



**HAL**  
open science

# Evaluating and Optimizing Surface Soil Moisture Drydowns in the ORCHIDEE Land Surface Model at In Situ Locations

Nina Raoult, Catherine Ottlé, Philippe Peylin, Vladislav Bastrikov, Pascal Maugis

► **To cite this version:**

Nina Raoult, Catherine Ottlé, Philippe Peylin, Vladislav Bastrikov, Pascal Maugis. Evaluating and Optimizing Surface Soil Moisture Drydowns in the ORCHIDEE Land Surface Model at In Situ Locations. *Journal of Hydrometeorology*, 2021, 22 (4), pp.1025-1043. 10.1175/JHM-D-20-0115.1 . hal-03249171

**HAL Id: hal-03249171**

**<https://hal.science/hal-03249171>**

Submitted on 9 Jun 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## 🔗 Evaluating and Optimizing Surface Soil Moisture Drydowns in the ORCHIDEE Land Surface Model at In Situ Locations✍️

NINA RAOULT,<sup>a</sup> CATHERINE OTTLÉ,<sup>a</sup> PHILIPPE PEYLIN,<sup>a</sup> VLADISLAV BASTRIKOV,<sup>a</sup> AND PASCAL MAUGIS<sup>a</sup>

<sup>a</sup> *Laboratoire des Sciences du Climat et de l'Environnement, IPSL, Gif-sur-Yvette, France*

(Manuscript received 14 May 2020, in final form 3 February 2021)

**ABSTRACT:** The rate at which land surface soils dry following rain events is an important feature of terrestrial models. It determines, for example, the water availability for vegetation, the occurrences of droughts, and the surface heat exchanges. As such, surface soil moisture (SSM) “drydowns,” i.e., the SSM temporal dynamics following a significant rainfall event, are of particular interest when evaluating and calibrating land surface models (LSMs). By investigating drydowns, characterized by an exponential decay time scale  $\tau$ , we aim to improve the representation of SSM in the ORCHIDEE global LSM. We consider  $\tau$  calculated over 18 International Soil Moisture Network sites found within the footprint of FLUXNET towers, covering different vegetation types and climates. Using the ORCHIDEE LSM, we compare  $\tau$  from the modeled SSM time series to values computed from in situ SSM measurements. We then assess the potential of using  $\tau$  observations to constrain some water, carbon, and energy parameters of ORCHIDEE, selected using a sensitivity analysis, through a standard Bayesian optimization procedure. The impact of the SSM optimization is evaluated using FLUXNET evapotranspiration and gross primary production (GPP) data. We find that the relative drydowns of SSM can be well calibrated using observation-based  $\tau$  estimates, when there is no need to match the absolute observed and modeled SSM values. When evaluated using independent data,  $\tau$ -calibration parameters were able to improve drydowns for 73% of the sites. Furthermore, the fit of the model to independent fluxes was only minutely changed. We conclude by considering the potential of global satellite products to scale up the experiment to a global-scale optimization.

**KEYWORDS:** Hydrology; Soil moisture; Data assimilation; Land surface model

### 1. Introduction

Understanding the dynamics of surface soil moisture (SSM) storage is vital for developing accurate land surface models (LSMs). Soil moisture availability regulates a number of key physical processes of the Earth system, including the partitioning of the energy fluxes at the land surface, the biogeochemical cycles, and streamflows. Since the persistence of the soil moisture anomalies, known as soil moisture memory, is longer than the memory of the atmosphere (days to months compared to hours to days), the residence times of soil moisture are also important for the prediction of heat waves (Lorenz et al. 2010), droughts (Nicholson 2000), floods (Bonan and Stillwell-Soller 1998), crop yield (de Wit and Van Diepen 2008), or wildfires (Krueger et al. 2015).

However, modeling soil moisture dynamics is complex since it exhibits large sensitivities to meteorological forcing data and land surface model parameterizations. Fortunately, there are now a large variety of soil moisture observational products available with which to confront the models and possibly improve their representation of soil moisture. These range from local in situ measurements making up the ISMN (International

Soil Moisture Network; Dorigo et al. 2011, 2013) to remotely sensed global products using active or passive microwave sensors, such as SMOS (Soil Moisture Ocean Salinity; Kerr et al. 2010), SMAP (Soil Moisture Active Passive; Entekhabi et al. 2010b), or the ESA CCI SM (European Space Agency Climate Change Initiative Soil Moisture) combined product (Dorigo et al. 2017).

Soil moisture observations and retrievals can be used not only to evaluate the different processes in the model but also to calibrate the associated parameters, using for example data assimilation (DA) techniques. DA refers to the act of combining models and observations, while using the available knowledge about their respective uncertainties (Tarantola 2005). This can be used to improve the values of the model state variables (e.g., De Lannoy and Reichle 2016; Kolassa et al. 2017) and/or parameters (e.g., Yang et al. 2016; Pinnington et al. 2018). When the misfit between modeled and observed soil moisture cannot be resolved through state or parameter estimation, i.e., the model cannot reproduce the data within the defined uncertainties after calibration, we gain insight into the model structural errors and possible directions for model improvement.

Soil moisture DA has been shown to improve the prediction of root-zone and surface soil moisture estimates as well as drought and flood modeling (Komma et al. 2008; Bolten et al. 2010; Wang et al. 2009; Liu et al. 2011; Draper et al. 2012; De Lannoy and Reichle 2016; Kolassa et al. 2017; Gruber et al. 2019). It has also been used to improve precipitation data (Román-Cascón et al. 2017) and infer soil textures (Pinnington et al. 2018). Due to its strong coupling with the carbon cycle, soil moisture observational products have also been assimilated in conjunction with other products such as FAPAR (fraction of absorbed photosynthetically active radiation; Wu et al. 2019), atmospheric CO<sub>2</sub> concentrations

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

✍️ Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JHM-D-20-0115.s1>.

Corresponding author: Nina Raoult, [nina.raoult@lsce.ipsl.fr](mailto:nina.raoult@lsce.ipsl.fr)

DOI: 10.1175/JHM-D-20-0115.1

© 2021 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy \(www.ametsoc.org/PUBSReuseLicenses\)](#).

(Wu et al. 2020) and leaf area index (Bonan et al. 2020) with the aim to improve simultaneously the modeling of carbon and water cycles.

There are a number of potential biases between the observations and model output that should be considered prior to parameter calibration. Bias in observations typically reflects instrumental inaccuracies, representational errors and, when considering satellite products, errors in the retrieval algorithms (De Lannoy et al. 2007), whereas biases in the model are due to imperfect model structure (e.g., a missing process), forcing data and initial conditions. Due to the large biases and discrepancies in the soil moisture climatologies caused by these errors, many studies perform bias removal prior to calibration. When performing state estimation, it is critical to remove all of these biases (Reichle and Koster 2004; Drusch et al. 2005; Crow et al. 2005; Draper et al. 2012). Parameter calibration, on the other hand, can provide a way to remove some of the biases between the model mean state variable and the observations. However, separating the biases that can be corrected through the parameter calibration from the other types of biases is challenging. Before applying a Bayesian parameter calibration, we still need to remove all biases that cannot be alleviated by the parameter calibration. Hence, in parameter estimation studies it is also common to remove biases prior to calibration (Scholze et al. 2016; Wu et al. 2020; de Rosnay et al. 2020). Another approach for dealing with soil moisture biases is to consider only temporal dynamics such as correlation and autocorrelation. These temporal dynamics, which tend to be less affected by biases, can be very informative about important model processes and phenomena controlling the residence time of soil moisture after a rain event (i.e., the soil moisture memory).

Soil moisture memory can be estimated using a number of different metrics, most of which are based on the autocorrelation of the soil moisture time series. These include the  $e$ -folding autocorrelation first proposed by Delworth and Manabe (1989) and used in many studies such as Koster and Suarez (2001) and Raoult et al. (2018). Autocorrelation metrics are applied to the anomalies of the soil moisture time series where anomalies are defined as deviations from a mean soil moisture state. However, in most cases, the soil moisture record is too short to accurately calculate the soil moisture mean state and so other metrics have been developed to measure soil moisture memory. One example is mean persistence time—the average time soil moisture spends above or below a certain threshold (Ghannam et al. 2016). Another is stored precipitation fraction, which is more dependent on precipitation events (McColl et al. 2017a). More recently, drydowns have been used to measure soil moisture memory.

Drydowns are characterized as the rate at which the soil dries during periods with no precipitation. After a precipitation event, part of the water input is intercepted by vegetation (if present), part runs off to surface water and the rest infiltrates into the soil. According to Laio et al. (2001), the drying rate of the soil typically follows three simplified different regimes:

- (i) While the soil moisture remains above field capacity, drainage and runoff dominate.
- (ii) When the soil moisture gets below field capacity but remains above a critical point, drying occurs by evapotranspiration at the rate of atmospheric demand and is

invariant with respect to soil moisture. This is known as “stage I” evapotranspiration.

- (iii) Below the critical point when plants start closing their stomata to reduce transpiration, the rate of evapotranspiration becomes limited by soil moisture availability. This is known as “stage II” evapotranspiration.

Generally, drainage and stage I evapotranspiration occur rapidly and so drydowns, the time series of soil moisture after the infiltration of precipitation into the soil, will, in most cases, be dominated by stage II evapotranspiration.

Drydowns have been the focus of many recent studies and drying rates have been shown to vary greatly between different models, in situ measurements and satellite retrievals. For example, Rondinelli et al. (2015) and Shellito et al. (2016) both showed that satellite data (SMOS and SMAP, respectively) dry out faster than in situ data and Shellito et al. (2018) further showed that soil moisture in satellite data can dry out faster than that simulated in a land surface model. Martínez-de la Torre et al. (2019), using evapotranspiration data to measure drydowns instead of soil moisture measurements, compared a number of land surface and hydrological models and found that drydown rates differed between them. Drying rates have also been shown to change seasonally and spatially within a given LSM (Salvia et al. 2018).

Many previous studies have estimated drying rates for the root-zone soil moisture (RZSM). This zone covers a depth of approximately one meter and relates to water availability for vegetation. However, more recently, studies have focused on surface soil moisture (SSM; McColl et al. 2017a,b), which represents the top soil moisture (i.e., water content in the first few centimeters of the soil layer). SSM and RZSM are not directly governed by the same processes: the surface being more influenced by direct evaporation and runoff, whereas the root zone is more directly influenced by transpiration and drainage. However, since all these processes and soil water transfers depend on soil properties and land coverage, both SSM and RZSM have been shown to be highly correlated in many cases. Nevertheless, they can be very different, as, for example, under dry conditions where there can be a decoupling between surface and root zone soil moisture. When comparing modeled and measured soil moisture, it is therefore important to account for representative depths. When in situ observations are used, the sensor depth is generally provided. However, for satellite products, the representative depth is less well defined, especially if various types of sensors or spectral bands with specific penetration/representative depths are involved.

In this study we will use the process-based ORCHIDEE LSM (Krinner et al. 2005) to investigate and calibrate SSM dynamics. We will focus on the use of in situ data to prepare for subsequent use of satellite SSM estimates. In light of the above, we aim to answer the following key questions:

- Can we identify SSM drydowns in the ORCHIDEE LSM in an automatic and consistent manner? How do these modeled drydowns compare to the ones observed in the in situ data?
- Which key hydrology, and possibly energy and carbon, parameters control drydown shapes? Can we improve these parameters through calibration?
- Is it better to constrain the calibration with  $\tau$  estimates derived from SSM observations or with the full time series of

(bias corrected) observations? How do these calibrations affect first the water fluxes and then the energy and carbon fluxes?

- What are the implications of these findings for the use of existing and upcoming soil moisture satellite products?

## 2. Methods and materials

### a. Model and data

#### 1) THE ORCHIDEE LSM

The ORCHIDEE LSM is the land component of the French Earth System Model (ESM) developed at the IPSL (Institut Pierre Simon Laplace). It simulates the interactions between the biosphere and the atmosphere by modeling the water, energy, and carbon cycles (Krinner et al. 2005). The model can be run over a wide range of time scales, from 30 min to thousands of years, and spatial scales, from single mesh to global. The version of ORCHIDEE used in this study (Boucher et al. 2020; Lurton et al. 2020) has been developed from the most recent phase of the Coupled Model Intercomparison Project (CMIP6). The only difference with the CMIP6 configuration of the model is the activation of the R-SOIL parameterization, which accounts for a supplementary resistance to soil evaporation. This is the version and configuration used in Raoult et al. (2018).

The model can either be run in coupled mode, where the atmospheric variables are calculated by another model, or run offline using meteorological forcing data to drive the model. When running the model offline, meteorological data measured at each location are used. These data are gap-filled (Vuichard and Papale 2015) to produce half-hourly or hourly data streams of air temperature, humidity, pressure, wind speed, precipitation rates, and shortwave and longwave incoming radiation. In ORCHIDEE, all the physical processes are treated at a half-hourly time step; however, some processes like phenology (senescence and mortality) are calculated over daily time steps.

ORCHIDEE's multilayered hydrological model discretizes the first 2 m of the soil column over 11 layers with increasing grid spacing following a geometric progression with common ratio 2 (de Rosnay et al. 2002). Soil moisture at different levels is obtained by resolving the Richards equation which represents the vertical water transfers in the unsaturated zone. With the finer discretization at the top of the column (i.e., depths of 1 mm, 4 mm, 1 cm, 2.2 cm, 4.5 cm, etc.), we are able to match the simulated soil moisture to the depth of the in situ sensor used to generate the observations. Although the vegetation in ORCHIDEE is represented by 15 plant functional types (PFTs), when calculating the soil hydrology, these PFTs are further aggregated into three groups: bare soil, short vegetation (grasses and crops), and forests. An independent soil water budget is calculated for each of these three groups. In this study, we consider the grid box weighted average SM profile (based on the fraction of vegetation present). The key water, energy, and carbon parameters of ORCHIDEE considered in this study are listed in Table 1 and further described in the online supplemental material.

#### 2) IN SITU DATA

Two databases of in situ data are used in this study. The first is from the International Soil Moisture Network (ISMN), which is an international cooperation that establishes and maintains a global in situ soil moisture database (Dorigo et al. 2011, 2013). In situ soil moisture measurements from operational networks and validation campaigns are collected and harmonized. The data are transformed into common volumetric soil moisture units and checked for outliers and implausible values (Dorigo et al. 2011). This database provides indispensable soil moisture observations needed for the calibration of LSMs like ORCHIDEE. To comply with the different measurement techniques and installation positions of the sensors, this database provides a measurement depth interval. At each site considered in this study (Table 3), the maximal depth of this interval is the depth used.

The second in situ database we use in this study is FLUXNET2015 (Pastorello et al. 2017). In this study we are running ORCHIDEE offline and so we need to prescribe the meteorological forcing data. Since soil moisture is highly influenced by meteorological forcing, especially precipitation data, it is important to use the most reliable local measurements. The FLUXNET2015 database provides in situ half-hourly or hourly meteorological forcing data. In addition, it provides net carbon flux measurements [further split into gross primary production (GPP) and ecosystem respiration (Resp)] and latent (LE) and sensible (HS) heat fluxes. In this study we use primarily the LE and GPP estimates to evaluate the optimized model. To select sites for our experiments (listed in Table 3) we consider all the ISMN sites which are found within 250 m of a FLUXNET tower from the FLUXNET2015 database. At each site, the overlap between both data records is used as the period of study. We want to ensure that the seasonal cycles are captured so we only keep sites with at least one year of near-continuous data.

We do acknowledge that the ISMN sites used may not be in the direct footprint of a FLUXNET tower, and the additional factors such as topology, vegetation distribution, soil heterogeneity, and prevailing winds may affect the measurements. However, this is more accurate than the alternatives such as global reanalysis forcing maps (e.g., CRUNCEP), which work on a coarser spatial resolution.

#### b. Selecting drydown events

Methods for identifying drydowns fall into two main categories. The first type of method uses precipitation data to determine dry periods. This is either done by defining two thresholds of significant rainfall events separated by at least a few days; one threshold marks the beginning of the period with a significant wetting of the soil and a second threshold marks the end of the dry period (e.g., Shellito et al. 2016; Salvia et al. 2018). Alternatively, a dry period can be defined as containing at least a given number of days without rainfall (e.g., Martínez-de la Torre et al. 2019). The second type of method analyses the soil moisture time series directly to find constantly decreasing dynamics, as in McColl et al. (2017b).

TABLE 1. List of the key water, energy, and carbon parameters of the ORCHIDEE LSM used in this study. The full list is used in the sensitivity analysis (section 4c) whereas the sublist (denoted by carets) is used during the calibration experiments. Shown are the parameter descriptions and default model values, i.e., the values used when performing a standard ORCHIDEE run. Note that a number of the parameters are PFT specific, while some others are specific to the soil texture used based on the soil composition of the site. Full parameter descriptions and equations can be found in the supplemental material.

Parameter	Description	Plant function type <sup>a</sup>						
		ENF temp	EBF	DBF/MF	ENF Bor	WSA/OSH	SAV	CRO
<b>Photosynthesis</b>								
a1	Empirical factors involved in the calculation of leaf-to-air vapor pressure difference	0.85	0.85	0.85	0.85	0.85	0.85	0.85
b1		0.14	0.14	0.14	0.14	0.14	0.14	0.14
$V_{cmax}^{\wedge}$	Maximum carboxylation rate ( $\mu\text{mol m}^{-2} \text{s}^{-1}$ )	40	50	45	35	50	60	60
$T_{min}$	Minimal photosynthesis temperature ( $^{\circ}\text{C}$ )	-4	-4	-4	-4	-4	-4	-4
$T_{max}$	Maximal photosynthesis temperature ( $^{\circ}\text{C}$ )	55	55	55	55	55	55	55
<b>Phenology</b>								
$LAI_{max}$	Maximum leaf area index ( $\text{m}^2 \text{m}^{-2}$ )	4	5	3.5	4	2	5	5
$L_{agecrit}$	Critical leaf age (days)	730	180	910	180	120	90	90
$K_{LAIhappy}$	Threshold of LAI below which plant uses carbohydrate reserves	0.5	0.5	0.5	0.5	0.5	0.5	0.5
$SLA^{\wedge}$	Specific leaf area ( $\text{m}^2 \text{g}^{-1}$ )	0.02	0.026	0.009 26	0.26	0.26	0.026	0.026
$\tau_{leafinit}$	Time to attain initial foliage (days)	10	10	10	10	10	10	10
$T_{senes}$	Temperature threshold for senescence ( $^{\circ}\text{C}$ )	—	—	12	—	-1.1375	5	5
<b>Autotrophic respiration</b>								
$Frac_{gresp}$	Fraction of GPP that is lost as growth respiration	0.28	0.28	0.35	0.35	0.28	0.28	0.28
$MR_{offset}$	Offset and slope of the linear relationship between temperature and maintenance respiration				0.14			
$MR_{slope}$		0.16	0.16	0.16	0.16	0.12	0.16	0.12
<b>Soil water availability</b>								
$z^{\wedge}$	Root profile ( $\text{m}^{-1}$ )	0.8	0.8	1.0	1.0	4.0	4.0	4.0
$\alpha^{\wedge}$	Controls water stress curve				1			
		Soil texture						
		Sandy loam		Loam			Clay loam	
<b>Soil hydrology</b>								
$n^{\wedge}$	Van Genuchten water retention curve coef $n$		1.86		1.56		1.31	
$A^{\wedge}$	Van Genuchten water retention curve coef $A$ ( $\text{mm}^{-1}$ )		0.0075		0.0036		0.0019	
$K_s^{\wedge}$	Hydraulic conductivity at saturation ( $\text{mm day}^{-1}$ )		1060.8		249.6		62.4	
$\theta_w^{\wedge}$	Volumetric water content at wilting point ( $\text{m}^3 \text{m}^{-3}$ )		0.0657		0.0884		0.1496	
$\theta_f^{\wedge}$	Volumetric water content at field capacity ( $\text{m}^3 \text{m}^{-3}$ )		0.1218		0.1654		0.2697	
$\theta_r^{\wedge}$	Residual volumetric water content ( $\text{m}^3 \text{m}^{-3}$ )		0.055		0.078		0.085	
$\theta_s^{\wedge}$	Saturated volumetric water content ( $\text{m}^3 \text{m}^{-3}$ )		0.41		0.42		0.41	
$p_{\%}^{\wedge}$	Percent of soil moisture above which transpiration is max				0.8			
$sl_r$	Slope coefficient for reinfiltration				0.1			

TABLE 1. (Continued)

Energy balance		
rs* <sup>A</sup>	Factor controlling soil resistance to evaporation	1
stc*	Factor controlling soil thermal conductivity (cnd)	1
hc*	Factor controlling heat capacity (pcapa)	1
z0	Reference rugosity length (m)	0.0625
wet	Wet soil heat capacity of soils	$3.03 \times 10^6$
dry	Dry soil heat capacity of soils	$1.80 \times 10^6$

<sup>a</sup> DBF, deciduous broadleaf forest; EBF, evergreen broadleaf forest; ENF, evergreen needleleaf forest; MF, mixed forest; WSA, woody savannah; OSH, open shrubland; SAV, savannah; CRO, cropland.

We choose to use precipitation data to identify drydowns for two reasons. First, we want to be able to compare modeled and observed SSM time series over the same period. The differences in the modeled and observed soil moisture temporal dynamics could mean that different drydown periods would be selected for each data stream if we used the second method. Second, by using FLUXNET sites, the precipitation data we have access to are measured locally and are therefore more reliable.

Using the FLUXNET data at each site, we first select 5-day periods with negligible rainfall ( $<1 \times 10^{-2}$  mm) for which we have at least 70% in situ SSM observations. Unlike Shellito et al. (2016), we do not include any criteria determining a threshold amount of precipitation accumulated over the previous day. We found that such criteria were more likely to exclude drydowns and that five dry days were sufficient to ensure an exponentially decaying drying of the soils.

Once a period of interest has been selected, an exponential model of the form

$$\theta(t) = \theta_A \exp\left(-\frac{t}{\tau}\right) + \theta_{eq} \quad (1)$$

is fitted to each drydown using a nonlinear least squares fitting algorithm, where  $\theta$  is surface soil moisture content,  $t$  is the time since the beginning of the drydown in days,  $\theta_A$  is the amplitude of the drydown,  $\tau$  is the temporal  $e$ -folding decay, and  $\theta_{eq}$  is the SSM equilibrium value. Modeled  $\theta$  approaches but never reaches  $\theta_{eq}$ . In fitting the exponential, bounds are set to keep realistic values. The amplitude of the drydown  $A$  is set to always be positive. When fitting the modeled drydowns,  $\theta_{eq}$  is set to be greater than  $\theta_r$ . For the ORCHIDEE run prior to calibration, the default value of  $\theta_r$  is used (see Table 1). During the optimization, this parameter is allowed to change, adjusting this lower bound. When fitting the observed drydowns,  $\theta_{eq}$  is set to be between the lowest value of the whole SSM time series and the lowest values of SSM during the drydown period considered (following Shellito et al. 2016).

Consistent with McColl et al. (2017b), the drydown is then only kept if the coefficient of determination ( $R^2$ ) of the exponential fit is at least 0.7. We also only consider drydowns with  $\tau$  values less than 50 days, following Martínez-de la Torre et al. (2019). This upper bound was implemented since drying events with  $\tau$  exceeding this value

would represent very slow drying or no drying at all which could be due to artificial water sourced from irrigation, too weak solar radiation in winter for soil evaporation, or no storage of soil moisture in regions of shallow groundwater table and slow drainage. These criteria are used when identifying drydowns in both the modeled and observed time series. No further filtering of drydowns is undertaken. Since in this study we only consider 18 sites, it would be possible to look at each drydown and discard the ones that do not “look right.” However, since we anticipate scaling the approach to a global scale, we choose instead to have a systematic procedure for identifying and fitting drydowns.

### 3. Optimization framework

#### a. Sensitivity analysis

Before calibrating a model, it is vital to identify the key internal parameters which have the most impact on the given model output. This is because (i) calibration methods can be costly and scale with the number of parameters used in the optimization, and (ii) attempting to calibrate an excessive number of parameters can lead to overfitting, and a severe degradation in model performance when the model is used in predictive mode, especially with respect to variables not used in the calibration. The first stage is therefore to carry out a parameter-sensitivity study by running the model in forward mode with many sets of feasible parameter values to select those parameters that will then be estimated in the calibration.

For this task, we used the Morris method (Morris 1991; Campolongo et al. 2007). The Morris method is a “one-at-a-time” method using randomized sampling matrices which allow direct observation of elementary effects. This guarantees that meaningful information can be extracted from each parameter, without mistakenly attributing effects to that parameter. This method requires a relatively low number of samples:  $(p + 1)n$ , where  $p$  is the number of parameters, and  $n$  is the number of trajectories generated. The method is qualitative; it provides a heuristic score to intuitively represent the relative sensitivity of parameters. While it does not inform us on how sensitive the parameters are, or on the interactions between parameters, and depends heavily on our choice of parameter intervals, this ranking can help highlight the important parameters and remove the ones with little to no effect.

### b. Principles of DA

DA refers to the act of optimally combining observations (whether measured directly or indirectly) with a numerical model, usually with the aim of reducing the difference between them. It can be used either to update the state variables of the model or improve the values of internal model parameters. In this study we will use DA for the latter. Using a Bayesian framework, and assuming the parameters, observations and model outputs are Gaussian distributed, we aim to minimize the cost function  $J(\mathbf{x})$ , where  $\mathbf{x}$  denotes the vector of parameters we wish to optimize, i.e., the most sensitive and uncertain model parameters:

$$J(\mathbf{x}) = [M(\mathbf{x}) - \mathbf{y}]^T \mathbf{R}^{-1} [M(\mathbf{x}) - \mathbf{y}] + (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b). \quad (2)$$

The first term of this function measures the mismatch between the observations  $\mathbf{y}$  and the model  $M(\mathbf{x})$ , where  $M$  is the composition of the model and observation operators; the model operator simulates the output for a given set of parameters  $\mathbf{x}$  and the observation operator ensures the simulated output is comparable to the observations treated (here the identity function). In this study,  $\mathbf{y}$  is taken to be either a time sequence of observed  $\tau$  values or the entire time series of SSM observations. The second term of the cost function measures the mismatch between the prior  $\mathbf{x}_b$  and parameter values  $\mathbf{x}$  (exponent T denotes the transpose).

These two terms are weighted by their error covariance matrices;  $\mathbf{R}$  and  $\mathbf{B}$  are the prior covariance matrices for the observation and parameter errors, respectively. The error covariance matrices are hard to estimate, therefore, as in most studies, these two matrices are kept diagonal, which offsets equifinality issues of correlated parameters. Following [Bastrikov et al. \(2018\)](#), we defined the observation errors (variances) as the mean-squared difference between the observations and the prior model simulations—this reflects not only the measurement errors but also significant model errors. When calibrating the  $\tau$  values, the errors are of order 100 days<sup>2</sup>. For the SSM values, these errors are of order 0.001 (m<sup>3</sup> m<sup>-3</sup>)<sup>2</sup>.

To check that the variances used for  $\mathbf{R}$  were sensible, after the optimization, we calculated the reduced  $\chi^2$ ; i.e., the normalized cost function at the minimum ([Tarantola 2005](#)). When this metric is significantly greater than 1, it implies an underestimation of the  $\mathbf{R}$  and/or  $\mathbf{B}$  matrices (note that  $\mathbf{B}$  matrices are likely to be less affected by diagonal elements). Therefore, for calibrations where the reduced  $\chi^2$  were significantly greater than 1, we performed the calibration again inflating variances. Fortunately, this was only the case for a couple of calibrations. (See supplemental material for more information on the reduced  $\chi^2$  and values used to define  $\mathbf{R}$  for each optimization).

The optimal set of parameters corresponds to the minimum of this function. To minimize the cost function, we use the DA system ORCHIDAS presented in [Bastrikov et al. \(2018; https://orchidas.lsce.ipsl.fr/; accessed September 2020\)](#) and the genetic random search algorithm ([Goldberg and Holland 1988](#); see supplemental material for the justification of algorithm

choice). This iterative method is a type of evolutionary algorithm that follows the principles of genetics and natural selection, where each parameter corresponds to a gene in a chromosome. At the optimum, assuming Gaussian prior errors and linearity of the model around the solution, we then approximate the posterior error covariance matrix  $\mathbf{A}$  using the following equation:

$$\mathbf{A} = [\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{B}^{-1}]^{-1}, \quad (3)$$

where  $\mathbf{H}$  is the Jacobian (the model sensitivity) at the minimum of the cost function  $J(\mathbf{x})$ . This posterior error covariance matrix can be used to compute the posterior uncertainties of each parameter.

### c. Bias correction

Extra care is needed when using soil moisture observations or retrievals in DA due to potential large biases between the observed and the modeled quantities. When using the full SSM time series to calibrate the model, we will bias correct the observations using the prior model. There are a number of different approaches used to rescale the observations to deal with these biases. These range from methods which simply rescale to the mean value, to full cumulative distribution function matching (CDF matching) methods, which match the statistical moments of the observational data to the model ones, typically mean, variance, skewness, and kurtosis ([Reichle and Koster 2004](#)). Commonly used by the land surface community is the simplified CDF matching approach proposed by [Scipal et al. \(2008\)](#), which corrects for the first two moments, i.e., the mean and the variance. This is the method we will focus on in this study. Note that this method does not account for systematic differences between the seasonal cycles, and while we do acknowledge such methods exist and are potentially more robust (e.g., [de Rosnay et al. 2020](#)), they are beyond the scope of this study.

All bias correction methods directly eliminate some of the information from measured data. Fortunately, metrics based solely on the temporal variation of the soil moisture tend to be little affected by bias correction: consider the following linear bias correction formula, where  $\xi$  is the modeled time series toward which we wish to rescale the observed time series  $\theta$ . The rescaled  $\theta'$ , where the prime denotes the rescaled value, can be written as

$$\theta'(t) = \frac{\sigma_\xi [\theta(t) - \bar{\theta}]}{\sigma_\theta} + \bar{\xi}. \quad (4)$$

Note that  $\sigma_\theta$  and  $\sigma_\xi$  denote the standard deviation of each dataset and  $\bar{\theta}$  and  $\bar{\xi}$  denote their mean values. By combining this definition with the drydown equation in Eq. (1) and rearranging terms, we see that  $\tau$  remains invariant under this bias correction formulation since the other parameters of the exponential equation absorb the changes:

$$\frac{\sigma_\xi [\theta(t) - \bar{\theta}]}{\sigma_\theta} + \bar{\xi} = \theta'_A \exp\left(\frac{t}{\tau'}\right) + \theta'_{\text{eq}}, \quad (5)$$

$$\theta(t) - \bar{\theta} = \frac{\sigma_{\theta}}{\sigma_{\xi}} \theta'_A \exp\left(\frac{t}{\tau}\right) + \frac{\sigma_{\theta}}{\sigma_{\xi}} (\theta'_{\text{eq}} - \bar{\xi}), \quad (6)$$

$$\theta(t) = \underbrace{\frac{\sigma_{\theta}}{\sigma_{\xi}} \theta'_A}_{\theta_A} \exp\left(\frac{t}{\tau}\right) + \underbrace{\frac{\sigma_{\theta}}{\sigma_{\xi}} (\theta'_{\text{eq}} - \bar{\xi})}_{\theta_{\text{eq}}} + \bar{\theta}. \quad (7)$$

This relationship can be affected if the bounds on  $\theta_{\text{eq}}$  do not allow this absorption. Fortunately, we have found this not to be the case in our study when validating this relationship at the in situ sites. We find it holds to within an order of  $10^{-4}$ , which can be explained by numerical rounding errors in the nonlinear least squares fitting algorithm.

*d. Additional metrics*

In this section, we introduce a few more statistical diagnostics used in the work, as well as outlining how these different metrics will be applied in this study.

In the first part of this work, we identify drydowns over a range of different sites. To partition the different sites, we use an aridity index defined as  $A = R_n/(\lambda P)$  (Budyko 1961) where  $\lambda$  is the latent heat of vaporization ( $\text{J kg}^{-1}$ ),  $R_n$  is the mean daily net radiation ( $\text{W m}^{-2}$ ), and  $P$  is the mean daily precipitation ( $\text{kg m}^{-2}$ ). The FLUXNET data are used to calculate this index; higher values indicate greater aridity. While we appreciate that there are more complex metrics out there to measure aridity, this simple metric is sufficient in partitioning the sites by climate.

The second part of the study is devoted to the calibration experiments. Prior to performing these experiments, we remove outliers from the calibration to ensure that these data points do not dominate the cost function. To find these outliers, first we consider the discrepancies between the observed and prior modeled values, where the values are either  $\tau$  or SSM daily values depending on the time series considered. We then calculate the lower ( $Q_1$ ) and upper quartiles ( $Q_3$ ) of this discrepancy time series. An outlier is defined as a data point where the discrepancy between the observed and prior modeled value is outside the range:

$$[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]. \quad (8)$$

Finally, when evaluating the calibration results, we use two further metrics to assess the fit of the model to the observations. These are the commonly used root-mean-square error (RMSE) and its variant, the unbiased RMSE (ubRMSE; Entekhabi et al. 2010a)

$$\text{ubRMSE} = \sqrt{E\{(\theta_{\text{est}} - E[\theta_{\text{est}}]) - (\theta_{\text{true}} - E[\theta_{\text{true}}])\}^2}. \quad (9)$$

This latter metric will be used when assessing the model–data fit between observed and modeled SSM due to the large biases between the two.

*e. Performed experiments*

In this study ORCHIDEE outputs are aggregated at daily time steps (from half-hourly data) over 18 of the ISMN sites

found within the footprint of FLUXNET towers (see section 2). The FLUXNET data are used to provide the meteorology at each site and ancillary files are used to prescribe vegetation cover and soil textures. A spinup procedure is applied to each site brings the prognostic variables including vegetation state, soil carbon pools, and soil moisture content to equilibrium. This is achieved by running the sites for several hundred years cycling the available meteorological forcing with present-day  $\text{CO}_2$  concentrations. Each spinup is followed by a historical simulation over the period of interest (i.e., column “Years” in Table 2) creating the prior runs. The prior runs are what we call the standard model outputs generated using the default parameter values of ORCHIDEE (listed in Table 1). These are the standard model runs prior to the calibration experiment. At each site, to calculate the simulated SSM in the prior runs, we sum the SSM values generated in the upper layers of the model down to the ORCHIDEE layer that matches most the observation sensor depth. Furthermore, we remove frozen soils and snow cover conditions from our analysis by masking parts of the time series for which the soil temperature is below  $0^\circ\text{C}$ . Drydowns do not occur when the soil is frozen, so this does not affect our results. The SSM time series are then used to identify drydowns in the ORCHIDEE land surface model.

For the sensitivity analysis, the 31 parameters listed in Table 1 are tested using 10 trajectories following the Morris algorithm, amounting to 320 model runs. The most sensitive parameters were kept for the optimizations. This resulted in 13 parameters: the 10 describing the soil hydrology plus 3 others from the carbon and energy parameters. Two optimizations were carried out at each of the 18 sites considered in the study:

- **Opt\_ $\tau$** : Time series of observed and modeled  $\tau$  values were created using the true in situ observations (i.e., no bias correction) and prior model. The length and variation of these time series are given in Table 2, and two example  $\tau$  time series are shown in the appendix. The difference between these two time series was considered and outliers were removed using Eq. (8). These amounted to 6% of all  $\tau$  over all sites to be rejected (on average one per site). The modeled  $\tau$  were then optimized against the observed  $\tau$  time series, using the observed  $\tau$  as the target.
- **Opt\_fullSSM**: The full SSM time series were used in this optimization. Here, however, the observed SSM time series against which we calibrate was bias corrected using Eq. (4) toward the prior run. Then, in a similar manner to the  $\tau$  time series, outliers were removed before the calibration; this equates to between 0.2% and 3.5% of the observed values from the SSM time series for the majority of sites, and up to 14% of the observed values from the Savannah sites (i.e., AU-DaS, SD-Dem, and SN-Dhr).

Opt\_ $\tau$  is the main optimization of the study testing the potential of using the  $\tau$  values as the observation operator in a calibration. Opt\_fullSSM follows a more classical approach to optimization and is used to provide a comparison to the former method. However, note that since in each case we have opted for the most practical approaches (e.g., using diagonal **R** matrices) the optimality of both calibration strategies is not

TABLE 2. List of sites used in this study. The experiment years correspond to the overlap between the FLUXNET and ISMN data record. The latitude and longitude correspond to the location of FLUXNET tower, all of the ISMN are within approximately 250 m of the given tower. We will use the FLUXNET site code to refer to each site henceforth. The soil percentages represent the fraction of different soil textures for the first 0.3 m taken from the ISMN metadata, which in turn are taken from the Harmonized World Soil Database v1.1 (FAO/IIASA/ISRIC/ISS-CAS/JRC 2012) and the depth refers to the depth of the sensor used to measure the observations. The PFT codes are given in Table 1. The latter half of the table shows the number of dry periods (d.p.) and drydowns ( $\tau$ ) identified at each site. First, we show the total number of dry periods identified in the FLUXNET precipitation record, i.e., periods of at least 5 days without rain. Second, we show the number (and percentage) of these dry periods that can be fit using an exponential drydown function [Eq. (1)] in both the modeled and observed time series. The standard deviation of the observed and modeled  $\tau$  values is shown in the following column. Finally, the number of drydowns in the year with the most drydowns is selected (denoted rep. for representative year).

Site		Total No.									
FLUXNET	ISMN	Years	PFT	Lat (°), lon (°)	Depth (m)	Soil % (sand, silt, clay)	d.p.	$\tau$	Std dev $\tau$ (obs, mod)	No. $\tau$ (rep. year)	
AU-DaS	Daly <sup>a</sup>	2011–14	SAV	−14.16, 131.39	0.55	65, 10, 21	32	8 (25%)	2.7, 7.7	4 (2012)	
AU-Gin	Gnangara <sup>a</sup>	2011–14	WSA	−31.38, 115.71	0.69	89, 7, 4	58	6 (10%)	4.0, 7.5	2 (2011)	
AU-Rob	RobsonsCreek <sup>a</sup>	2014–14	EBF	−17.12, 145.63	0.19	48, 25, 28	19	1 (5%)	—	1 (2014)	
AU-Tum	Tumbarumba <sup>a</sup>	2011–14	EBF	−35.66, 148.15	0.14	53, 12, 34	74	2 (2.7%)	3.8, 3.9	1 (2011)	
DE-RuS	Selhausen <sup>b</sup>	2013–14	CRO	50.87, 6.452	0.19	36, 40, 28	35	12 (34%)	6.9, 3.6	7 (2014)	
FI-Sod	SOD013 <sup>c</sup>	2011–14	ENF	67.36, 26.64	0.05	88, 9, 4	25	20 (80%)	3.9, 3.8	7 (2013)	
SD-Dem	SD-DEM <sup>d</sup>	2008–09	SAV	13.28, 0.48	0.05	90, 6, 5	17	12 (71%)	6.1, 1.7	6 (2008)	
SN-Dhr	DAHRA <sup>e</sup>	2010–13	SAV	15.40, −15.43	0.05	90, 5, 6	29	13 (45%)	8.9, 4.3	4 (2011)	
US-ARM	ARM-1 <sup>a</sup>	2010–12	CRO	36.61, −97.49	0.19	29, 41, 29	34	16 (47%)	5.9, 4.2	6 (2011)	
US-GLE	GLEES <sup>a</sup>	2011–14	ENF	41.37, −106.24	0.10	33, 44, 32	18	17 (94%)	4.8, 4.4	8 (2013)	
US-Ha1	HarvardForest <sup>a</sup>	2011–12	DBF	42.54, −72.17	0.11	86, 10, 4	26	3 (12%)	1.2, 0.2	2 (2012)	
US-Me2	Metolius <sup>a</sup>	2011–14	ENF	44.45, −121.56	0.17	75, 12, 11	43	25 (58%)	5.5, 2.9	7 (2013)	
US-MMS	MorganMonroe <sup>a</sup>	2011–13	DBF	39.32, −86.41	0.09	40, 32, 31	62	5 (8.1%)	3.0, 1.1	3 (2013)	
US-Ne2	NebField3 <sup>a</sup>	2011–13	CRO	41.16, −96.47	0.13	30, 41, 34	41	17 (41%)	5.7, 3.8	6 (2012)	
US-PFa	ParkFalls <sup>a</sup>	2011–14	MF	45.95, −90.27	0.24	84, 11, 5	38	14 (37%)	3.8, 2.2	5 (2013)	
US-Ton	TonziRanch <sup>f</sup>	2001–12	WSA	38.43, −120.97	0.02	39, 29, 34	120	78 (65%)	10.3, 7.3	11 (2007)	
US-UMB	UMBS <sup>a</sup>	2011–14	DBF	45.56, −84.71	0.29	84, 11, 5	58	26 (45%)	5.3, 4.5	11 (2012)	
US-Whs	LuckyHills <sup>a</sup>	2012–14	OSH	31.74, −110.05	0.31	44, 29, 28	26	18 (69%)	4.5, 6.2	8 (2012)	
Total							755	293 (39%)	6.79, 5.81	99	

<sup>a</sup> COSMOS.

<sup>b</sup> TERENO.

<sup>c</sup> FMI.

<sup>d</sup> CARBOAFRICA.

<sup>e</sup> DAHRA.

<sup>f</sup> AMERIFLUX.

equally guaranteed, and so they are not perfectly comparable. For both optimizations, the first 70% of points are used for the calibration and the rest are used in the evaluation. We chose to select the first 70% of point instead of considering a set number of years due to the fact that the number of drydown events can vary year to year.

## 4. Results

### a. Identifying drydowns

In Table 3 we show the number of dry periods identified at each site and the percentage for which we were able to fit  $\tau$  for both the modeled and observed time series. In fact, for 89.76% of the dry periods considered, we were able to fit an exponential curve to the modeled time series. In comparison, we were only able to determine a  $\tau$  value for 42.95% of the dry periods identified in the observed time series and 39% both for the modeled and observed time series simultaneously. This is

due to a number of reasons including gaps in the observations, cases where the local meteorology is out of phase with the FLUXNET data, and noise in the observed in situ SSM measurements. The observed soil moisture time series tend to be noisier than the modeled time series (Fig. 1). However, by requiring the coefficient of determination ( $R^2$ ) of the fitted exponential to be at least 0.7, we ensure the behavior of the drydown is captured by  $\tau$ . Since we start the drydown periods the day after a rainfall event has occurred (i.e., the first dry day), we occasionally miss the peak of the exponential slope. In some cases, missing this peak when fitting the observed time series could also be explained by time lags with local meteorology at the FLUXNET tower. In any case, missing the peak is not an issue since we are interested in stage II evapotranspiration, i.e., when evapotranspiration is limited by soil moisture availability. More importantly, by ensuring at least 5-day drydowns, we make sure we fit this part of the curve.

For the remainder of the study, we only consider drydowns that had a successful  $\tau$  fit to them in both observed and

TABLE 3. Identified  $\tau$  values partitioned by different properties of the measurement site including vegetation (PFTs—definitions explained in Table 2), aridity index, sand fraction, and sensor depth. The total number of  $\tau$  values for each category is shown by  $i$ ;  $\tau_m$  and  $\tau_o$  denote the modeled and observed  $\tau$  values, respectively; and  $\tilde{x}$  and  $\bar{x}$  denote the median and mean of each set. The last column in each case shows the RMSE between the modeled and observed  $\tau$  values.

PFT	$i$	$\tilde{\tau}_m$	$\tilde{\tau}_o$	$\bar{\tau}_m$	$\bar{\tau}_o$	RMSE ( $\tau_m, \tau_o$ )
Vegetation type						
DBF	28	7.64	4.24	8.28	6.56	7.99
EBF	2	20.87	20.19	20.87	20.19	17.43
ENF	22	5.17	5.69	6.52	7.68	7.09
WSA/OSH	21	9.66	8.32	12.13	11.46	12.93
SAV	14	5.56	8.23	7.0	14.67	15.52
CRO	12	8.89	8.24	9.8	11.17	9.31
Aridity index						
0–1	17	4.78	4.29	6.9	5.28	7.81
1–2	38	7.93	5.15	8.76	8.16	7.91
2–3	18	5.81	9.87	7.44	14.35	11.8
3–4	8	5.74	22.62	11.75	20.97	20.87
4–5	12	10.31	5.19	11.44	6.87	11.1
Sand fraction						
0.2–0.4	41	6.05	5.05	7.05	9.75	8.96
0.4–0.6	13	10.96	3.51	13.12	8.03	12.92
0.6–0.8	11	7.07	6.6	9.53	9.71	10.63
0.8–1.0	37	7.09	7.95	9.0	10.46	11.73
Sensor depth (m)						
0.0–0.2	63	5.95	6.56	7.73	9.89	9.77
0.2–0.4	30	8.95	7.14	9.77	10.0	12.06
0.4–0.6	4	9.44	5.96	11.53	5.55	9.09
0.6–0.8	2	30.44	13.92	30.44	13.92	19.07

simulated time series since we want to be able to compare and later calibrate our model. For each site, we identify a number of drydowns varying in length. The longest drydown periods are found in arid and semiarid regions where there is a long dry season. These drydowns are associated with the largest  $\tau$  values (example drydown G in Fig. 1). On average, we find between 1 and 10 drydowns per year, with a mode of the distribution at 5. The number of drydowns is relatively uniform over the different climates, be they boreal or semiarid. For boreal sites, time series are shorter due to the masked snow-covered periods. Similarly, the semiarid sites experience less rainfall events but more clearly defined dry periods, which are more likely to be eligible in our analysis. The number of drydowns per year is comparable to McColl et al.’s (2017b) study.

*b. Comparing observed and modeled drydowns*

The  $\tau$  values for both the observed and the modeled time series tend to be similar in value. Nevertheless, we do observe some differences. For example, the observations can suggest a more gradual decrease in moisture than in the model (e.g., D in Fig. 1a), i.e., a shallower slope where the water is preserved longer in the soils. We also find opposite cases where soils dry out faster in the observation than in the model (e.g., F in Fig. 1b). Big disparities between observed and modeled  $\tau$  values can be used to identify different reactions to rainfall

events. For SD-Dem (Fig. 1a), we note that the model is more sensitive to rainfall events, increasing the amount of SSM when the observations display no change. For two examples in Fig. 1, model–observation difference is mostly of the same sign for each site over all displayed drydowns. However, in Fig. 2 we see that even at the same site, the model can dry out faster than the observations for some drydowns and slower for others.

We observe a greater spread in  $\tau$  values from the observations than from the modeled time series (Fig. 2). This is also visible when considering the variation of  $\tau$  at a given site (Table 2); the range of  $\tau$  values calculated at a given site is larger for the observed SSM time series than the modeled SSM time series. The standard deviation of observed  $\tau$  is the largest for savannah sites, where there can be a large  $\tau$  value associated with long dry periods and smaller  $\tau$  values calculated during the rainy seasons. This trend is not observed for the modeled  $\tau$ .

In Fig. 2, we show all identified drydowns in a given representative year for all sites in this study. With 54% of  $\tau$  values found above the 1:1 line in Fig. 2, and the rest below, we see a mix of differences between the modeled and observed  $\tau$  values. The model generally dries out slower than the observations for longer dry periods (78% of  $\tau$  fit over dry periods exceeding 20 days), however, for the longest periods (lasting nearly half a year) the model dries out faster. The model also tends to dry out faster than the observations for shorter dry periods (60% of  $\tau$  values fit over dry periods less than 20 days).

In Table 3, we further show these drydowns classified according to vegetation cover (PFT), soil texture (% of sand present), climate (through the aridity index), and sensor depth used at the measurement site, as listed in Table 2. First consider Table 3, where identified  $\tau$  values are split by vegetation type. Omitting the  $\tau$  values found over the EBF sites since sample size is too small over this vegetation type for a proper analysis, we see that for the observed  $\tau$  values, the forest sites (DBF and ENF) have slightly lower values than the grass/savannah sites, by a few days. However, this is not observed in the modeled  $\tau$  values resulting in higher RMSE over the grass/savannah sites than over the forested ones. In contrast, Martínez-de la Torre et al. (2019) found that tree sites did yield higher values of  $\tau$  than the grass sites. However, by considering evapotranspiration from the FLUXNET towers, Martínez-de la Torre et al.’s (2019)  $\tau$  values apply to the root zone, whereas, in our study, we consider SSM from the ISMN and so our recession times apply to the SSM. Since McColl et al. (2017b) also considered drydowns in the SSM and found no trends between  $\tau$  values and vegetation types, it is likely that the influence of vegetation on recession times is more important in the root zone than over the first few centimeters of the soil column.

There appears to be an impact of sensor depth on the value of  $\tau$ . Sites where the measurements are taken at larger depths have larger  $\tau$  values, meaning the soil takes longer to dry out. However, given the small sample size of sites and indeed drydowns at the large depths, these findings are not statistically significant.

The decay parameter  $\tau$  does not appear correlated to soil type, consistent with Martínez-de la Torre et al. (2019) but not

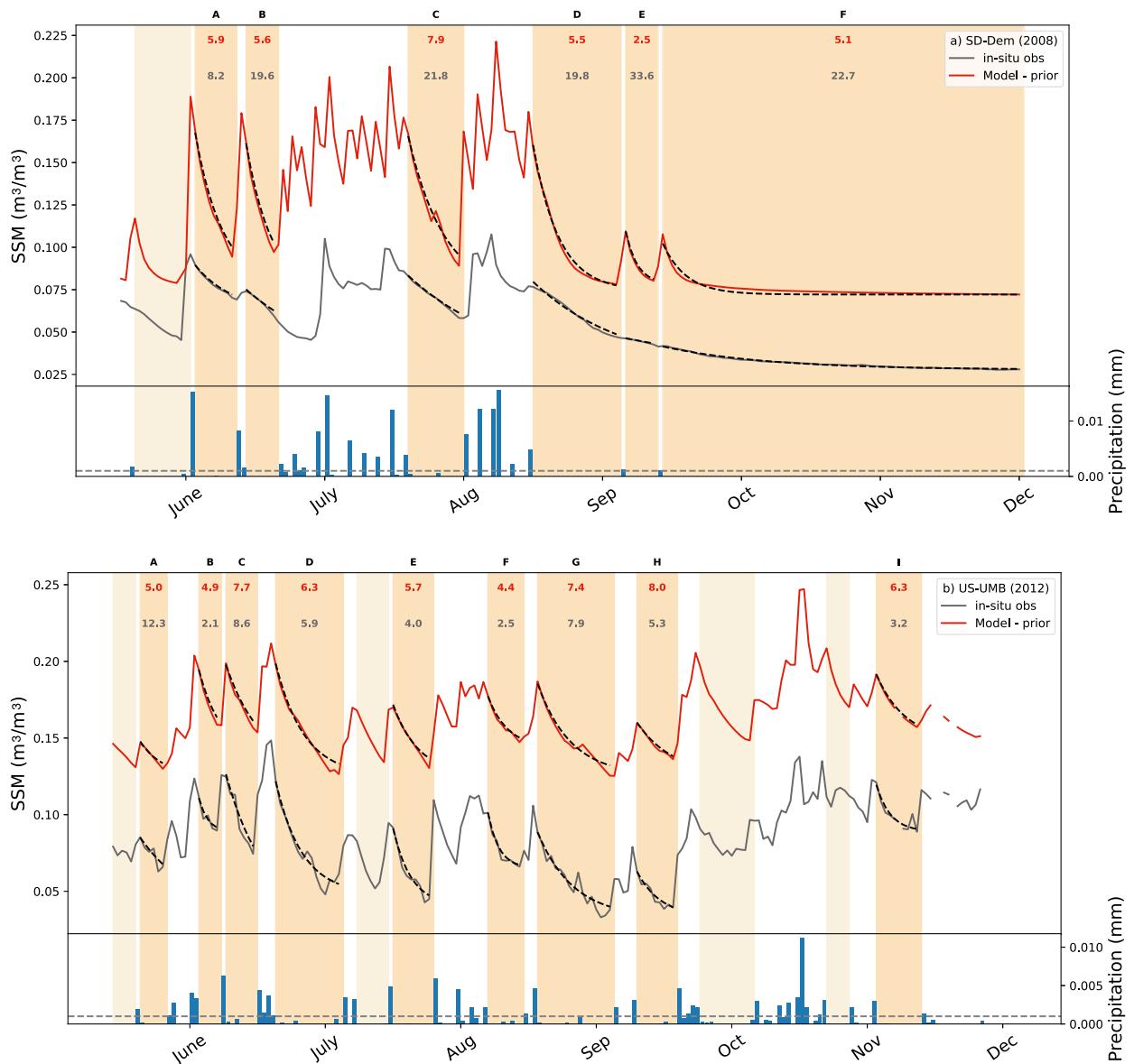


FIG. 1. Examples of surface soil moisture (SSM) temporal evolutions used to identify drydowns at (a) an arid site in the Sudan (SD-Dem) and (b) a temperate site in the United States (US-UMB). In each case, raw time series are shown (no bias correction), with rainfall shown in the bottom panel and the SSM time series for the in situ observations (gray) and modeled values (red) shown in (a). Shaded in orange are periods identified as having at least 5 days without significant rainfall ( $<0.001$  mm shown by the dotted gray line). The periods retained are shown by a darker shade of orange. Values of  $\tau$  (days), which determine the shape of each drydown [see Eq. (1)], can be found at the top of each drydown, and the exponential fit to each drydown shown by a dotted black curve. Each drydown is labeled with a letter to help cross referencing.

with McColl et al. (2017b). The latter found a clear decrease in drydown rates with increasing sand fraction explained by sandy soils having larger pores with lower suction caused by surface tension forces and therefore releases water more readily to the atmosphere via transpiration. In addition, McColl et al. (2017b) found a clear anticorrelation of  $\tau$  with aridity, whereas the present modeled values show an opposite trend, namely, a positive correlation. These discrepancies between McColl et al. (2017b) (treating SMAP retrievals) and our study (treating

in situ observations) may be due to the different types of observational data considered, especially since satellite data have been shown to dry out faster than in situ data. It is also due to the different sample sizes used to determine these relationships; by using satellite retrievals, McColl et al. (2017b) consider a much larger set of sites from which to draw conclusions. Overall, due to the limited number of eligible sites and drydowns in our study (only 18 sites covering a range of vegetation, climates, and soil textures), we cannot generalize to a global scale the (de)

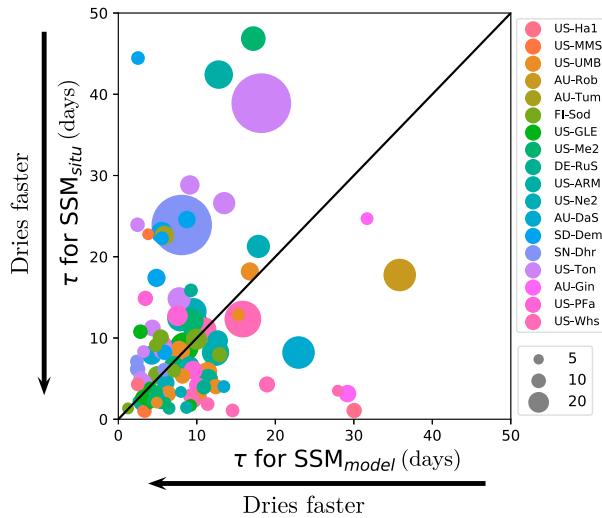


FIG. 2. Scatterplot showing the modeled vs observed  $\tau$  values. Each  $\tau$  value is associated with a drydown identified during a representative year for each of the 18 sites used in this study. The length (in days) of each drydown period (i.e., number of days with no rain) over which the exponential is fitted is represented by the size of the dot.

correlations between  $\tau$  and the classification categories found here without more data.

c. Identifying key parameters

Parameters used in the upcoming ORCHIDEE calibration were selected using a Morris sensitivity experiment (Fig. 3). Unsurprisingly, we find that the  $\tau$  values are sensitive to most of the soil hydrology parameters, notably to  $n$  and, to a lesser extent, to  $A$ . These parameters control the water retention curve and so directly influence the water content of the soil. The only soil hydrology parameter for which we observe no sensitivity is  $sl_R$ . This parameter controls the fraction of runoff

which could be stored at the surface and reinfiltreated later in flat terrain conditions. It is likely that after the precipitation events no runoff was generated over the flat sites considered in this study. Furthermore, this parameter is only important for one time step after the precipitation event and so will not greatly affect the shape of the drydown.

All  $\tau$  values over all sites show a strong sensitivity to the phenology parameters such as  $V_{cmax}$  and SLA, especially at the boreal site. The size of the leaf will determine transpiration rates which are intrinsically linked to soil moisture. The savannah site shows the least sensitivity to the phenology parameters between all of the sites, however, this may be more linked to a poor representation of the complex phenology in savannah regions in the ORCHIDEE model.

Finally, we also find that  $\tau$  is highly sensitive to the soil resistance to evaporation, since it controls the rate at which soil moisture is directly transferred to the atmosphere. We therefore choose to optimize over the 10 key soil hydrology parameters, plus  $V_{cmax}$ , SLA, and  $rs^*$  since these are the highest ranked nonhydrology parameters over the three sites. These identified parameters are listed in Table 1.

d. Optimization results

1) CHANGES TO SIMULATED  $\tau$  AND SSM

The first optimizations focus on calibrating all  $\tau$  at each given site (Fig. 4a). In this Opt\_ $\tau$  calibration, the RMSE is reduced at all sites. For sites with high prior errors like AU-Rob, AU-DaS, FI-Sod, and US-Whs, the percentage reduction in RMSE is over 80%. Sites which have a smaller initial error, like US-Ha1, US-MMS, and US-Me2 do not improve as much, showing a reduction closer to 10%.

This behavior is generally mirrored when the sites are confronted by independent data, i.e., the evaluation drydowns. Sites experiencing a large reduction during the calibration show a similarly large reduction in the evaluation. For sites with a small reduction in the calibration (e.g., US-Me2, US-MMS), the fit during the evaluation remains the same, or, in some cases (e.g.,

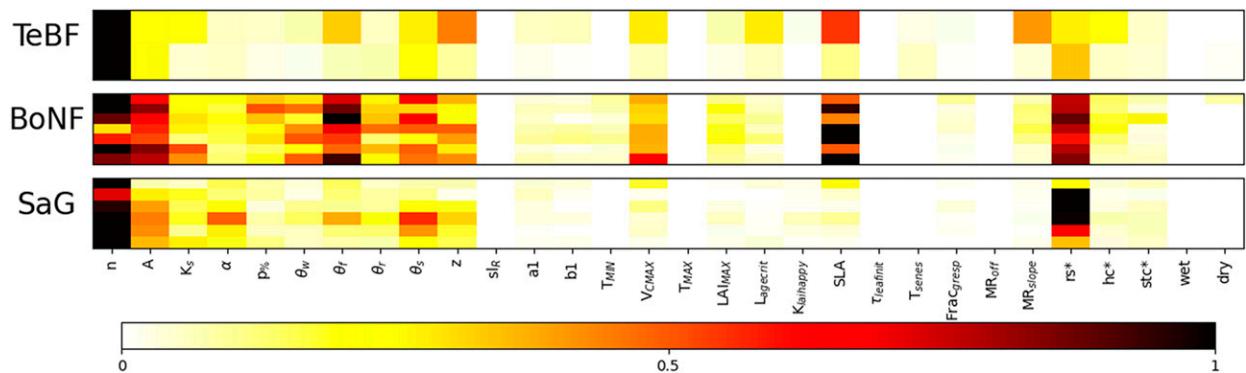


FIG. 3. Morris sensitivity scores used to identify the key parameters to which the shapes of drydowns are sensitive are shown in this figure. Three sites covering a range of vegetation and climate are shown: (top) temperate broadleaf forest (TeBF; US-Ha1), (middle) boreal needleleaf forest (BoNF; FI-Sod), and (bottom) semiarid grassland (SaG; SD-Dem). For each site, the year with the most drydowns was selected and the response of each drydown during that year is shown here. For each drydown, the sensitivity scores are normalized to [0, 1] with 1 as most sensitive and 0 as least sensitive. Full descriptions of each tested parameter and site can be found in Tables 1 and 2, respectively.

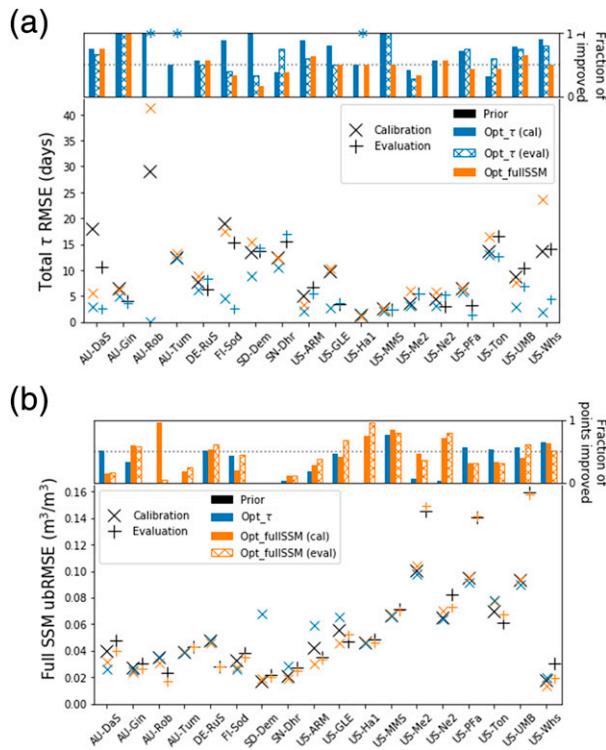


FIG. 4. (a) Changes to the modeled  $\tau$  values and (b) changes to the full modeled SSM time series. In each case, the top panels show the fraction of individual events that are closer to the observed values, where for (a) the events are the drydowns and for (b) the events are the daily SSM values. At 0, all modeled events are farther from the observations than before and at 1, all modeled events were improved. The bottom panel shows the RMSE and ubRMSE for all the  $\tau$  and raw (i.e., nonbias corrected) SSM values, respectively. In each case, two parts of the time series are considered: a calibration part (first 70% of the time series) and an evaluation part (last 30% of the time series). This is with the exception of three sites in the  $\tau$  experiment where, since there were less than three drydown events total, the full time series was used for the calibration and no events were saved for the evaluation. These are marked by an asterisk.

DE-RuS, US-Ne2), degrades slightly. One notable exception is US-GLE, which has a large reduction in RMSE during the calibration but a slight degradation during the evaluation. For this site, the prior RMSE of the evaluation set starts very low. In contrast, for US-Ton, the calibration does not improve much but the evaluation shows a marked improvement. In this case, the prior RMSE of the evaluation set is larger and the posterior RMSE is similar to that of the calibration set prior.

In total, 11/15 of the sites improve, or remain the same during the evaluation. Overall, counting all sites, we observe a total reduction of 48% in RMSE during the calibration and a total reduction of 27% during the evaluation. This is compared to a 3% increase in total RMSE when using the Opt\_fullSSM parameters on the calibration set.

At least half of the  $\tau$  values in the time series get closer to the observed estimates for 15/18 sites during the calibration and for 11/15 sites during the evaluation (Fig. 4a, top). For the latter,

at a handful of sites, this result does not correspond to the observed change in total  $\tau$  RMSE. DE-RuS and SN-Dhr show a degradation in fit to all  $\tau$  taken together, but improve over half of the drydown events evaluated against. FI-Sod, on the other hand, shows one of the largest reductions in total RMSE but improves only 40% of the individual events. This is due to one or more drydown events dominating the cost function.

For 10 out of the 18 sites, the Opt\_fullSSM calibration is also able to improve at least half the recession events. The fraction of  $\tau$  values improved is not as high as in the Opt\_ $\tau$  optimization. This is to be expected since the Opt\_fullSSM optimization calibrates other parts of the time series as well as the drydowns whereas the Opt\_ $\tau$  calibration focuses only on the drydown values. However, even though at most sites approximately half the  $\tau$  values are improved, the total  $\tau$  RMSE generally increases, or stays the same, meaning an overall degradation, or no change, in matching the modeled  $\tau$  values to the observed ones. Only 7/18 sites show an improvement and only 4 of these show an improvement greater than 4%, i.e., AU-DaS, FI-Sod, US-ARM, and US-UMB. The largest degradations can be found over the sites where the Opt\_ $\tau$  calibration improved the most i.e., AU-Rob and US-Whs. However, in the case of AU-Rob, only one  $\tau$  is being optimized so the Opt\_ $\tau$  optimization overcalibrates to the specifics of that one drydown. To be more robust, it would be better to ensure that time series of  $\tau$  values have more events. This highlights one of the issues using a short data record, with only one year of in situ data, the number of identified drydown events against which to calibrate is limited. A short data record is also an issue when using Opt\_fullSSM since multiple years are needed to properly calculate and remove the mean-state bias. Moving to satellite retrievals will increase the length of data records with which we can confront the model.

Overall we have shown the ability to optimize drydowns. To improve drydowns, it is better to optimize  $\tau$  only (Opt\_ $\tau$ ) than to use whole time series (Opt\_fullSSM). Optimizing the full time series can still improve drydowns in some cases but to a lesser extent; in most cases, however, the total  $\tau$  RMSE remains close to the prior showing little change.

In Fig. 4 we also consider the effect of Opt\_ $\tau$  and Opt\_fullSSM on the whole non-bias-corrected SSM time series at each site using the ubRMSE diagnosis. Overall for Opt\_fullSSM, the total reduction over all sites of SSM ubRMSE is of 7.8% during the calibration and 8.3% during the evaluation. In contrast, for Opt\_ $\tau$ , the total reduction of the calibration set is only 3.1% and we observe significant degradation for four of the sites (i.e., SD-Dem, SN-Dhr, US-ARM, and US-GLE). However, we do also observe cases where the Opt\_ $\tau$  outperforms the Opt\_fullSSM (e.g., AU-DaS and US-UMB).

Although Opt\_fullSSM does not show any severe degradations in fit, the improvement is limited with many sites remaining close to unchanged during both the calibration and evaluation stages. There are two main reasons why the Opt\_fullSSM results are mixed. First, when calibrating the model, the cost function uses a metric resembling an RMSE, whereas in the analysis, we are considering the ubRMSE. Second, prior to the Opt\_fullSSM calibration, the observations

were bias corrected. In Fig. 4b, however, the raw observations are used. At sites where the Opt\_fullSSM calibration does not perform as well as the Opt\_ $\tau$  experiments, information for SSM improved modeling is lost when the optimization is performed on bias-corrected SSM observations. This emphasizes further an advantage of the Opt\_ $\tau$  approach: by removing the need for bias correction in the Opt\_ $\tau$  calibration, the assimilation valorizes the information carried by the first two moments of the observations. Alternatively, the Opt\_fullSSM calibration might be improved with more complex and more robust methods when rescaling the data to deal with biases; methods such as the inclusion of seasonality or using longer data periods to calculate the mean state.

For half of the sites, the fractions of daily points in the SSM time series that are improved by the Opt\_ $\tau$  optimization are much lower than in the Opt\_fullSSM calibration. This is to be expected, since the full SSM time series is made up from more than just drydown events. Similarly, the fraction of daily points improved overall during either optimization is less than the fraction of the drydown events improved. This is because there are significantly more points in the full time series (365 per year) compared to the drydown time series (closer to 5 points a year).

Overall, the SSM ubRMSE remains close to the prior for a number of sites after to Opt\_fullSSM assimilation, whereas the results using parameters from the Opt\_ $\tau$  assimilation are mixed for this diagnosis. This shows that different parameters or even the structure of the code may need to be considered in order to minimize this error. Since we performed a sensitivity analysis to find the most sensitive parameters to use during the optimization, it is more likely that model structural errors are at play and that parameterizations of different soil moisture processes need to be reassessed. Another reason for irreducible  $\tau$  discrepancies could be related to the forcing data: local meteorology may slightly differ between the SSM measurement sites and their neighboring FLUXNET towers.

## 2) OPT\_ $\tau$ POSTERIOR PARAMETER VALUES

In addition to considering the fit between model and observations, it is important to assess the effect of calibration on the parameters themselves (Fig. 5). The volumetric water content parameters, which provide thresholds for the simulated SSM time series, show little change after optimization but do have a high error reduction in posterior parameter uncertainty. We find that the van Genuchten parameters  $A$  and  $n$  also yield a relatively high error reduction. However, parameter uncertainty is not greatly reduced for the hydraulic conductivity at saturation  $K_s$ , nor is it greatly reduced either for the  $V_{cmax}$  parameter.

We generally find that soil resistance to evaporation—through parameter  $rs^*$ —is increased after assimilation. Since we found that the model generally dries out faster than the observations, this result is consistent with our prior analysis. We also find  $rs^*$  to be the parameter with the highest reduction in error highlighting its importance in accurately modeling drydowns.

The value of root profile  $z$  is reduced for most sites, especially for deciduous forests. This again may be linked to the fact

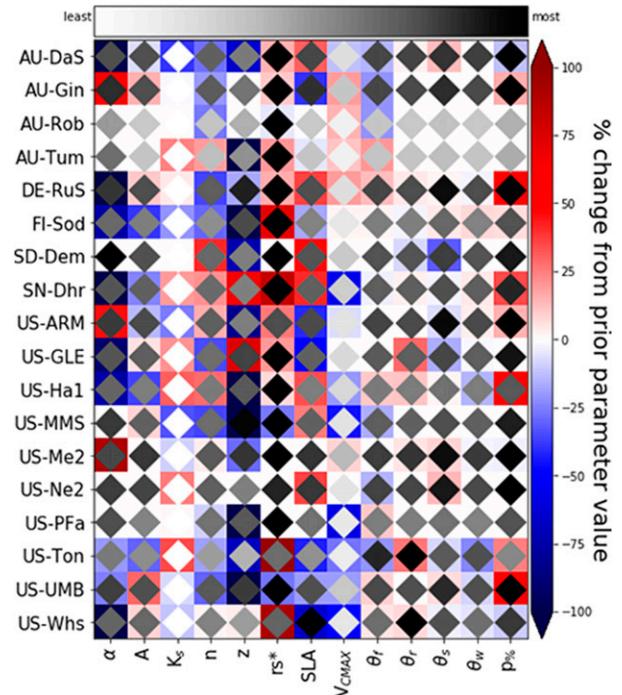


FIG. 5. Posterior parameter values and their uncertainties for the Opt\_ $\tau$  assimilation for each of the sites in this study (parameter and site descriptions can be found in Tables 1 and 2, respectively). The outer color (from blue to red) signifies how the parameter value has changed as a percentage from its prior value. The inner color represents the posterior uncertainty of that parameter, where the greater the reduction in error, the darker the diamond. This is calculated using prior minus the posterior error all divided by the prior error and is normalized for each site.

that the model was drying out faster than the observations. By reducing the root profile, the roots do not reach as deep in the soil and therefore have access to less water. More water is kept in the root zone and transferred to the surface layers by vertical diffusion (because of the larger hydraulic conductivity). For sites with predominant grass cover, the prior SLA parameter was too low whereas for the boreal evergreen forests, this parameter is set too high. The SLA parameter directly affects the LAI and thus the transpiration. Finally, the water stress parameter  $\alpha$  also comes out as one of the most important parameters. For half of the sites,  $\alpha$  is reduced suggesting that the plants are less sensitive to water stress at these sites than currently modeled in ORCHIDEE.

By calibrating drydowns, since we are directly considering one feature of the SSM time series, there is potential to better understand the processes involved and to pinpoint exactly how the parameters affect the processes. For example, the suggested changes to  $rs^*$  parameter highlight the importance of this parameterization in modeling drying rates in the soil.

## 3) IMPACT OF OPT\_ $\tau$ ON CARBON AND ENERGY FLUXES

There is a strong coupling between water and carbon cycles, and a number of parameters optimized in this study will affect

TABLE 4. Relative change of flux RMSE after calibration on  $\tau$  and full SSM. Bold indicates improved or unchanged RMSE. Values higher than 1 imply a degradation.

Site	Opt_ $\tau$			Opt_fullSSM		
	GPP	Resp	LE	GPP	Resp	LE
AU-DaS	1.24	1.23	1.35	1.32	1.18	1.1
AU-Gin	1.04	1.03	<b>0.9</b>	1.02	1.04	<b>0.91</b>
AU-Rob	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>	1.16	1.07	1.17
AU-Tum	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>	1.03	<b>0.97</b>	2.37
DE-RuS	1.01	1.02	1.04	1.06	1.06	1.21
FI-Sod	1.04	1.02	<b>0.97</b>	3.38	1.9	2.23
SD-Dem	1.44	1.6	1.4	2.21	1.27	1.18
SN-Dhr	1.07	<b>0.99</b>	1.03	1.1	1.03	1.05
US-ARM	1.06	1.07	<b>0.91</b>	1.01	1.04	<b>0.96</b>
US-GLE	1.01	1.01	1.04	2.93	1.63	1.62
US-Ha1	<b>0.99</b>	<b>0.99</b>	<b>0.97</b>	<b>0.79</b>	1.13	<b>0.94</b>
US-MMS	1.05	1.14	1.05	1.55	1.19	<b>0.82</b>
US-Me2	<b>0.94</b>	0.97	0.92	1.08	1.02	1.25
US-Ne2	<b>0.98</b>	1.01	1.0	1.1	1.14	1.16
US-PFa	1.07	1.06	1.2	1.25	1.13	0.73
US-Ton	<b>0.99</b>	<b>0.99</b>	<b>0.98</b>	<b>0.97</b>	<b>0.99</b>	<b>0.9</b>
US-UMB	1.05	1.08	1.94	1.15	1.18	<b>0.95</b>
US-Whs	<b>0.97</b>	<b>1.0</b>	<b>0.59</b>	<b>0.74</b>	<b>0.93</b>	<b>0.92</b>

other processes in the model. Therefore, it is important to consider the impact of calibrating the model using solely hydrologic fluxes on the other fluxes. In this section, we compare the fluxes (LE, GPP, Resp) from the calibrated model to the FLUXNET data (see Table 2) that are in the vicinity of the SSM measurements.

For over half of the sites, using the parameters from Opt\_ $\tau$  assimilation maintains or improves at least one of the three investigated fluxes (Table 4), particularly the evapotranspiration LE which is more related to hydrologic variables. All the fluxes are also improved for a handful of sites when using the Opt\_fullSSM results. However, for other sites the fit is degraded by up to a factor between 2 and 3 (e.g., AU-Tum, FI-Sod, SD-Dem, and US-GLE). The amplitude of such a worsening is reduced when it happens after Opt\_ $\tau$  assimilation.

The time series themselves (Fig. 6) show that the Opt\_ $\tau$  optimization induces minor changes, whereas Opt\_fullSSM assimilation induces more significant ones. For example, at FI-Sod site, the Opt\_fullSSM parameters lead to a decrease in LAI which in turn reduces the presence of vegetation. In contrast, the fluxes hardly change when the Opt\_ $\tau$  parameters are used, even though these parameters resulted in one of the greatest RMSE reductions when fitting the drydowns. For the SD-Dem site, the production peak is delayed when using the Opt\_fullSSM parameters whereas for Opt\_ $\tau$ , it is the magnitude of the peak which is reduced.

Overall, the changes to the fluxes are minimal for the Opt\_ $\tau$  optimization. However, this is encouraging. With a complex process-based model like ORCHIDEE, it is likely that improving some parts of the model may degrade other simulated quantities because of potential compensating errors. The fact that Opt\_ $\tau$  does not significantly degrade the fit against these fluxes, and for some sites Opt\_fullSSM does, is an important result. This suggests that the Opt\_fullSSM optimization may lead to

suboptimal parameters. It should also be noted that the ORCHIDEE model has been previously optimized, calibrating these fluxes using FLUXNET data from a number of the same sites (Kuppel et al. 2014). Therefore the initial fit to the observations was already very good (with the exception of savannah sites like US-Whs, which remains poorly represented by the model).

## 5. Discussion

### a. Scaling up to a global calibration

The ORCHIDEE LSM is a global model and a component of a larger ESM. It can be run at different resolutions, from individual pixels, as in this study, to small catchment areas up to fully global runs. As such, it is important to be able to calibrate the model at a larger scale covering a greater selection of vegetation, climates, and soil types. The global coverage satellite products can provide an opportunity to calibrate the model at this scale. However, extra care will be needed when considering satellite data since soil moisture memory metrics are dependent on the sampling frequency of the observations (McColl et al. 2017b; Ruscica et al. 2020). If  $\tau$  is less than the revisit time of the sensor, then the drydown will not be captured. Combined products such as ESA CCI SM (v4.4; Dorigo et al. 2017) do provide daily SSM values, by merging different satellites. However, such a product runs the risk of introducing conflicting sources of information, since the merge currently depends on an auxiliary model. Alternatively, if we use raw products like SMOS SSM or SMAP we will have less coverage, possibly less information content, and the risk of missing fast drydowns from our analysis.

We briefly tested our drydown identification approach on the ESA CCI SM (v4.4) satellite product but could only identify a handful of drydowns using the FLUXNET precipitation data. Instead we would need to use a precipitation product of the same resolution. During our previous study (Raoult et al. 2018), we found that the ESA CCI SM product and soil moisture simulated in ORCHIDEE using the satellite-era Coupled ECMWF Reanalysis (CERA-SAT) forcing data (Schepers et al. 2018) were highly correlated (with average correlations of 0.7 in tropics and southern latitudes). Such a dataset could be used to determine periods without precipitation; however, further study will be needed to really understand the comparable strengths and weaknesses of different reanalysis products.

There will be many advantages in moving to satellite data. With a lot more data, it should be easier to identify more trends linked to soil textures and types of climate. We will be able to include nonvegetated areas in the analysis, which currently are poorly represented in the FLUXNET database, to understand better bare soil evaporation. In addition, satellite SM observations represent only the top few centimeters of the soil column. Therefore we will be able to use a relatively consistent depth for the whole world matching the instruments theoretical global mean sensing depth of 5 cm.

We will also be able to consider the calibration of multiple data streams using a range of other satellite products. Simultaneous calibration of SSM drydowns in conjunction with other data streams will provide valuable insights into the links between the carbon, energy, and water cycles. Using, for example, LAI or

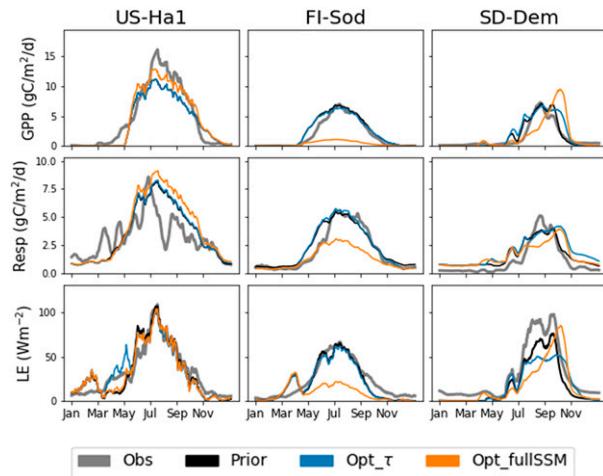


FIG. 6. Seasonal time series of (top) GPP, (middle) Resp, and (bottom) LE for three representative sites. In each case, the prior model run (black) and the posterior model runs from the Opt\_ $\tau$  assimilation (blue) and the Opt\_fullSSM assimilation (yellow) are compared to the FLUXNET observations (gray).

solar-induced fluorescence (SIF) data to constrain plant activity, and therefore transpiration, could help investigate the impact of droughts on vegetation. Alternatively, SSM drydowns could be calibrated alongside land surface temperature to constrain the energy budget. By including more data streams into the calibration, we ensure that the optimizations have less of a deteriorating effect on the other model outputs.

### b. Limitations

Our analysis is subject to several limitations. The first are linked to the methodology. Outliers—here defined by points which differ too much between the model and the observations—are currently removed in order not to dominate the cost function. A more robust way of dealing with outliers would be to have a dynamic observation error which would change at each time step. That way, outliers could be included in the analysis but with a higher uncertainty weight during the optimization.

For the optimizations, we have used a diagonal  $\mathbf{R}$  matrix. While this is commonly used in land surface model parameter estimation studies, this is not ideal since observations tend to be correlated. This will be especially true in the case of Opt\_fullSSM optimization since SSM observations are strongly autocorrelated, whereas the Opt\_ $\tau$  optimization deals with independent events. Fortunately, when checking the reduced  $\chi^2$  fit at each site we find that this metric is generally close to 1, meaning that we did not over/under use the information content brought by the observations in the calibration. Nevertheless, we did find that the calibration on  $\tau$  much more often leads to large  $\chi^2$  values, whereas calibration on SSM leads to small  $\chi^2$  values, highlighting the differences between both optimization strategies, especially with respect to the optimality of their setup. Alternatively, there are a number of methods that can be used to compensate for these correlations; for example, we can choose to inflate the variances (Chevallier 2007), or using diagnostics such as the one proposed by Desrozières et al. (2005) to help define the

off-diagonal elements. These methods, while out of the scope of this study, will be considered in future optimizations.

Second, there are limitations linked to the data used. The in situ measurements assessed in this study, while being grouped by the ISMN, use different probes and different measurement protocols which result in probes installed at different depths. ORCHIDEE's multilayered soil column has allowed us to match as closely as possible the sensor's depth. However, when using cosmic ray neutron sensors to measure soil moisture, as used by the COSMOS network, the sampling depth of the soil moisture measurement changes with soil moisture (Antoniu et al. 2019). Unfortunately, the estimate of this sampling depth is not provided in the ISMN database so instead we matched to the interval's maximal depth. If we had had access to this depth, we could have used a weighted depth approach to calculate the average soil moisture corresponding to the estimated sampling depth (as in Cooper et al. 2020). In comparison, the estimated sensing depth of satellite retrievals is more uniform and will be better for proper comparison and analysis between sites. Furthermore, the number of years used in this study was limited by the small overlap between the FLUXNET2015 and ISMN data records. For both types of optimizations considered, a longer data stream would be preferable. A longer record would provide a clearer seasonality over the sites, needed for bias correction (important for Opt\_fullSSM) and would ensure drydowns at different times of the year are captured (important for Opt\_ $\tau$ ).

Finally, compared to forest sites, which have one clear dominating ORCHIDEE PFT, savannah sites are not explicitly described in ORCHIDEE. As such ancillary files are used to describe the different vegetation fractions at these sites. Generally the largest fraction of a prescribed PFT is less than half (e.g., 40% of C4 crops). During the calibration, the parameters for this PFT are optimized but the rest of the vegetation remains unchanged. In future experiments, it may be necessary to do simultaneous calibration of the different types of vegetation present to improve all the different parameters.

## 6. Conclusions

We have shown that the drydowns in the ORCHIDEE LSM can be fitted using an exponential curve. This curve is characterized by  $\tau$ , which is invariant under simple bias correction algorithms as performed prior to calibration experiments. The  $\tau$  values from the model were compared to  $\tau$  calculated using in situ SSM data. The values were generally comparable but a greater range of  $\tau$  values were found in observed time series and overall the model was found to dry out faster than the observations. Calibrating an ensemble of parameters (13 parameters here covering a range of hydrology and photosynthesis parameters) with the objective to fit  $\tau$  was shown to be promising. At least half of the drydowns were improved at each site and the ubRMSE of the overall SSM time series was maintained or improved at half of the sites. In contrast, a standard calibration of the whole SSM time series was generally not able to improve drydowns. The latter calibration also had a more significantly deteriorating effect on noncalibrated fluxes such as GPP and LE, whereas these other fluxes were only

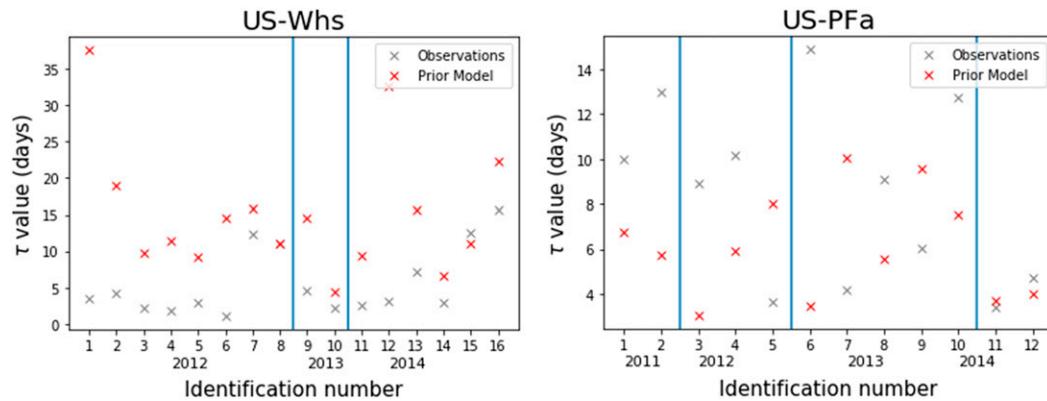


FIG. A1. Example of  $\tau$  events calculated at two contrasting sites.

minutely affected by the  $\tau$  calibration. Both optimizations, however, do not improve much the overall simulated SSM, with the  $\tau$  calibration performing less well than the full SSM optimization. Furthermore, there is scope to improve both calibration strategies, through for example the inclusion of off-diagonal elements in the  $\mathbf{R}$  matrices, which may change the results.

Both the absolute values of SSM and the drydown velocity (measured by  $\tau$ ) reflect integrated variables related to observation scale and support volume, soil texture, vegetation, and potential evaporation. While in situ derived  $\tau$  will still be biased relative to gridscale  $\tau$  values, as an intrinsic temporal metric, using  $\tau$ , however, removes the need of mean-state bias correction prior to comparison or model calibration. The objective of the ORCHIDEE model is not to represent the absolute values of SSM, which are not comparable to point measurements since, in most cases, these measurements are not representative of the model scale. Rather ORCHIDEE seeks to represent some physical behaviors that are crucial for simulating the temporal dynamics of model state variable responses (linked to rainfall and drought events). These results give us a potential avenue in using  $\tau$  observations to calibrate the model. The next step is to use this  $\tau$  metric at a global scale by optimizing against SSM satellite data and possibly simultaneously with other satellite products linked to the water and carbon cycles, such as surface temperature or vegetation optical depth.

**Acknowledgments.** N. Raoult is funded by the H2020 MULTIPLY project, and this research has been supported by the European Commission, Horizon 2020 Framework Programme (VERIFY, Grant 776810). This work used eddy covariance data acquired and shared by the FLUXNET community, including these networks: AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada, GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOS-Siberia, and USCCC. The ERA-Interim reanalysis data are provided by ECMWF and processed by LSCE. The FLUXNET eddy covariance data processing and harmonization was carried out by the European Fluxes Database Cluster, AmeriFlux Management Project, and Fluxdata project of FLUXNET, with the support of

CDIAC and ICOS Ecosystem Thematic Center, and the OzFlux, ChinaFlux and AsiaFlux offices. This work also used in situ soil moisture data acquired and shared the ISMN community, including these networks: AMERIFLUX, CARBOAFRICA, COSMOS, DAHRA, FMI, and TERENO. Finally, we thank the ORCHIDEE team for developing and providing the latest version of the ORCHIDEE model. We would also like to thank Clement Albergel and three anonymous reviewers for their helpful comments that improved the manuscript.

**Data availability statement.** The in situ and satellite data used in this study are freely available upon registration from their dedicated websites; FLUXNET2015 data (Pastorello et al. 2017) from <https://fluxnet.fluxdata.org/data/fluxnet2015-dataset/>, ISMN data Dorigo et al. (2011, 2013) from <https://ismn.geo.tuwien.ac.at/>, and ESA CCI SM data (Dorigo et al. 2017) from <https://www.esa-soilmoisture-cci.org/>. ORCHIDEE data are available on request.

## APPENDIX

### Drydown Time Series

Drydown events are independent. Since they are calculated from rainfall events, they can occur at regular intervals. Figure A1 shows events for two contrasting sites. The first example is over US-Whs, a semiarid site, where the observations dry out faster than the observations. The second example is for US-PFa, a temperate site, where the model dries out faster than the observations. These  $\tau$  values are used in the calibration—we used the observed values as the target for the modeled value.

## REFERENCES

- Antoniou, V., and Coauthors, 2019: COSMOS-UK user guide: Users' guide to sites, instruments and available data (version 2.10). NERC/Centre for Ecology & Hydrology, 58 pp.
- Bastrikov, V., N. MacBean, C. Bacour, D. Santaren, S. Kuppel, and P. Peylin, 2018: Land surface model parameter optimisation

- using in situ flux data: Comparison of gradient-based versus random search algorithms (a case study using ORCHIDEE v1.9.5.2). *Geosci. Model Dev.*, **11**, 4739–4754, <https://doi.org/10.5194/gmd-11-4739-2018>.
- Bolten, J. D., W. T. Crow, X. Zhan, T. J. Jackson, and C. A. Reynolds, 2010: Evaluating the utility of remotely sensed soil moisture retrievals for operational agricultural drought monitoring. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **3**, 57–66, <https://doi.org/10.1109/JSTARS.2009.2037163>.
- Bonan, B., C. Albergel, Y. Zheng, A. Barbu, D. Fairbairn, S. Munier, and J.-C. Calvet, 2020: An ensemble square root filter for the joint assimilation of surface soil moisture and leaf area index within the land data assimilation system LDAS-Monde: Application over the Euro-Mediterranean region. *Hydrol. Earth Syst. Sci.*, **24**, 325–347, <https://doi.org/10.5194/hess-24-325-2020>.
- Bonan, G. B., and L. M. Stillwell-Soller, 1998: Soil water and the persistence of floods and droughts in the Mississippi River basin. *Water Resour. Res.*, **34**, 2693–2701, <https://doi.org/10.1029/98WR02073>.
- Boucher, O., and Coauthors, 2020: Presentation and evaluation of the IPSL-CM6A-LR climate model. *J. Adv. Model. Earth Syst.*, **12**, 1–52, <https://doi.org/10.1029/2019MS002010>.
- Budyko, M. I., 1961: The heat balance of the Earth's surface. *Sov. Geogr.*, **2**, 3–13, <https://doi.org/10.1080/00385417.1961.10770761>.
- Campolongo, F., J. Cariboni, and A. Saltelli, 2007: An effective screening design for sensitivity analysis of large models. *Environ. Modell. Software*, **22**, 1509–1518, <https://doi.org/10.1016/j.envsoft.2006.10.004>.
- Chevallier, F., 2007: Impact of correlated observation errors on inverted CO<sub>2</sub> surface fluxes from OCO measurements. *Geophys. Res. Lett.*, **34**, L24804, <https://doi.org/10.1029/2007GL030463>.
- Cooper, E., E. Blyth, H. Cooper, R. Ellis, E. Pinnington, and S. J. Dadson, 2020: Using data assimilation to optimize pedo-transfer functions using large-scale in-situ soil moisture observations. *Hydrol. Earth Syst. Sci. Discuss.*, <https://doi.org/10.5194/hess-2020-359>.
- Crow, W. T., R. D. Koster, R. H. Reichle, and H. O. Sharif, 2005: Relevance of time-varying and time-invariant retrieval error sources on the utility of spaceborne soil moisture products. *Geophys. Res. Lett.*, **32**, L24405, <https://doi.org/10.1029/2005GL024889>.
- De Lannoy, G., and R. Reichle, 2016: Assimilation of SMOS brightness temperatures or soil moisture retrievals into a land surface model. *Hydrol. Earth Syst. Sci.*, **20**, 4895–4911, <https://doi.org/10.5194/hess-20-4895-2016>.
- De Lannoy, G. J., R. H. Reichle, P. R. Houser, V. R. Pauwels, and N. E. Verhoest, 2007: Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter. *Water Resour. Res.*, **43**, W09410, <https://doi.org/10.1029/2006WR005449>.
- Delworth, T., and S. Manabe, 1989: The influence of soil wetness on near-surface atmospheric variability. *J. Climate*, **2**, 1447–1462, [https://doi.org/10.1175/1520-0442\(1989\)002<1447:TIOSWO>2.0.CO;2](https://doi.org/10.1175/1520-0442(1989)002<1447:TIOSWO>2.0.CO;2).
- de Rosnay, P., J. Polcher, M. Bruen, and K. Laval, 2002: Impact of a physically based soil water flow and soil-plant interaction representation for modeling large-scale land surface processes. *J. Geophys. Res.*, **107**, 4118, <https://doi.org/10.1029/2001JD000634>.
- , J. Muñoz-Sabater, C. Albergel, L. Isaksen, S. English, M. Drusch, and J.-P. Wigneron, 2020: SMOS brightness temperature forward modelling and long term monitoring at ECMWF. *Remote Sens. Environ.*, **237**, 111424, <https://doi.org/10.1016/j.rse.2019.111424>.
- Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 3385–3396, <https://doi.org/10.1256/qj.05.108>.
- de Wit, A. J., and C. A. Van Diepen, 2008: Crop growth modelling and crop yield forecasting using satellite-derived meteorological inputs. *Int. J. Appl. Earth Obs. Geoinf.*, **10**, 414–425, <https://doi.org/10.1016/j.jag.2007.10.004>.
- Dorigo, W., and Coauthors, 2011: International Soil Moisture Network: A data hosting facility for global in situ soil moisture measurements. *Hydrol. Earth Syst. Sci.*, **15**, 1675–1698, <https://doi.org/10.5194/hess-15-1675-2011>.
- , and Coauthors, 2013: Global automated quality control of in situ soil moisture data from the international soil moisture network. *Vadose Zone J.*, **12**, vzj2012.0097, <https://doi.org/10.2136/vzj2012.0097>.
- , and Coauthors, 2017: ESA CCI soil moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sens. Environ.*, **203**, 185–215, <https://doi.org/10.1016/j.rse.2017.07.001>.
- Draper, C., R. Reichle, G. De Lannoy, and Q. Liu, 2012: Assimilation of passive and active microwave soil moisture retrievals. *Geophys. Res. Lett.*, **39**, L04401, <https://doi.org/10.1029/2011GL050655>.
- Drusch, M., E. F. Wood, and H. Gao, 2005: Observation operators for the direct assimilation of TRMM Microwave Imager retrieved soil moisture. *Geophys. Res. Lett.*, **32**, L15403, <https://doi.org/10.1029/2005GL023623>.
- Entekhabi, D., R. H. Reichle, R. D. Koster, and W. T. Crow, 2010a: Performance metrics for soil moisture retrievals and application requirements. *J. Hydrometeorol.*, **11**, 832–840, <https://doi.org/10.1175/2010JHM1223.1>.
- , and Coauthors, 2010b: The Soil Moisture Active Passive (SMAP) mission. *Proc. IEEE*, **98**, 704–716, <https://doi.org/10.1109/JPROC.2010.2043918>.
- FAO/IIASA/ISRIC/ISS-CAS/JRC, 2012: Harmonized World Soil Database (version 1.2). FAO and IIASA, <http://web.archive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/index.html?sb=1>.
- Ghannam, K., T. Nakai, A. Paschalis, C. A. Oishi, A. Kotani, Y. Igarashi, T. Kumagai, and G. G. Katul, 2016: Persistence and memory timescales in root-zone soil moisture dynamics. *Water Resour. Res.*, **52**, 1427–1445, <https://doi.org/10.1002/2015WR017983>.
- Goldberg, D. E., and J. H. Holland, 1988: Genetic algorithms and machine learning. *Mach. Learn.*, **3**, 95–99, <https://doi.org/10.1023/A:1022602019183>.
- Gruber, A., G. De Lannoy, and W. Crow, 2019: A Monte Carlo based adaptive Kalman filtering framework for soil moisture data assimilation. *Remote Sens. Environ.*, **228**, 105–114, <https://doi.org/10.1016/j.rse.2019.04.003>.
- Kerr, Y. H., and Coauthors, 2010: The SMOS mission: New tool for monitoring key elements of the global water cycle. *Proc. IEEE*, **98**, 666–687, <https://doi.org/10.1109/JPROC.2010.2043032>.
- Kolassa, J., R. Reichle, and C. S. Draper, 2017: Merging active and passive microwave observations in soil moisture data assimilation. *Remote Sens. Environ.*, **191**, 117–130, <https://doi.org/10.1016/j.rse.2017.01.015>.
- Komma, J., G. Blöschl, and C. Reszler, 2008: Soil moisture updating by ensemble Kalman filtering in real-time flood

- forecasting. *J. Hydrol.*, **357**, 228–242, <https://doi.org/10.1016/j.jhydrol.2008.05.020>.
- Koster, R. D., and M. J. Suarez, 2001: Soil moisture memory in climate models. *J. Hydrometeorol.*, **2**, 558–570, [https://doi.org/10.1175/1525-7541\(2001\)002<0558:SMMICM>2.0.CO;2](https://doi.org/10.1175/1525-7541(2001)002<0558:SMMICM>2.0.CO;2).
- Krinner, G., and Coauthors, 2005: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochem. Cycles*, **19**, GB1015, <https://doi.org/10.1029/2003GB002199>.
- Krueger, E. S., T. E. Ochsner, D. M. Engle, J. Carlson, D. Twidwell, and S. D. Fuhlendorf, 2015: Soil moisture affects growing-season wildfire size in the southern great plains. *Soil. Sci. Soc. Amer. J.*, **79**, 1567–1576, <https://doi.org/10.2136/sssaj2015.01.0041>.
- Kuppel, S., P. Peylin, F. Maignan, F. Chevallier, G. Kiely, L. Montagnani, and A. Cescatti, 2014: Model-data fusion across ecosystems: From multisite optimizations to global simulations. *Geosci. Model Dev.*, **7**, 2581–2597, <https://doi.org/10.5194/gmd-7-2581-2014>.
- Laio, F., A. Porporato, L. Ridolfi, and I. Rodriguez-Iturbe, 2001: Plants in water-controlled ecosystems: Active role in hydrologic processes and response to water stress: II. Probabilistic soil moisture dynamics. *Adv. Water Resour.*, **24**, 707–723, [https://doi.org/10.1016/S0309-1708\(01\)00005-7](https://doi.org/10.1016/S0309-1708(01)00005-7).
- Liu, Q., and Coauthors, 2011: The contributions of precipitation and soil moisture observations to the skill of soil moisture estimates in a land data assimilation system. *J. Hydrometeorol.*, **12**, 750–765, <https://doi.org/10.1175/JHM-D-10-05000.1>.
- Lorenz, R., E. B. Jaeger, and S. I. Seneviratne, 2010: Persistence of heat waves and its link to soil moisture memory. *Geophys. Res. Lett.*, **37**, L09703, <https://doi.org/10.1029/2010GL042764>.
- Lurton, T., and Coauthors, 2020: Implementation of the CMIP6 forcing data in the IPSL-CM6A-LR model. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001940, <https://doi.org/10.1029/2019MS001940>.
- Martínez-de la Torre, A., E. M. Blyth, and E. L. Robinson, 2019: Evaluation of drydown processes in global land surface and hydrological models using flux tower evapotranspiration. *Water*, **11**, 356, <https://doi.org/10.3390/w11020356>.
- McColl, K. A., S. H. Alemohammad, R. Akbar, A. G. Konings, S. Yueh, and D. Entekhabi, 2017a: The global distribution and dynamics of surface soil moisture. *Nat. Geosci.*, **10**, 100–104, <https://doi.org/10.1038/ngeo2868>.
- , W. Wang, B. Peng, R. Akbar, D. J. Short Gianotti, H. Lu, M. Pan, and D. Entekhabi, 2017b: Global characterization of surface soil moisture drydowns. *Geophys. Res. Lett.*, **44**, 3682–3690, <https://doi.org/10.1002/2017GL072819>.
- Morris, M. D., 1991: Factorial sampling plans for preliminary computational experiments. *Technometrics*, **33**, 161–174, <https://doi.org/10.1080/00401706.1991.10484804>.
- Nicholson, S., 2000: Land surface processes and Sahel climate. *Rev. Geophys.*, **38**, 117–139, <https://doi.org/10.1029/1999RG900014>.
- Pastorello, G., D. Papale, H. Chu, C. Trotta, D. Agarwal, E. Canfora, D. Baldocchi, and M. Torn, 2017: A new data set to keep a sharper eye on land-air exchanges. *Eos, Trans. Amer. Geophys. Union*, **98**, <https://doi.org/10.1029/2017EO071597>.
- Pinnington, E., T. Quaipe, and E. Black, 2018: Impact of remotely sensed soil moisture and precipitation on soil moisture prediction in a data assimilation system with the JULES land surface model. *Hydrol. Earth Syst. Sci.*, **22**, 2575–2588, <https://doi.org/10.5194/hess-22-2575-2018>.
- Raoult, N., B. Delorme, C. Ottlé, P. Peylin, V. Bastrikov, P. Maugis, and J. Polcher, 2018: Confronting soil moisture dynamics from the ORCHIDEE land surface model with the ESA-CCI product: Perspectives for data assimilation. *Remote Sens.*, **10**, 1786, <https://doi.org/10.3390/rs10111786>.
- Reichle, R. H., and R. D. Koster, 2004: Bias reduction in short records of satellite soil moisture. *Geophys. Res. Lett.*, **31**, L19501, <https://doi.org/10.1029/2004GL020938>.
- Román-Cascón, C., and Coauthors, 2017: Correcting satellite-based precipitation products through SMOS soil moisture data assimilation in two land-surface models of different complexity: API and SURFEX. *Remote Sens. Environ.*, **200**, 295–310, <https://doi.org/10.1016/j.rse.2017.08.022>.
- Rondinelli, W. J., B. K. Hornbuckle, J. C. Patton, M. H. Cosh, V. A. Walker, B. D. Carr, and S. D. Logsdon, 2015: Different rates of soil drying after rainfall are observed by the SMOS satellite and the South Fork in situ soil moisture network. *J. Hydrometeorol.*, **16**, 889–903, <https://doi.org/10.1175/JHM-D-14-0137.1>.
- Ruscica, R., J. Polcher, M. Salvia, A. Sörensson, M. Piles, E. Jobbágy, and H. Karszenbaum, 2020: Spatio-temporal soil drying in southeastern south America: The importance of the effective sampling frequency and observational errors on drydown time scale estimates. *Int. J. Remote Sens.*, **41**, 7958–7992, <https://doi.org/10.1080/01431161.2020.1767825>.
- Salvia, M., R. Ruscica, A. Sörensson, J. Polcher, M. Piles, and H. Karszenbaum, 2018: Seasonal analysis of surface soil moisture dry-downs in a land-atmosphere hotspot as seen by LSM and satellite products. *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symp.*, Valencia, Spain, IEEE, 5521–5524, <https://doi.org/10.1109/IGARSS.2018.8518548>.
- Schepers, D., E. de Boissésón, R. Eresmaa, C. Lupu, and P. de Rosnay, 2018: CERA-SAT: A coupled satellite-era reanalysis. *ECMWF Newsletter*, No. 155, ECMWF, Reading, United Kingdom, 32–37, <https://www.ecmwf.int/node/18210>.
- Scholze, M., T. Kaminski, W. Knorr, S. Blessing, M. Vossbeck, J. Grant, and K. Scipal, 2016: Simultaneous assimilation of SMOS soil moisture and atmospheric CO<sub>2</sub> in-situ observations to constrain the global terrestrial carbon cycle. *Remote Sens. Environ.*, **180**, 334–345, <https://doi.org/10.1016/j.rse.2016.02.058>.
- Scipal, K., M. Drusch, and W. Wagner, 2008: Assimilation of a ERS scatterometer derived soil moisture index in the ECMWF numerical weather prediction system. *Adv. Water Resour.*, **31**, 1101–1112, <https://doi.org/10.1016/j.advwatres.2008.04.013>.
- Shellito, P. J., and Coauthors, 2016: SMAP soil moisture drying more rapid than observed in situ following rainfall events. *Geophys. Res. Lett.*, **43**, 8068–8075, <https://doi.org/10.1002/2016GL069946>.
- , E. E. Small, and B. Livneh, 2018: Controls on surface soil drying rates observed by SMAP and simulated by the Noah land surface model. *Hydrol. Earth Syst. Sci.*, **22**, 1649–1663, <https://doi.org/10.5194/hess-22-1649-2018>.
- Tarantola, A., 2005: *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, 354 pp.
- Vuichard, N., and D. Papale, 2015: Filling the gaps in meteorological continuous data measured at FLUXNET sites with ERA-Interim reanalysis. *Earth Syst. Sci. Data*, **7**, 157–171, <https://doi.org/10.5194/essd-7-157-2015>.

- Wang, D., Y. Chen, and X. Cai, 2009: State and parameter estimation of hydrologic models using the constrained ensemble Kalman filter. *Water Resour. Res.*, **45**, W11416, <https://doi.org/10.1029/2008WR007401>.
- Wu, M., M. Scholze, M. Voßbeck, T. Kaminski, and G. Hoffmann, 2019: Simultaneous assimilation of remotely sensed soil moisture and FAPAR for improving terrestrial carbon fluxes at multiple sites using CCDAS. *Remote Sens.*, **11**, 27, <https://doi.org/10.3390/rs11010027>.
- , ——, T. Kaminski, M. Voßbeck, and T. Tagesson, 2020: Using SMOS soil moisture data combining CO<sub>2</sub> flask samples to constrain carbon fluxes during 2010–2015 within a Carbon Cycle Data Assimilation System (CCDAS). *Remote Sens. Environ.*, **240**, 111719, <https://doi.org/10.1016/j.rse.2020.111719>.
- Yang, K., and Coauthors, 2016: Land surface model calibration through microwave data assimilation for improving soil moisture simulations. *J. Hydrol.*, **533**, 266–276, <https://doi.org/10.1016/j.jhydrol.2015.12.018>.