Global Optimization of Soil Texture Maps from Satellite-Observed Soil Moisture Drydowns and Its Implementation in Noah-MP Land Surface Model

Qing He[1,2], Hui Lu[2], Kun Yang[2,3], Taikan Oki[1], Jianhong Zhou[2], Long Zhao[4], Panpan Yao[5], Jie He[6], Aihui Wang[7] and Yawei Xu[2]

[1]River and Environmental Engineering Laboratory, Department of Civil Engineering, The University of Tokyo, Tokyo, Japan

[2]Ministry of Education Key Laboratory for Earth System Modeling and the Department of Earth System Science, Tsinghua University, Beijing, 100084, China

[3] Tibetan Plateau Data Center, State Key Laboratory of Tibetan Plateau Earth System and Resource Environment, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, China

[4]Chongqing Jinfo Mountain Karst Ecosystem National Observation and Research Station, School of Geographical Sciences, Southwest University, Chongqing 400715, China

[5]State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Science, Beijing, 100101, China

[6]National Tibetan Plateau Data Center, Institute of Tibetan Plateau Research, Chinese Academy of Science, Beijing, 100101, China

[7]Nansen-Zhu International Research Centre, Institute of Atmospheric Sciences, Chinese Academy of Science, Beijing, China

Corresponding author: Hui Lu (luhui@tsinghua.edu.cn)

Key Points:

- Satellite soil moisture data is used to calibrate two major soil texture datasets at the global scale.
- Soil moisture simulations using the updated soil maps outperform those using the baseline soil texture products.
- Soil moisture improvement is more significant in simulations considering soil organic carbon.

Abstract

Soil moisture (SM) plays an important role in regulating regional weather and climate. However, the simulations of SM in current land surface models (LSMs) contain large biases and model spreads. One primary reason contributing to such model biases could be the misrepresentation of soil texture in LSMs, since current available large-scale soil texture data are often generated from extrapolation algorithm based on a scarce number of in-situ geological measurements. Fortunately, recent advancements in satellite technology provide a unique opportunity to constrain the soil texture datasets by introducing observed information at large spatial scales. Here, two major soil texture baseline datasets (Global Soil Datasets for Earth system science, GSDE and Harmonized World Soil Data from Food and Agriculture Organization, HWSD) are optimized with satellite-estimated soil hydraulic parameters. The optimized soil maps show increased (decreased) sand (clay) content over arid regions. The soil organic carbon content increases globally especially over regions with dense vegetation cover. The optimized soil texture datasets are then used to run simulations in one example LSM, i.e., Noah LSM with Multiple Parameters. Results show that the simulated SM with satellite-optimized soil texture maps is improved at both grid and in-situ scales. Intercase comparison analyses show the SM improvement differs between simulations using different soil maps and soil hydraulic schemes. Our results highlight the importance of incorporating observation-oriented calibration on soil texture in current LSMs. This study also joins the call for a better soil profile representation in the next generation of Earth System Models.

Plain Language Summary

Soil moisture (SM) is important for weather and climate but is often poorly simulated by Land Surface Models (LSMs). One possible reason could be the inappropriate representation of soil texture maps utilized in LSMs since current gridded soil texture maps are often derived from a limited number of in-situ measurements. In this study, we leverage the benefits of modern satellite products and land surface theories to improve sever major global soil texture maps, and use the calibrated soil maps to improve soil moisture simulation in one example LSM. Results show increased sand content over arid areas while the results for clay content show the opposite pattern. The soil organic carbon (SOC) result shows an overall increase over the entire globe but is more evident in dense vegetation land covers. The model simulated SM using the calibrated soil maps generally outperforms those with the baseline soil maps. The improvement is more significant in the experiment with soil maps considering SOC. Our results here provide successful evidence for constraining soil texture data from large-scale observations. We also show that observation-oriented calibration on soil texture maps is necessary for a better land surface simulation, which is critically important for the development of Earth System Models.

1 Introduction

Soil moisture (SM) is a crucial component in terrestrial water cycles (Oki and Kanae, 2006). It can have substantial influence on regional weather and climate through its role as the intermediate medium for land-atmosphere water and energy flux exchanges (Seneviratne et al., 2010). Persistent negative anomaly in SM could induce severe droughts, causing great loss in agriculture production and socio-economic activities (Miralles et al., 2019; Mukherjee et al., 2018). It can also cause hazardous extreme events that threaten human life (Miralles et al., 2014; Seneviratne et al., 2010; Zhou et al., 2019). Accurate predictions of SM is thus essentially useful for decision-makers to take prevention and mitigation measures for such natural disasters in time.

SM is usually simulated by land surface models (LSMs) and contributes to the boundary layer condition in Earth System Models (ESMs) through regulations on surface water flux exchanges. However, evaluation studies have shown that the soil moisture simulation in current LSMs contains significant biases, for example, wide SM spreads were found within a diverse ensemble of LSMs (Boone et al., 2004; Dirmeyer et al., 2016, 2006b, 2006a). The large SM biases could then be amplified through land-atmosphere interaction processes and result in substantial air temperature and precipitation biases in ESMs (Dong et al., 2022). In previous studies, successful efforts have been addressed to improve SM simulations in LSMs, including calibrating the models' soil hydrothermal parameterization schemes (Zheng et al., 2015; Koster et al., 2017, 2018), essential land surface parameters such as Leaf Area Index (LAI), surface albedo (Kumar et al., 2019; Malik et al., 2012; Yin et al., 2016) etc. However, these improvement studies are model specific, that is to say, SM biases cannot be improved if the models already use the optimal parameterization schemes or land surface parameters. The systematical SM biases (i.e., the long-term equilibrium SM) determined by soil texture between models should be improved through the calibration of soil texture datasets utilized in current LSMs (Teuling et al., 2009; He et al., 2023).

Soil texture is defined as a certain soil composition of sand, clay and loam content, and determines the soil hydrothermal properties such as saturated soil matric potential, saturated soil hydraulic conductivity and diffusivity etc. Generally, there are two options to define soil texture in LSMs: one is to use the model-prescribed look-up table, in which a certain value obtained from in-situ geological surveys (e.g., particle size analyses (PCA) (Black, 1965; Gee and Bauder, 2018)) is assigned for each soil type; The other is to incorporate gridded soil maps externally, which is favored in recent land surface simulation studies sinc it is spatially more representative compared to the look-up table approach (Xu et al., 2023). However, the gridded soil maps themselves contain biases, since they are often generated by extrapolating the point-scale measurements into polygons (Dai et al., 2019; Shangguan et al., 2014). Such mismatch in spatial scales might lead to unreal soil texture results, particularly when the number of in-situ measurements is limited. However, since the large-scale soil texture observation is unavailable, it is difficult to improve the quality of current gridded soil maps.

Recent advancements in remote sensing technologies have provided unique opportunities to observe Earth's land surface from large spatial scales. Development of satellite-observed land surface products such as land surface temperature (Wan, 2013), soil moisture (Entekhabi et al., 2010; Kerr et al., 2001; Njoku et al., 2003), and vegetation indices (e.g., LAI, Fractional Vegetation Cover (FVC), etc.) (Cohen et al., 2006; Garrigues et al., 2008; Xiao et al., 2016) facilitates improvement of land surface modeling by providing real-time and realistic land surface information (He et al., 2021; Kolassa et al., 2020; Kumar et al., 2019, 2015; Li et al., 2019). Data assimilation (DA) technology is one typical way of incorporating satellite datasets to improve land

surface simulations (Gettelman et al., 2022; Z.-L. Yang et al., 2020). While this approach is effective in improving models' biases, to maintain the persistent improvement of LSM simulations, the DA approach requires continuous input of satellite signals. The improvement thus cannot be extended to future predictions when satellite measurements are unavailable. An alternative and much simpler approach for satellite-based model improvement is using satellite information to calibrate LSMs' physical schemes and parameters (Yang et al., 2009; Balsamo et al., 2018; Lu et al., 2020). Efforts using this approach have demonstrated that the parameter calibration approach can also substantially improve land surface modeling biases (Chen et al., 2010; Kolassa et al., 2020; Sun et al., 2021), and more importantly, it can complement the DA-based model improvements (Koster et al., 2018). The model improvement from the calibration approach can be efficiently transferred into future scenarios compared to the DA approach.

In land surface modeling studies, empirical functions (i.e., pedo-transfer functions, PTFs, e.g., Wösten et al., 2001) are often used to link soil texture data and soil hydraulic parameters (e.g., wilting point, field capacity, saturated point, etc.) However, since the acquisition of soil hydraulic parameters could be more labor-intensive and more costing than soil texture measurements, the typical way is to first collect soil texture information, then estimate the soil hydraulic parameters by applying PTF equations. Instead of directly obtaining them from laboratory measurements or field observations, recent studies have shown that some soil hydraulic parameters can be estimated from soil moisture data only (McColl et al., 2017; Akbar et al., 2018). With satellite observations of soil moisture being increasingly available and the empirical functions being known, the newly developed soil moisture – soil hydraulic parameter estimation approach brings a new opportunity to derive soil texture from soil hydraulic parameters in turn. Such an approach could also provide new insight into improvement on currently existing soil texture products.

In this light, this study aims to use soil hydraulic parameters (i.e., soil wilting point θ_w and critical point θ_c) estimated from long-term satellite soil moisture dataset to calibrate two major soil texture baseline datasets, i.e., Global Soil Datasets for Earth system science (GSDE) developed by Land-Atmosphere Interaction Laboratory of Sun Yat-Sen University (Shangguan et al., 2014), and Harmonized World Soil Data developed by Food and Agriculture Organization (FAO) (Nachtergaele et al., 2009). The calibration is realized by applying an optimization procedure, where the soil hydraulic parameters θ_w and θ_c calculated from baseline soil maps and PTF functions are compared with those from satellite estimations. Two PTF schemes with and without considering soil organic carbon (SOC) are used to investigate the influence of SOC on the calibration results. The baseline and calibrated soil maps are then incorporated into an example land surface model (i.e., Noah LSM with Multiple Parameters, Noah-MP (Niu et al., 2011; Yang et al., 2011)) to conduct simulations. The simulated SM with baseline and calibrated soil maps are then compared and validated with observational data at both regional and in-situ scales. In doing so, this study could offer an observation-based reference for improving SM simulations in LSMs. This study also provides a method for calibrating gridded soil texture maps with satellite observations, which joins the call for updating the characterization of the soil profile for the development of next generation LSMs (as well as ESMs).

2 Data and Methods

This section contains information of the methodology and primary datasets this study used and is divided into three parts based on their purposes: Section 2.1 describes the methods for obtaining soil hydraulic parameters, i.e., soil wilting point θ_w and critical point θ_c, from soil

moisture time series. The satellite soil moisture dataset used in this study will also be introduced; Section 2.2 describes the main algorithm for constraining soil texture datasets using the parameters obtained from Section 2.1. Two PTF schemes with and without SOC effects, and two global baseline soil texture datasets are introduced; In Section 2.3, experiment designs for incorporating the calibrated soil texture datasets in Noah-MP LSM are described. Meteorological forcing data and validation datasets are introduced in this part.

2.1 Soil Hydraulic Parameters from Satellite Soil Moisture Data

As discussed above, recent studies have shown that soil hydraulic parameters can be estimated solely from surface soil moisture data (McColl et al., 2017). This method will be adopted in our study to derive important soil hydraulic parameters (i.e., θ_w and θ_c). The basic hypothesis of the method in (McColl et al., 2017) is that as intrinsic properties of soil, the soil hydraulic parameters (also defined as SM thresholds) are reflected in the SM temporal changes. This can be explained by the surface water loss function ($L(\theta)$) shown in Figure 1a. First, when a precipitation event occurs, SM will be immediately saturated after the rainfall event, at this time surface water decreases as drainage loss until it reaches the field capacity θ_{fc}. Then evapotranspiration (ET) occurs at its maximum rate (i.e., potential ET); SM decreases as potential ET loss until it reaches the soil critical point θ_c. Since the potential ET process occurs very rapidly, the two thresholds, θ_{fc} and θ_c, are considered to be consistent in some cases. Afterward, SM continues decreasing as the water-limited rate of ET loss (typically this rate can be defined as a first-order relation between ET and SM) until all water loss processes cease. SM at this time is defined as θ_w. Converting $L(\theta)$ to the time domain an exponential curve showing the relationship between SM and time occurs ($\theta(t)$, Figure 1b). This curve is defined as the SM drydown event where SM changes are consistently negative.

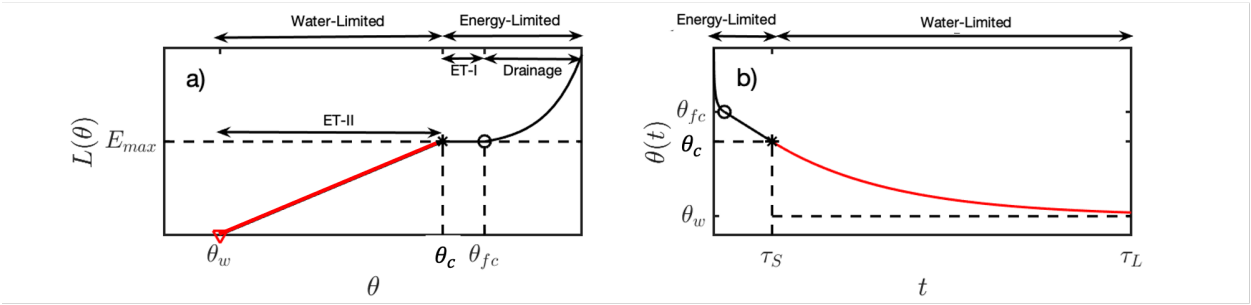

Figure 1. Diagram for surface water loss function (a) and soil moisture drydown curve (b). The x-axis in (a) refers to soil moisture ($m^3\ m^{-3}$), and y-axis refers to surface water loss rate ($L(\theta)$, in unit of LT^{-1}); E_{max} is the maximum evapotranspiration rate (LT^{-1}). Figure 1b is the projection of 1a in the time domain, where the x-axis refers to time (e.g., days) and y-axis refers to soil moisture content ($m^3\ m^{-3}$). θ_w, θ_c, and θ_{fc} refers to soil wilting point, critical point, and field capacity, respectively. Figures adaptated from (He et al., 2023).

Recall that θ_c and θ_w refers to the start and end point of the water-limited surface water loss process, respectively. In this light, it is reasonable to expect θ_c corresponds to the initial SM value (θ_{dd0}), and θ_w refers to the minimum SM value (θ_{min}) of the water-limited drydown curve

(solid red line in Figure 1b). To characterize this part of drydown curve, the linear relation between the rate of water loss and SM can be described as:

$$\frac{d\theta(t)}{dt} = k\theta(t), \quad (1)$$

where, k refers to the slope of the curve (d^{-1}). Integrating both sides of (1) as a function of time, an empirically derived equation can be derived, as:

$$\theta(t) = \Delta\theta \exp\left(-\frac{t}{\tau_L}\right) + \widehat{\theta_w} \quad (2)$$

where, $\Delta\theta$ refers to the soil moisture change during each soil drying event ($m^3\ m^{-3}$); $\widehat{\theta_w}$ refers to the estimated minimum soil moisture value ($m^3\ m^{-3}$); t refers to the time duration of the soil moisture drydown event (days); and τ_L refers to the time scale when SM returns back to an equilibrium state, i.e., soil moisture memory time (days). Details of the derivation from equations (1) to (2) can be found in (McColl et al., 2017).

To obtain estimations of θ_c and θ_w, the water-limited soil moisture drydown event is first identified as the observed SM timeseries with consistent soil moisture decrement. θ_c thus corresponds to the observed initial SM value of each drydown event. Then equation (2) is used to fit the identified drydown event to derive θ_w. We note that the estimated parameters can be different for each identified drydown event (i.e., there are temporal variations in θ_c and θ_w estimations). We here use their statistical medians as "representative" estimations for the parameters at each grid. The analyses in this study are conducted by using 19-year (2002-2020) remote sensing soil moisture data, i.e., Neural Network Soil Moisture (NNSM (Yao et al., 2021)). Compared to other remote sensing SM datasets such as Soil Active and Passive (SMAP) and Soil Moisture and Ocean Salinity (SMOS), NNSM provides a much longer temporal span and larger spatial coverage over frozen and heavy-vegetated areas with high data accuracy (ubRMSE ≤ 0.04 $m^3\ m^{-3}$). NNSM has the daily temporal resolution and the spatial resolution of 36km. The identified drydown events containing less than 3 observation samples and events with the determination coefficient (R^2) of the fitting less than 0.7 are filtered, consistent with (McColl et al., 2017).

2.2 Soil Texture Optimization through SCE-UA Algorithm

Once θ_w and θ_c are estimated from satellite observation, and relations between them and soil texture have been known through PFT functions, the most straightforward way to derive soil texture data is to directly solve the empirical equations. However, in doing so the underlying hypothesis would be to assume satellite estimations are the truth. However, theoretically the truth could never be known for any variable or system even though with abundant observation records. To avoid such a problem, we here apply another method, i.e., SCE-UA (Shuffled Complex Evolution – University of Arizona (Duan et al., 1994, 1992, 1993)) algorithm to obtain soil texture

from the satellite-estimated θ_w and θ_c without explicitly solving the PTF equations. The basic working flow of this algorithm is depicted in Figure 2.

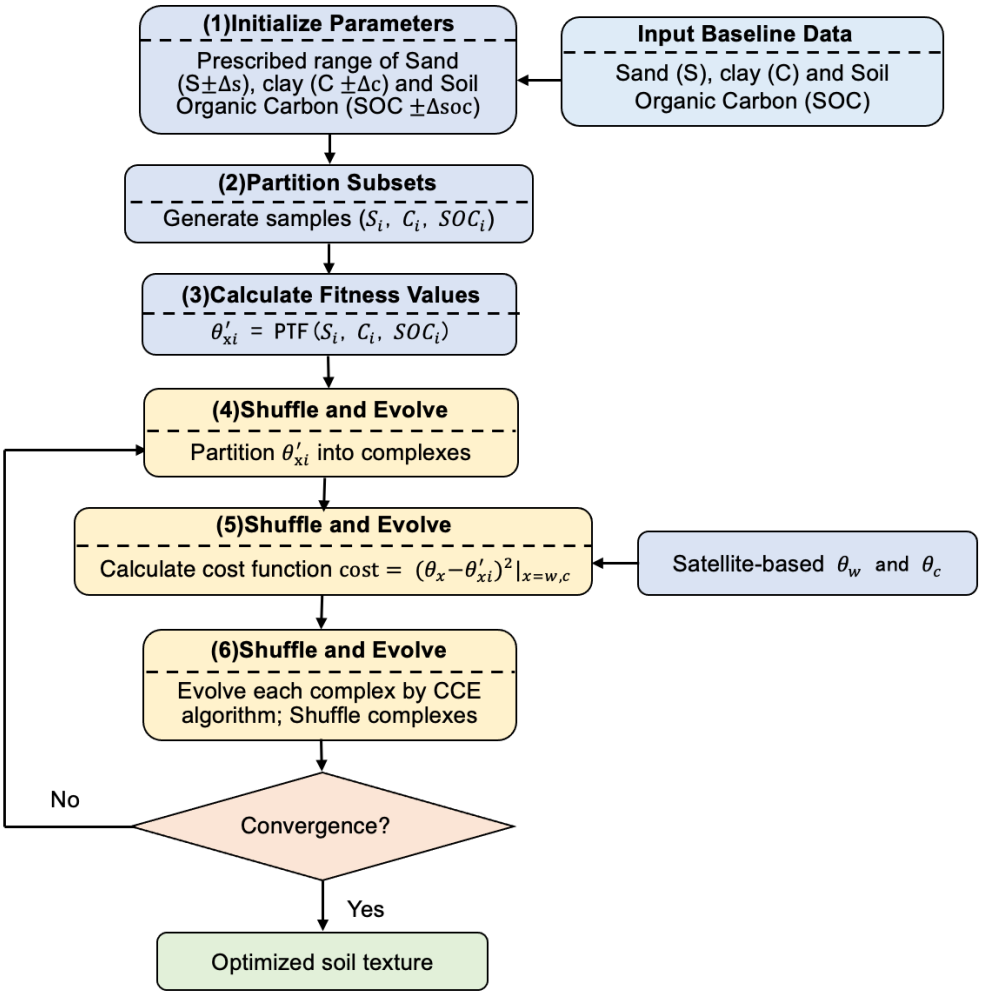

Figure 2. Working flow of soil texture optimization by using SCE-UA algorithm in this study. Parameters in each block are: (1) S, C, and SOC are for soil sand content (%), soil clay content (%) and soil organic carbon content (%) from the input baseline soil texture data; Δs, Δc and Δsoc are the absolute value of prescribed variation range for sand, clay and soil organic carbon respectively; (2)S_i, C_i and S_i indicate the sand, clay and SOC content generated by SCE-UA program; i refers to the sample index; (3)PTF indicates the PedoTransfer Function (in this study, both SR06 and CH78 schemes are used), θ'_{xi} indicates the soil hydraulic parameters calculated using the soil composition samples generated by SCE-UA program; $x = w, c$ indicates both θ_c and θ_w will be calculated; i refers to the sample index.

The basic working flow of applying SCE-UA in this study is that for a certain grid, a variation range of each soil texture component will be first prescribed (Step 1 in Figure 2). In this study, the variation range of $\pm 20\%$ for sand, $\pm 10\%$ for clay, and $\pm 5\%$ for SOC is defined. The SCE-UA algorithm will then generate a sample pool based on the input baseline soil texture data (Step 2). For each sample, fitness values of wilting point and critical point is calculated through

PTF (θ'_{wi} and θ'_{ci}, $i = 1, n$ where n refers to the total number of soil texture combinations, Step 3). Then the generated θ'_{wi} and θ'_{ci} are shuffled (Step 4) and the cost function between the fitness values and satellite estimation of θ_w and θ_c is calculated (Step 5). Then the evolution procedure is conducted using the Constraint-handling using Constrained Evolutionary (CCE) optimization algorithm (Step 6) to ensure the cost function is minimum between the fitness values and satellite estimations. The soil texture combination corresponding to the time when the cost function is minimum is chosen as the optimal soil texture composition for this grid.

Two baseline soil texture datasets, i.e., GSDE that includes soil organic carbon (SOC) and HWSD that includes only the basic components of soil texture (i.e., sand, clay, and loam) are chosen in this study because we want to investigate the effect of different PTF schemes on SM improvement. SOC has attracted increasing interests in recent land surface modeling studies (Chen et al., 2016; Lawrence and Slater, 2008) since it can have significant influence on soil hydraulic properties through its effect on soil structure and adsorption properties (Rawls et al., 2004, 2003). Two PTF schemes with and without consideration of SOC effect (i.e., SR06 (Saxton and Rawls, 2006) and CH78 (Clapp and Hornberger, 1978), correspondingly) are therefore chosen to apply for GSDE and HWSD, respectively. Descriptions of the two PTF schemes are listed in Table S1. The original spatial resolution of both GSDE and HWSD is 0.25 degree.

Three experiment designs for soil texture optimization in this study are shown in Table 1. Opti_exp1 uses sand, clay and SOC from GSDE as baseline soil texture data and SR06 PTF scheme to account for the SOC effect; Opti_exp3 uses sand and clay from the HWSD dataset and CH78 as the PTF scheme that does not account for SOC effect. However, since different soil texture datasets are used, comparing results between these two experiments cannot give an explicit indications of how SOC affects soil texture optimizations. As such, an additional experiment, Opti_exp2, which uses only soil sand and clay content from GSDE and the non-SOC effect PTF scheme is also conducted. Comparing results from Opti_exp2 and Opti_exp1 therefore indicates effects of SOC on soil texture optimization, and comparing results from Opti_exp2 and Opti_exp3 can inform the global spatial patterns of soil texture before and after optimization in two different datasets.

Table 1. Experiment Designs for Global Soil Texture Optimization

Experiment Name	Baseline Soil Data Name	PTF Scheme	Optimized Soil Data Name	Prescribed Ranges†
Opti_exp1	GSDE_default	SR06	GSDEoc_sce	ΔSand: ±20% ΔClay: ±10% ΔSOC: ±5% Iteration: 3000
Opti_exp2	GSDE_default	CH78	GSDEnoc_sce	
Opti_exp3	HWSD_default	CH78	HWSDnoc_sce	

†The lower and upper bound for Sand, Clay and SOC are 6% to 98%, 3% to 58%, 0 to 15% respectively. Optimization results exceed this range is regarded as ineffective and will be masked out. The analysis in the main context is conducted on 3000 iteration steps, i.e., for each experiment there are 3000 suits of soil

sand, clay (and SOC) combinations. The opitimization result will then be chosen as the one combination that can produce the closest θ_w and θ_c to the satellite estimations.

2.3 Simulation Designs in Noah-MP LSM

As aforementioned in our research purposes, the final goal of optimizing soil texture data is to improve SM simulation in LSMs. Here we use Noah-MP (Niu et al., 2011; Yang et al., 2011) as an example LSM to evaluate how the optimized soil texture data can help improve SM simulations. For each optimization experiment listed in Table 1, we design two simulation experiments – one uses the baseline soil texture dataset, and the other one uses the optimized soil texture – to run simulations in Noah-MP (Table 2). A high-quality meteorological dataset based on remote sensing data, reanalysis data and observation data of over 700 stations (CMFD, China Meteorological Forcing Dataset) is chosen to drive the model runs, meaning the focus area of the model simulation part is China, different from the global analyses in the other part of this study. The simulation period is from Jan 01 2008 to Dec 31 2010, in which the first two years are run for the model's spin-up. The model is run at the 3-hourly time step and at a spatial resolution of 0.1 degree. All the soil texture datasets used in the model simulation are downscaled into this spatial resolution.

To evaluate how much the simulated SM can be improved from the optimized soil texture data, the difference between SM from simulations using baseline soil texture and NNSM, and the difference between SM simulation using optimized soil texture and NNSM are compared. Since NNSM also serves as the primary satellite dataset to extract soil hydraulic parameters, the other remote sensing SM product that is independent of the optimization procedure, namely, ITPLDAS (soil moisture dataset of China based on microwave data assimilation (Yang et al., 2020; Yang et al., 2016)), is also used for evaluation. Additionally, the Root-Mean-Square-Error (RMSE) between SM from simulations using baseline and optimized soil texture data are also compared by using 732 in-situ SM observations (Wang and Shi, 2019).

Table 2. Experiment Designs for Noah-MP Model Simulations over China

	Soil Texture	PTF Scheme	Forcing	Time period
Exp1	GSDE_default	SR06	CMFD	2008.01.01 – 2010. 12.31
Exp2	GSDEoc_sce	SR06		
Exp3	GSDE_default	CH78		
Exp4	GSDEnoc_sce	CH78		
Exp5	HWSD_default	CH78		
Exp6	HWSDnoc_sce	CH78		

3 Results

3.1 Global Patterns of Soil Hydraulic Parameters from Satellite Estimations

Global medians of the two soil thresholds, namely, θ_w and θ_c estimated from satellite-identified drydown events are shown in Figure 3a and Figure 3b, respectively. Both θ_w and θ_c show clear geographical gradient globally, for example, they both show high values in climatologically humid regions (e.g., eastern U.S., southern Sahara, and southern China) whereas show low values in climatologically arid regions (e.g., western U.S., central Asia, and central Australia). In terms of magnitude, θ_c with the mode of ~0.23m^3m^{-3}, is overall higher than θ_w, which has a mode of 0.13m^3m^{-3} (*x-axis* value corresponding to the peak of PDF function in Figure 3). The results above suggest that our estimations of θ_w and θ_c using NNSM dataset in this study are reasonable, where their global patterns compare consistently to the soil drydown time in several previous studies (McColl et al., 2019, 2017).

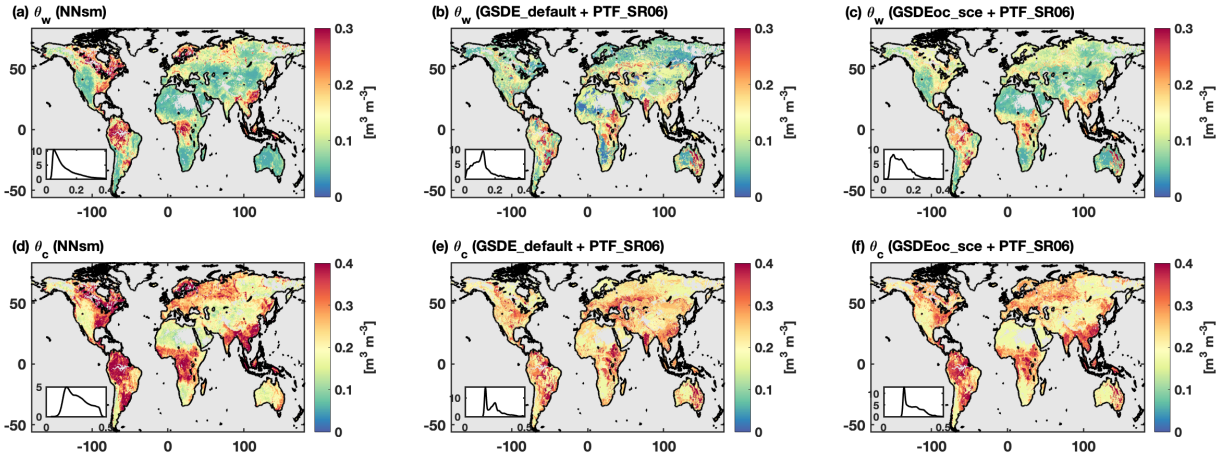

Figure 3. Global distribution of θ_w (top) and θ_c (bottom) from satellite estimations (a and c), baseline soil texture map (b and e) and optimized soil texture map (c and f). Insets refer to the probability density distribution of each map.

Compared to satellite estimations, the soil thresholds calculated from one of the baseline soil texture datasets (i.e., GSDE, using SR06 PTF) without any satellite correction show much less spatial gradient, for example, both θ_w and θ_c are relatively uniform (e.g., North America), showing scattered hotspots across climate zones (e.g., Amazon, eastern Africa, and northwestern India) (Figure 3, b and e). Results are similar when using different combinations of baseline soil texture data and PTF functions (Figure S1). This uniform-with-hotspots pattern could be a product of the extrapolation algorithm of the baseline soil texture dataset, where the limited point-scale geological survey samples can barely represent the soil characteristics to a much larger spatial extent, which indicates that the spatial pattern of SM simulations might be biased if baseline soil texture dataset were directly used in LSMs without corrections from satellite information. In addition, the uniform distribution of soil thresholds in baseline soil texture also results in

overestimation in the magnitude of θ_w and θ_c over arid areas while underestimation over humid areas, which may add additional biases to SM simulations.

The soil thresholds from baseline soil texture with satellite corrections, that is, θ_w and θ_c corresponding to the minimum value of the cost function in the SCE-UA procedure, compare much closer to satellite estimations in both spatial pattern and magnitude, than to those from baseline soil texture only (Figure 3, c and f, and Figure S1 for results using different baseline soil texture datasets and PTF functions). The East-West gradient over North America is reproduced, and the left-skewed PDF of θ_w is largely corrected. A similar correction is also made to the dual-peak distribution of θ_c. The results suggest that the SCE-UA procedure can efficiently reproduce θ_w and θ_c that are closest to satellite drydown estimations, and the soil texture components derived under this condition can also serve as an improved dataset to represent soil characteristics at large spatial scales.

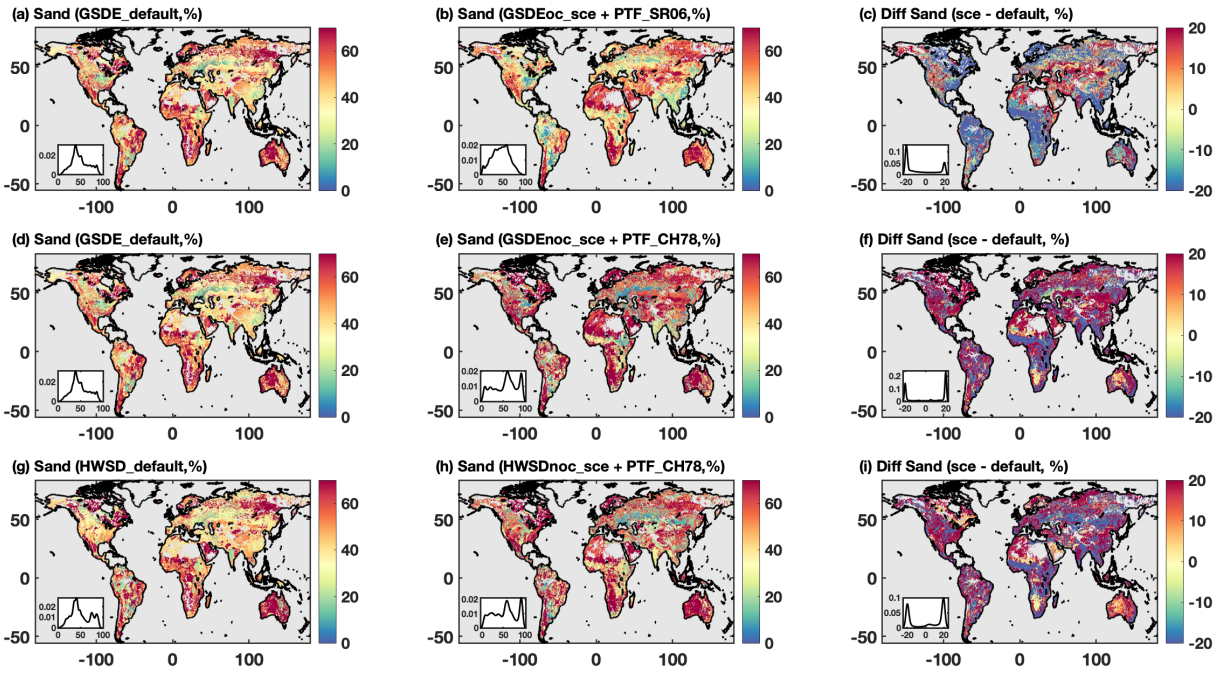

Figure 4. Global patterns of sand content from baseline soil maps (left column), optimized soil maps (middle column) and their difference (right column). For the middle and right columns, From top to bottom are results from GSDE soil map with SR06 PTF scheme, GSDE soil map with CH78 PTF scheme and HWSD soil map with CH78 PTF scheme. Insets refer to probability density distribution.

3.2 Global Patterns of Soil Texture Optimized from SCE-UA Algorithm

Before discussing their spatial patterns, the sensitivity of the optimization results to one important parameter of the SCE-UA algorithm, i.e., the iteration steps *n*, is analyzed. We conducted 4 optimization experiments with *n* equaling 300, 3000, 10000 and 30000 to conduct analyses. Figure S2 shows that the optimization results of sand, clay and SOC are insensitive to the iteration steps. The optimized sand and clay content both remain consistent among the four

experiments. The SOC content in the four experiments is also highly consistent, only except that the experiment with 300 iteration steps shows minorly lower SOC content. The analyses in the context below are based on optimization results with iteration steps of 3000.

Soil texture derived from the SCE-UA procedure with satellite information (we will refer to them as "optimized" dataset hereafter) as well as comparison with their baseline soil texture datasets are shown in Figures 4 – 6. Overall, the soil texture components from the SCE-UA algorithm show a similar geographical pattern to the baseline maps, e.g., sand content from both optimized and baseline datasets are high in climatologically arid areas (e.g., great deserts such as California, Sahara and central Australia), while low in climatologically humid areas (Figure. 4, a – f), and vice versa for the soil clay component (Figure 5, a – f). Soil organic carbon, however, is overall evenly distributed across climate zones, but shows close relations to vegetation density. For example, in areas with dense vegetation coverage, SOC is comparably high (Figure 6, a – b). The above results suggest that the SCE-UA procedure performs reasonably in reproducing the overall geographical pattern of global soil component content.

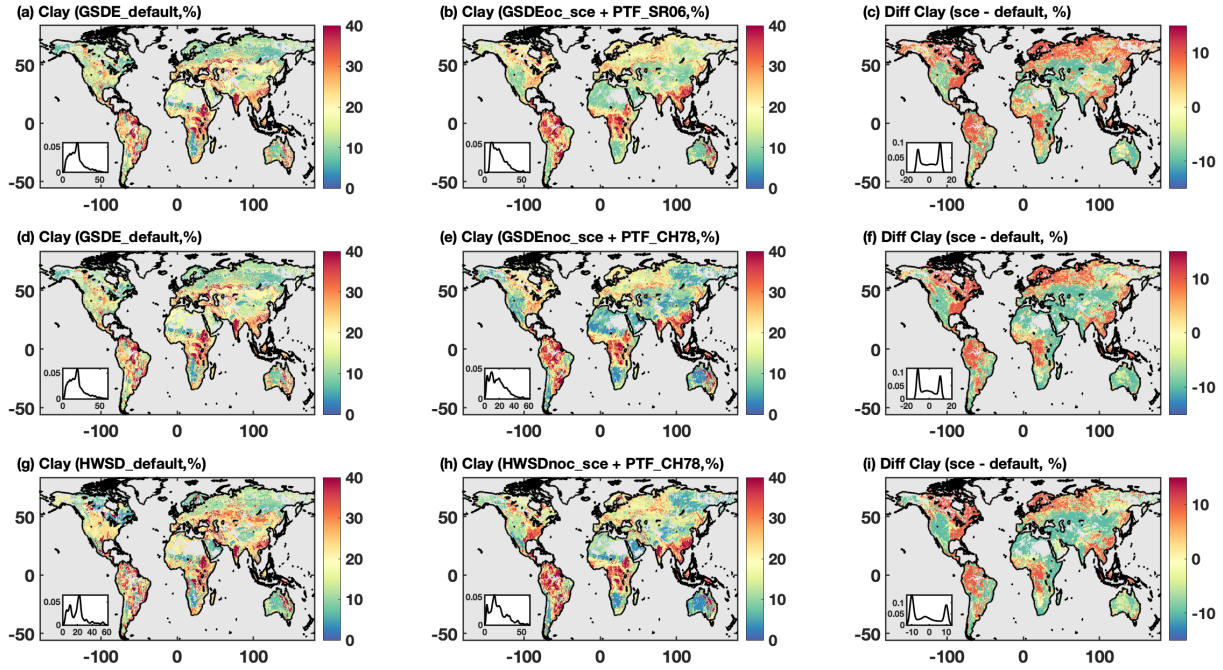

Figure 5. Same as Figure 4 but for the soil clay content.

However, the optimized soil texture datasets differ from the baseline soil maps in magnitude (Figures 4 – 5, g – i and Figure 6c). Each soil content from the optimized dataset shows up to 20%, 10%, and 5% changes in absolute values respectively (the three numbers correspond to the prescribed range in Table 1). Similar to their spatial patterns, the magnitude changes in sand and clay content show dependence on climate zones, that is, the sand content from the optimized dataset increases over arid regions whereas decreases over humid regions (the other way around for the clay content changes), while SOC shows overall increase globally, although the increase is comparably more over the densely covered vegetation areas. The spatial dependence of the sand and clay changes corresponds well to the satellite-baseline soil thresholds difference (Section III-

A), further verifying that the optimized soil texture results do contain satellite information from satellite observations, such that they could perform better in representing large-scale soil characteristics (compared to the baseline soil maps). We should also note that such spatial dependence of sand changes is also related to PTF functions, for example, the decrease of sand content in humid areas in the cases Opti_exp2 and Opti_exp3 is not as much as that in Opti_exp1 (even showing an increase in some areas such as eastern U.S. and southern China). This indicates that appropriately choosing PTF functions is essential to obtain reasonable soil sand estimations.

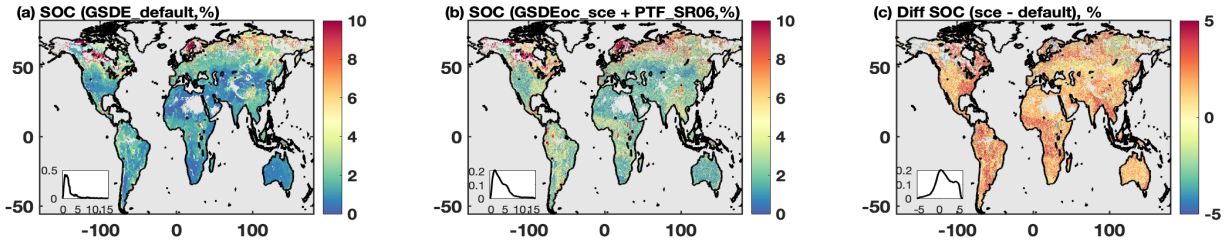

Figure 6. Global patterns of soil organic carbon from baseline soil maps (a), optimized soil maps (b) and their difference (c). Insets refer to probability density distribution.

3.3 Soil Moisture Simulations from Noah-MP LSM

JJA soil moisture simulations from Noah-MP LSM using baseline and optimized soil texture maps are then compared with contemporary NNSM SM product and shown in Figure 7. SM simulated using baseline soil maps (i.e., exp1, exp3 and exp5) show overall positive biases (biases of ~0.1 m^3 m^{-3} compared to NNSM, inset PDF distribution) over the entire study area. The positive biases are especially large in southeastern China (humid climate) while substantial negative biases are observed in the northwestern part (arid climate), with maximum biases larger than 0.15 m^3 m^{-3} (in absolute value). By comparison, SM using optimized soil data (i.e., exp2, exp4 and exp6) show reduced biases, with the right-skewed PDF distribution corrected left-forward. Noticeably, the substantial positive and negative biases over southeastern and northwestern regions are largely corrected, with the maximum biases constrained within 0.1 m^3 m^{-3} respectively. Scattered hotspots with extremely large SM biases are still observed, possibly due to local land surface conditions (e.g., vegetation properties) that may not be accounted for by soil characteristics.

In addition to the overall SM improvement in experiments using the optimized soil texture dataset, the SM improvements are also case-dependent. In exp4 and exp6, where different baseline soil data and PTF functions without considering SOC are used, the negative SM biases – meaning less soil moisture retained by the soil in the simulations compared to satellite observation – over southeastern regions do not show as much improvement as in exp2. The SM simulation difference between exp2 and the other two experiments corresponds reasonably to the discrepancy in the spatial distribution of soil sand content – the decrement of soil sand percentage in exp2 will consequently enhance the soil water-holding capability, thus that more water can be retained in the soil column, while in the other two cases the soil sand percentage is increased, resulting in

degradation of the soil water-holding capacity. Detailed inter-case comparisons will be discussed in Section 3.4.

Since NNSM serves as input (although not directly) to the optimization procedure, using it as a reference to validate SM simulations might lead to the comparison itself being biased. To further validate the simulated SM independently, a comparison using SM from ITPLDAS is conducted. Results show that SM biases over southeastern (negative biases) and northwestern regions (positive biases, more inland compared to Figure 7) are still substantial between simulations using baseline soil maps and ITPLDAS, and biases are both reduced, although relatively moderate when compared to SM comparison with NNSM (Figure S3).

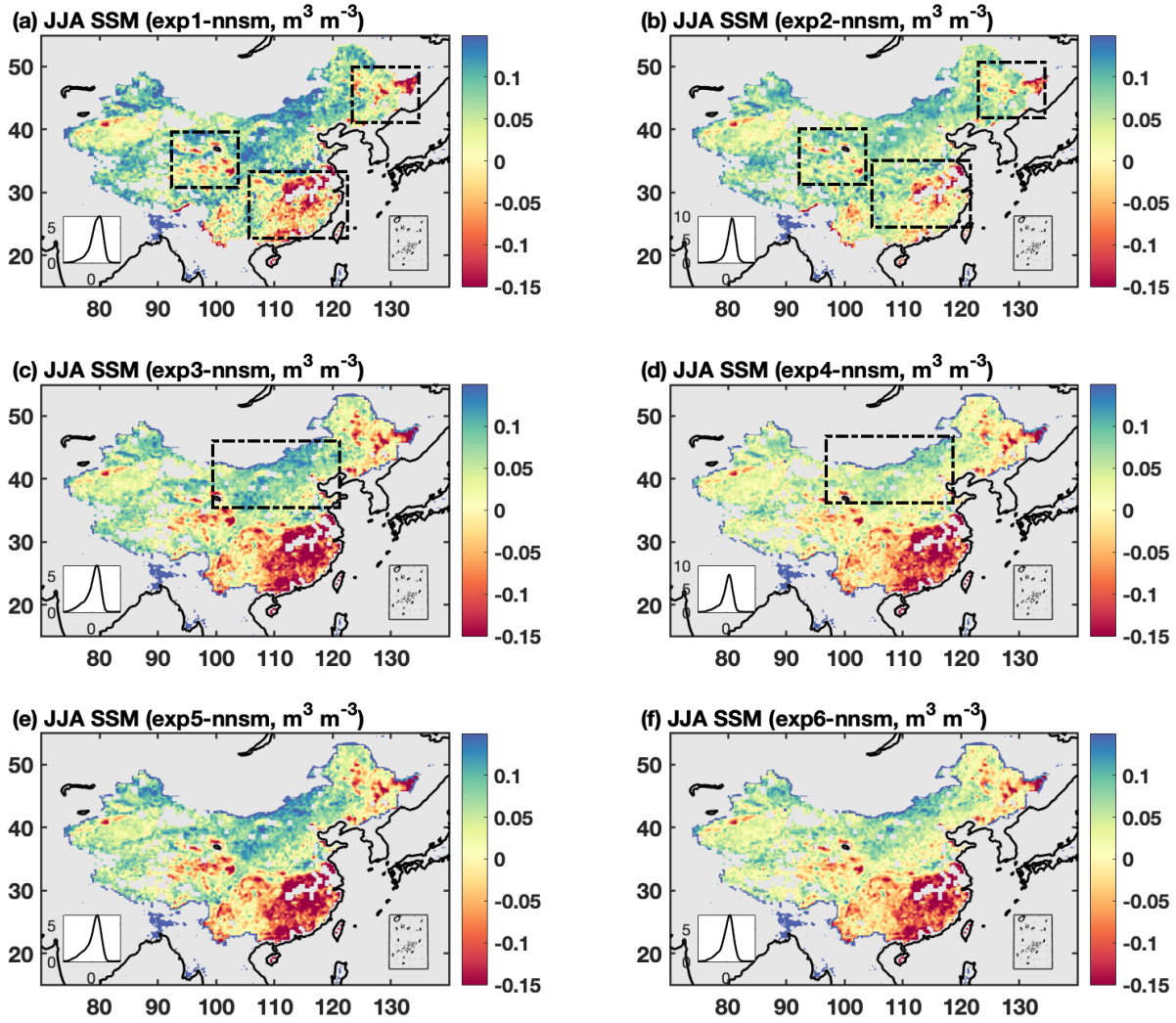

Figure 7. Difference of JJA soil moisture from simulations with baseline soil maps and NNsm (left column), and simulations with optimized soil maps and NNsm (right column). Panels from top to bottom refers to simulations using GSDE soil map with SR06 PTF scheme, GSDE soil map with CH78 PTF scheme, and HWSD soil map with CH78 PTF scheme. Inserted boxes with dash lines indicate areas where soil moisture simulations are substantially improved using the optimized

soil map results. Insets on the leftbottom corner indicate the probability density distribution. Areas with water bodies are masked.

Another thing noticed in particular when compared with ITPLDAS is that large biases in northeastern areas also exist in exp1, exp3 and exp5, and using the optimized soil texture dataset does not visibly help to reduce the biases. The relative moderate improvement over southeastern and northeastern areas could be again because over these areas where the land surface is covered by dense vegetation, the soil texture may play a secondary role in regulating SM dynamics. However, we should also note that the different performance of how soil texture can improve SM in comparison with ITPLDAS may also be caused by the behavior of the reference SM data itself. ITPLDAS is an assimilation system where the SiB2 LSM is incorporated (K. Yang et al., 2020), therefore the output SM product may unavoidably be influenced by the prescribed model parameters (e.g., land surface parameters such as vegetation height, leaf reflectance, surface roughness, etc.), especially over areas with high vegetation density. For example, SiB2 only uses 9 vegetation types (Table 2 in (Sellers et al., 1996)) while the vegetation category in Noah-MP is 27 (this option is based on USGS's land cover classification; the other option is based on MODIS product, which has 20 categories). This means the representation of vegetation spatial heterogeneity in SiB2 is much coarser, which may yield biases in the ITPLDAS product when compared with simulations from Noah-MP. Even for similar vegetation types, their parameters differ a lot. For example, the canopy base and top height for Broadleaf Evergreen Forests (BEF) are 1m and 35m in SiB2 (Table 5 in (Sellers et al., 1996)), while 8m and 20m in Noah-MP. Similar differences are also found in other vegetation biomes. These differences indicate that improving soil texture only may not necessarily help to reduce the simulated SM biases. Instead, additional efforts should be addressed to improve the characterization of other important land surface parameters.

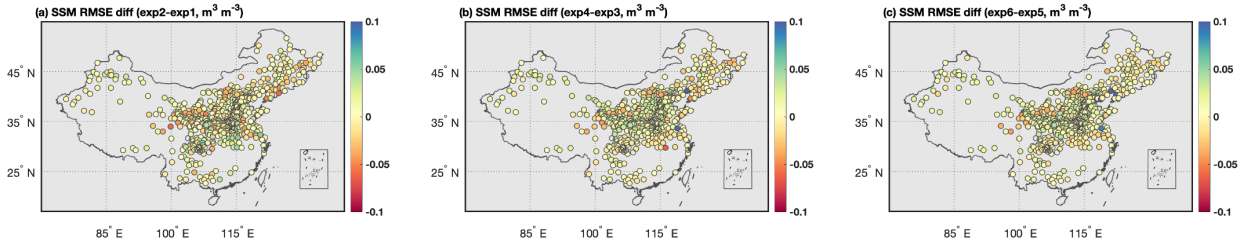

Figure 8. RMSE difference of simulated soil moisture with in-situ observations between cases using optimized soil maps and default soil maps. Panel (a) is for RMSE difference between the simulation using baseline and optimized GSDE soil map with SR06 scheme; (b) is for RMSE difference between simulations using baseline and optimized GSDE soil maps with CH78 scheme; and (c) is for RMSE difference between simulations using baseline and optimized HWSD soil maps with CH78 scheme. Red color means the soil moisture simulations using optimized soil maps outperform those with baseline soil maps.

We further compared the SM simulation with 732 in-situ observations of a half-monthly time scale. Nearly at half of the stations, RMSEs of SM simulated with the optimized soil texture dataset are improved (Figure 8). Even though that at the other half of the stations the SM RMSEs are slightly increased in simulation with optimized soil texture, this result further proves that the SM can be improved using the optimized soil texture data at the in-situ scale. The RMSE increase

at the other half station could possibly be explained by the scale mismatch between the model's grid-based result and the point scale in-situ observations.

3.4 Intercomparison of Results between Different Optimization and Simulation Cases

In previous sections we describe the overall behaviors of soil texture, soil thresholds and SM improvement in different optimization/simulation cases and show that they are comparably consistent. However, there are still some inter-case differences that may be caused by the different baseline soil data or PTF functions used, and therefore need to be examined in detail. In this section, we will elucidate how the optimized soil texture, soil thresholds, and simulated SM results behave between different optimization and simulation cases.

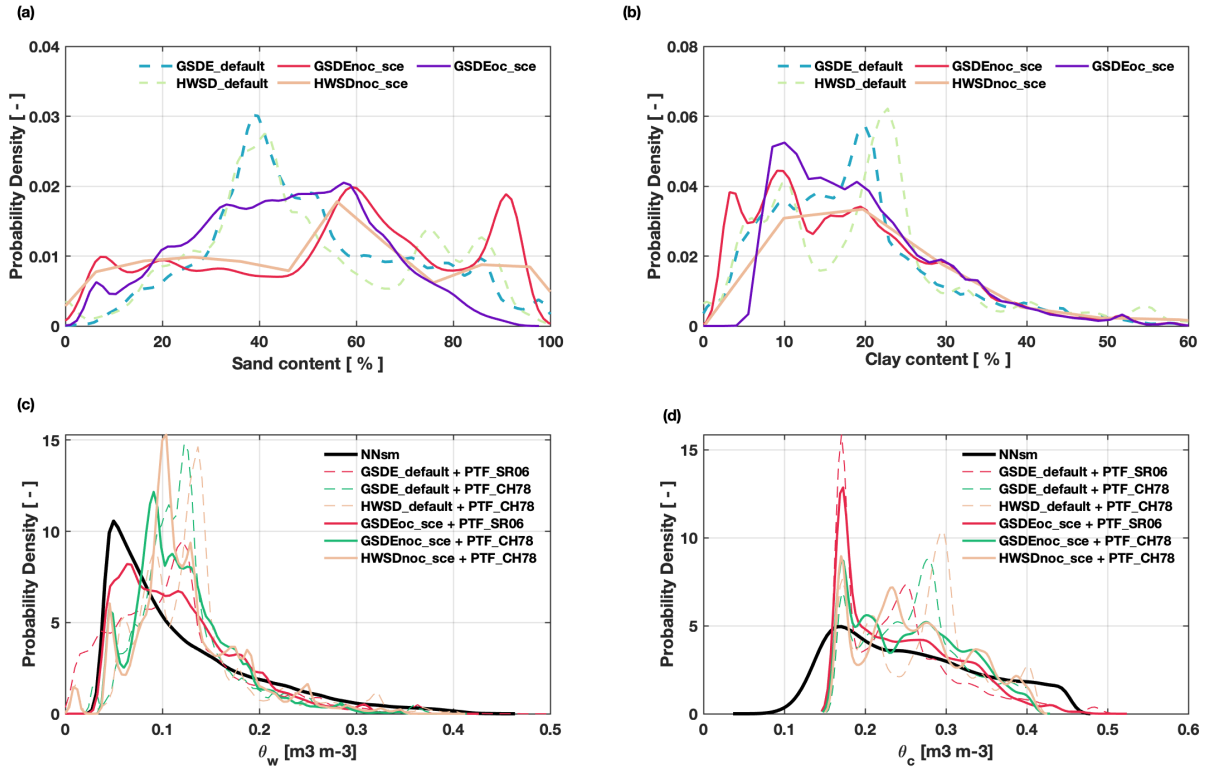

Figure 9. Intercomparison of soil texture (a and b) and soil thresholds (c and d) for difference cases (global). For (a) and (b), dash lines indicate probability density function (PDF) of soil texture from default soil maps and solid lines indicate result from optimized soil maps. For (c) and (d), dash lines indicate soil thresholds calculated using default soil maps and PTF functions, black solid lines indicate estimations from NNsm product, and solid lines in color indicate results from optimized soil maps and PTF functions. See legend in each panel for explanation of lines in different color.

Figure 9 and Figure 10 show the intercomparison of PDF distribution of soil texture data, soil thresholds and SM simulations between different cases. Compared to all other cases, the optimization case using the baseline soil data of GSDE and PTF function considering SOC effect (Opti_exp1) and the corresponding simulation case (exp2) shows the closest performance of soil thresholds and SM simulation compared to satellite observations respectively (Figure 9, c – e, solid

red lines). The optimization case using GSDE soil data but the PTF function without considering the SOC effect shows the second closest result to satellite observations in terms of SM simulation and soil thresholds (Figure 9, c – e, green solid lines). The comparison between the above two cases suggests that the choice of PTF is essential in the soil texture optimization procedure, and a PTF considering the SOC effect may help improve the soil texture optimization performance. Since there is no directly available satellite-observed soil texture data for comparison, the above analyses may also suggest the soil sand and clay distribution from the optimized soil texture using GSDE and SOC-based PTF – that is, unimodal distribution of both sand and clay but with sand content more frequently occurring between 30%-60% while clay content 8%-30% (Figure 9, a – b, solid violet lines) – should outperform other results and serve as a better reference for soil texture assessment studies.

The optimization case uses the same non-PTF scheme as in Opti_exp2 but the HWSD soil data show the most inferior performance of soil thresholds estimation as well as SM simulation compared to satellite observations (Figure 9, c – e, orange solid lines; only among cases using optimized soil data). Comparison between cases using soil data without optimization, however, shows consistent results (Figure 9, c – e, dash lines). The comparison here reflects inconsistent behaviors (thus large uncertainties) of current soil texture datasets, and our results show that utilization of GSDE soil texture data could produce more realistic soil thresholds and SM simulation in LSM.

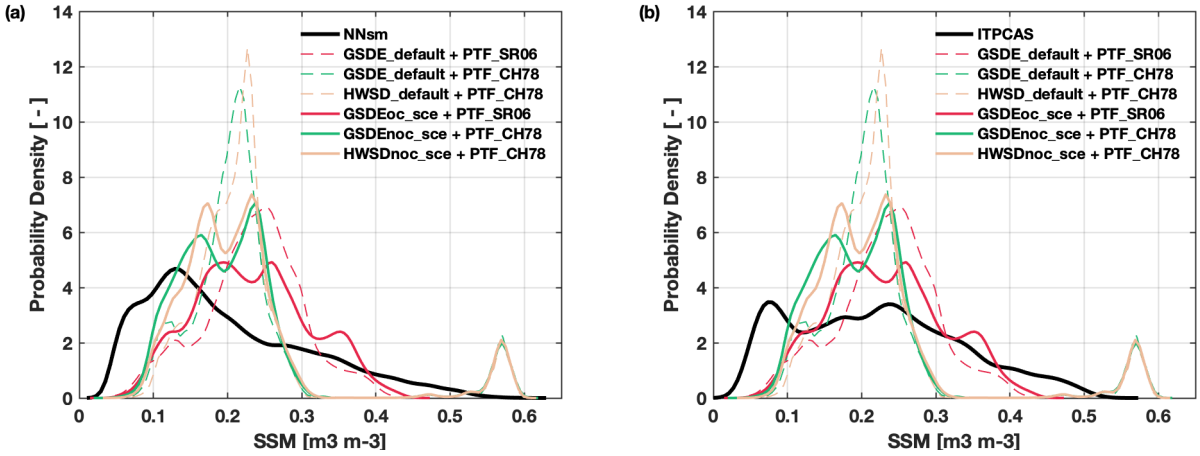

Figure 10. Intercomparison of soil moisture simulations between different cases. Dash lines in color indicates results simulated with default soil maps and PTF functions; solid lines in color indicate results simulated with optimized soil maps and PTF functions. Panel (a) and (b) are the same except for the black solid line in (a) is soil moisture from NNsm while in (b) is from ITPLDAS.

4 Discussions

4.1 Limited ET Improvement in Noah-MP Optimized Experiment

In addition to SM, ET serves as another hot but challenging topic in land surface modeling studies since one important role of LSMs is to provide boundary conditions (i.e., flux) to the upper

layer Atmospheric Circulation Models (ACMs) or the coupled Global Climate Models (GCMs). In theory, the incorporation of observation-constrained soil texture data in LSMs should lead to improved ET simulation results. Therefore, here we compare the simulated ET results with one observation-based ET product, i.e., Penman-Monteith-Leuning (PML) version 2 ET product (Zhang et al., 2019) in order to see how soil texture influences the ET results. In addition to the total ET results, different components of ET, i.e., bare soil evaporation (E_b) and plant transpiration (E_t) are also compared respectively, since in LSMs the plant transpiration processes are not solely determined by soil parameters (e.g., they are also controlled by vegetation parameters such as stomatal resistance, plant hydraulic parameters etc.).

Results show that the total ET result is only minorly different between experiments with baseline and optimized soil texture data, and both of them compare underestimated to PMLv2 (Figure 11). Focusing on ET only, the result indicates the incorporation of observation-constrained soil texture only has limited influence on the model's total ET simulation. However, the comparison of E_b and E_t shows that E_b in experiment with optimized soil texture is substantially improved by ~40mm/yr (measured by medians) compared to PMLv2 bare soil product, whereas E_t simulations are almost consistent in both experiments. This result suggests that the soil texture does indeed have an impact on the ET, but the impact is limited to the bare soil evaporation process only. Figure 11 also shows an overall underestimation of E_t/ET ratio in all simulation cases. The results are consistent with several previous studies (Dong et al., 2022; Zhou et al., 2023), in which the E_t/ET ratio biases are found to be dominated by the wrong SM-ET coupling relationship. As

such, to further improve the total ET simulation, the representation of the model's vegetation parameters as well as the SM-ET coupling regime should be corrected.

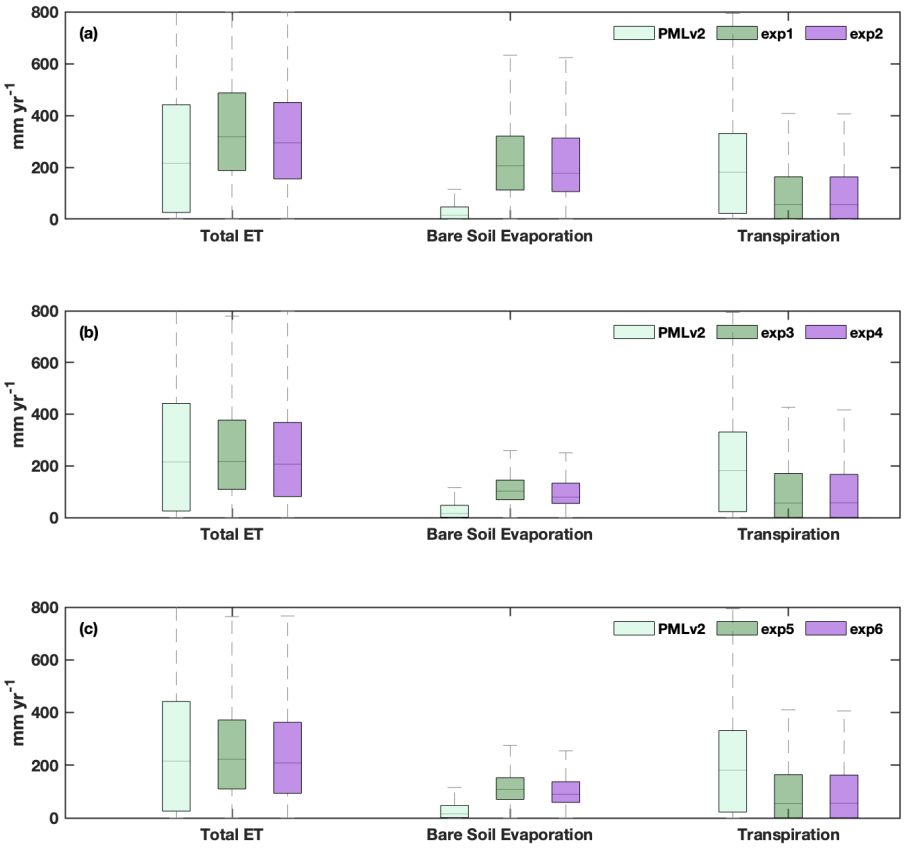

Figure 11. Comparison of annual accumulated total ET , bare soil evaporation, and plant transpiration with PMLv2 between different simulation cases over China. Panel (a) is for simulations using GSDE baseline and optimized soil maps and SR06 PTF schemes; (b) is for simulations using GSDE baseline and optimized soil maps and CH78 PTF schemes; and (c) is for simulations using HWSD baseline soil maps and CH78 PTF schemes.

4.2 Limitations of This Study

While the method this study provides has many advantages and does indeed improve the land surface simulation results, several limitations of the method should be noted as precautions for further application. First, the two soil thresholds, θ_w and θ_c, may contain high uncertainty over extremely humid and arid areas, respectively. In humid regions, the soil could be persistently wet so that it may never dry down to its wilting point; similarly the soil in arid regions can be consistently dry so that it never rises up to reach θ_c. Therefore over these areas the derived soil texture results may still be subject to large biases, although using long-term (i.e., 20-year) soil moisture products may help reduce such uncertainty.

Second, to provide a strict constraint on the soil hydraulic parameters – soil texture relationship, more soil hydraulic parameters should be incorporated with the optimization

549 procedure, e.g., incorporating all the 7 parameters on the lefthand side of the PTF functions (Table
550 S1). However, only two of them were applied here because of the non-availability of other
551 parameters when this study was conducted. We did consider using the 20-year maximum satellite
552 SM as the saturated point (θ_{sat}) for further analyses. The idea was finally abandoned since satellite
553 sensors can hardly detect the moment when SM achieves its saturated point. Moreover, recently
554 we also noticed that there are some studies providing large-scale datasets of saturated soil
555 hydraulic conductivity (κ_{sat}, cited though still not published when this manuscript was drafted).
556 While their quality may need to be validated, these parameters mentioned above can indeed
557 provide insight for improvement on the soil texture calibration work in the future.

558 Finally, we note that only two soil texture datasets and two PTF functions are tested within
559 the optimization framework here. The inter-case comparison analyses already show the
560 optimization results can be sensitive to the baseline soil texture products as well as the PTF
561 functions. Therefore, additional optimization cases including more soil texture products and PTF
562 functions are necessary for a comprehensive evaluation of the optimized soil texture results.

563 4.3 Implications for Soil Texture Calibration Using Satellite Datasets

564 As we noted in the introduction part, soil texture serves as the fundamental parameter in
565 LSMs, while current soil texture datasets may be prone to large biases when applied to the large
566 spatial extent since they are often extrapolated from point-scale in-situ measurements, and
567 therefore need constraint by large-scale soil information. Satellite-observed soil moisture products
568 indeed provide the large-scale observation-based constraint on soil texture, yet seldom has been
569 done to bridge the missing link, likely because the issue of scale gaps has not gained sufficient
570 attention within the land surface modeling community in the past. Nowadays as we are in the big
571 data era where spatial heterogeneity of model outputs can be especially important for local
572 environmental solutions, such scale gap problem has to be considered carefully in the model
573 development.

574 So far, the only study (to our best knowledge) on incorporating satellite soil moisture to
575 soil texture calibration in LSMs is Zhao et al. 2021, in which the calibrated soil texture is derived
576 by converting the best-screened soil type from the Noah-MP look-up table that can produce the
577 closest soil moisture simulation to SMAP through FAO soil fraction triangle. Using GSDE soil
578 texture as the reference dataset ("truth"), the calibrated soil texture outperforms its baseline soil
579 texture dataset (i.e., GLDAS-Noah soil texture, on which the Noah-MP soil type look-up table is
580 based). Our study resembles Zhao et al. 2021 in that we both apply an "effect-to-cause" procedure
581 to derive soil texture, which can be seen as a backward process compared to the traditional "from
582 soil texture to soil hydraulic parameters to soil moisture" approach. Such a backward approach
583 should be valued in future large-scale soil texture calibration studies since the availability of the
584 effect variables can be more easily accessed with the development of new land surface theories
585 and data products.

586 However, we should note that although Zhao et al. 2021 does serve as a pioneering study
587 that motivates the conceptualization of our work, it does also contain several limitations, which
588 may impede the applications of the derived soil texture dataset. The biggest disadvantage is that
589 the result is very much limited by the soil type look-up table, where only 12 available soil types
590 are presented for each model grid. That is to say the soil texture fraction at each grid can only be

591 selected from a fixed number of soil sand, clay and loam combinations. In the real world, it is
592 certainly possible that there can be other soil fraction combinations beyond the 12 available
593 choices. In comparison, the optimization procedure our study used here can have as many soil
594 fraction combinations as possible (within the variation upper and lower bounds), therefore the
595 derived results may be closer to the real soil conditions.

596 **5 Conclusions**

597 This study set out to improve soil moisture simulation in Noah-MP LSM by using soil
598 texture datasets constrained by satellite-derived soil thresholds (i.e., θ_w and θ_c). Soil moisture
599 thresholds are first estimated from satellite-identified soil drydown events. The SCE-UA algorithm
600 and two PTF schemes are then chosen to conduct optimization procedure based on two baseline
601 soil texture datasets, i.e., GSDE and HWSD. The baseline and optimized soil texture datasets are
602 then incorporated with Noah-MP LSM to conduct SM simulations. The model simulation results
603 show that using the optimized soil texture data can indeed improve SM simulations in Noah-MP,
604 and utilization of GSDE soil data optimized with SOC-based PTF scheme can produce the most
605 significant SM improvement.

606 The inter-case comparison shows that there is large uncertainty in the current global soil
607 texture database. Since soil texture often serves as the prescribed parameter in LSMs, such
608 uncertainty will lead to systematical model biases, that is, even though all the other physical
609 parameterizations or parameters were perfect, the model simulations could still be unrealistic due
610 to the biased representation of soil texture. Our study, by using 20-yr NNSM SM data, shows that
611 satellite soil moisture data has the potential to constrain such uncertainty. The improved soil
612 texture in this study could therefore serve as a new database to be employed in LSMs for a more
613 accurate land surface simulation. However, we recognize this study could only serve as a testbed.
614 To build a more robust satellite-optimized soil texture database, further experiments using different
615 satellite products (e.g., SMAP, SMOS, CCI etc.), global soil database (e.g., Data and Information
616 System of International Geosphere-Biosphere Programme， IGBP-DIS), and PTF schemes are
617 required and inter-case comparison analyses are highly expected.

618 Our study also provides a method to improve land surface simulations from the perspective
619 of land-atmosphere interactions. Previous LSM improvement studies prefer incorporating the
620 “state” variable (e.g., land surface temperature, leaf area index, soil moisture etc.) to improve land
621 surface simulations. However, such improvement may be limited (especially for improvement in
622 flux simulations) since the state variables could only poorly capture the near-surface atmosphere’s
623 impact on or response to the land surface. By comparison, soil moisture drydown used in this study
624 directly reflects the coupling between SM and the key flux variable ET, therefore it provides more
625 “dynamic” information to be incorporated into LSM. Although only the bare soil evaporation is
626 significantly improved in the results, this study does show that incorporating land-atmosphere
627 coupling information should serve as an essential part in LSM improvement studies. Recent
628 research also shows that the wrong representation of the SM-ET coupling regime is the dominant
629 reason for land surface simulation biases (Dong et al. 2022; Zhou et al. 2023), which further
630 confirms our conclusion here. Joint efforts considering both state variables and land-atmosphere

interactions could therefore complement each other and facilitate the better LSM simulations in future.

Acknowledgments

This study is supported by the Second Tibetan Plateau Scientific Expedition and Research Program (STEP, grant No. 2019QZKK0206) and the International Parternership Program of Chinese Academy of Sciences (182211KYSB20200015). Q.He and T.Oki also appreciates funding from the Japan Society for the Promotion of Science (KAKENHI), Grant 21H05002 and the Environment Research and Technology Development Fund of the Environmental Restoration and Conservation Agency of Japan, Grant JPMEERF20202005.

Open Research

HWSD soil texture data is available from https://daac.ornl.gov/SOILS/guides/HWSD.html .

GSDE soil texture data is available from http://globalchange.bnu.edu.cn/research/soilw .

Half-monthly in-situ soil moisture measuremtnes at 10cm soil layer is available at https://figshare.com/articles/dataset/Half-monthly_in-situ_SM_measurements_at_10cm_depth/22757663 .

The below data are available at National Tibetan Plateau Data Center (English version available):

NNsm soil moisture data is available at https://data.tpdc.ac.cn/zh-hans/data/c26201fc-526c-465d-bae7-5f02fa49d738 . CMFD meteorological forcing data is available at https://data.tpdc.ac.cn/en/data/8028b944-daaa-4511-8769-965612652c49/ . PMLv2 ET data is available at https://data.tpdc.ac.cn/en/data/48c16a8d-d307-4973-abab-972e9449627c/ .

ITPLDAS soil moisture data is available at https://data.tpdc.ac.cn/zh-hans/data/add77205-ad37-4b93-bfe8-25618529229a .

References

Akbar, R., Short Gianotti, D.J., McColl, K.A., Haghighi, E., Salvucci, G.D., Entekhabi, D., 2018. Estimation of Landscape Soil Water Losses from Satellite Observations of Soil Moisture. J. Hydrometeor. 19, 871–889. https://doi.org/10.1175/JHM-D-17-0200.1

Balsamo, G., Agusti-Panareda, A., Albergel, C., Arduini, G., Beljaars, A., Bidlot, J., Blyth, E., Bousserez, N., Boussetta, S., Brown, A., Buizza, R., Buontempo, C., Chevallier, F., Choulga, M., Cloke, H., Cronin, M.F., Dahoui, M., De Rosnay, P., Dirmeyer, P.A., Drusch, M., Dutra, E., Ek, M.B., Gentine, P., Hewitt, H., Keeley, S.P.E., Kerr, Y., Kumar, S., Lupu, C., Mahfouf, J.-F., McNorton, J., Mecklenburg, S., Mogensen, K., Muñoz-Sabater, J., Orth, R., Rabier, F., Reichle, R., Ruston, B., Pappenberger, F., Sandu, I., Seneviratne, S.I., Tietsche, S., Trigo, I.F., Uijlenhoet, R., Wedi, N., Woolway, R.I., Zeng, X., 2018. Satellite and In Situ Observations for Advancing Global Earth Surface Modelling: A Review. Remote Sensing 10, 2038. https://doi.org/10.3390/rs10122038

Black, C.A. (Ed.), 1965. Methods of Soil Analysis: Part 1 Physical and Mineralogical Properties, Including Statistics of Measurement and Sampling, Agronomy Monographs. American Society of Agronomy, Soil Science Society of America, Madison, WI, USA. https://doi.org/10.2134/agronmonogr9.1

Boone, A., Habets, F., Noilhan, J., Clark, D., Dirmeyer, P., Fox, S., Gusev, Y., Haddeland, I., Koster, R., Lohmann, D., Mahanama, S., Mitchell, K., Nasonova, O., Niu, G.-Y., Pitman, A., Polcher, J., Shmakin, A.B., Tanaka, K., van den Hurk, B., Vérant, S., Verseghy, D., Viterbo, P., Yang, Z.-L., 2004. The Rhône-Aggregation Land Surface Scheme Intercomparison Project: An Overview. J. Climate 17, 187–208. https://doi.org/10.1175/1520-0442(2004)017<0187:TRLSSI>2.0.CO;2

Chen, L., Li, Y., Chen, F., Barr, A., Barlage, M., Wan, B., 2016. The incorporation of an organic soil layer in the Noah-MP land surface modeland its evaluation over a boreal aspen forest. Atmos. Chem. Phys. 16, 8375–8387. https://doi.org/10.5194/acp-16-8375-2016

Chen, Y., Yang, K., Zhou, D., Qin, J., Guo, X., 2010. Improving the Noah Land Surface Model in Arid Regions with an Appropriate Parameterization of the Thermal Roughness Length. J. Hydrometeor. 11, 995–1006. https://doi.org/10.1175/2010JHM1185.1

Clapp, R.B., Hornberger, G.M., 1978. Empirical equations for some soil hydraulic properties. Water Resour. Res. 14, 601–604. https://doi.org/10.1029/WR014i004p00601

Cohen, W.B., Maiersperger, T.K., Turner, D.P., Ritts, W.D., Pflugmacher, D., Kennedy, R.E., Kirschbaum, A., Running, S.W., Costa, M., Gower, S.T., 2006. MODIS land cover and LAI collection 4 product quality across nine sites in the western hemisphere. IEEE Trans. Geosci. Remote Sensing 44, 1843–1857. https://doi.org/10.1109/TGRS.2006.876026

Dai, Y., Shangguan, W., Wei, N., Xin, Q., Yuan, H., Zhang, S., Liu, S., Lu, X., Wang, D., Yan, F., 2019. A review of the global soil property maps for Earth system models. SOIL 5, 137–158. https://doi.org/10.5194/soil-5-137-2019

Dirmeyer, P.A., Gao, X., Zhao, M., Guo, Z., Oki, T., Hanasaki, N., 2006a. GSWP-2: Multimodel Analysis and Implications for Our Perception of the Land Surface. Bull. Amer. Meteor. Soc. 87, 1381–1398. https://doi.org/10.1175/BAMS-87-10-1381

Dirmeyer, P.A., Koster, R.D., Guo, Z., 2006b. Do Global Models Properly Represent the Feedback between Land and Atmosphere? J. Hydrometeor 7, 1177–1198. https://doi.org/10.1175/JHM532.1

Dirmeyer, P.A., Wu, J., Norton, H.E., Dorigo, W.A., Quiring, S.M., Ford, T.W., Santanello, J.A., Bosilovich, M.G., Ek, M.B., Koster, R.D., Balsamo, G., Lawrence, D.M., 2016. Confronting Weather and Climate Models with Observational Data from Soil Moisture Networks over the United States. Journal of Hydrometeorology 17, 1049–1067. https://doi.org/10.1175/JHM-D-15-0196.1

Dong, J., Lei, F., Crow, W.T., 2022. Land transpiration-evaporation partitioning errors responsible for modeled summertime warm bias in the central United States. Nat Commun 13, 336. https://doi.org/10.1038/s41467-021-27938-6

Duan, Q., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resour. Res. 28, 1015–1031. https://doi.org/10.1029/91WR02985

Duan, Q., Sorooshian, S., Gupta, V.K., 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. Journal of Hydrology 158, 265–284. https://doi.org/10.1016/0022-1694(94)90057-4

Duan, Q.Y., Gupta, V.K., Sorooshian, S., 1993. Shuffled complex evolution approach for effective and efficient global minimization. J Optim Theory Appl 76, 501–521. https://doi.org/10.1007/BF00939380

Entekhabi, D., Njoku, E.G., O'Neill, P.E., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Entin, J.K., Goodman, S.D., Jackson, T.J., Johnson, J., Kimball, J., Piepmeier, J.R., Koster, R.D., Martin, N., McDonald, K.C., Moghaddam, M., Moran, S., Reichle, R., Shi, J.C., Spencer, M.W., Thurman, S.W., Tsang, L., Van Zyl, J., 2010. The Soil Moisture Active Passive (SMAP) Mission. Proc. IEEE 98, 704–716. https://doi.org/10.1109/JPROC.2010.2043918

Garrigues, S., Lacaze, R., Baret, F., Morisette, J.T., Weiss, M., Nickeson, J.E., Fernandes, R., Plummer, S., Shabanov, N.V., Myneni, R.B., Knyazikhin, Y., Yang, W., 2008. Validation and intercomparison of global Leaf Area Index products derived from remote sensing data: GLOBAL LAI PRODUCTS INTERCOMPARISON. J. Geophys. Res. 113, n/a-n/a. https://doi.org/10.1029/2007JG000635

Gee, G.W., Bauder, J.W., 2018. Particle-size Analysis, in: Klute, A. (Ed.), SSSA Book Series. Soil Science Society of America, American Society of Agronomy, Madison, WI, USA, pp. 383–411. https://doi.org/10.2136/sssabookser5.1.2ed.c15

Gettelman, A., Geer, A.J., Forbes, R.M., Carmichael, G.R., Feingold, G., Posselt, D.J., Stephens, G.L., van den Heever, S.C., Varble, A.C., Zuidema, P., 2022. The future of Earth system prediction: Advances in model-data fusion. Sci. Adv. 8, eabn3488. https://doi.org/10.1126/sciadv.abn3488

He, Q., Lu, H., Yang, K., 2023. Soil Moisture Memory of Land Surface Models Utilized in Major Reanalyses Differ Significantly From SMAP Observation. Earth's Future 11, e2022EF003215. https://doi.org/10.1029/2022EF003215

He, Q., Lu, H., Yang, K., Zhao, L., Zou, M., 2021. Improving Land Surface Temperature Simulation of NOAH-MP on the Tibetan Plateau, in: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. Presented at the IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Brussels, Belgium, pp. 6217–6220. https://doi.org/10.1109/IGARSS47720.2021.9555059

Kerr, Y.H., Waldteufel, P., Wigneron, J.-P., Martinuzzi, J., Font, J., Berger, M., 2001. Soil moisture retrieval from space: the Soil Moisture and Ocean Salinity (SMOS) mission. IEEE Trans. Geosci. Remote Sensing 39, 1729–1735. https://doi.org/10.1109/36.942551

Kolassa, J., Reichle, R.H., Koster, R.D., Liu, Q., Mahanama, S., Zeng, F., 2020. An Observation-Driven Approach to Improve Vegetation Phenology in a Global Land Surface Model. J Adv Model Earth Syst 12. https://doi.org/10.1029/2020MS002083

Koster, R.D., Liu, Q., Mahanama, S.P.P., Reichle, R.H., 2018. Improved Hydrological Simulation Using SMAP Data: Relative Impacts of Model Calibration and Data Assimilation. Journal of Hydrometeorology 19, 727–741. https://doi.org/10.1175/JHM-D-17-0228.1

Koster, R.D., Reichle, R.H., Mahanama, S.P.P., 2017. A Data-Driven Approach for Daily Real-Time Estimates and Forecasts of Near-Surface Soil Moisture. Journal of Hydrometeorology 18, 837–843. https://doi.org/10.1175/JHM-D-16-0285.1

Kumar, S.V., M. Mocko, D., Wang, S., Peters-Lidard, C.D., Borak, J., 2019. Assimilation of Remotely Sensed Leaf Area Index into the Noah-MP Land Surface Model: Impacts on Water and Carbon Fluxes and States over the Continental United States. Journal of Hydrometeorology 20, 1359–1377. https://doi.org/10.1175/JHM-D-18-0237.1

Kumar, S.V., Peters-Lidard, C.D., Santanello, J.A., Reichle, R.H., Draper, C.S., Koster, R.D., Nearing, G., Jasinski, M.F., 2015. Evaluating the utility of satellite soil moisture retrievals over irrigated areas and the ability of land data assimilation methods to correct for unmodeled processes. Hydrol. Earth Syst. Sci. 19, 4463–4478. https://doi.org/10.5194/hess-19-4463-2015

Lawrence, D.M., Slater, A.G., 2008. Incorporating organic soil into a global climate model. Clim Dyn 30, 145–160. https://doi.org/10.1007/s00382-007-0278-1

Li, C., Lu, H., Leung, L.R., Yang, K., Li, H., Wang, W., Han, M., Chen, Y., 2019. Improving Land Surface Temperature Simulation in CoLM Over the Tibetan Plateau Through Fractional Vegetation Cover Derived From a Remotely Sensed Clumping Index and Model-Simulated Leaf Area Index. J. Geophys. Res. Atmos. 124, 2620–2642. https://doi.org/10.1029/2018JD028640

Lu, H., Zheng, D., Yang, K., Yang, F., 2020. Last-decade progress in understanding and modeling the land surface processes on the Tibetan Plateau. Hydrol. Earth Syst. Sci. 24, 5745–5758. https://doi.org/10.5194/hess-24-5745-2020

Malik, M.J., van der Velde, R., Vekerdy, Z., Su, Z., 2012. Assimilation of Satellite-Observed Snow Albedo in a Land Surface Model. Journal of Hydrometeorology 13, 1119–1130. https://doi.org/10.1175/JHM-D-11-0125.1

764 McColl, K.A., He, Q., Lu, H., Entekhabi, D., 2019. Short-Term and Long-Term Surface Soil Moisture Memory
765 Time Scales Are Spatially Anticorrelated at Global Scales. J. Hydrometeor. 20, 1165–1182.
766 https://doi.org/10.1175/JHM-D-18-0141.1
767 McColl, K.A., Wang, W., Peng, B., Akbar, R., Short Gianotti, D.J., Lu, H., Pan, M., Entekhabi, D., 2017. Global
768 characterization of surface soil moisture drydowns: SURFACE SOIL MOISTURE DRYDOWN
769 ANALYSIS. Geophys. Res. Lett. 44, 3682–3690. https://doi.org/10.1002/2017GL072819
770 Miralles, D.G., Gentine, P., Seneviratne, S.I., Teuling, A.J., 2019. Land-atmospheric feedbacks during droughts and
771 heatwaves: state of the science and current challenges: Land feedbacks during droughts and heatwaves.
772 Ann. N.Y. Acad. Sci. 1436, 19–35. https://doi.org/10.1111/nyas.13912
773 Miralles, D.G., Teuling, A.J., van Heerwaarden, C.C., Vilà-Guerau de Arellano, J., 2014. Mega-heatwave
774 temperatures due to combined soil desiccation and atmospheric heat accumulation. Nature Geosci 7, 345–
775 349. https://doi.org/10.1038/ngeo2141
776 Mukherjee, S., Mishra, A., Trenberth, K.E., 2018. Climate Change and Drought: a Perspective on Drought Indices.
777 Current Climate Change Reports 4, 145–163. https://doi.org/10.1007/s40641-018-0098-x
778 Nachtergaele, F., Velthuizen, H. van, Batjes, N.H., Fischer, G., 2009. The Harmonized World Soil Database
779 (Techinical Report).
780 Niu, G.-Y., Yang, Z.-L., Mitchell, K.E., Chen, F., Ek, M.B., Barlage, M., Kumar, A., Manning, K., Niyogi, D.,
781 Rosero, E., Tewari, M., Xia, Y., 2011. The community Noah land surface model with
782 multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale
783 measurements. J. Geophys. Res. 116, D12109. https://doi.org/10.1029/2010JD015139
784 Njoku, E.G., Jackson, T.J., Lakshmi, V., Chan, T.K., Nghiem, S.V., 2003. Soil moisture retrieval from AMSR-E.
785 IEEE Trans. Geosci. Remote Sensing 41, 215–229. https://doi.org/10.1109/TGRS.2002.808243
786 Oki, T., Kanae, S., 2006. Global Hydrological Cycles and World Water Resources. Science 313, 1068–1072.
787 https://doi.org/10.1126/science.1128845
788 Rawls, W.J., Nemes, A., Pachepsky, Ya., 2004. Effect of soil organic carbon on soil hydraulic properties, in:
789 Developments in Soil Science. Elsevier, pp. 95–114. https://doi.org/10.1016/S0166-2481(04)30006-1
790 Rawls, W.J., Pachepsky, Y.A., Ritchie, J.C., Sobecki, T.M., Bloodworth, H., 2003. Effect of soil organic carbon on
791 soil water retention. Geoderma 116, 61–76. https://doi.org/10.1016/S0016-7061(03)00094-6
792 Saxton, K.E., Rawls, W.J., 2006. Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic
793 Solutions. Soil Sci. Soc. Am. J. 70, 1569–1578. https://doi.org/10.2136/sssaj2005.0117
794 Sellers, P.J., Tucker, C.J., Collatz, G.J., Los, S.O., Justice, C.O., Dazlich, D.A., Randall, D.A., 1996. A Revised
795 Land Surface Parameterization (SiB2) for Atmospheric GCMs. Part II: The Generation of Global Fields of
796 Terrestrial Biophysical Parameters from Satellite Data. Journal of Climate 9, 706–737.
797 https://doi.org/10.1175/1520-0442(1996)009<0706:ARLSPF>2.0.CO;2
798 Seneviratne, S.I., Corti, T., Davin, E.L., Hirschi, M., Jaeger, E.B., Lehner, I., Orlowsky, B., Teuling, A.J., 2010.
799 Investigating soil moisture–climate interactions in a changing climate: A review. Earth-Science Reviews
800 99, 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004
801 Shangguan, W., Dai, Y., Duan, Q., Liu, B., Yuan, H., 2014. A global soil data set for earth system modeling. J. Adv.
802 Model. Earth Syst. 6, 249–263. https://doi.org/10.1002/2013MS000293
803 Sun, J., Chen, Y., Yang, K., Lu, H., Zhao, L., Zheng, D., 2021. Influence of organic matter on soil hydrothermal
804 processes in the Tibetan Plateau: Observation and parameterization. Journal of Hydrometeorology.
805 https://doi.org/10.1175/JHM-D-21-0059.1
806 Teuling, A.J., Uijlenhoet, R., van den Hurk, B., Seneviratne, S.I., 2009. Parameter Sensitivity in LSMs: An Analysis
807 Using Stochastic Soil Moisture Models and ELDAS Soil Parameters. Journal of Hydrometeorology 10,
808 751–765. https://doi.org/10.1175/2008JHM1033.1
809 Wan, Z., 2013. Collection-6 MODIS Land Surface Temperature Products Users' Guide.
810 Wang, A., Shi, X., 2019. A Multilayer Soil Moisture Dataset Based on the Gravimetric Method in China and Its
811 Characteristics. Journal of Hydrometeorology 20, 1721–1736. https://doi.org/10.1175/JHM-D-19-0035.1
812 Wösten, J.H.M., Pachepsky, Ya.A., Rawls, W.J., 2001. Pedotransfer functions: bridging the gap between available
813 basic soil data and missing soil hydraulic characteristics. Journal of Hydrology 251, 123–150.
814 https://doi.org/10.1016/S0022-1694(01)00464-4
815 Xiao, Z., Wang, T., Liang, S., Sun, R., 2016. Estimating the Fractional Vegetation Cover from GLASS Leaf Area
816 Index Product. Remote Sensing 8, 337. https://doi.org/10.3390/rs8040337
817 Yang, K., Chen, Y., He, J., Zhao, L., Lu, H., Qin, J., Zheng, D., Li, X., 2020. Development of a daily soil moisture
818 product for the period of 2002–2011 in Chinese mainland. Sci. China Earth Sci. 63, 1113–1125.
819 https://doi.org/10.1007/s11430-019-9588-5

Yang, K., Chen, Y.-Y., Qin, J., 2009. Some practical notes on the land surface modeling in the Tibetan Plateau. Hydrol. Earth Syst. Sci. 13, 687–701. https://doi.org/10.5194/hess-13-687-2009

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

Yang, Z.-L., Niu, G.-Y., Mitchell, K.E., Chen, F., Ek, M.B., Barlage, M., Longuevergne, L., Manning, K., Niyogi, D., Tewari, M., Xia, Y., 2011. The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins. J. Geophys. Res. 116, D12109. https://doi.org/10.1029/2010JD015139

Yang, Z.-L., Zhao, L., He, Y., Wang, B., 2020. Perspectives for Tibetan Plateau data assimilation. National Science Review 7, 495–499. https://doi.org/10.1093/nsr/nwaa014

Yao, P., Lu, H., Shi, J., Zhao, T., Yang, K., Cosh, M.H., Gianotti, D.J.S., Entekhabi, D., 2021. A long term global daily soil moisture dataset derived from AMSR-E and AMSR2 (2002–2019). Sci Data 8, 143. https://doi.org/10.1038/s41597-021-00925-8

Yin, J., Zhan, X., Zheng, Y., Hain, C.R., Ek, M., Wen, J., Fang, L., Liu, J., 2016. Improving Noah land surface model performance using near real time surface albedo and green vegetation fraction. Agricultural and Forest Meteorology 218–219, 171–183. https://doi.org/10.1016/j.agrformet.2015.12.001

Zhang, Y., Kong, D., Gan, R., Chiew, F.H.S., McVicar, T.R., Zhang, Q., Yang, Y., 2019. Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017. Remote Sensing of Environment 222, 165–182. https://doi.org/10.1016/j.rse.2018.12.031

Zheng, D., van der Velde, R., Su, Z., Wang, X., Wen, J., Booij, M.J., Hoekstra, A.Y., Chen, Y., 2015. Augmentations to the Noah Model Physics for Application to the Yellow River Source Area. Part I: Soil Water Flow. Journal of Hydrometeorology 16, 2659–2676. https://doi.org/10.1175/JHM-D-14-0198.1

Zhou, J., Yang, K., Crow, W.T., Dong, J., Zhao, L., Feng, H., Zou, M., Lu, H., Tang, R., Jiang, Y., 2023. Potential of remote sensing surface temperature- and evapotranspiration-based land-atmosphere coupling metrics for land surface model calibration. Remote Sensing of Environment 291, 113557. https://doi.org/10.1016/j.rse.2023.113557

Zhou, S., Williams, A.P., Berg, A.M., Cook, B.I., Zhang, Y., Hagemann, S., Lorenz, R., Seneviratne, S.I., Gentine, P., 2019. Land–atmosphere feedbacks exacerbate concurrent soil drought and atmospheric aridity. Proc Natl Acad Sci USA 201904955. https://doi.org/10.1073/pnas.1904955116

Figure 1. The figure caption should begin with an overall descriptive statement of the figure followed by additional text. They should be immediately after each figure. Figure parts are indicated with lower-case letters (**a, b, c…**). For initial submission, please place both the figures and captions in the text near where they are cited rather than at the end of the file (not both). At revision, captions can be placed in-text or at the end of the file, and figures should be uploaded separately. Each figure should be one complete, cohesive file (please do not upload sub-figures or figure parts in separate files).

Table 1. Start this caption with a short description of your table. Format tables using the Word Table commands and structures. Additional information on table formatting can be found in our Style Guide, Table Formatting. Do not create tables using spaces or tabs characters. Large tables presenting rich data should be presented as separate excel or .csv files, not as part of the main text.