/Centre for Ecology and Hydrology, 2006.

Finkelstein, P. L., and Sims, P. F.: Sampling error in eddy correlation flux measurements. J. Geophys. Res., 106(D4), 3503-3509, 2001.

5   Franz, T. E., Zreda, M., Rosolem, R., Ferre, T. P. A.: Field validation of a cosmic-ray neutron sensor using a distributed sensor network. Vadose Zone Journal 11 (4), doi:10.2136/vzj2012.0046, 2012.

Goulden, M. L., Munger, J.W., Fan, S.M., Daube, B.C., and Wofsy, S.C.:, Measurements of carbon sequestration by long-term eddy covariance: Methods and a critical evaluation of accuracy, Global Change Biol., 2(3), 169–182, doi:10.1111/j.1365-2486.1996.tb00070.x, 1996.

10   Goulden, M. L., Winston, G. C., McMillan, A.M.S., Litvak, M. E., Read, E. L., Rocha, A. V., and Elliot, J. R.: An eddy covariance mesonet to measure the effect of forest age on land-atmosphere exchange, Global Change Biol., 12(11), 2146–2162, doi:10.1111/j.1365-2486.2006.01251.x, 2006.

Goulden, M.L., Anderson, R.G., Bales, R.C., Kelly, A.E., Meadows, M., and Winston, G.C.: Evapotranspiration along an elevation gradient in California's Sierra Nevada. Journal of Geophysical Research, 117, G03028, doi:10.1029/2012JG002027, 15   2012.

Goulden, M.L.: Measurement of Energy, Carbon and Water Exchange Along California Climate Gradients. Website: http://www.ess.uci.edu/~california/, last visited: July 2016. 2015.

Gupta, H.V., and Sorooshian, S.: Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. Water Resources Research, 34(4), 751-763, 1998.

20   Gupta, H. V., Bastidas, L. A., Sorooshian, S., Shuttleworth, W. J., and Yang, Z. L.: Parameter estimation of a land surface scheme using multicriteria methods. J. Geophys. Res., 104(D16), 19491-19503, 1999.

Gupta, H. V., King, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol., 377, 80-91, 2009.

Hawdon, A., D. McJannet, and J. Wallace.: Calibration and correction procedures for cosmic-ray neutron soil moisture probes 25   located across Australia, Water Resources Research, 50(6), 5029-5043. doi:10.1002/2013WR015138, 2014.

Hadka, D., and Reed P.: Borg: an auto-adaptive many-objective evolutionary computing framework. Evolut. Compt., 21(2), 231-259, 2013.

Henderson-Sellers, A., McGuffie, K., Pitman, A.J.: The Project for Intercomparison of Land-Surface Parametrization Schemes (PILPS): 1992 to 1995. Clim. Dyn. 12, 849–859, 1996.

30   Ishida, T., and Makino, T., and Wang, C. J., 2000. Dielectric-relaxation spectroscopy of kaolinite, montmorillonite, allophane, and imogolite under moist conditions. Clays Clay Miner, 48, 75-84.

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, doi:10.1002/2015WR017169, 2015.

Kollat, J.B., Reed, P M., and Wagener, T.: When are multiobjective calibration trade-offs in hydrological models meaningful? Water Resour. Res., 48, W03520, doi:10.1029/2011WR011534.

Koster, R.D., Dirmeyer, P.A., Guo, Z.C., Bonan, G., Chan, E., Cox, P., Gordon, C.T., Kanae, S., Kowalczyk, E., Lawrence, D., Liu, P., Lu, C.H., Malyshev, S., McAvaney, B., Mitchell, K., Mocko, D., Oki, T., Oleson, K., Pitman, A., Sud, Y.C., Taylor, C.M., Verseghy, D., Vasic, R., Xue, Y.K., Yamada, T.: Regions of strong coupling between soil moisture and precipitation. Science 305, 1138–1140, 2004.

Koster, R.D., Guo, Z., Yang, R., Dirmeyer, P.A., Mitchell, K., Puma, M.J.: On the nature of soil moisture in land surface models. J. Climate 22 (16), 4322–4335, 2009.

Kurc, S. A. and E. E. Small: Soil moisture variations and ecosystem-scale fluxes of water and carbon in semiarid grassland and shrubland, Water Resources Research, 43, W06416, doi:10.1029/2006WR005011, 2007.

Larson, K.M., Small, E.E., Gutmann, E.D., Bilich, A.L., Braun, J.J., Zavorotny, V.U.: Use of GPS receivers as a soil moisture network for water cycle studies. Geophysical Research Letters, 35, L24405, doi:10.1029/2008GL036013, 2008.

Larson, K.M., Braun, J.J., Small, E.E., Zavorotny, V.U., Gutmann, E.D., Bilich, A.L.: GPS Multipath and Its Relation to Near-Surface Soil Moisture Content. IEEE J. Sel. Topics in Appl., Earth Obs. And Rem. Sens., 3(1), 91-99, 2010.

Lee, D. H., and Abriola, L. M.: Use of Richards equation in land surface parameterizations. J. Geophys. Res., 104(D22), 27519-27526, 1999.

Manabe, S.: Climate and ocean circulation. Part I: the atmospheric circulation and the hydrology of the earth's surface. Mon. Weather Rev. 97, 739–774, 1969.

Mittelbach, H. and Seneviratne, S. I.: A new perspective on the spatio-temporal variability of soil moisture: temporal dynamics versus time-invariant contributions. Hydrol. Earth Syst. Sci., 16, 2169-2179, doi:10.5194/hess-16-2169-2012, 2012.

Morris, M.D.: Factorial Sampling Plans for Preliminary Computational Experiments. Technometrics 33(2), doi:10.2307/1269043, 1991.

Mualem, Y.: A new model for predicting the hydraulic conductivity of unsaturated porous media, Water Resour. Res., 12(3), 513–522, doi:10.1029/WR012i003p00513, 1976.

ORNL-DAAC, 2015. AmeriFlux Web Page. http://ameriflux.ornl.gov.

Pianosi, F., Sarrazin, F., and Wagener, T.: A Matlab toolbox for Global Sensitivity Analysis. Environmental Modelling and Software 70, 80-85, doi:10.1016/j.envsoft.2015.04.009, 2015.

Pitman, A.J.: The evolution of, and revolution in, land surface schemes designed for climate models. Int. J. Climatol. 23, 479–510, 2003.

Qu, W., Bogena, H. R., Huisman, J. A., Vanderborght, J., Schuh, M., Priesack, E., Vereecken, H.: Predicting subgrid variability of soil water content from basic soil information, Geophysical Research Letters, 2015, 42, 1-8.

Rahman, M, and Rosolem, R.: The effects of chalk representation in land surface modelling. Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-244, accepted, 2017.Richter, H., Western, A. W., and Chiew, F. H. S.: The Effect of Soil and Vegetation Parameters in the ECMWF Land Surface Scheme. J. Hydrometeor., 5, 1131–1146, doi: 10.1175/JHM-362.1, 2004.

Ritsema, C.J., Kuipers, H., Kleiboer, L., Van Den Elsen, E., Oostindie, K., Wesseling, J.G., Wolthuis, J.W. and Havinga, P.: A new wireless underground network system for continuous monitoring of soil water contents. Water resources research, 45(4), 2009.

Robinson, D. A., Campbell, C. S., Hopmans, J. W., Hornbuckle, B. K., Jones, S. B., Knight, R., Ogden, F., Selker, J., Wendroth, O.: Soil moisture measurement for ecological and hydrological watershed-scale observatories: A review. Vadose Zone Journal 7 (1), 358{389. URL http://dx.doi.org/doi:10.2136/vzj2007.0143, 2008.

Rosolem, R., W. J. Shuttleworth, X. Zeng, S. R. Saleska, and T. E. Huxman: Land surface modeling inside the Biosphere 2 tropical rain forest biome, J. Geophys. Res., 115, G04035, doi: 10.1029/2010JG001443, 2010.

Rosolem, R., Gupta, H. V., Shuttleworth, W. J., Gonalves de Gonalves, L. G., Zeng, X.. Towards a comprehensive approach to parameter estimation in land surface parameterization schemes. Hydrological Processes 27 (14), 2075{2097, 2013b.

Rosolem, R., W. J. Shuttleworth, M. Zreda, T. E. Franz, X. Zeng, and Kurc. S. A.: The Effect of Atmospheric Water Vapor on Neutron Count in the Cosmic-Ray Soil Moisture Observing System. J. Hydrometeor., 14, 1659–1671, doi: 10.1175/JHM-D-12-0120.1.

R Rosolem, T Hoar, A Arellano, JL Anderson, WJ Shuttleworth, X Zeng, TE Franz: Assimilation of near-surface cosmic-ray neutrons improves summertime soil moisture profile estimates at three distinct biomes in the USA. Hydrol. Earth Syst. Sci. Discuss., 11(5), 5515-5558, doi: 10.5194/hessd-11-5515-2014, 2014.

Running, S.W., Baldocchi, D.D., and Turner, D.: A global terrestrial monitoring network, scaling tower fluxes with ecosystem modelling and EOS satellite data. Remote Sensing, 70, 108-127, 1999.

Skott, R.L., Cable, W.L., Hultine, K.R.: The ecohydrologic significance of hydraulic redistribution in a semiarid savanna. Water Resources Research, 44 (2), 1990.

Sellers, P. J., W. J. Shuttleworth, J. L. Dorman, A. Dalcher, and Roberts, J.M.: Calibrating the Simple Biosphere Model for Amazonian tropical forest using field and remote sensing data. Part I: Average calibration with field data. J. Appl. Meteor., 28, 727–759, 1989.

Sellers, P. J., and Hall, F. G.: FIFE in 1992: Results, scientific gains, and future research directions. J. Geophys. Res., 97(D17), 19091-19109, doi:10.1029/92JD02173, 1992.

Sellers, P.J., Dickinson, R.E., Randall, D.A., Betts, A.K., Hall, F., Berry, J., Collatz, G., Denning, A., Mooney, H., Nobre, C., Sato, N., Field, C., Henderson-Sellers, A.: Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. Science 275 (5299), 502–509, 1997.

Seneviratne, S. I. and Corti, T. and Davin, E. L. and Hirschi, M. and Jaeger, E. B. and Lehner, I. and 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.

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

Smirnova, T. G., Brown, J. M., and Benjamin, S. G.: Performance of different soil model configurations in simulating ground surface temperature and surface fluxes. Monthly Weather Review, 125, 1871-1884, 1997.

Smith, A. B., J. P.Walker, A. W.Western, R. I.Young, K. M.Ellett, R. C.Pipunic, R. B.Grayson, L.Siriwardena, F. H. S.Chiew, and H.Richter: The Murrumbidgee soil moisture monitoring network data set, Water Resour. Res., 48, W07701, doi:10.1029/2012WR011976, 2012.

Stöckli, R., Lawrence, D.M., Niu, G.Y., Oleson, K.W., Thornton, P.E., Yang, Z.L., Bonan, G.B., Denning, A.S., Running, S.W.: Use of FLUXNET in the community land model development. J. Geophys. Res 113 (G1), G01025, 2008.

Teuling, A. J., and P. A. Troch: Improved understanding of soil moisture variability dynamics. Geophys. Res. Lett., 32, L05404, doi:10.1029/2004GL021935, 2005.

Teuling, A. J., Uilenhoet, R., Hupet, F., Van Loon, E. E., and Troch, P. A.: Estimating spatial mean root-zone soil moisture from point-scale observations. Hydrol. Earth Syst. Sci., 10, 755-767, 2006.

Teuling, A. J., Uilenhoet, R., Van den Hurk, B., and Seneviratne, S. I.: Parameter Sensitivity in LSMs: An Analysis Using Stochastic Soil Moisture Models and ELDAS Soil Parameters. J. Hydrometeorol., 10(3), 751-765, doi:10.1175/2008JHM1033.1, 2009

Vachaud, G., Passerat de Silans, P., Balabanis, P., Vauclin, M.: Temporal stability of spatially measured soil water probability density function, Soil Sci. Soc. Am. J., 49, 822–828, 1985.

Van Dam, J. C., and Feddes, R. A.: Numerical simulation of infiltration, evaporation, and shallow groundwater levels with the Richards equation. J. Hydrol., 233, 72-85.

Vanderlinden, K., H. Vereecken, H. Hardelauf, M. Herbst, G. Martínez, M. H. Cosh, and Pachepsky, Y. A.: Temporal Stability of Soil Water Contents: A Review of Data and Analyses. Vadose Zone J., 11., doi:10.2136/vzj2011.0178, 2012.

Van Genuchten, M. T.: A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci. Soc. America J., 44, 892–898, 1980.

Vereecken, H., Huisman, J.A., Bogena, H., Vanderborght, J., Vrugt, J.A., and Hopmans, J.W.: On the value of soil moisture measurements in vadose zone hydrology: A review. Water Resources Research, 44, doi:10.1029/2008WR006829, 2008.

Vereecken, H., Huisman, J.A., Hendricks Franssen, H.J., Bruggemann, N., Bogena, H.R., Kollet, S., Javaux, M., van der Kruk, J., and Vanderborght, J: Soil hydrology: recent methodological advaces, challenges, and perspectives. Water Resources Research, 51, 10.1002/2014WR016852, 2015.Appendices

Vinnikov, K. Y., Robock, A., Shuang, W., Schlosser, A.: Scales of temporal and spatial variability of midlatitude soil moisture. Journal of Geophyisical Research 101, 7163{7174. URL http://dx.doi.org/10.1029/95JD02753, 1996.

Vivoni, E.R., Moreno, H.A., Mascaro, G., Rodriguez, J.C., Watts, C.J., Garatuza-Payan, J. and Scott, R.L.: Observed relation between evapotranspiration and soil moisture in the North American monsoon region. Geophysical Research Letters, 35(22), 2008.

Wood, E.F., et al.: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, Water Resour. Res., 47, W05301, doi:10.1029/2010WR010090, 2011.

Wösten, J. H. M.: Pedotransfer functions to evaluate soil quality. In Gregorich, E. G. and Carter, M. R. (eds.) Soil Quality for Crop Production and Ecosystem Health. Elsevier, Amsterdam. pp. 221–246, 1997.

Zreda, M., Desilets, D., Ferre, T. P. A., Scott, R. L.: Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic- ray neutrons. Geophysical Research Letters 35, 1-5, doi:10.1029/2008GL035655, 2008.

5   Zreda, M., Shuttleworth, W. J., Zeng, X., Zweck, C., Desilets, D., Franz, T., Rosolem, R.: Cosmos: the cosmic-ray soil moisture observing system. Hydrology and Earth System Sciences 16, 4079-4099, doi:10.5194/hess-16-4079-2012, 2012.

Table 1: Site characteristics. Altitude from COSMOS website, land cover percentages from AmeriFlux and publications. Harmonised World Soil Database (HWSD) data was used to define default model parameter values (here only soil categories shown).

Site	Altitude (meter above sea level; COSMOS)	Land cover (%)		HWSD dominant soil type	Site soil info (AmeriFlux, COSMOS, literature)	Data sources
		Dominant	**Remaining**			
UM	220	100% broadleaf		Loamy sand	Deep well drained soils	COSMOS, ORNL-DAAC (2015)
DC	1300	46% shrubs	46% bare, 8% needleleaf	Sandy loam		Anderson and Goulden (2012)
SO	1160	63% needleleaf	37% shrubs	Loam	0-10 cm thick organic litter layer, bedrock at 1-2 m	COSMOS, Goulden et al. (2012)
KE	1531	100% C4-grass		Loam	Coarse-loamy, limestone fragments	ORNL-DAAC (2015)
ME	1253	100% needleleaf		Loamy sand	Sandy, minimal organic	COSMOS, ORNL-DAAC (2015)
SR	989	76% bare	24% shrubs	Loam	Silty clay loam	COSMOS, Cavanaugh et al. (2011)
CS	457	100% shrubs		Loam		Anderson and Goulden (2012)
MM	288	100% broadleaf		Loam	Well drained clay loam	COSMOS, ORNL-DAAC (2015)
TR	177	60% C3-grass	40% broadleaf	Loam	Sandy clay loam, clay hardpan at 30-40 cm	COSMOS, ORNL-DAAC (2015), Chen et al. (2008)
AR	314	100% C4-grass		Loam	Sandy	COSMOS, ORNL-DAAC (2015)
WR	371	100% needleleaf trees		Loam	5-10 cm organic layer, silty sand	COSMOS, ORNL-

					DAAC (2015)
MO	219	100% broadleaf trees	Loam	Silty loam	COSMOS, ORNL-DAAC (2015)

Table 2: Calibrated parameter definitions and calibration ranges.

JULES parameter name	Unit	Role	Range	
			Minimum	**Maximum**
b	-	Mualem-Van Genuchten parameter (b=1/(n-1))	0.63	24.43
sathh	m	Mualem-Van Genuchten parameter (sathh=α^{-1})	0.09	28.01
K_{sat}	mm s^{-1}	Mualem saturation hydraulic conductivity	$3 \cdot 10^{-5}$	$4.3 \cdot 10^{-1}$
θ_{crit}	m^{3}m^{-3}	Critical soil moisture content	$0.1 \cdot \theta_{sat}$	$0.9 \cdot \theta_{sat}$
θ_{wilt}	m^{3}m^{-3}	Wilting soil moisture content	$0.1 \cdot \theta_{crit}$	$0.9 \cdot \theta_{crit}$
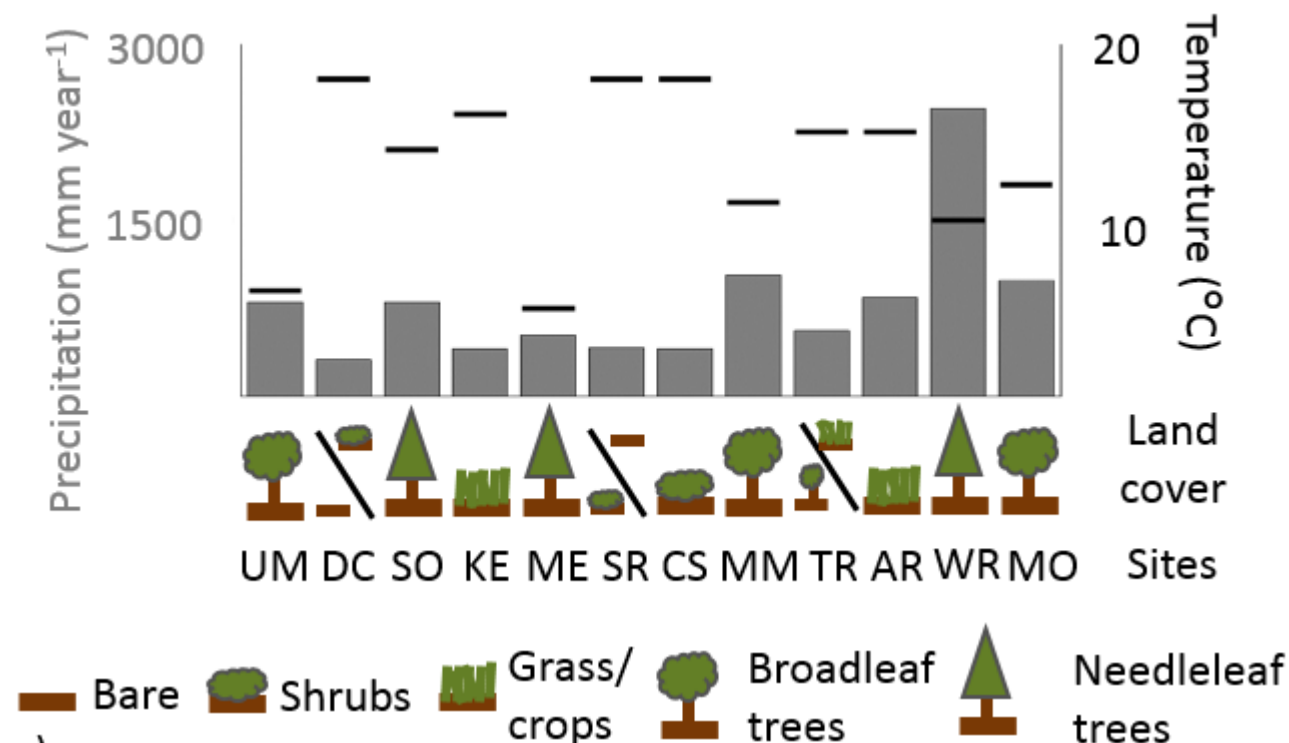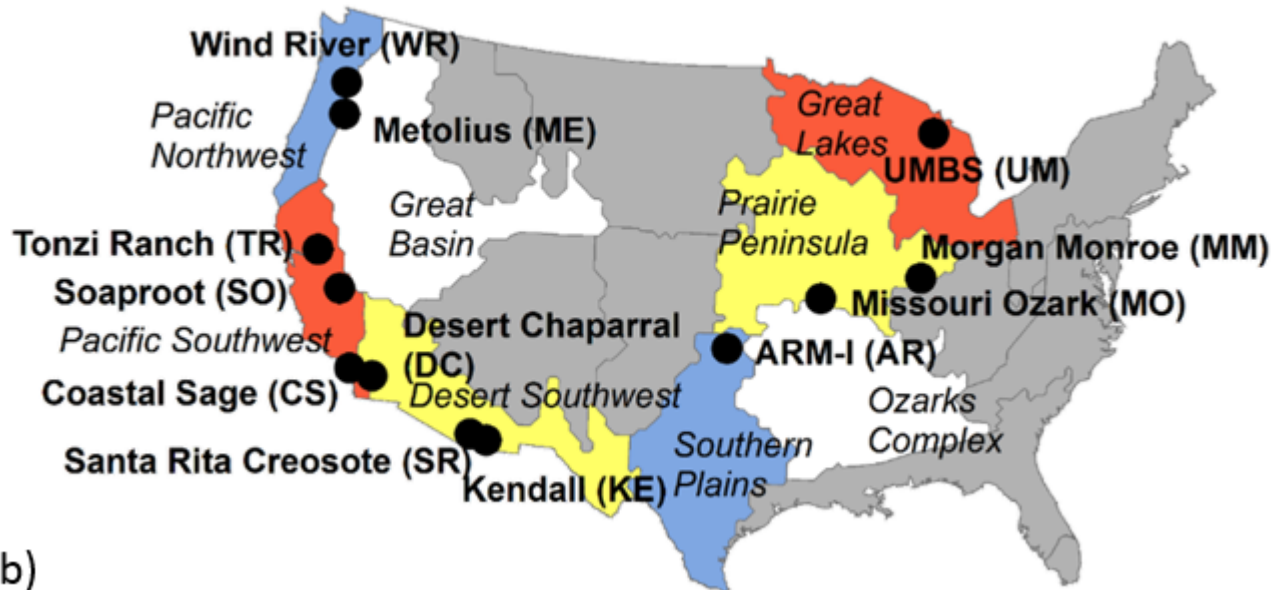
Figure 1: Figure a shows the yearly mean precipitation, air temperature, and dominant land cover types for the twelve AmeriFlux/COSMOS sites used. At sites DC, SR, and TR two different land cover types were shown because they covered similar areas in size. The map below (b) shows the locations of the twelve sites within eight NEON Ecoclimatic Domains. Data sources: COSMOS, ORNL-DAAC (2015), Goulden et al. (2012), Anderson and Goulden (2012), Scott et al. (1990), Chen et al. (2008).
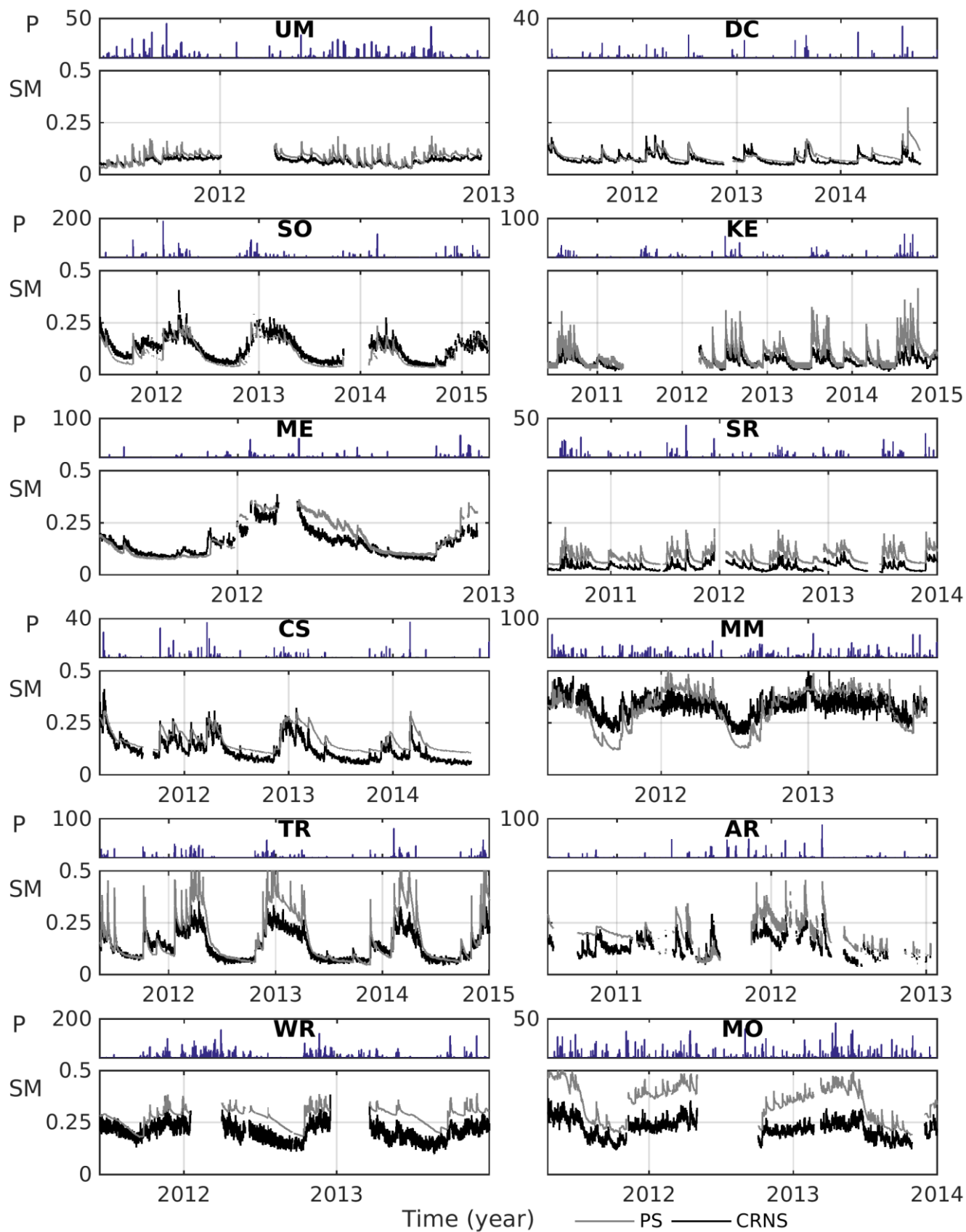
Figure 2: PS and CRNS observed soil moisture (SM; m^3m^{-3}) time series for the twelve study sites. Notice the PS soil moisture time series have been linearly interpolated from individual measurement depths to the corresponding JULES soil layers. CRNS soil moisture was obtained using COSMIC while assuming vertically homogeneous soil moisture. Daily precipitation (mm day^{-1}) is also shown here for each site.
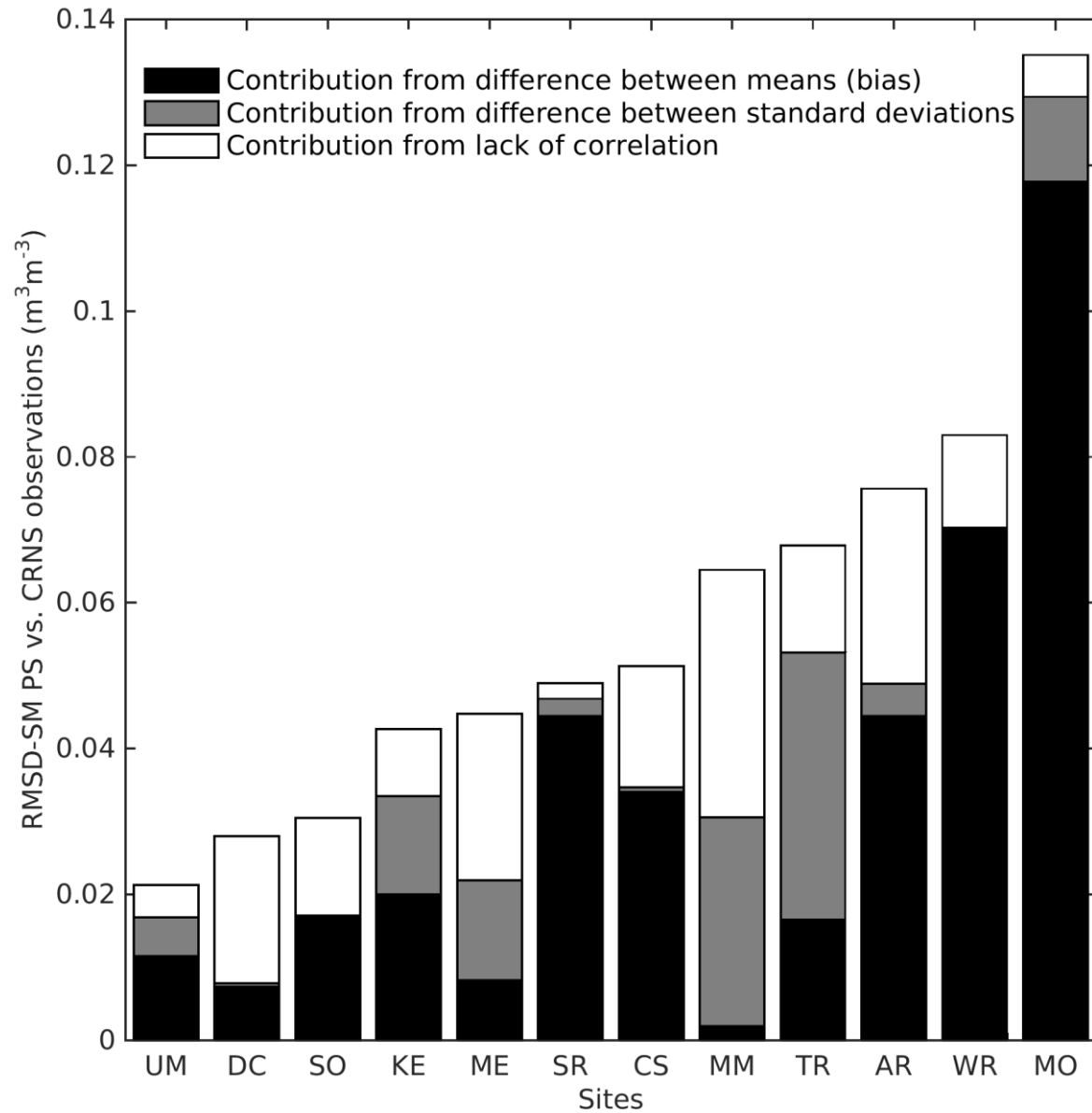
Figure 3: Root Mean Squared Deviation (RMSD) between observed PS and CRNS soil moisture (SM; m³m⁻³). MSD decomposition (Gupta et al., 2009) was calculated and the fractions were then applied to the RMSD values. Sites ranked from most similar to most different.
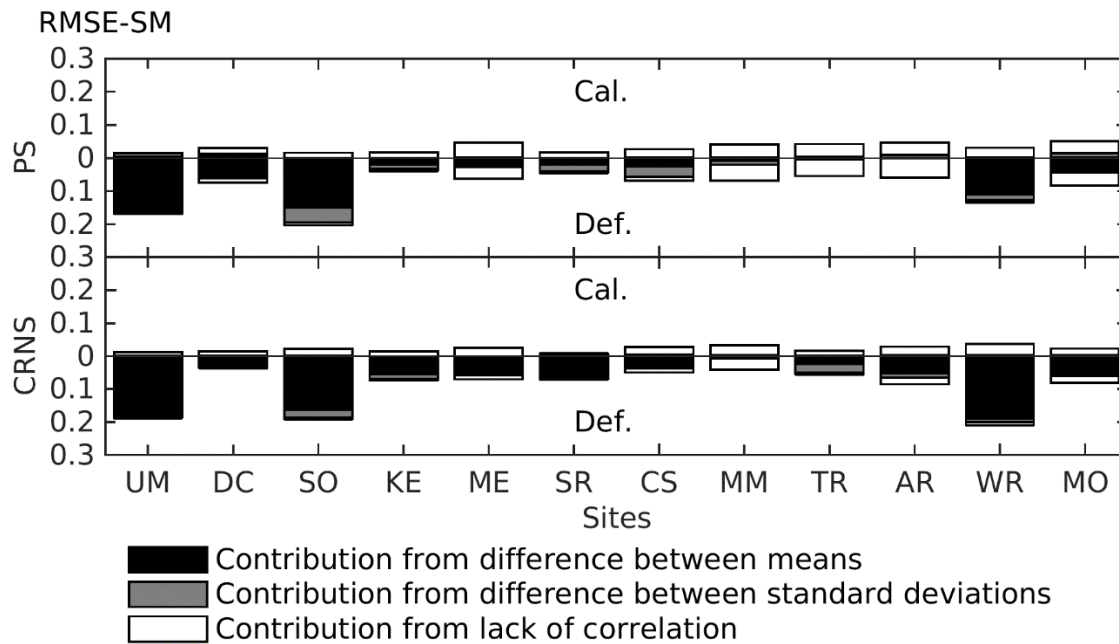
Figure 4: Objective function (Root Mean Square Error; RMSE (m³m⁻³)) values between observed PS/CRNS soil moisture and JULES simulated soil moisture. For each calibration the RMSE of the default run is shown from the horizontal axis down, and the result after calibration is shown in the upward directions. The different error contributions from the MSE decomposition are shown as stacked bar plots.
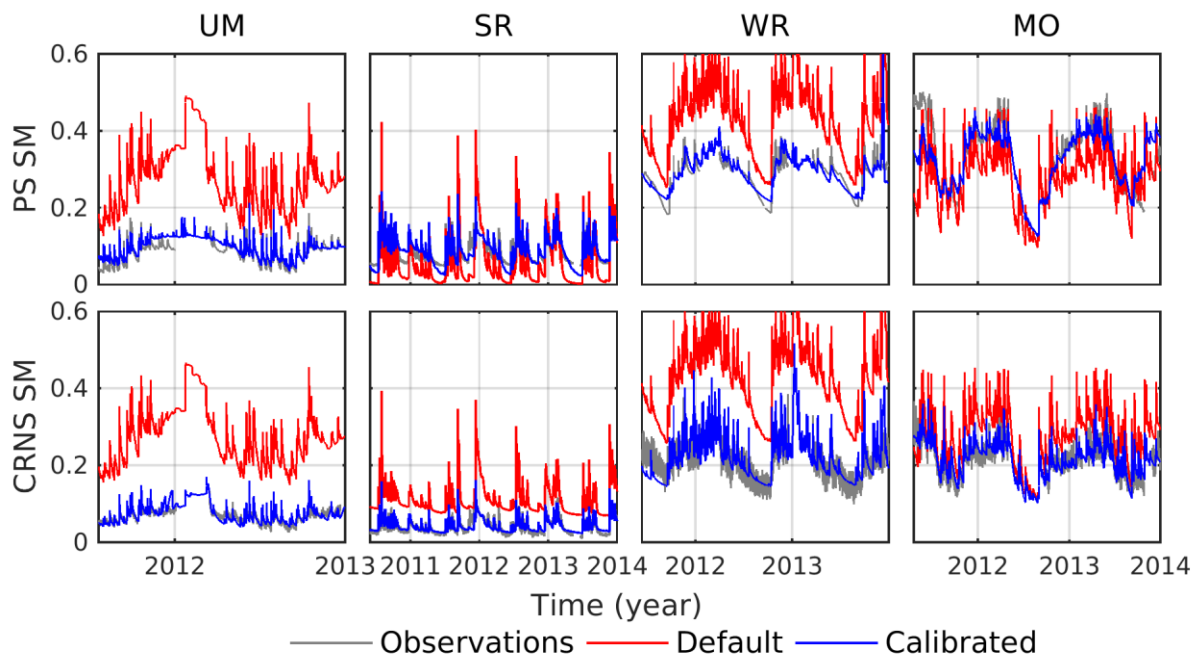
Figure 5: Hourly soil moisture time series of JULES default and calibrated runs against observations (PS or CRNS) for four of twelve sites.
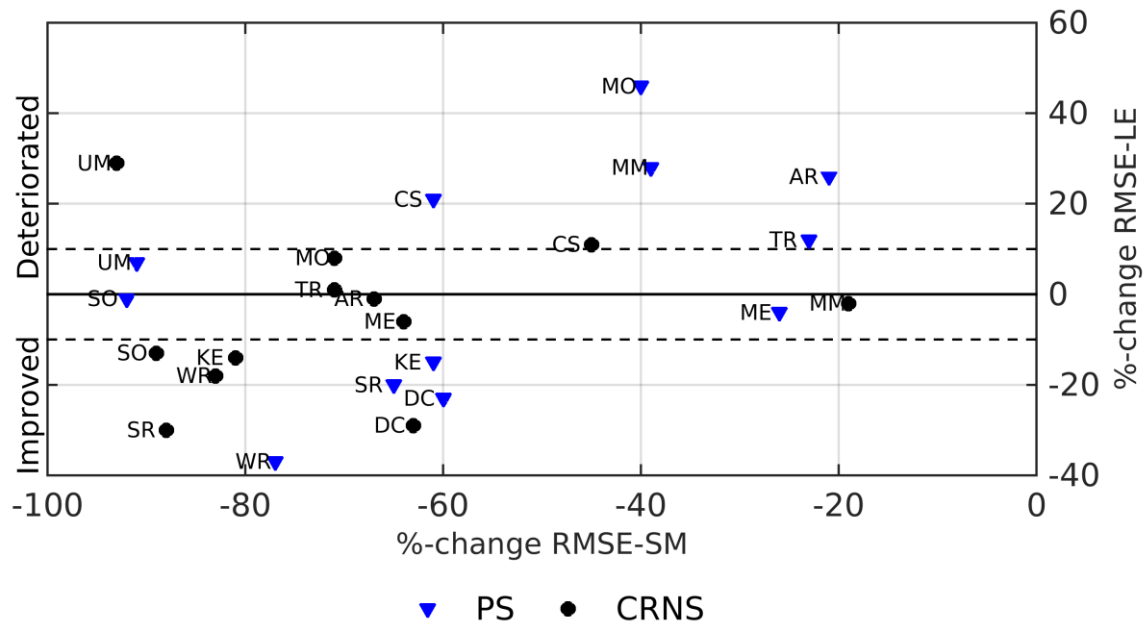
Figure 6: Relative change in Root Mean Square Error (RMSE) values after single-objective calibration against PS soil moisture and CRNS neutron count data. Change in RMSE between observed and simulated latent heat flux (LE) is plotted against change in RMSE between observed and simulated soil moisture.
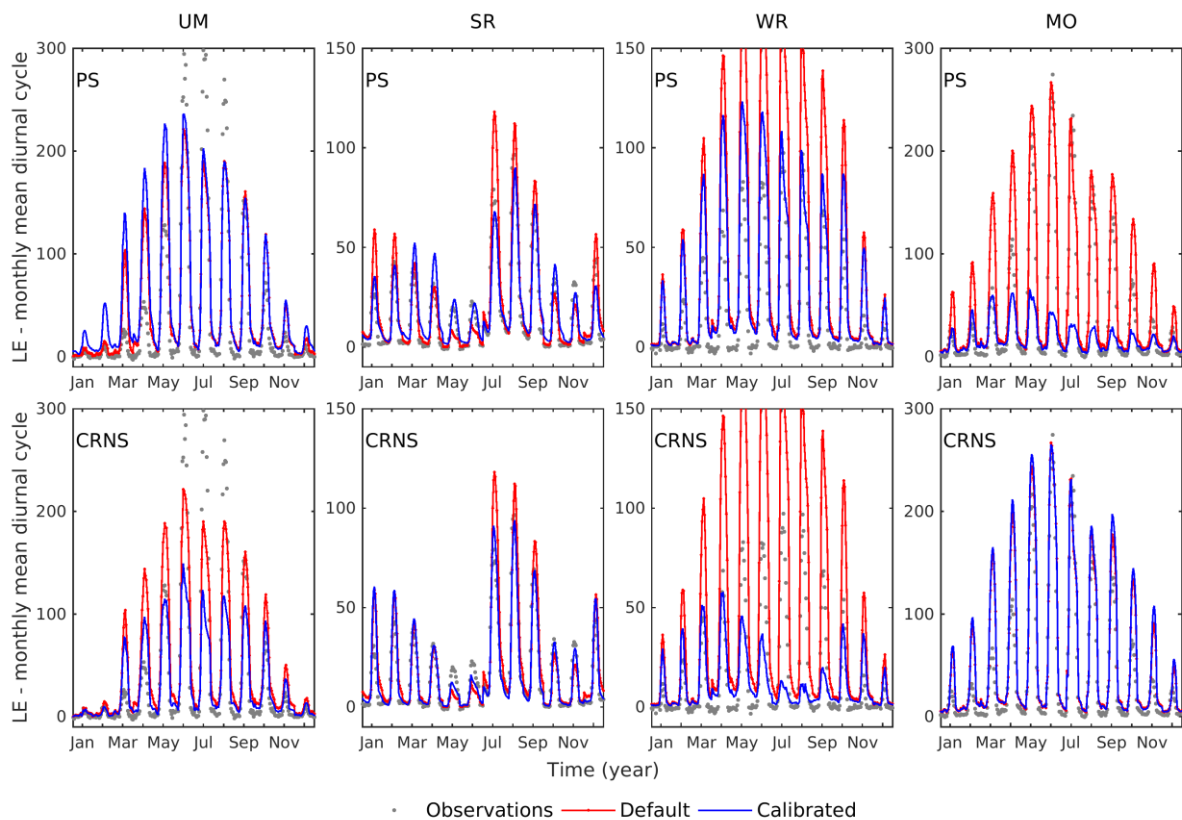
Figure 7: Monthly mean diurnal latent heat flux (LE) cycles. The upper row contains PS calibrated solutions, the lower row shows the CRNS calibrated solutions. Monthly mean diurnal latent heat flux cycles. The upper row contains PS calibrated solutions, the lower row shows the CRNS calibrated solutions.
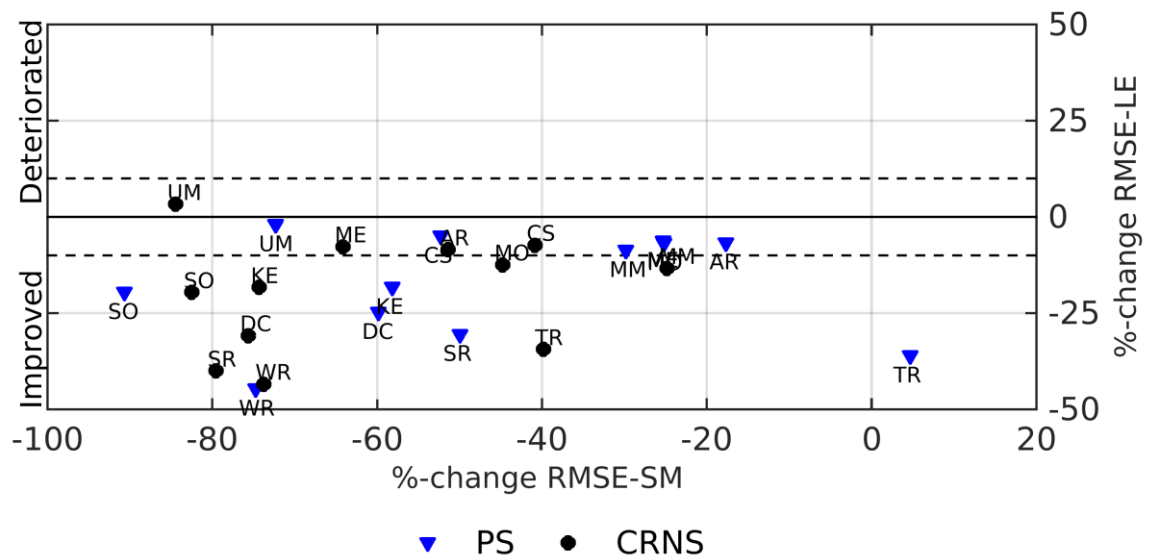
Figure 8: Relative change in Root Mean Square Error (RMSE) values after two-objective calibration against (1) PS soil moisture and latent heat flux and (2) against CRNS neutron count data and latent heat flux. Change in RMSE between observed and simulated latent heat flux (LE) is plotted against change in RMSE between observed and simulated soil moisture.
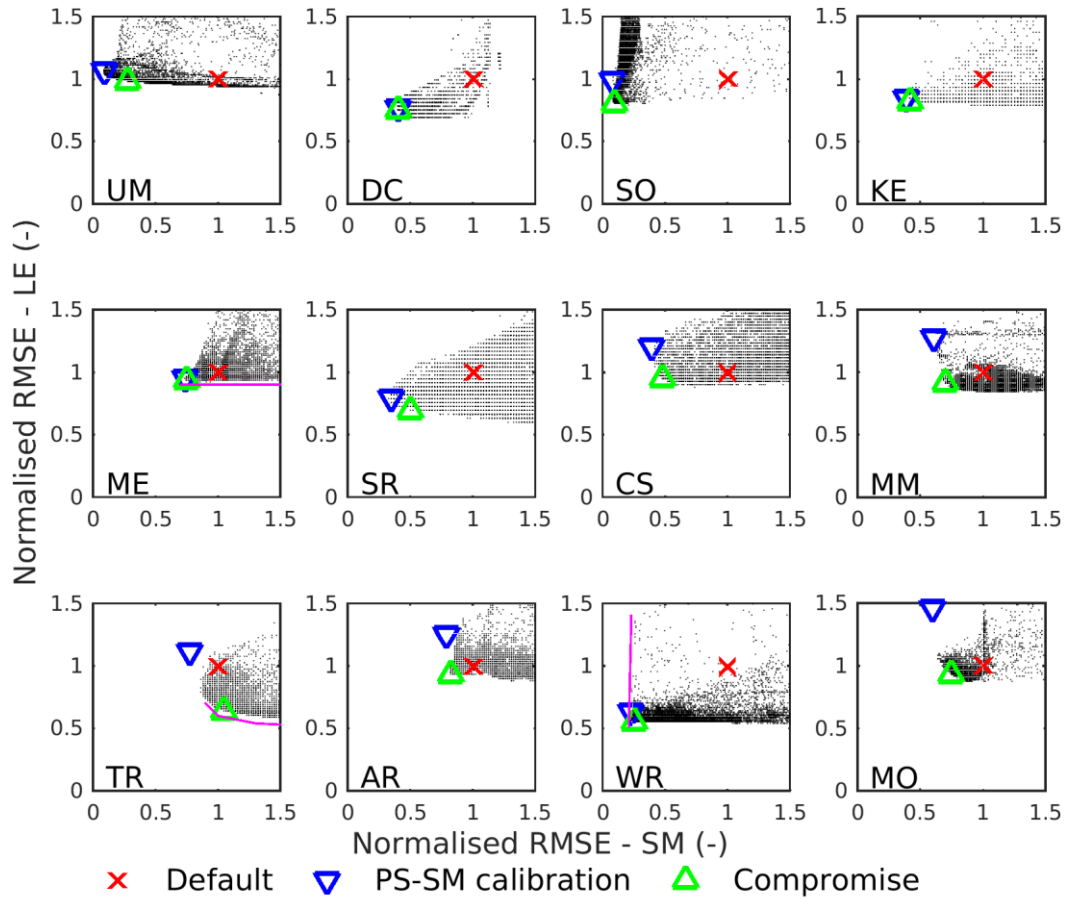
Figure 9: All runs from the two-objective calibrations against PS soil moisture and day time hourly latent heat flux (LE) at each site. Default run, PS-SM single objective calibrated run, and compromise solution runs are shown on top. All values were normalised to the values of the Default run. The plots were zoomed in to the area of 0 to 1.5 times the Default values.
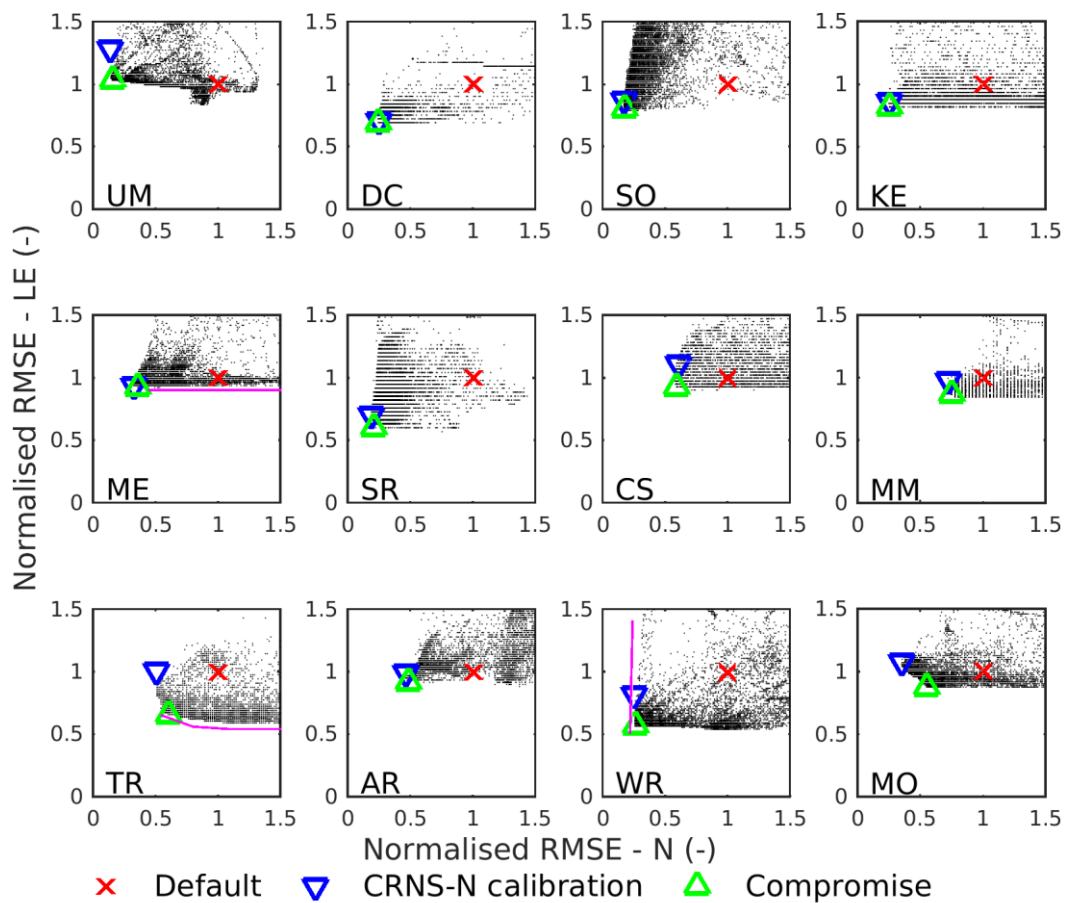
Figure 10: Same as Figure 9 but for the two-objective calibrations against CRNS neutron counts (representing soil moisture) and day time hourly latent heat flux (LE) at each site.

Appendices

Appendix 1: Additional Data and Methods tables and figures

Table A1.1: PS types, installation depths, number of profiles. In the last column it is shown to which JULES layer the observations were linearly interpolated.

Site	Type	Installation layers (cm below surface)	Number of profiles	Interpolated to JULES layer (cm)
UMBS	Campbell Scientific CS615 / CS616 (reflectometer)	0-30 average	1	0-10
Desert Chaparral UCI	Campbell Scientific CS616 (reflectometer)	0-30 average	4	0-10
Soaproot	Campbell Scientific CS615 (reflectometer)	0-30 average	4	0-10
Kendall	Stevens Water Hydra Probe	5,15	1	0-10
Metolius	Campbell Scientific CS615 (reflectometer)	0-30 average	1	0-10
Santa Rita Creosote	Time Domain Resistivity	2.5,12.5	6	0-10
Coastal Sage UCI	Campbell Scientific CS616 (reflectometer)	0-30 average	4	0-10
Morgan Monroe	Campbell Scientific CS615 (reflectometer)	0-30 average	1	0-10
Tonzi Range	ThetaProbe (ML2)	0,(20)	2	0-10,(10-35)
ARM-1	Decagon Echo2 EC-20 (capacitance)	10,25	2	10-35
Wind River	Campbell Scientific CS615 (reflectometer)	0-30 average	1	0-10
Mozark	Delta-T	10	1	0-10

Table A1.2: Data used per site for downward longwave radiation (L$_{win}$), net radiation (R$_{net}$), and atmospheric pressure (P$_{atm}$). A is AmeriFlux Level 2, COSMOS is COSMOS website, and NLDAS is National Land Data Assimilation Systems Forcing data Phase 2 (http://disc.sci.gsfc.nasa.gov/uui/datasets?keywords=NLDAS).

Site	L$_{w,in}$	R$_{net}$	P$_{atm}$
UM	A		A
DC		A	COSMOS
SO	NLDAS		A
KE		A	A
ME	A		A
SR	A		A
CS	NLDAS		COSMOS
MM	A		A
TR	NLDAS		A
AR	A		A
WR	A		A
MO	A		A

Appendix 2: Parameter sensitivity analysis

5 We investigated if a change in some parameters had a relatively small effect on soil moisture, while at the same time having a large effect on latent heat flux. In such case calibration could yield better soil moisture, but the parameter value might be inappropriate for latent heat flux. The inverse (influential on SM but not on LE) could also occur. We explored this by performing a sensitivity analysis with Morris' method (Morris, 1991) as implemented in the SAFE Toolbox (Pianosi et al., 2015) on the exact same parameter value ranges as used during the calibration. We computed the sensitivity indices (mean and

10 standard deviations of the elementary effects) on the RMSE values (i.e. our Objective Functions) of simulated vs. observed PS soil moisture, simulated vs. observed CRNS neutron counts, and simulated vs. observed latent heat flux (day time only).

The results (shown for four sites in Figure A2.2 of Appendix 2) were consistent across most sites: all three Objective Functions were most sensitive to changes in parameter b and least sensitive to the wilting point multiplier. Finch and Haria (2006) also found JULES parameter b to be most influential on soil moisture and latent heat, at a UK chalk site. The critical point soil

15 moisture content was influential with respect to soil moisture / neutron counts but had at all sites less effect on latent heat flux. This could relate to the self-adjusting behaviour of the wilting point and critical point soil moisture described in Section 3.3. The lack of effect from the wilting point soil moisture content can probably be attributed to the use of the multiplier, despite being a common approach (Prihodko et al., 2008; Rosolem et al., 2012); a certain multiplier value could be good in combination with a certain value for the critical point but not for a different critical point value.

20 We have evaluated the differences between default and calibrated parameters and their degree of physical realism in Supplement 3. If parameter values obtained after calibration were not physically feasible (e.g. representing a sandy soil while there was a clay soil) then, if model structure is assumed to represent biophysical processes sufficiently well, that could yield undesirable results. The wide ranges of parameter values, especially for the water entry pressure (sathh) and the saturation

hydraulic conductivity. However, physically realistic parameters also often provide unrealistic results (Gupta et al., 1998).
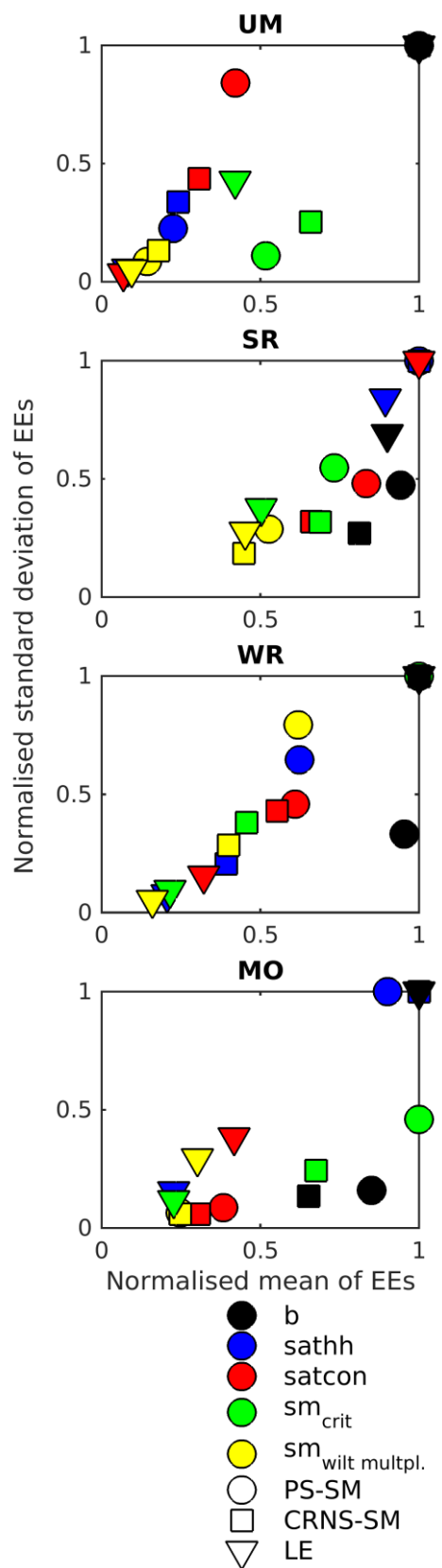
Figure A2.1: Morris sensitivity indices for the five calibrated parameters at four sites: UM, SR, WR, and MO. The three Objective Functions are the RMSEs between simulated and observed PS soil moisture (PS-SM), CRNS neutron counts (CRNS-N), and latent heat flux (LE). The mean elementary effects (EEs), representing the main effects, are displayed along the horizontal axis. The standard deviations of the elementary effects, representing parameter interactions, are shown along the vertical axis. Values were normalised to the most influential parameter for each Objective Function.