Water Resources Research

RESEARCH ARTICLE
10.1029/2022WR032336

Key Points:

- State-of-the-art digital soil maps improve the agreement between soil moisture modeling and observations
- Including vertical soil heterogeneity improves the accuracy of simulated soil moisture
- Simulating soil moisture using soil hydraulic properties computed from contemporary pedotransfer functions generally outperforms classic lookup tables

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

C. Xu,
chengcheng.xu@duke.edu

Citation:

Xu, C., Torres-Rojas, L., Vergopolan, N., & Chaney, N. W. (2023). The benefits of using state-of-the-art digital soil properties maps to improve the modeling of soil moisture in land surface models. *Water Resources Research*, *59*, e2022WR032336. https://doi.org/10.1029/2022WR032336

Received 10 MAR 2022
Accepted 9 MAR 2023

The Benefits of Using State-Of-The-Art Digital Soil Properties Maps to Improve the Modeling of Soil Moisture in Land Surface Models

Chengcheng Xu[1], Laura Torres-Rojas[1], Noemi Vergopolan[2,3], and Nathaniel W. Chaney[1]

[1]Department of Civil and Environmental Engineering, Duke University, Durham, NC, USA, [2]Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, NJ, USA, [3]NOAA Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA

Abstract This study assesses the added value of using emerging maps of soil properties to improve surface soil moisture simulations using the HydroBlocks land surface model with different soil hydraulic parameterization schemes. Simulations were run at an hourly 30-m resolution between 2012 and 2019 and evaluated against U.S. Climate Reference Network measurements. The results show that state-of-the-art soil properties maps (POLARIS and SoilGrids250m V2.0) improve the accuracy of simulated surface soil moisture when compared to the STATSGO-derived CONUS-SOIL map. Contemporary pedotransfer functions (multi-linear regression and Artificial Neural Networks-based) also improve model performance in comparison to the lookup table-derived soil parameterization schemes. The addition of vertical heterogeneity to the soil properties further improves the mean Kling-Gupta efficiency by 0.04 and lowers the mean Root mean square error by 0.003 over the CONUS. This study demonstrates that land surface modeling can be improved by using state-of-the-art maps of soil properties, accounting for the vertical heterogeneity of soils, and advancing the use of contemporary pedotransfer functions.

Plain Language Summary This work studied if advanced soil properties maps can improve soil moisture modeling using a Land Surface Model. The model was run over 7 years and was compared to site measurements. The results showed that using contemporary soil properties maps and pedotransfer functions to estimate soil properties improved model performance, especially for soils with different layers. This study demonstrated that improving hydraulic modeling is possible by using soil properties maps and contemporary pedotransfer functions in the setting of different vertical layers.

1. Introduction

A wide range of hydrological, meteorological, and biogeochemical processes in the Earth system influence and are influenced by soil moisture dynamics (Falloon et al., 2011; Sanchez-Mejia & Papuga, 2014; Seneviratne et al., 2010). Indeed the monitoring and managing of soil moisture has significant applications in many fields, such as precision agriculture, weather forecasting, crop yield, runoff control, soil erosion, and geotechnical construction (Adamchuk et al., 2004; Dirmeyer & Halder, 2016; Keesstra et al., 2016; Vergopolan et al., 2021). Not surprisingly then, there remains a continued interest to accurately track and predict its spatiotemporal dynamics (Beck et al., 2021; Vergopolan et al., 2020). The three primary approaches that are currently used to monitor soil moisture include in situ observations (W. A. Dorigo et al., 2011; W. Dorigo et al., 2021), satellite remote sensing (Wagner et al., 1999; Wigneron et al., 2007), and land surface models (LSMs; Maurer et al., 2002).

The LSM approach is especially appealing given that it does not require observations of soil moisture and instead it is modeled using more readily available topographic, meteorological, soils, and vegetation data (Chaney et al., 2015); however, its reliability is strongly tied to the uncertainties of these data (Baroni et al., 2017; Chaney et al., 2015; Koster et al., 2009; Loosvelt et al., 2011; Teuling et al., 2009). The behavior of soil moisture modeling is especially sensitive to soil hydraulic properties (e.g., porosity) (Arsenault et al., 2018; Cai et al., 2014); this is complicated further by the large difficulty in prescribing soil hydraulic parameters since they are difficult to obtain from in situ measurements and even more difficult to directly retrieve from remote sensing (Hogue et al., 2006). Instead, a suite of methods is used to assemble soil properties maps for land surface models including leveraging spatial data sets of more readily available soil information (e.g., soil texture) and to then use pedotransfer functions (PTFs) to estimate soil hydraulic properties. To this day, the most common approach in LSMs is to derive soil hydraulic properties via lookup tables; spatial maps of soil classes are prescribed to the model and the look-up tables are used to assign soil hydraulic properties through soil classes (e.g., USDA soil order). However, these lookup tables are known to be deficient and are not able to fully leverage the plethora of environmental data and observations of soil hydraulic properties (Kishné et al., 2017).

With the increasing availability of environmental information and soil observations, the emergence of digital soil mapping products (DSM—georeferenced rasterized maps of soil classes and properties; (McBratney et al., 2003)) enables the creation of improved soil hydraulic properties in LSMs with the following advantages: (a) improved accuracy and spatial resolution of continental to global maps of soil texture, organic matter, and bulk density, among others (e.g., 250-m SoilGrids, 30-m POLARIS; (Chaney, Wood, et al., 2016; Chaney et al., 2019; Hengl et al., 2017; Poggio et al., 2021)), (b) contemporary PTFs that are able to fully leverage state-of-the-art DSM information to improve estimates of soil hydraulic properties, and (c) enhanced accuracy of vertical soil properties in LSMs. More detail on each of these advantages is provided below.

The emergence of hyper-resolution LSMs enables the framework to address land heterogeneity and hydraulic processes at finer spatial resolutions (Chaney, Metcalfe, Wood, 2016; Vergopolan et al., 2022; Wood et al., 2011, p. 20), however, these efforts necessitate higher fidelity and spatial resolution of the model inputs; especially soil hydraulic parameters when accurate soil moisture modeling is required (Pinheiro & van Lier, 2021). In comparison to baseline soil data in LSMs (e.g., 1 km; Miller & White, 1998), state-of-the-art soil properties maps provide an increasingly detailed and accurate representation of soil properties (e.g., 30-m, 250-m; (Chaney et al., 2019; Poggio et al., 2021, p. 2)). These new DSM products have been developed through a series of efforts, such as expanding the use of soil profile data (e.g., 240000 standardized soil profile data globally); choosing model features from 400 global environmental covariates; and aggregating and harmonizing legacy soil surveys (Chaney et al., 2019; Poggio et al., 2021). All these products are made with the goal of improving the spatial estimates of soil properties (Chaney et al., 2019; Poggio et al., 2021).

PTFs in turn can leverage the soil properties from DSM products (e.g., sand, silt, clay, organic matter, and bulk density) to predict soil hydraulic properties, which are often not measured nor retrievable from remote sensing data (Schaap et al., 1998, 2001; Wösten et al., 1999). Lookup tables can be seen as the simplest form of class-based PTFs (Van Looy et al., 2017) that ignore inner-class variability (Schaap et al., 2001); contemporary PTFs that could be used as continuous PTFs and replace look-up tables including multi-linear regression (e.g., Saxton & Rawls, 2006) and Artificial Neural Networks (ANNs; Minasny & McBratney, 2002), which have been used in cutting-edge maps of soil hydraulic properties (Chaney et al., 2019; Simons et al., 2020; Tóth et al., 2017). Despite many studies evaluating the performance of PTFs across scales and study areas (Nemes et al., 2003; Schaap & Leij, 1998; Weihermüller et al., 2021), the assessment of implementing state-of-the-art soil properties maps with hydraulic properties in soil moisture modeling remains unclear, particularly for large-scale soil properties maps (e.g., continental-scale). Consequently, quantifying the benefits of such soil hydraulic parameterization schemes is important. Besides, future PTFs and DSM products can be revised and improved accordingly.

Another benefit of using DSM products in LSMs is the improved representation of vertical heterogeneity. Modern DSM products often map soil properties over multiple depth intervals. Such heterogeneity of soil properties affects soil moisture estimates by influencing soil infiltration (Beven & Germann, 1982) and redistribution of soil-water content (i.e., evaporation, root water uptake; Liu et al., 2020). Contemporary LSMs are able to include vertically layered soil profiles (Clark & Gedney, 2008; Collins et al., 2011; Essery et al., 2003; Niu et al., 2011); however, until now, data has generally lacked adequate parameterizations of these profiles. Hence, there is a need to investigate the added value of vertical soil properties available in state-of-the-art DSM products.

This study provides insights into advancing surface soil moisture estimates through the use of contemporary soil properties maps by evaluating the added value of soil parameters derived from three soil properties maps (POLARIS, SoilGrids 250m V2.0, and CONUS-SOIL) (Chaney et al., 2019; Miller & White, 1998; Poggio et al., 2021, p. 202). To this end, a set of experiments were conducted to quantify the dependence of LSM-simulated surface soil moisture on soil parameters via either directly providing soil hydraulic parameters as model inputs, or determining preliminary soil texture classes and later estimating soil hydraulic parameters using a lookup table or contemporary PTFs, with the aim of examining the impact of contemporary PTFs on the modeling of soil moisture.

Table 1
An Overview of the Soil Parameterization Schemes Used in This Study

	Soil mineral and organic components	Vertical characteristic	Acronym	Soil depths (cm)		Soil hydraulic parameterization
1	POLARIS	Vertically homogeneous	HPT	i.	30–60	Lookup table
2	SoilGrids250m V2.0	Vertically homogeneous	HST	i.	30–60	Lookup table
3	CONUS-SOIL	Vertically homogeneous	HCT	i.	30–60	Lookup table
4	POLARIS	Vertically heterogeneous	VPT	i.	0–5	Lookup table
				ii.	5–15	
				iii.	15–30	
				iv.	30–60	
				v.	60–100	
				vi.	100–200	
5	SoilGrids250m V2.0	Vertically homogeneous	HSS	i.	30–60	Saxton and Rawls PTFs
6	POLARIS	Vertically homogeneous	HPS	i.	30–60	Saxton and Rawls PTFs
7	POLARIS	Vertically homogeneous	HP	i.	30–60	SSURGO and NeuroTheta PTFs
8	SoilGrids250m V2.0	Vertically heterogeneous	VSS	i.	0–5	Saxton and Rawls PTFs
				ii.	5–15	
				iii.	15–30	
				iv.	30–60	
				v.	60–100	
				vi.	100–200	
9	POLARIS	Vertically heterogeneous	VPS	i.	0–5	Saxton and Rawls PTFs
				ii.	5–15	
				iii.	15–30	
				iv.	30–60	
				v.	60–100	
				vi.	100–200	
10	POLARIS	Vertically heterogeneous	VP	i.	0–5	SSURGO and NeuroTheta PTFs
				ii.	5–15	
				iii.	15–30	
				iv.	30–60	
				v.	60–100	
				vi.	100–200	

Note. CONUS-SOIL did not include information on organic components.

2. Data and Methods

2.1. Soil Parameterization

This study employed 10 separate soil parameterization schemes in the HydroBlocks LSM. The model was used to simulate the top 5 cm of soil water content to assess the influence of soil properties on the modeling of soil moisture. An overview of the different soil data sets is provided below as well as in Table 1.

2.1.1. POLARIS

Probabilistic remapping of Soil Survey Geographic Database (POLARIS) is a 30-m probabilistic map of soil classification and properties database over the CONUS with data at six GlobalSoilMap standard depth increments (0–5 cm, 5–15 cm, 15–30 cm, 30–60 cm, 60–100 cm, and 100–200 cm) (Chaney, Wood, et al., 2016; Chaney et al., 2019). POLARIS developed built-in soil properties using the Soil Survey Geographic Database (SSURGO) and the National Cooperative Soil Survey Soil Characterization Database (SCD). POLARIS soil hydraulic properties are a combination of hydraulic properties reported in these data and derived hydraulic parameters using the NeuroTheta PTFs, which were assembled using the approach described in Minasny and McBratney (2002). The resulting POLARIS soil properties maps include soil texture fractions (percentage of sand, silt, and clay), organic matter, pH, saturated hydraulic conductivity, Brooks-Corey, and Van Genuchten water retention curve parameters, bulk density, and saturated water content with associated uncertainties (Chaney et al., 2019). POLARIS offers soil hydrological properties for direct use in hydrologic simulation.

2.1.2. SoilGrids250m V2.0

SoilGrids250m V2.0 is a global digital soil class and properties mapping system with a 250 m spatial resolution (Hengl et al., 2017; Poggio et al., 2021, p. 2). The newly developed product (SoilGrids version 2.0) maps and assesses soil properties with associated uncertainties for soil organic carbon content, total nitrogen, coarse fragments, pH (water), cation exchange capacity, bulk density, and texture fractions over six GlobalSoilMap standard depth intervals (Poggio et al., 2021). Since soil hydraulic properties are not included in the SoilGrids system, recent studies have created maps of soil hydraulic parameters generated from SoilGrids, enabling soil moisture estimates in LSMs (Kearney & Maino, 2018; Montzka et al., 2017).

2.1.3. CONUS-SOIL

The multilayer soil characteristic data set over the conterminous United States (CONUS-SOIL) is a grid-based multilayer soil data set derived from the first version of the Natural Resources Conservation Service (NRCS) State Soil Geographic (STATSGO) database (Miller & White, 1998). CONUS-SOIL rasterized STATSGO at a 1-km spatial resolution and standardized vertical soil properties to 11 layers (from 0 to 250 cm). CONUS-SOIL soil properties include soil texture fractions, depth-to-bedrock, bulk density, porosity, rock fragment volume, particle-size fractions, available water capacity, and hydrologic soil group over the CONUS (Miller & White, 1998). This study uses CONUS-SOIL as the baseline data set, as it is the default setting of soil properties (i.e., STATSGO-FAO) in many implementations of LSMs over CONUS (e.g., NLDAS-2, VIC model; (Liang et al., 1994; Xia et al., 2012)).

2.1.4. Lookup Table Derived Soil Hydraulic Parameters (Textured Soils)

A lookup table is the default method used in LSMs to estimate soil hydraulic parameters. This study uses the default soil parameters lookup table utilized by the Noah-MP land surface model (Niu et al., 2011). The given lookup table is applied to each soil properties data set to derive the soil hydraulic properties. The lookup table-derived soil parameters in this study consist of POLARIS texture (PT), SoilGrids250m V2.0 texture (ST), and CONUS-SOIL texture (CT) in both vertically heterogeneous (V) and homogeneous (H) configurations. When necessary, we first determine soil texture classes from soil properties maps and then use the default lookup table approach to connect soil texture class with soil hydraulic parameters for each grid cell. In this work, the lookup table assigns LSM required soil hydraulic parameters based on nineteen CONUS-SOIL categories (12 USDA soil texture orders plus organic material, water, bedrock, other, playa, lava, and white sand).

2.1.5. PTFs Derived Soil Hydraulic Parameters From Textured Parameterization

Two sets of PTFs are used in this study to investigate the impact of PTF selection on soil moisture modeling. The first one is the NeuroTheta PTFs; this approach uses an ensemble of Artificial Neural Networks to estimate Brooks-Corey water retention curve parameters from sand, clay, bulk density, θ_s, θ_{33}, and θ_{1500} at each soil series in POLARIS. The estimates are then mapped in space as weighted averages of each pixel's predicted soil series. NeuroTheta was designed for POLARIS and thus cannot be readily applied to the other data in this study. Note that in POLARIS, saturated soil hydraulic conductivity was calculated using SSURGO's reported values for soil series (after aggregating their corresponding components). NeuroTheta was developed from soil properties provided by the National Cooperative Soil Survey soil characterization database (SCD 2016), UNSODA (Nemes et al., 2001), and GRIZZLY (Haverkamp et al., 1997) data sets. The second PTF used in this study is Saxton and Rawls PTFs (2006); this approach uses a set of multi-linear regressions that estimate soil hydraulic properties from measured soil properties (e.g., soil texture, organic matter content). These statistical relationships were developed from soil properties retrieved from the USDA soil database (Saxton & Rawls, 2006).

2.2. In Situ Surface Soil Moisture Measurements

The U.S. Climate Reference Network (USCRN) is a network of climate monitoring stations over the CONUS, Alaska, and Hawaii. Since 2009, soil moisture and temperature networks have been added to the USCRN at

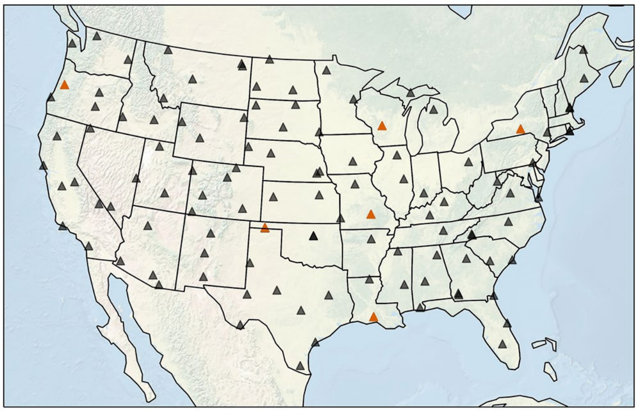

Figure 1. Spatial distribution of 114 USCRN measurement sites over the CONUS. Each triangle represents a USCRN site. Orange sites are chosen to display the time series performance of models; these six sites display a range of environmental conditions and are geographically dispersed over CONUS.

114 sites across the CONUS (Figure 1). USCRN provides hourly, daily, and monthly data on soil temperature and moisture per site using three distinct devices (Bell et al., 2013; Diamond et al., 2013). To enhance data quality, USCRN employs triple sensor redundancy for soil moisture measurements. Most soil probes were positioned at five standard World Meteorological Organization depths over permeable areas. To evaluate the HydroBlocks simulations, the daily mean of triplicate sets of surface 5 cm soil moisture measurements were employed to evaluate model performance in this study. It is important to ensure that the results of the study are robust and not influenced by any correlation among the soil properties maps and USCRN measurements. To this aim, the study examined their correlation and found that the co-located point measurements and soil parameters did not show excessively high correlations. The study also included a quality check of USCRN site measurements, which can be found in the Supporting Information S1 (Figure S1).

2.3. HydroBlocks Land Surface Model

HydroBlocks is a field-scale resolving LSM that quantifies water, energy, and carbon exchanges at the land-atmosphere interface (Chaney, Metcalfe, Wood, 2016; Chaney et al., 2021). As a semi-distributed LSM, HydroBlocks clusters areas with similar hydrological behavior into Hydrologic Response Units (HRUs) and updates the hydrological states at each HRU (Chaney et al., 2021; Vergopolan et al., 2021). The community Noah LSM with multi-parameterization options (Noah-MP) LSM (Niu et al., 2011) is run as a single column in an HRU framework in HydroBlocks. This way, HydroBlocks uses the same core-physics of Noah-MP LSM; and consequently, it can be parameterized with the same lookup table approach and calculate soil water contents using the Richards equation for each soil layer as implemented in Noah-MP LSM (Niu et al., 2011; Richards, 1931).

For this study, HydroBlocks was driven by a stack of high-resolution environmental data sets to give a more detailed picture of sub-grid land heterogeneity. These products include the one arcsec (~30 m) USGS National elevation data set (NED), the one arcsec (~30 m) National Land Cover Database (NLCD), and the 1/32° (~3 km) Princeton CONUS Forcing (PCF) data set that provides meteorological forcing at 1-hr temporal resolution (Gesch et al., 2002; Homer et al., 2015; Pan et al., 2016).

2.4. Hydrological Modeling Experiments

In this study, the HydroBlocks LSM was used to simulate the temporal and spatial distribution of surface soil moisture (top 5 cm) using different parameterizations produced from soil properties maps. Soil moisture deeper than 5 cm was also simulated but not included in the evaluation. Table 1 provides an overview of how the experimental soil information was parameterized. The study used 10 distinct soil parameterization schemes that can be classified into three major categories: (a) different soil properties data sets, (b) vertically homogeneous or heterogeneous parameterization schemes, and (c) parameterizations produced from the lookup table PTF (categorical PTF) and non-lookup table PTFs (multi-linear regression and Artificial Neural Network-based regression).

The vertically homogenous soil profiles replicated Noah-MP LSM's default setup (Niu et al., 2011). In this setup, only the mean values of soil parameters (30–60 cm) were used to represent the entire soil column (0–200 cm), indicating that the soil properties were assumed to be the same throughout the entire soil column. Since hydrological modeling is influenced by organic matter and clay contents, the chosen depth interval is within the root zone where organic matter and clay contents are within reasonable ranges. In addition, this setup was chosen to provide a point of comparison with the other parameterization schemes that consider vertical soil heterogeneity. As with the vertically heterogeneous experiments, the HydroBlocks LSM incorporated vertically discretized soil parameters according to the GlobalSoilMap depth intervals (Arrouays et al., 2014), to be consistent with the studied soil maps.

The HydroBlocks LSM was configured to operate at 114 study stations that were co-located with USCRN sites across the CONUS to enable a direct comparison between the simulated soil moisture and measurements at each individual station. Each study domain had between 200 and 300 HRUs representing the land heterogeneity patterns. HydroBlocks was run at a 30 m spatial resolution with an hourly time step from 2012 to 2019. Given that the vertically heterogeneous soil parameterization required a longer time to reach convergence, a 3-year spin-up period was chosen. Each study domain consisted of a 0.25° cell with an additional 0.1° buffer from each USCRN site location. To provide the most objective evaluation, no model calibration was performed in this work. This approach aims to examine the impacts of different soil parameterization schemes on soil moisture modeling at various locations and environmental conditions and the influence of vertical soil heterogeneity on soil moisture modeling without any bias from model calibration.

2.5. Evaluation Analysis

The 5 cm HydroBlocks simulated soil moisture was evaluated using the Kling-Gupta efficiency (KGE), Root mean square error (RMSE), and unbiased Root mean square error (ubRMSE) in comparison to daily soil moisture measurements from the USCRN. To prevent misinterpretation caused by frozen soils, we masked out data when modeled soil temperature was below 2°C. The quality assessment at the USCRN sites was used to remove time steps that met the following criteria: (a) dielectric value outside the range from 0.1 to 70; (b) temperature range outside −30°C–65°C; (c) probe temperatures below 0.5°C.

RMSE (Equation 1) and ubRMSE (Equation 2) are prevalent metrics to evaluate model performance for soil moisture simulations or retrievals (Chew & Small, 2020; Gruber et al., 2020; Vergopolan et al., 2020). ubRMSE removes the mean bias (Equation 3) from RMSE. Seasonal mean bias is averaged in four seasons: December to February (winter), March to May (spring), June to August (summer), and September to November (autumn). Lower RMSE and ubRMSE values indicate a more precise model fit to soil moisture measurements.

$$\mathrm{RMSE} = \sqrt{E\left((\theta_s - \theta_m)^2\right)} \tag{1}$$

$$\mathrm{ubRMSE} = \sqrt{E(((\theta_s - E(\theta_s)) - (\theta_m - E(\theta_m)))^2)} \tag{2}$$

$$b = E(\theta_s) - E(\theta_m) \tag{3}$$

$$\mathrm{RMSE}^2 = \mathrm{ubRMSE}^2 + b^2 \tag{4}$$

E: Expectation operator
s: Simulations
m: Measurements

KGE is a goodness-of-fit metric primarily designed for hydrological modeling (Gupta et al., 2009; Kling et al., 2012). KGE combines the linear Pearson correlation coefficient (ρ), the bias ratio (β) defined by the ratio of estimated and measured means, and the variability ratio (α) as the ratio of the estimated and measured coefficients of variation:

$$\mathrm{KGE} = 1 - \sqrt{(\rho - 1)^2 + (\beta - 1)^2 + (\alpha - 1)^2} \tag{5}$$

$$\alpha = \frac{\sigma_s / \mu_s}{\sigma_m / \mu_m} \tag{6}$$

$$\beta = \frac{\mu_s}{\mu_m} \tag{7}$$

$$\rho = \frac{\sum (s_i - \mu_s)(m_i - \mu_m)}{\sqrt{\sum (s_i - \mu_s)^2 \sum (m_i - \mu_m)^2}} \tag{8}$$

μ_s, μ_m : Simulation, measurement distribution mean
σ_s, σ_m : Simulation, measurement standard deviation
s_i, m_i : Simulated, measured soil moisture

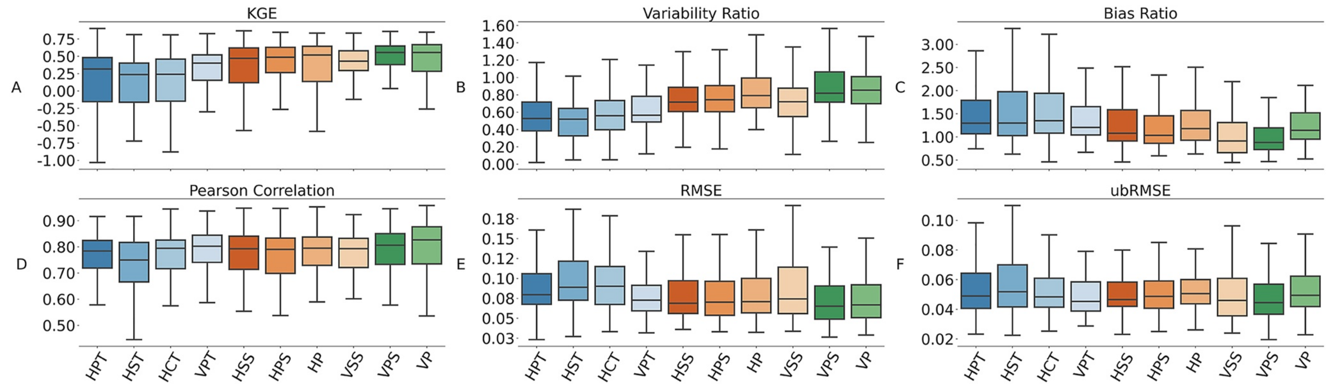

Figure 2. Box plots of (a) Kling-Gupta efficiency scores, (b) variability ratios (α), (c) bias ratios (β), (d) Pearson correlation coefficients (ρ), (e) Root mean square error (m^3/m^3), and (f) ubRMSE (m^3/m^3) for (i) HPT (steel blue), (ii) HST (sky blue), (iii) HCT (light sky blue), (iv) VPT (light cyan), (v) HSS (saddle brown), (vi) HPS (sandy brown), (vii) HP (peach puff), (viii) VSS (linen), (ix) VPS (forest green), (x) VP (lime green) soil parameterization schemes over the CONUS from 2014 to 2019.

KGE ranges from negative infinity (very poor fit) to 1 (perfect fit). A KGE value of one indicates ideal simulations with respect to the measurements. However, given the requirements of optimal values for α, β, and ρ all equal to one, a KGE value of one is very difficult to obtain. α below one means that the model tends to underestimate the normalized simulated temporal variability (e.g., coefficient of variation) of surface soil moisture in the time series. β over one shows that the model overestimates surface soil moisture. A strong positive (or strong negative) ρ implies that simulations and measurements are (strongly) related. KGE and its constituent parts (α, β, and ρ) allowed us to assess the temporal variation in model performance and errors, hence facilitating the detection of model weakness. In this work, metrics between simulated and measured soil moisture were computed only when simulations and measurements were simultaneously available.

3. Results

3.1. Comparison of Soil Properties Maps in Soil Moisture Simulations

HydroBlocks simulations were conducted using 10 alternative soil parameterization schemes (as shown in Table 1) at 114 locations over the CONUS. These simulations were compared to the in situ observations using the model performance indicators described previously (KGE and its components α, β, and ρ; RMSE; ubRMSE). Figure 2e shows that the best-performing parameterization schemes were those that utilized vertically heterogeneous soil properties derived from the POLARIS soil textural components with the Saxton and Rawls PTFs (2006) (VPS) and with soil properties maps (VP), as indicated by the lowest RMSE (0.065 and 0.067, respectively). Vertically homogeneous PTFs-determined parameterization schemes (e.g., HSS, HPS) and POLARIS soil properties maps (HP) demonstrated moderate performance. Experiments driven by vertically homogenous lookup table-derived soil parameterization schemes (HPT, HST, and HCT) showed the worst RMSE ranging from 0.080 to 0.090. Differences in ubRMSE were not substantial (ranging from 0.045 to 0.052) (Figure 2f). The RMSE ranking (Figure 2e) indicated that state-of-the-art soil properties maps (including prediction of both soil physical properties and hydraulic properties), contemporary PTFs, and vertically heterogeneous soil profiles are useful alternatives for lookup tables in soil moisture modeling across the CONUS.

The KGE ranking (Figure 2a) of model performance is consistent with the RMSE ranking. When using the median KGE as the metric for evaluation, the vertically heterogeneous soil properties map (VP) and the contemporary PTF-derived soil parameterization scheme (VPS) performed the best, with a KGE of 0.56. They were followed by the vertically homogeneous soil properties map (HP; 0.52 KGE). VP outperformed others primarily due to a better temporal variability ratio α, demonstrating that models driven by vertical soil properties map can more accurately capture the temporal variation of surface soil moisture (Figure 2b). All parameterization schemes obtained from the contemporary PTFs, that is HPS, HSS, VSS, and VPS, outperformed others in terms of median bias ratio β. Both α and β of vertically homogenous lookup table-derived soil parameterization schemes (HPT, HST, and HCT) were less satisfactory. The replacement of contemporary PTFs (VSS, HSS, VPS, and HPS) with the lookup table (VPT, HPT, HST, and HCT) improved the averaged β by 0.31; and the mean α difference between the vertically homogenous lookup table-derived soils and soil properties maps (HP and VP) was 0.28.

This implies that lookup table parameterizations tend to have a wetter bias and underestimate temporal range of water content change, especially for the vertically homogenous setting.

The analysis of the spatial distribution of each metric offers an understanding of site-specific performance. The site-specific maps of KGE and its components (α, β, and ρ) from each soil parameterization scheme are presented in Figures 3–5. These results also describe the CONUS-wide median metrics and their distribution (insets in the left corner). Figure 3 shows that the experiments exhibited similar β spatial patterns: (a) most sites were wet biased, resulting in a CONUS-wide median of β higher than 1, especially for lookup table-derived soil parameterization schemes (Figures 3a–3d); (b) wet biased sites were primarily located in the western US and along the coasts (blue sites); (c) most parameterizations showed better performance over the central US (white or light-colored sites). The Saxton and Rawls PTFs (2006) derived soil parameterization schemes (HPS, HSS, VSS, and VPS) aided the simulated soil moisture to fluctuate closer to measurements, resulting in a median β that was better than the soil properties maps (VP and HP) and the lookup table-derived soils (VPT, HPT, HCT, and HST).

The temporal variability ratio α of simulated soil moisture with respect to site-level measurements is depicted in Figure 4. The vertical soil properties map (VP) had the best CONUS-wide median α at 0.85. All soil parameterization schemes mostly underestimated the temporal amplitude of surface soil moisture variations ($\alpha < 1$), particularly for the lookup table-derived soil parameterization schemes (HPT, HST, HCT, and VPT), where there was a high prevalence of sites with worse α and positive skewness (Figures 4a–4d). Figure 5 shows maps of Pearson correlation coefficients ρ to quantify the strength and direction of the correlated relationship between simulations and measurements. The POLARIS vertical soil properties map (VP) had the best CONUS-wide median ρ (0.83). Vertical soil texture with contemporary PTFs derived parameterization (VPS) minimized the occurrence of poorly correlated sites ($\rho < 0.4$). All soil parameterization schemes exhibited strong and good linear correlations ($\rho > 0.75$), indicating that most simulations can often capture the wet and dry trend of surface soils.

KGE maps (Figure 6) were computed using the temporal variability ratio α, bias ratio β, and correlation ratio ρ. KGE scores were generally lower for the vertically homogeneous lookup table-derived soils (KGE < 0.32). In comparison, adding in the heterogeneous soils improved model performance at many sites with low KGE scores (KGE < −1). For example, when HPT (Figure 6a) was substituted with VPT (Figure 6d), the red sites in Oregon (John Day site and Riley site), Arizona (Tucson site and Eglin site), and Georgia (Watkinsville site) exhibited an increase in KGE. Changes from lookup table-derived soils to soil properties maps also improve site KGE scores. However, there are occasional exceptions. As an example, the Moose site in Wyoming (near the eastern border of Idaho) showed negative KGE with all soil parameterization schemes. The poor performance was a result of a wet bias of surface soil moisture; this could be due to persistent issues in the soil hydraulic properties or due to other factors such as model forcing data bias or ground-truth measurement errors.

To illustrate the differences in spatial characteristics of soil hydraulic properties, Figure 7 displays saturated hydraulic conductivity (Ksat) patterns for the different parameterization schemes at the Lafayette USCRN site in Louisiana. At this site, silty loam is the dominant soil texture class from 30 to 60 cm; however, the estimated saturated hydraulic conductivity varied for each soil parameterization scheme (Figures 7m–7p). The Saxton and Rawls PTFs (2006) derived soils (VSS and VPS) limited the variation of saturated hydraulic conductivities, while vertical soil properties map and lookup table-derived soils (VP and VPT) led to the central area (blue areas) having higher saturated hydraulic conductivities. The Saxton and Rawls PTFs (2006) derived soil parameterizations estimated that the saturated hydraulic conductivity for each layer was roughly an order of magnitude lower than that of the VPT (lookup table-derived). Since the properties were generated from a lookup table using coarsely categorized soil types (soil order), soil classes-derived saturated hydraulic conductivity maps constrained the complexity of soil heterogeneity, oversimplying the considerable intra-variabilities within the same soil class.

3.2. Seasonal Evaluation of Soil Moisture Dynamics

Soil moisture modeling may exhibit seasonal responses to different soil parameterization schemes. In this study, the model performance results did not show significant discrepancies between seasons. Figure 8 displays the seasonal distribution of simulated daily averaged and measured volumetric soil water contents at the top 5 cm over the CONUS. Each subplot in Figure 8 shows CONUS-wide simulations (each point is a multi-year simula-

10.1029/2022WR032336

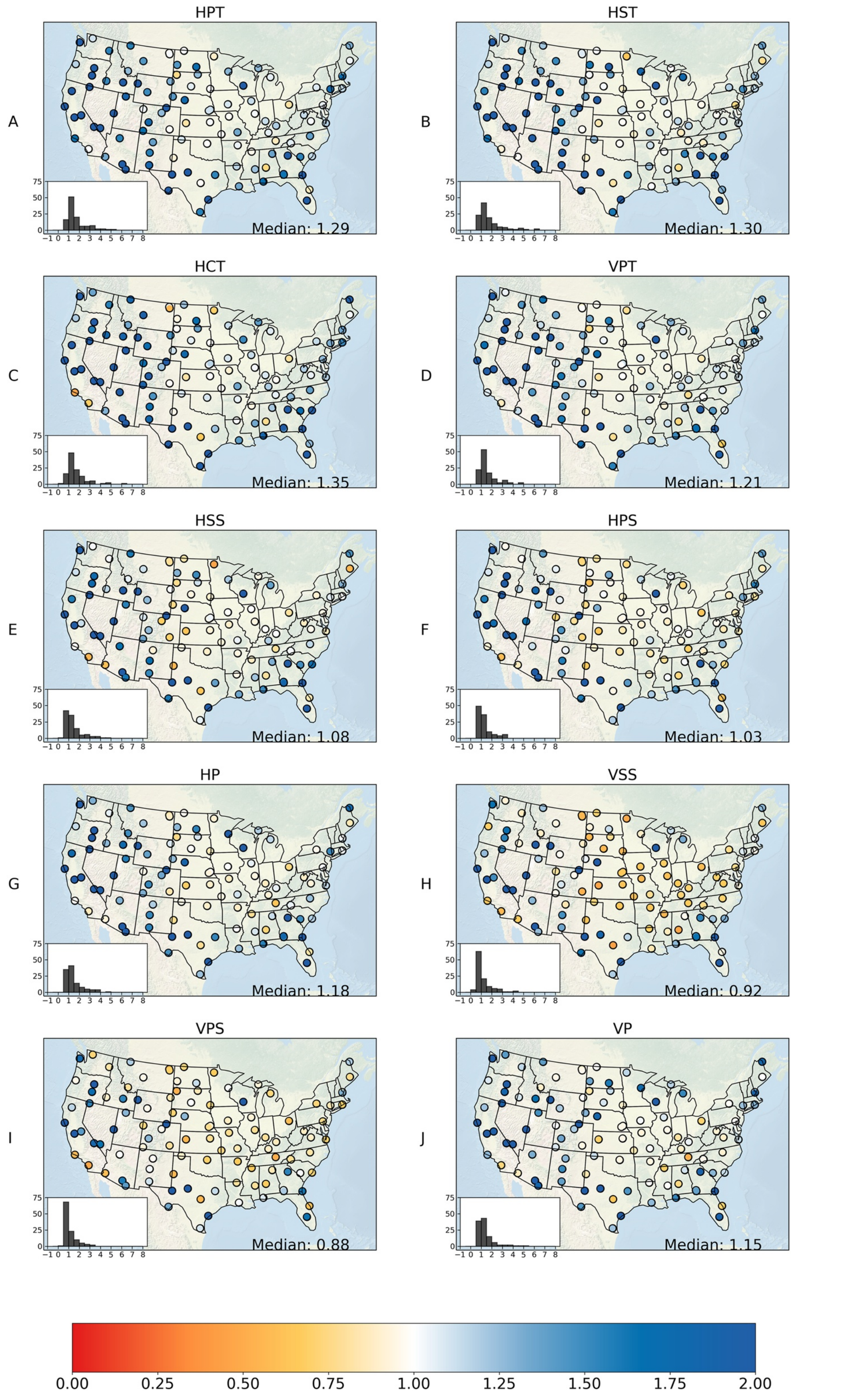

Figure 3. Spatial distribution of bias ratios (β) from (a) HPT, (b) HST, (c) HCT, (d) VPT, (e) HSS, (f) HPS, (g) HP, (h) VSS, (i) VPS, (j) VP simulated soil moisture with HydroBlocks (2014–2019) compared to the USCRN in situ surface 5 cm soil moisture contents. In each subplot, the left corner displays the frequency of bias ratios (β).

10.1029/2022WR032336

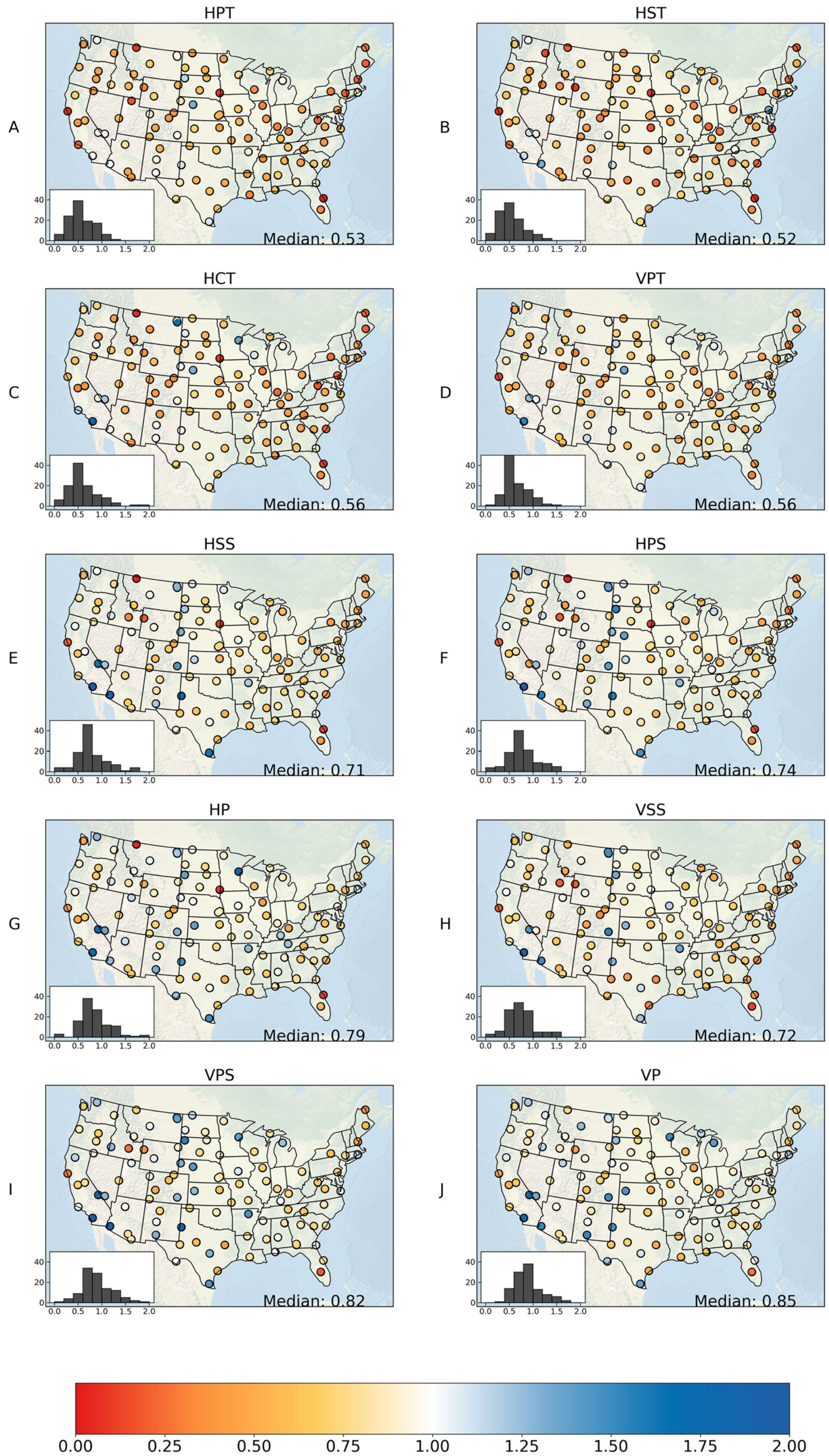

Figure 4. Spatial distribution of variability ratios (α) from (a) HPT, (b) HST, (c) HCT, (d) VPT, (e) HSS, (f) HPS, (g) HP, (h) VSS, (i) VPS, (j) VP HydroBlocks simulation of soil moisture (2014–2019) compared to the USCRN in situ surface 5 cm soil moisture contents. In each subplot, the left corner displays the frequency of variability ratios (α).

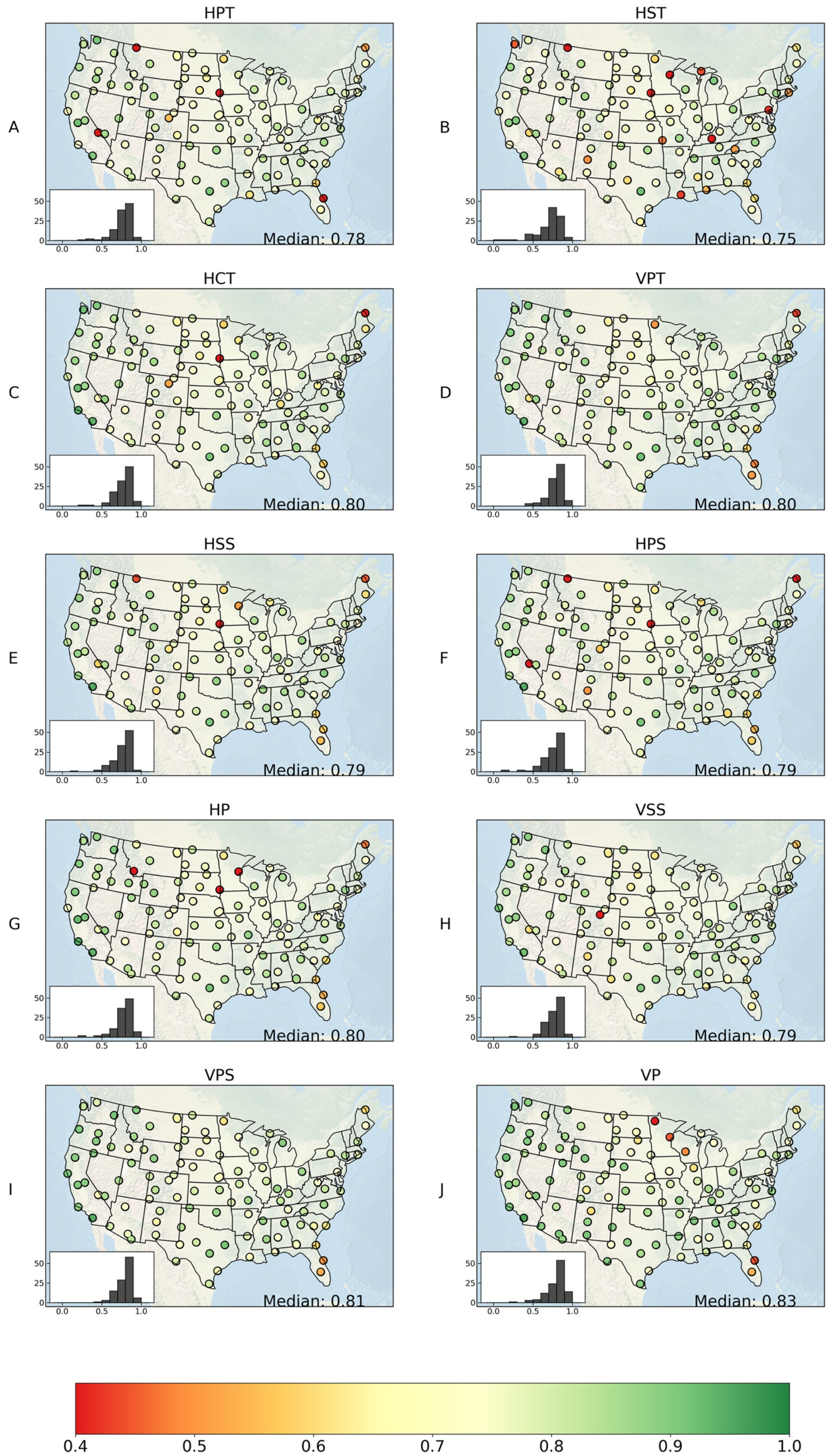

Figure 5. Spatial distribution of Pearson correlation coefficients (ρ) from (a) HPT, (b) HST, (c) HCT, (d) VPT, (e) HSS, (f) HPS, (g) HP, (h) VSS, (i) VPS, (j) VP soil moisture HydroBlocks simulation (2014–2019) compared to the USCRN in situ surface 5 cm soil moisture contents. In each subplot, the left corner displays the occurrence of Pearson correlation coefficients (ρ).

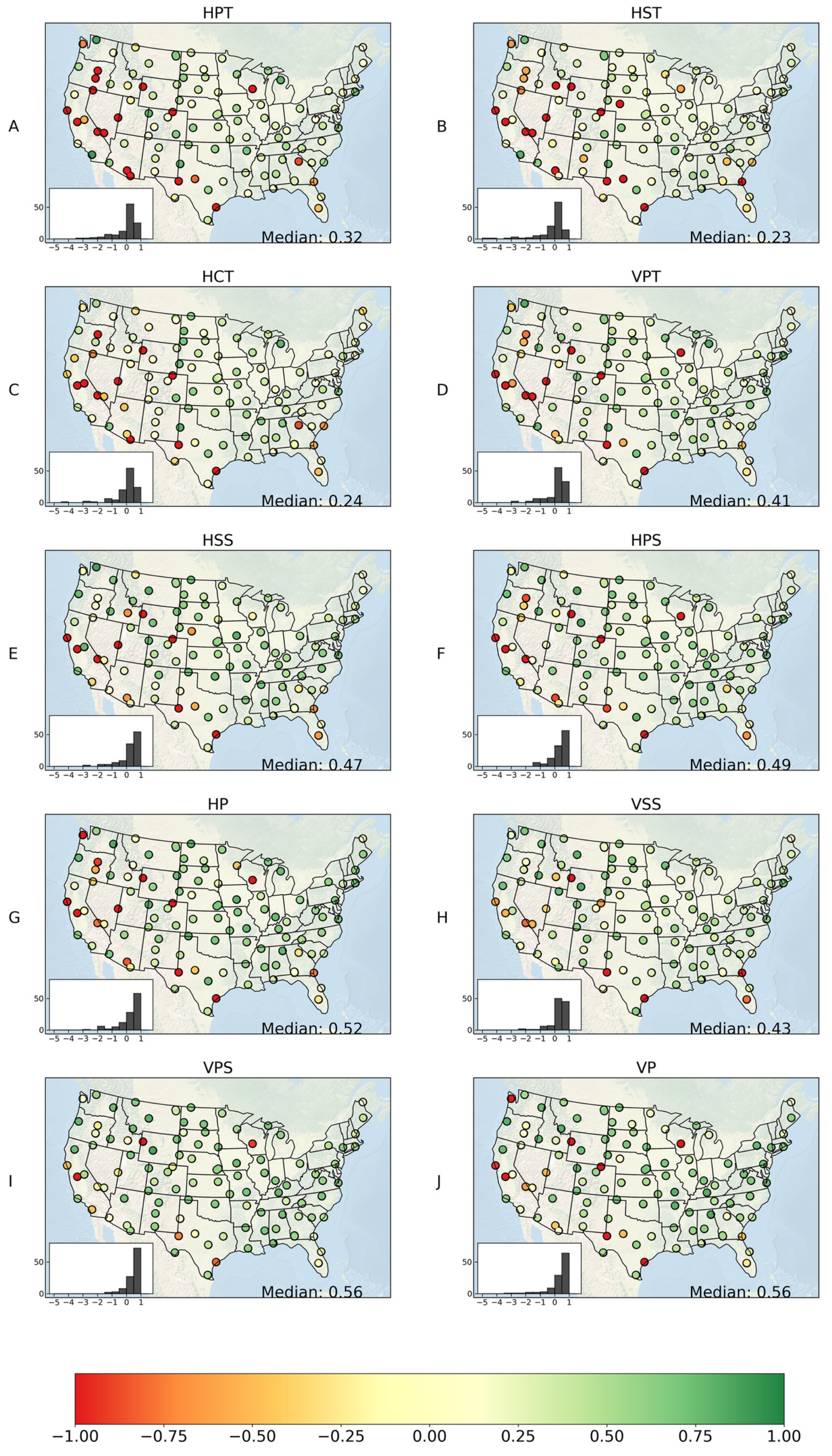

Figure 6. Spatial distribution of Kling-Gupta efficiency (KGE) scores from (a) HPT, (b) HST, (c) HCT, (d) VPT, (e) HSS, (f) HPS, (g) HP, (h) VSS, (i) VPS, (j) VP HydroBlocks simulation of soil moisture (2014–2019) compared to the USCRN in situ surface 5 cm soil moisture contents. In each subplot, the left corner displays the frequency of KGE scores.

10.1029/2022WR032336

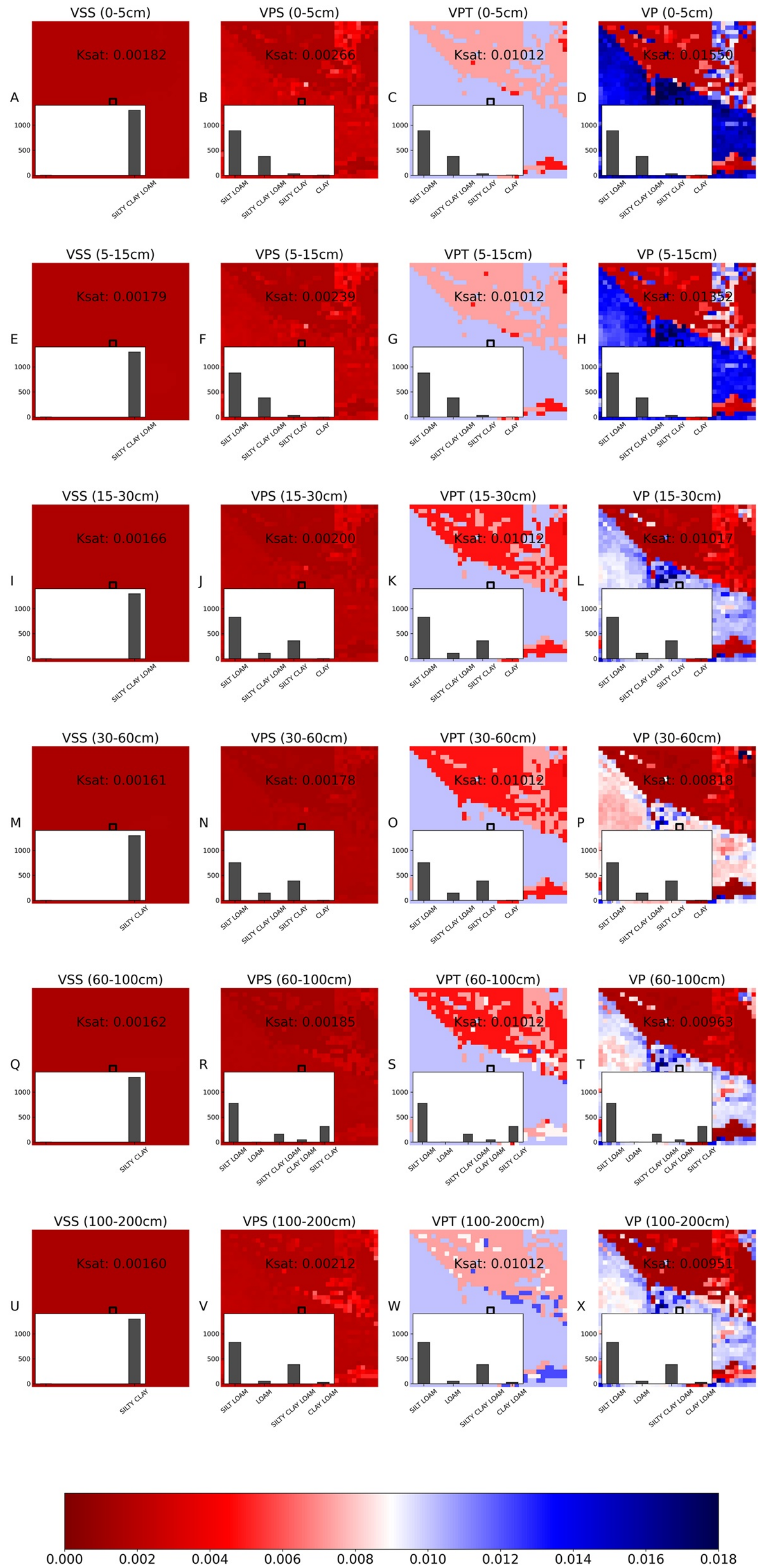

Figure 7. Spatial distribution of saturated hydraulic conductivity for (a–d) 0–5 cm, (e–h) 5–15 cm, (i–l) 15–30 cm, (m–p) 30–60 cm, (q–t) 60–100 cm, (u–x) 100–200 cm from VSS, VPS, VPT, and VP soil parameterization schemes. The size of each map is about 540 m by 540 m. The central square covers the USCRN site (Lafayette, LA) with a size of 30 m by 30 m. The values of saturated hydraulic conductivity (unit: m/h) are shown in the upper center. The occurrence of soil order has been shown in the lower left corners.

10.1029/2022WR032336

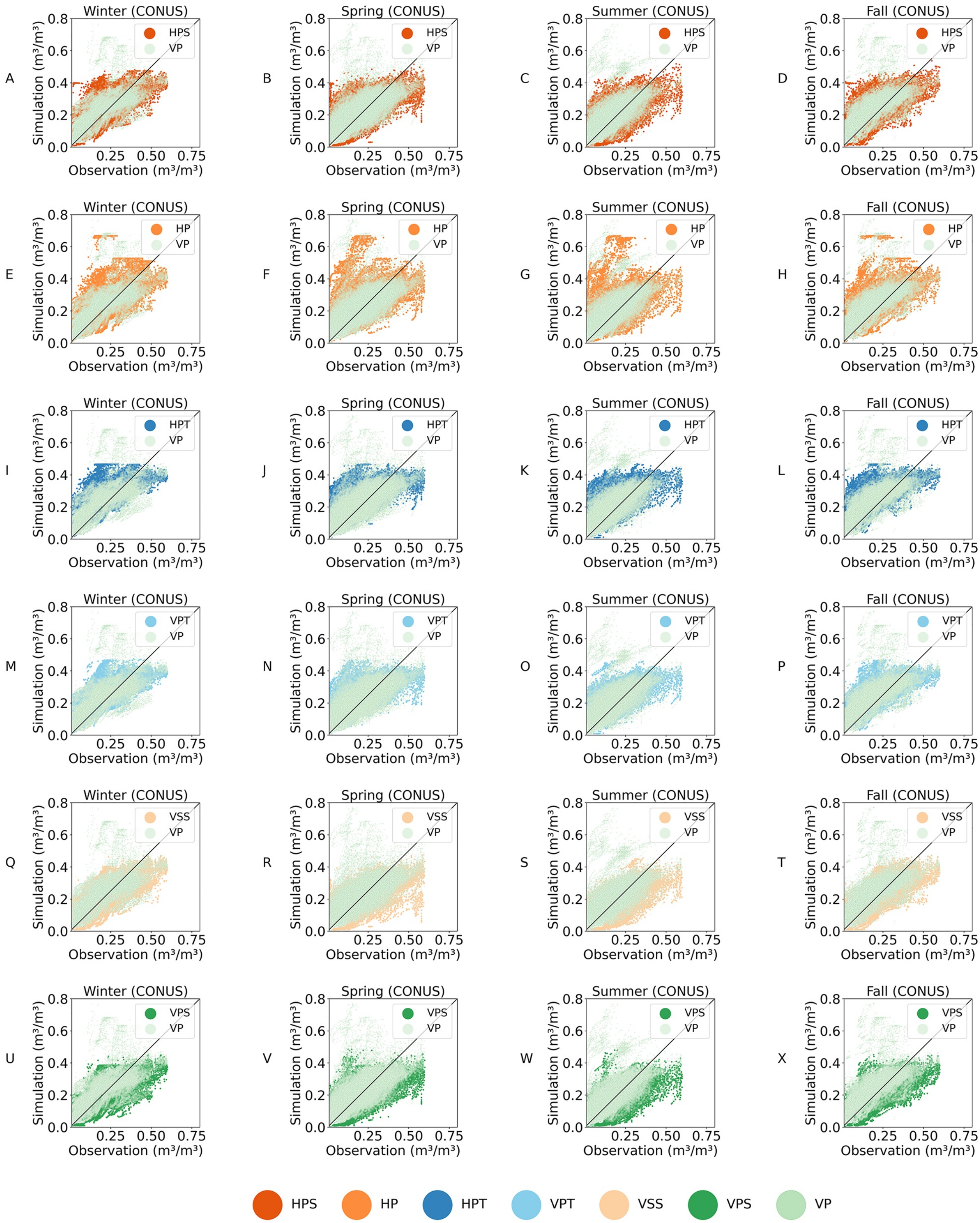

Figure 8. Seasonal comparison of the HydroBlocks simulated daily soil moisture using the HPS (sandy brown), HP (peach puff), HPT (steel blue), VPT (light cyan), VSS (linen), and VPS (forest green) soil parameterization schemes and USCRN measurements (m^3/m^3) from 2014 to 2019 over the CONUS. Each point represents a daily averaged surface soil moisture from the corresponding soil parameterization scheme. Each plot contains CONUS-wide simulations. Each plot consists of scatter plots with the superimposed VP (lime green) on top of the other soil parameterization schemes.

tion) with the top-performing soil parameterization scheme (vertical soil properties map (VP); lime green points) superimposed on top. The simulations using the vertical soil properties map exhibited the strongest seasonal correlations. However, there were some wet outliers among seasons. The use of contemporary PTFs reduced a few of the prevalent wet outliers (e.g., VPS).

To visualize the HydroBlock's temporal response to different soil parameterizations, Figure 9 shows the time series of HydroBlocks simulated surface volumetric soil moisture content using VSS, VPS, VPT, and VP at multiple USCRN measurements from 2014 to 2019. These soil parameterization schemes were chosen because they represented a range of variations of soil parameterization schemes, including contemporary PTFs (VSS and VPS), lookup table-derived soils (VPT), and soil properties maps with soil hydraulic properties (VP). The USCRN measurements are shown via red dots. All the chosen soil parameterization schemes were skillful in replicating the temporal trends of surface soil moisture at each site. Seasonal dry, wet, and transitional periods can be observed. However, significant positive biases were present, such as at the Necedah site in Wisconsin. The overestimation at the Necedah site could be caused by many reasons, such as forcing data bias or measurement bias; however, the choice of PTFs and soil textural information decreased the overestimation (Figures 9e–9h). The Supporting Information S1 also includes maps of RMSE and ubRMSE (Figures S2 and S3 in Supporting Information S1) and simulation results using 0–5 cm soil properties in line with the default setting in Noah-MP LSM (Figures S4–S14 in Supporting Information S1).

4. Discussion

This study is unique as it evaluates the performance of cutting-edge DSM products (i.e., POLARIS and SoilGrids250m V2.0) in soil moisture modeling on the CONUS-scale. The results of this study provide insights for the land surface modeling regarding the added value of contemporary PTFs, the importance of soil texture maps, the role of vertical heterogeneity of soil properties, and the need to move away from using lookup tables in LSMs.

4.1. The Impact of Vertically Heterogeneous Soil Profiles

The importance of a vertical representation of soil layers for simulating water movement in the vadose zone was investigated by comparing heterogeneous and the corresponding homogeneous soil parameterizations. Variations in soil mineral components, pore spaces, and soil hydraulic properties were observed across different depths at USCRN sites (Wilson et al., 2016). This supports the notion that there is a need to represent these vertical variations of soil properties. The results show that most soil schemes improved CONUS-wide median KGE, RMSE, and ubRMSE (Figure 2) by accounting for the vertical complexity of soil properties. The more significant benefit resides in the improved estimation of the temporal change of surface soil moisture (better α; Figure 4). Consequently, land surface modeling should definitely move away from vertically homogeneous settings for soil properties.

The relationships between soil properties and environmental characteristics generally weaken in the deeper layers of soils (Keskin & Grunwald, 2018). One issue is that deeper layers of soils are generally more difficult to sample than superficial ones. Consequently, the available information at these layers is not as reliable or accurate. Moreover, when evaluating ground-truth soil profiles or network monitoring soil hydrology processes, more attention is paid to the top layers than the deeper layers. Furthermore, state-of-the-art soil properties maps often utilize remote sensing information from the surface or top of the canopy as soil-forming covariates in their predictive models, which makes it more challenging to sense environmental characteristics in the root zone. To tackle this issue, researchers have suggested using depth function approaches (Kidd et al., 2015), adding vertically changing spatial covariates in the predictive models (Ma et al., 2021), and continuing to share open source soil profile data with the community (Batjes et al., 2020). Geostatistical models have been used to account for three-dimensional spatial correlations of soil properties (Poggio & Gimona, 2017). Going forward, research should focus on assessing deeper soil moisture to understand how vertical soil heterogeneity partitions soil water storage, modifies water movement, and affects water redistribution in three-dimensional space.

4.2. Moving Beyond Lookup Tables

Although lookup tables remain the prevalent technique to estimate soil hydraulic parameters in LSMs, their deficiency in comparison to the state-of-the-art soil hydraulic property maps and contemporary PTFs has been demonstrated. This is mainly due to the intra-variability of soil properties within soil classes being disregarded

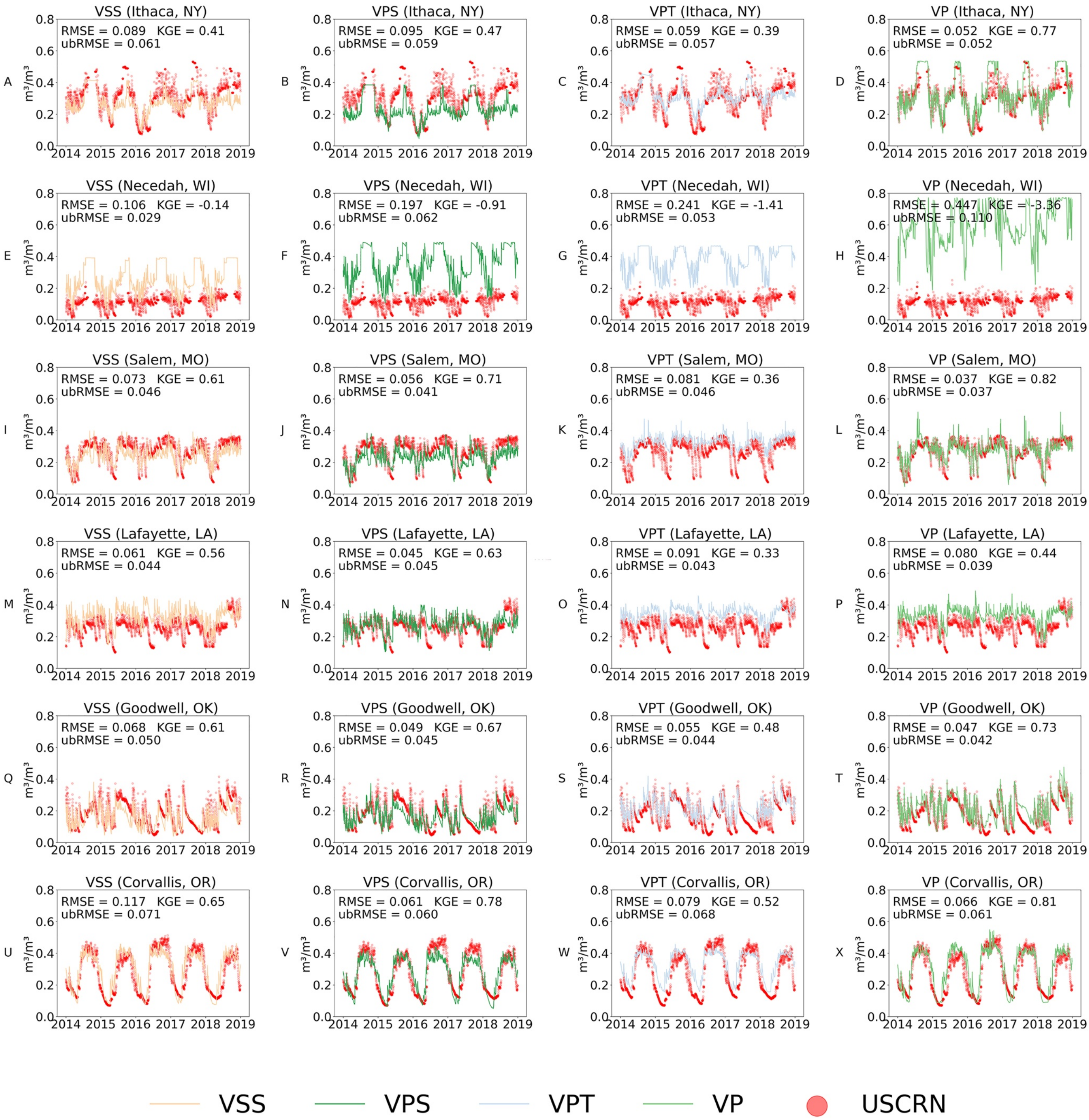

Figure 9. Time series of HydroBlocks simulated surface 5 cm volumetric soil moisture content (m^3/m^3), using POLARIS and Saxton and Rawls PTFs (2006) relevant parameterization schemes (VSS, VPS, VPT, and VP) against USCRN measurements from 2014 to 2019 over the (a–d) Ithaca site, New York; (e–h) Necedah site, Wisconsin; (i–l) Salem site, Missouri; (m–p) Lafayette site, Louisiana; (q–t) Goodwell site, Oklahoma; (u–x) Corvallis site, Oregon. Solid lines represent simulated surface 5 cm soil moisture content. Red dots are USCRN-measured surface 5 cm soil moisture content (m^3/m^3). Kling-Gupta efficiency, ubRMSE, and Root mean square error computed during the study period per site are shown at the top.

(Kishné et al., 2017). Within the same textural classification at an individual study area, the spatial variability of soil hydraulic characteristics (i.e., hydraulic conductivity, bulk density, soil matric potential) can vary substantially (Figure 7; Wilson et al., 2016; Kishné et al., 2017). Solutions to this issue have been suggested: (a) updating the existing lookup tables (Kishné et al., 2017), (b) developing new lookup tables based on finer categorized soil classes, (c) calibrating LSM or input parameters based on localized conditions (Yang et al., 2021).

Replacing class-based lookup tables with multi-linear or non-linear regression-based PTFs (e.g., Random Forests) can be a good substitute for predicting soil hydraulic properties. Our results demonstrated that the multi-linear regression PTFs—Saxton and Rawls PTFs (Saxton & Rawls, 2006) derived soil hydraulic parameters led to improvements in the simulation of surface soil moisture by reducing the overall bias (Figures 2 and 3). In comparison, the POLARIS water retention curve parameters developed from Artificial Neural Network-based PTFs—NeuroTheta PTFs (Minasny & McBratney, 2002) displayed larger biases. Variations in PTF behaviors are not only related to the statistical techniques used, but also to the soil sampling locations, the quality of soil samples, and the measured soil properties used to train the PTF models. The Saxton and Rawls PTFs were developed using only soil texture and organic matter from the USDA soil database as inputs. However, the NeuroTheta PTFs were trained using data from NRCS SCD (SCD 2016), UNSODA (Nemes et al., 2001), and GRIZZLY (Haverkamp et al., 1997) data sets, which includes soil samples from more extensive geological locations. Moreover, these PTFs were also trained using soil properties with higher uncertainties, such as bulk density, field capacity, and permanent wilting point. The selection of appropriate PTFs is essential to accurately predict soil moisture behavior, as variations in PTFs can be attributed to a variety of factors.

Various efforts have been made to improve the performance of PTFs, such as the inclusion of novel soil hydrology-related predictors (Fatichi et al., 2020), topographic and land use features (Van Looy et al., 2017), and localized predictors (Al Majou et al., 2008). Different statistical approaches, including regression techniques (Cosby et al., 1984), Machine Learning methods (Rawls & Pachepsky, 2002), and Neural Networks (Schaap et al., 2001), have been used to enhance the performance of soil moisture modeling. However, the complexity of these approaches limits their use in large-scale studies. Linear regression methods can be implemented into LSM parameterization options, while Deep Learning-based PTFs or locally calibrated PTFs can better be incorporated into soil properties maps and used as a ready-to-use soil data set in LSMs. It remains unclear how robust and reliable these methodologies are when used in large-scale studies. Different PTFs exhibit evident variability between model outputs. For instance, Weihermüller et al. (2021) found that Rosetta, Wösten, and Tóth PTFs were the most reliable for the Mualem van Genuchten soil hydraulic properties and the PTFs of Cosby for the Brooks Corey functions when utilizing the large-scale model HYDRUS 1D Model (Weihermüller et al., 2021). Further studies are needed to determine how to apply PTFs efficiently. Despite this, their findings are still valuable since they can offer guidelines for choosing PTFs for large-scale LSMs.

4.3. Opportunity for Improvement of Soil Properties Maps

The results of this study have shown that vertical soil properties maps and contemporary PTFs can lead to appreciable improvements in the prediction of soil moisture. However, more can be done to improve these soil properties maps. It is worthwhile to explore new approaches to assemble another iteration of POLARIS. The existing POLARIS soil series have high uncertainties that reside in the weighted random allocation algorithm—soil series were determined by the normalized proportion of occurrence for the soil component within the corresponding SSURGO map unit (Chaney, Wood, et al., 2016). In the future product, we will leverage the availability of over 294,746 descriptions of soil pedons from the National Soil Information System (NASIS) as model inputs (Fortner, 2008) instead of polygon-based SSURGO to reduce predictive uncertainties. The future POLARIS soil classification will be performed based on the delineation of eight-digit Hydrologic Unit Code (HUC8) Watersheds and cluster HUC8 subbasins with neighboring subbasins to form predictive units, leveraging the existing USGS hydrological boundaries. The development of feature engineering, cutting-edge remote sensing, and airborne products also demonstrated their ability to improve the prediction of soil types and classification (Fink & Drohan, 2016; Mulder et al., 2011; Reinhardt & Herrmann, 2019; Vaudour et al., 2019). By wisely selecting key features, utilizing recursive feature reduction, and conducting de-correlation analysis, the potential of model over-fitting and redundancy between input features can be reduced, leading to more effective models. By taking these steps, future DSM products have the potential to become more robust and useful.

4.4. Uncertainty Quantification

The present study uses deterministic estimates of soil hydraulic properties to perform the HydroBlocks simulations. A complete evaluation would involve performing ensemble simulations at each site using a parameter ensemble to understand whether the remaining errors in the simulated soil moisture are still driven by the soil parameter uncertainty (and whether the POLARIS parameter distributions contain the "true" parameter values).

POLARIS also provides information on the distributions of each parameter per layer and variable per 30-m pixel over CONUS. However, there is no information regarding how the variables correlate with each other, nor their vertical correlation. As such, although approaches could be implemented, they would not be sufficiently robust at this time and would likely lead to randomly generated profiles of soil hydraulic properties that are not self-consistent with observations. To make such an analysis possible, one option for the next generation of POLARIS will be to provide the distributions of the parameters per layer and pixel and a variable/layer covariance matrix. This covariance matrix could be calculated when assembling the data set from the soil pedons and PTFs, and it would provide sufficient information to assemble parameter ensembles for a comprehensive uncertainty quantification.

5. Conclusions

Soil moisture varies strongly in space and time, reflecting a complex interaction between climate, landscape, and particularly soil properties. This work aimed to assess the impact of cutting-edge soil properties maps, contemporary PTFs, and vertical soil heterogeneity on hydrological representation. Based on an evaluation of surface soil moisture modeling over a multi-year (2014–2019) basis across the CONUS, different soil parameterization schemes were able to reproduce the main features of soil moisture seasonal cycle and interannual fluctuation. The use of POLARIS and SoilGrids250m V2.0 soil properties maps and their derived parameters with the Saxton and Rawls PTFs (2006) were found to improve soil moisture modeling, and their associated vertically heterogeneous soil parameterization schemes further improved performance in comparison to lookup table-derived schemes, particularly with vertically homogeneous setting for soil columns. The study highlighted the importance of accurately predicting soil properties (specifically soil hydraulic properties) and the need for detailed soil heterogeneity at finer spatial resolution. It also emphasized the added value of robust vertically heterogeneous soil properties and contemporary soil properties products for LSM's soil parameterizations. In addition, the study suggests that moving beyond lookup tables is crucial for soil parameterization in LSMs to improve the representation of soil moisture variability.

Data Availability Statement

POLARIS, SoilGrids250m V2.0, and CONUS-SOIL soil data used in this work can be found at the following site: http://hydrology.cee.duke.edu/POLARIS/ (Chaney et al., 2019; Chaney, Wood, et al., 2016); https://soilgrids.org/ (Poggio et al., 2021, p. 2); http://www.soilinfo.psu.edu/index.cgi?soil_data&conus (Miller & White, 1998). The HydroBlocks model (Chaney et al., 2021) code used in this study is preserved at https://zenodo.org/record/4071692. The USCRN soil moisture is available online at https://www.ncei.noaa.gov/access/crn/qcdatasets.html (Bell et al., 2013).

Acknowledgments
This study was supported by funding from NOAA Grant NA19OAR4590200 (Modernizing Observation Operator and Error Assessment for Assimilating In-situ and Remotely Sensed Snow/Soil Moisture Measurements into NWM) and USDA grant S-001298-02 / 2020-69012-31914 (Artificial Intelligence for Enhancing Sustainability of Water, Nutrient, Salinity, and Pest Management in the Western USA).

References

Adamchuk, V. I., Hummel, J. W., Morgan, M. T., & Upadhyaya, S. K. (2004). On-the-go soil sensors for precision agriculture. *Computers and Electronics in Agriculture*, *44*(1), 71–91. https://doi.org/10.1016/J.COMPAG.2004.03.002

Al Majou, H., Bruand, A., & Duval, O. (2008). The use of in situ volumetric water content at field capacity to improve the prediction of soil water retention properties. *Canadian Journal of Soil Science*, *88*(4), 533–541. https://doi.org/10.4141/CJSS07065

Arrouays, D., Grundy, M. G., Hartemink, A. E., Hempel, J. W., Heuvelink, G. B. M., Hong, S. Y., et al. (2014). GlobalSoilMap. Toward a fine-resolution global grid of soil properties. *Advances in Agronomy*, *125*, 93–134. https://doi.org/10.1016/B978-0-12-800137-0.00003-0

Arsenault, K. R., Nearing, G. S., Wang, S., Yatheendradas, S., & Peters-Lidard, C. D. (2018). Parameter sensitivity of the Noah-MP land surface model with dynamic vegetation. *Journal of Hydrometeorology*, *19*(5), 815–830. https://doi.org/10.1175/JHM-D-17-0205.1

Baroni, G., Zink, M., Kumar, R., Samaniego, L., & Attinger, S. (2017). Effects of uncertainty in soil properties on simulated hydrological states and fluxes at different spatio-temporal scales. *Hydrology and Earth System Sciences*, *21*(5), 2301–2320. https://doi.org/10.5194/hess-21-2301-2017

Batjes, N. H., Ribeiro, E., & Van Oostrum, A. (2020). Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019). *Earth System Science Data*, *12*(1), 299–320. https://doi.org/10.5194/essd-12-299-2020

Beck, H. E., Pan, M., Miralles, D. G., Reichle, R. H., Dorigo, W. A., Hahn, S., et al. (2021). Evaluation of 18 satellite- and model-based soil moisture products using in situ measurements from 826 sensors. *Hydrology and Earth System Sciences*, *25*(1), 17–40. https://doi.org/10.5194/hess-25-17-2021

Bell, J. E., Palecki, M. A., Baker, C. B., Collins, W. G., Lawrimore, J. H., Leeper, R. D., et al. (2013). U.S. climate reference network soil moisture and temperature observations. [Dataset]. Journal of Hydrometeorology, *14*, 977–988. https://doi.org/10.1175/JHM-D-12-0146.1

Beven, K., & Germann, P. (1982). Macropores and water flow in soils. *Water Resources Research*, *18*(5), 1311–1325. https://doi.org/10.1029/WR018i005p01311

Cai, X., Yang, Z. L., David, C. H., Niu, G. Y., & Rodell, M. (2014). Hydrological evaluation of the Noah-MP land surface model for the Mississippi River Basin. *Journal of Geophysical Research: Atmospheres*, *119*(1), 23–38. https://doi.org/10.1002/2013JD020792

Chaney, N. W., Metcalfe, P., & Wood, E. F. (2016). HydroBlocks: A field-scale resolving land surface model for application over continental extents. *Hydrological Processes*, *30*(20), 3543–3559. https://doi.org/10.1002/hyp.10891

Chaney, N. W., Minasny, B., Herman, J. D., Nauman, T. W., Brungard, C. W., Morgan, C. L. S., et al. (2019). POLARIS soil properties: 30-m probabilistic maps of soil properties over the contiguous United States. [Dataset]. Water Resources Research, *55*, 2916–2938. https://doi.org/10.1029/2018WR022797

Chaney, N. W., Roundy, J. K., Herrera-Estrada, J. E., & Wood, E. F. (2015). High-resolution modeling of the spatial heterogeneity of soil moisture: Applications in network design. *Water Resources Research*, *51*(1), 619–638. https://doi.org/10.1002/2013WR014964

Chaney, N. W., Torres-Rojas, L., Vergopolan, N., & Fisher, C. K. (2021). HydroBlocks v0.2: Enabling a field-scale two-way coupling between the land surface and river networks in Earth system models. [Software]. Geoscientific Model Development, *14*, 6813–6832. https://doi.org/10.5194/gmd-14-6813-2021

Chaney, N. W., Wood, E. F., McBratney, A. B., Hempel, J. W., Nauman, T. W., Brungard, C. W., & Odgers, N. P. (2016). POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. [Dataset]. Geoderma, *274*, 54–67. https://doi.org/10.1016/j.geoderma.2016.03.025

Chew, C., & Small, E. (2020). Description of the UCAR/CU soil moisture product. *Remote Sensing*, *12*, 1558. https://doi.org/10.3390/RS12101558

Clark, D. B., & Gedney, N. (2008). Representing the effects of subgrid variability of soil moisture on runoff generation in a land surface model. *Journal of Geophysical Research*, *113*(D10), 10111. https://doi.org/10.1029/2007JD008940

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., et al. (2011). Development and evaluation of an Earth-System model–HadGEM2. *Geoscientific Model Development*, *4*(4), 1051–1075. https://doi.org/10.5194/gmd-4-1051-2011

Cosby, B. J., Hornberger, G. M., Clapp, R. B., & Ginn, T. R. (1984). A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. *Water Resources Research*, *20*(6), 682–690. https://doi.org/10.1029/WR020i006p00682

Diamond, H. J., Karl, T. R., Palecki, M. A., Baker, C. B., Bell, J. E., Leeper, R. D., et al. (2013). U.S. climate reference network after one decade of operations status and assessment. *Bulletin of the American Meteorological Society*, *94*(4), 485–498. https://doi.org/10.1175/BAMS-D-12-00170.1

Dirmeyer, P. A., & Halder, S. (2016). Sensitivity of numerical weather forecasts to initial soil moisture variations in CFSv2. *Weather and Forecasting*, *31*(6), 1973–1983. https://doi.org/10.1175/WAF-D-16-0049.1

Dorigo, W., Himmelbauer, I., Aberer, D., Schremmer, L., Petrakovic, I., Zappa, L., et al. (2021). The international soil moisture network: Serving Earth system science for over a decade. *Hydrology and Earth System Sciences*, *25*(11), 5749–5804. https://doi.org/10.5194/hess-25-5749-2021

Dorigo, W. A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., et al. (2011). The international soil moisture network: A data hosting facility for global in situ soil moisture measurements. *Hydrology and Earth System Sciences*, *15*(5), 1675–1698. https://doi.org/10.5194/hess-15-1675-2011

Essery, R. L. H., Best, M. J., Betts, R. A., Cox, P. M., & Taylor, C. M. (2003). Explicit representation of subgrid heterogeneity in a GCM land surface scheme. *Journal of Hydrometeorology*, *4*(3), 530–543. https://doi.org/10.1175/1525-7541(2003)004<0530:EROSHI>2.0.CO;2

Falloon, P., Jones, C. D., Ades, M., & Paul, K. (2011). Direct soil moisture controls of future global soil carbon changes: An important source of uncertainty. *Global Biogeochemical Cycles*, *25*(3), GB3010. https://doi.org/10.1029/2010GB003938

Fatichi, S., Or, D., Walko, R., Vereecken, H., Young, M. H., Ghezzehei, T. A., et al. (2020). Soil structure is an important omission in Earth System Models. *Nature Communications*, *11*(1), 522. https://doi.org/10.1038/s41467-020-14411-z

Fink, C. M., & Drohan, P. J. (2016). High resolution hydric soil mapping using LiDAR digital terrain modeling. *Soil Science Society of America Journal*, *80*(2), 355–363. https://doi.org/10.2136/sssaj2015.07.0270

Fortner, J. R. (2008). National cooperative soil survey national soil information system brief history and status of soil survey. Retrieved from http://soils.usda.gov/

Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M., & Tyler, D. (2002). The national elevation dataset. *Photogrammetric Engineering & Remote Sensing*, *68*(1).

Gruber, A., De Lannoy, G., Albergel, C., Al-Yaari, A., Brocca, L., Calvet, J. C., et al. (2020). Validation practices for satellite soil moisture retrievals: What are (the) errors? *Remote Sensing of Environment*, *244*, 111806. https://doi.org/10.1016/J.RSE.2020.111806

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Haverkamp, R., Zammit, C., Boubkraoui, F., Rajkai, K., Arrúe, J. L., & Heckmann, N. (1997). *GRIZZLY, Grenoble soil catalogue: Soil survey of field data and description of particle-size, soil water retention and hydraulic conductivity functions* (Vol. 9). Laboratoire d'Etude Des Transferts En Hydrologie et Environnement (LTHE).

Hengl, T., De Jesus, J. M., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One*, *12*(2), e0169748. https://doi.org/10.1371/journal.pone.0169748

Hogue, T. S., Bastidas, L. A., Gupta, H. V., & Sorooshian, S. (2006). Evaluating model performance and parameter behavior for varying levels of land surface model complexity. *Water Resources Research*, *42*(8), W08430. https://doi.org/10.1029/2005WR004440

Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., et al. (2015). Completion of the 2011 national land cover database for the conterminous United States – Representing a decade of land cover change information. *Photogrammetric Engineering & Remote Sensing*, *81*(5). https://doi.org/10.1016/S0099-1112(15)30100-2

Kearney, M. R., & Maino, J. L. (2018). Can next-generation soil data products improve soil moisture modelling at the continental scale? An assessment using a new microclimate package for the R programming environment. *Journal of Hydrology*, *561*, 662–673. https://doi.org/10.1016/J.JHYDROL.2018.04.040

Keesstra, S., Pereira, P., Novara, A., Brevik, E. C., Azorin-Molina, C., Parras-Alcántara, L., et al. (2016). Effects of soil management techniques on soil water erosion in apricot orchards. *Science of the Total Environment*, *551–552*, 357–366. https://doi.org/10.1016/J.SCITOTENV.2016.01.182

Keskin, H., & Grunwald, S. (2018). Regression kriging as a workhorse in the digital soil mapper's toolbox. *Geoderma*, *326*, 22–41. https://doi.org/10.1016/j.geoderma.2018.04.004

Kidd, D., Webb, M., Malone, B., Minasny, B., & McBratney, A. (2015). Eighty-metre resolution 3D soil-attribute maps for Tasmania, Australia. *Soil Research*, *53*(8), 932. https://doi.org/10.1071/SR14268

Kishné, A. S., Yimam, Y. T., Morgan, C. L. S., & Dornblaser, B. C. (2017). Evaluation and improvement of the default soil hydraulic parameters for the Noah Land Surface Model. *Geoderma*, *285*, 247–259. https://doi.org/10.1016/j.geoderma.2016.09.022

Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, *424–425*, 264–277. https://doi.org/10.1016/j.jhydrol.2012.01.011

Koster, R. D., Guo, Z., Yang, R., Dirmeyer, P. A., Mitchell, K., & Puma, M. J. (2009). On the nature of soil moisture in land surface models. *Journal of Climate*, *22*(16), 4322–4335. https://doi.org/10.1175/2009JCLI2832.1

Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*, *99*(D7), 14415–14428. https://doi.org/10.1029/94JD00483

Liu, X., Chen, F., Barlage, M., & Niyogi, D. (2020). Implementing dynamic rooting depth for improved simulation of soil moisture and land surface feedbacks in Noah-MP-Crop. *Journal of Advances in Modeling Earth Systems*, *12*(2), e2019MS001786. https://doi.org/10.1029/2019MS001786

Loosvelt, L., Pauwels, V. R. N., Cornelis, W. M., Lannoy, G. J. M. D., & Verhoest, N. E. C. (2011). Impact of soil hydraulic parameter uncertainty on soil moisture modeling. *Water Resources Research*, *47*(3), W03505. Article 3. https://doi.org/10.1029/2010WR009204

Ma, Y., Minasny, B., McBratney, A., Poggio, L., & Fajardo, M. (2021). Predicting soil properties in 3D: Should depth be a covariate? *Geoderma*, *383*, 114794. https://doi.org/10.1016/j.geoderma.2020.114794

Maurer, E. P., Wood, A. W., Adam, J. C., Lettenmaier, D. P., & Nijssen, B. (2002). A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. *Journal of Climate*, *15*(22), 3237–3251. https://doi.org/10.1175/1520-0442(2002)015<3237:ALTHBD>2.0.CO;2

McBratney, A. B., Mendonça Santos, M. L., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, *117*(1–2), 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4

Miller, D. A., & White, R. A. (1998). A conterminous United States multilayer soil characteristics dataset for regional climate and hydrology modeling. [Dataset]. Earth Interactions, *2*, 2. https://doi.org/10.1175/1087-3562(1998)002<0002:cusms>2.0.co;2

Minasny, B., & McBratney, A. B. (2002). The Neuro-m method for fitting neural network parametric pedotransfer functions. *Soil Science Society of America Journal*, *66*(2), 352–361. https://doi.org/10.2136/sssaj2002.3520

Montzka, C., Herbst, M., Weihermüller, L., Verhoef, A., & Vereecken, H. (2017). A global data set of soil hydraulic properties and sub-grid variability of soil water retention and hydraulic conductivity curves. *Earth System Science Data*, *9*(2), 529–543. https://doi.org/10.5194/essd-9-529-2017

Mulder, V. L., De Bruin, S., Schaepman, M. E., & Mayr, T. R. (2011). The use of remote sensing in soil and terrain mapping—A review. *Geoderma*, *162*(1–2), 1–19. https://doi.org/10.1016/j.geoderma.2010.12.018

Nemes, A., Schaap, M. G., Leij, F. J., & Wösten, J. H. M. (2001). Description of the unsaturated soil hydraulic database UNSODA version 2.0. *Journal of Hydrology*, *251*(3–4), 151–162. https://doi.org/10.1016/S0022-1694(01)00465-6

Nemes, A., Schaap, M. G., & Wösten, J. H. M. (2003). Functional evaluation of pedotransfer functions derived from different scales of data collection. *Soil Science Society of America Journal*, *67*(4), 1093–1102. https://doi.org/10.2136/SSSAJ2003.1093

Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research*, *116*(D12), D12109. https://doi.org/10.1029/2010JD015139

Pan, M., Cai, X., Chaney, N. W., Entekhabi, D., & Wood, E. F. (2016). An initial assessment of SMAP soil moisture retrievals using high-resolution model simulations and in situ observations. *Geophysical Research Letters*, *43*(18), 9662–9668. https://doi.org/10.1002/2016GL069964

Pinheiro, E. A. R., & van Lier, Q. J. (2021). Propagation of uncertainty of soil hydraulic parameterization in the prediction of water balance components: A stochastic analysis in kaolinitic clay soils. *Geoderma*, *388*, 114910. https://doi.org/10.1016/j.geoderma.2020.114910

Poggio, L., De Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., & Rossiter, D. (2021). SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. [Dataset]. Soil, *7*, 217–240. https://doi.org/10.5194/SOIL-7-217-2021

Poggio, L., & Gimona, A. (2017). Assimilation of optical and radar remote sensing data in 3D mapping of soil properties over large areas. *Science of the Total Environment*, *579*, 1094–1110. https://doi.org/10.1016/j.scitotenv.2016.11.078

Rawls, W. J., & Pachepsky, Y. A. (2002). Soil consistence and structure as predictors of water retention. *Soil Science Society of America Journal*, *66*(4), 1115–1126. https://doi.org/10.2136/sssaj2002.1115

Reinhardt, N., & Herrmann, L. (2019). Gamma-ray spectrometry as versatile tool in soil science: A critical review. *Journal of Plant Nutrition and Soil Science*, *182*(1), 9–27. https://doi.org/10.1002/jpln.201700447

Richards, L. A. (1931). Capillary conduction of liquids through porous mediums. *Journal of Applied Physics*, *1*(5), 318–333. https://doi.org/10.1063/1.1745010

Sanchez-Mejia, Z. M., & Papuga, S. A. (2014). Observations of a two-layer soil moisture influence on surface energy dynamics and planetary boundary layer characteristics in a semiarid shrubland. *Water Resources Research*, *50*(1), 306–317. https://doi.org/10.1002/2013WR014135

Saxton, K. E., & Rawls, W. J. (2006). Soil water characteristic estimates by texture and organic matter for hydrologic solutions. *Soil Science Society of America Journal*, *70*(5), 1569–1578. https://doi.org/10.2136/sssaj2005.0117

Schaap, M. G., & Leij, F. J. (1998). Database-related accuracy and uncertainty of pedotransfer functions. *Soil Science*, *163*(10), 765–779. https://doi.org/10.1097/00010694-199810000-00001

Schaap, M. G., Leij, F. J., & van Genuchten, M. T. (1998). Neural network analysis for hierarchical prediction of soil hydraulic properties. *Soil Science Society of America Journal*, *62*(4), 847–855. https://doi.org/10.2136/sssaj1998.03615995006200040001x

Schaap, M. G., Leij, F. J., & Van Genuchten, M. T. (2001). Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology*, *251*(3–4), 163–176. https://doi.org/10.1016/S0022-1694(01)00466-8

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., et al. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, *99*(3), 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004

Simons, G., Koster, R., & Droogers, P. (2020). 213 REPORT AUTHORS DATE HiHydroSoil v2.0-high resolution soil maps of global hydraulic properties. Retrieved from www.futurewater.eu/hihydrosoil

Teuling, A. J., Uijlenhoet, R., van den Hurk, B., & Seneviratne, S. I. (2009). Parameter sensitivity in LSMs: An analysis using stochastic soil moisture models and ELDAS soil parameters. *Journal of Hydrometeorology*, *10*(3), 751–765. https://doi.org/10.1175/2008JHM1033.1

Tóth, B., Weynants, M., Pásztor, L., & Hengl, T. (2017). 3D soil hydraulic database of Europe at 250 m resolution. *Hydrological Processes*, *31*(14), 2662–2666. https://doi.org/10.1002/hyp.11203

Van Looy, K., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., et al. (2017). Pedotransfer functions in Earth system science: Challenges and perspectives. *Reviews of Geophysics*, *55*(4), 1199–1256. https://doi.org/10.1002/2017RG000581

Vaudour, E., Gomez, C., Loiseau, T., Baghdadi, N., Loubet, B., Arrouays, D., et al. (2019). The impact of acquisition date on the prediction performance of topsoil organic carbon from Sentinel-2 for croplands. *Remote Sensing*, *11*(18), 2143. https://doi.org/10.3390/rs11182143

Vergopolan, N., Chaney, N. W., Beck, H. E., Pan, M., Sheffield, J., Chan, S., & Wood, E. F. (2020). Combining hyper-resolution land surface modeling with SMAP brightness temperatures to obtain 30-m soil moisture estimates. *Remote Sensing of Environment*, *242*, 111740. https://doi.org/10.1016/j.rse.2020.111740

Vergopolan, N., Sheffield, J., Chaney, N. W., Pan, M., Beck, H. E., Ferguson, C. R., et al. (2022). High-resolution soil moisture data reveal complex multi-scale spatial variability across the United States. *Geophysical Research Letters*, *49*(15), e2022GL098586. https://doi.org/10.1029/2022GL098586

Vergopolan, N., Xiong, S., Estes, L., Wanders, N., Chaney, N. W., Wood, E. F., et al. (2021). Field-scale soil moisture bridges the spatial-scale gap between drought monitoring and agricultural yields. *Hydrology and Earth System Sciences*, *25*(4), 1827–1847. https://doi.org/10.5194/HESS-25-1827-2021

Wagner, W., Lemoine, G., & Rott, H. (1999). A method for estimating soil moisture from ERS Scatterometer and soil data. *Remote Sensing of Environment*, *70*(2), 191–207. https://doi.org/10.1016/S0034-4257(99)00036-X

Weihermüller, L., Lehmann, P., Herbst, M., Rahmati, M., Verhoef, A., Or, D., et al. (2021). Choice of pedotransfer functions matters when simulating soil water balance fluxes. *Journal of Advances in Modeling Earth Systems*, *13*(3), e2020MS002404. https://doi.org/10.1029/2020MS002404

Wigneron, J. P., Kerr, Y., Waldteufel, P., Saleh, K., Escorihuela, M. J., Richaume, P., et al. (2007). L-band Microwave Emission of the Biosphere (L-MEB) Model: Description and calibration against experimental data sets over crop fields. *Remote Sensing of Environment*, *107*(4), 639–655. https://doi.org/10.1016/J.RSE.2006.10.014

Wilson, T. B., Baker, C. B., Meyers, T. P., Kochendorfer, J., Hall, M., Bell, J. E., et al. (2016). Site-specific soil properties of the US climate reference network soil moisture. *Vadose Zone Journal*, *15*(11), 1–14. https://doi.org/10.2136/vzj2016.05.0047

Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., et al. (2011). Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resources Research*, *47*(5), W05301. https://doi.org/10.1080/00397910600978218

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

Xia, Y., Mitchell, K. E., Ek, M., Sheffield, J., Cosgrove, B., Wood, E. F., et al. (2012). Continental-scale water and energy flux analysis and validation for the North American land data assimilation system project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of Geophysical Research*, *117*, D03109. https://doi.org/10.1029/2011JD016048

Yang, Y., Guan, K., Peng, B., Pan, M., Jiang, C., & Franz, T. E. (2021). High-resolution spatially explicit land surface model calibration using field-scale satellite-based daily evapotranspiration product. *Journal of Hydrology*, *596*, 125730. https://doi.org/10.1016/J.JHYDROL.2020.125730