



**Report 341**  
*February 2020*

# Emulation of Community Land Model Version 5 (CLM5) to Quantify Sensitivity of Soil Moisture to Uncertain Parameters

Xiang Gao, Alexander Avramov, Eri Saikawa, and C. Adam Schlosser

MIT Joint Program on the Science and Policy of Global Change combines cutting-edge scientific research with independent policy analysis to provide a solid foundation for the public and private decisions needed to mitigate and adapt to unavoidable global environmental changes. Being data-driven, the Joint Program uses extensive Earth system and economic data and models to produce quantitative analysis and predictions of the risks of climate change and the challenges of limiting human influence on the environment—essential knowledge for the international dialogue toward a global response to climate change.

To this end, the Joint Program brings together an interdisciplinary group from two established MIT research centers: the Center for Global Change Science (CGCS) and the Center for Energy and Environmental Policy Research (CEEPR). These two centers—along with collaborators from the Marine Biology Laboratory (MBL) at

Woods Hole and short- and long-term visitors—provide the united vision needed to solve global challenges.

At the heart of much of the program's work lies MIT's Integrated Global System Model. Through this integrated model, the program seeks to discover new interactions among natural and human climate system components; objectively assess uncertainty in economic and climate projections; critically and quantitatively analyze environmental management and policy proposals; understand complex connections among the many forces that will shape our future; and improve methods to model, monitor and verify greenhouse gas emissions and climatic impacts.

This report is intended to communicate research results and improve public understanding of global environment and energy challenges, thereby contributing to informed debate about climate change and the economic and social implications of policy alternatives.

—*Ronald G. Prinn and John M. Reilly,*  
*Joint Program Co-Directors*

# Emulation of Community Land Model Version 5 (CLM5) to Quantify Sensitivity of Soil Moisture to Uncertain Parameters

Xiang Gao<sup>1,2</sup>, Alexander Avramov<sup>3</sup>, Eri Saikawa<sup>3</sup>, and C. Adam Schlosser<sup>2</sup>

**Abstract:** The amount of water in the soil is a critical determinant in many complex processes of the Earth system. Model-simulated soil moisture has been widely used to understand these processes attributed to its large spatial and long temporal coverage at any desirable location and time. However, it is known that land surface models are strongly limited in their ability to reproduce observed soil moisture, often with biases in the mean, dynamic range, and time variability. In this study, we presented a cost-effective application of variance-based sensitivity analysis to quantify the relative contribution of different parameters and their interactions to the overall uncertainty in the modeled surface and root zone soil moisture from the Community Land Model 5.0 (CLM5). We focus on four parameters associated with the hydraulic property of mineral soil (saturated hydraulic conductivity, porosity, saturated soil matric potential, and shape-parameter) and organic matter fraction. A Gaussian process emulator is used to estimate the soil moisture across the five-dimensional parameter uncertainty space, based on a small number of CLM5 simulations at combinations of parameter values sampled with Maximin Latin hypercube. The procedure is exemplified for four seasons (DJF, MAM, JJA, and SON) across various sites of distinct soil and vegetation types in the continental US. Our results have shown that the emulator captures well the behavior of CLM5 across the entire parameter uncertainty space for different soil textures and seasons, with high correlations and low RMSEs between the emulator-predicted and CLM5-simulated soil moisture as well as small emulator uncertainty. We found that the large portion of the variances of both surface and root zone soil moisture is described by uncertainty in five parameters (excluding their interactions) and is dominated by the uncertainty in porosity and shape parameter for almost all the sites and seasons. Generally, the lower the fraction of sand is, the stronger (weaker) the individual parameter effects (the interaction effects) are. However, the relative importance of porosity versus shape parameter varies strongly with variables (surface versus root zone), soil textures (sites), and seasons. Over the majority of sites, the variance in surface soil moisture is attributed distinctly more to the uncertainty in shape parameter, while the uncertainty in porosity is more important in the variance of root zone soil moisture. Also, both individual parameter effects and interaction effects for root zone soil moisture demonstrate less variability across different soil textures and seasons than for surface soil moisture. These sensitivity results clearly indicate which parameters should be focused on to improve the model simulations of surface versus root zone soil moisture for different soil textures and seasons, which serves as a useful guidance to achieve improved modeling of soil moisture on a large scale.

1. INTRODUCTION.....	2
2. DATASETS AND MODEL.....	3
3. METHODS .....	5
4. RESULTS .....	9
5. SUMMARY AND CONCLUSIONS.....	18
6. REFERENCES .....	21
APPENDIX .....	23

1 Corresponding author (Email: [xgao304@mit.edu](mailto:xgao304@mit.edu)).

2 Joint Program on the Science and Policy of Global Change, Massachusetts Institute of Technology, MA, USA

3 Department of Environmental Sciences, Emory University

## 1. Introduction

The state and amount of water in the soil is a critical determinant in many complex Earth System processes. Soil moisture serves as the reservoir for the land surface hydrologic cycle and a boundary condition for the atmosphere. It regulates the partitioning of land surface heat fluxes, affects the status of overlying vegetation, modulates the thermal properties of the soil, and controls exchange of trace gases at the Earth's surface. Knowledge of the temporal and spatial variation of soil moisture is essential for climate predictability on seasonal to annual time scales (van den Hurk *et al.*, 2012; Sospedra-Alfonso and Merryfield, 2018), flood and drought forecasts (Sheffield *et al.*, 2014; Wanders *et al.*, 2014), and climate impact studies (Seneviratne *et al.*, 2010).

Soil moisture can be estimated in three ways: in-situ measurement, satellite remote sensing, and model-based simulation. Each of these techniques has its own specific properties and limitations. In-situ measurements are generally sparse both in time and space because soil moisture is difficult to measure *in situ*—particularly under a sustained, coordinated large-scale effort. There is no global in situ observation network for soil moisture. Regional and national networks exist of varying density and quality (e.g. Soil Climate Analysis Network (SCAN); U.S. Climate Reference Network (USCRN)), but many stations do not possess long continuous histories of operation. A critical shortcoming of these point-based measurements is the lack of representativeness of the surrounding area due to spatial variability of soil moisture which generally increases with extent scale (Famiglietti *et al.*, 2008). Such heterogeneity impedes meaningful assessment of area-representative soil moisture from a single point. Area-average soil moisture of a desired precision is attainable with sufficient point measurements made over the area (Zreda *et al.*, 2012), but is costly and impractical. As a result, in situ measurement is not suitable for large-scale atmospheric and other applications. A non-invasive technique that reduces the scale-related representativeness of point-based in situ measurement is the COsmic-ray Soil Moisture Observing System (COSMOS), which is designed to improve the availability of continental-scale soil moisture measurements. COSMOS consists of a network of portable probes that provide intermediate scale average soil moisture by measuring cosmic-ray neutrons above the land surface (Zreda *et al.*, 2008). Each probe measures average soil water content within a diameter of a few hectometers (~ 660m at sea level) and to a depth of a few decimeters (Zreda *et al.*, 2008), thereby averaging soil moisture heterogeneities. The measurement takes minutes to hours, permitting long-term monitoring of undisturbed soil moisture conditions. The horizontal footprint depends on the atmospheric pressure and humidity. It increases approximately by 25% between sea level and 3000 m of

altitude and decreases by 10 % between dry air and saturated air (Zreda *et al.*, 2012). The effective measurement depth depends strongly on soil moisture, ranging from 76cm in dry soils (zero water content) to 12cm in saturated soils (0.4 m<sup>3</sup>m<sup>-3</sup>). The precision of soil moisture measurement can be improved by increasing integration time. The unique features of the cosmic-ray neutron probe make it ideal for providing measurements with the precision and the scale appropriate for intermediate- to large-scale meteorological, hydrological, and ecological applications.

Satellite remote sensing, mostly by microwave sensors, can provide near-surface soil moisture of global coverage at coarse-scale, moderate temporal resolution. Currently several satellite missions provide global surface soil moisture products, including the Soil Moisture Active Passive (SMAP) (Entekhabi *et al.*, 2010b), the Soil Moisture and Ocean Salinity (SMOS) (Kerr *et al.*, 2012), the METOP-A/B Advanced Scatterometer (ASCAT) (Wagner *et al.*, 2013), the Special Sensor Microwave/Imager (SSM/I) mission, the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) (Njoku *et al.*, 2003) and Advanced Microwave Scanning Radiometer 2 (AMSR2) mission (Parinussa *et al.*, 2015). Satellite retrievals of soil moisture suffer from several limitations (Entekhabi *et al.*, 2004), including a shallow vertical penetration depth of centimeters, limited capability to penetrate vegetation or snow, sensitivity to surface roughness, discontinuous temporal coverage, and the short life span of satellite missions. The retrieved upper few centimeters of soil moisture represents the fast manifold of the soil moisture reservoir that provides little memory to the climate system and is thus of limited value for the numerical weather and seasonal climate prediction. Some sort of model has to be employed to propagate this information down into the soil to derive the slow manifold of the subsurface soil moisture reservoir relevant for predictability.

Global long-term analyses of surface and subsurface soil moisture can be produced with the Land Surface Model (LSM) either coupled to an atmospheric General Circulation Model (GCM), or driven by observations-based near-surface meteorological forcing. This approach has been widely employed to understand many complex Earth System processes in the past and future, mostly attributed to its continuous spatial and temporal coverage of flexible resolutions at any desirable location and time. However, the simulated soil moisture is heavily dependent on the character of the chosen LSM and the quality of the meteorological data to which the model is exposed (Guo *et al.*, 2006). Soil moisture parameterizations can vary considerably among models and result in a wide range in their fidelity (Dirmeyer *et al.*, 2004). These models need to be assessed in terms of their validity and accuracy in representing soil moisture and if necessary, be optimized with improved parameterizations before being used in various applications.

Model calibration/optimization usually involves attributing the uncertainty in model simulations (i.e. soil moisture) to various processes and the poorly constrained model parameters that describe these processes via sensitivity analysis. Assessing the effect of parameter uncertainty in complex models is often limited by computational resource constraints. The most commonly used sensitivity analysis approach used in complex models is single parameter perturbation or one-at-a-time (OAT) approach, which quantifies the departure of the model output from some baseline calculated using “default” parameters as a result of a perturbation in a single model parameter. However, this approach is inadequate because it ignores interactions between parameters with all sensitivity information calculated at one point in a parameter space. Recently, variance-based sensitivity analysis has been increasingly used to understand the sensitivity of complex models at the process level by quantifying the relative contribution of different model parameters and their interactions to the overall uncertainty in the model simulation. Variance-based methods require a very large number of model runs for complete specification of the model output throughout the parameter uncertainty space. Such Monte Carlo simulation is not feasible for complex models such as LSMs. The alternative is the emulation in which the complex model is replaced by a computationally-efficient statistical surrogate model. The emulator estimates the output of the model at a large number of unsampled parameter combinations across multi-dimensional parameter uncertainty space, using information from a small number of model simulations at chosen parameter values so that variance-based sensitivity analysis becomes feasible.

The aim of this study is to demonstrate the use of the emulation approach to quantify variance-based sensitivity of simulated soil moisture from a complex LSM to uncertain parameters and further to calibrate/optimize these parameters for improved soil moisture prediction. In this paper we focus on the emulation and sensitivity analysis, and leave parameter calibration as the subject of a subsequent paper. We use the Community Land Model 5.0 (CLM5), the latest in a series of land models developed and embedded in the Community Earth System Model (CESM), and focus on key parameters that control vertical water flow in multi-layer soil column. We use a Gaussian process emulator to estimate the soil moisture across multi-dimensional parameter space, based on information from a small number of CLM5 simulations at parameter values chosen using a Maximin Latin hypercube space-filling design, quantify the uncertainty and carry out sensitivity analysis. We also evaluate the SMAP L3 surface soil moisture product against the COSMOS measurements with various performance metrics and choose the sites with the high-quality soil moisture observations. The entire procedure is exemplified for seasonal (DJF, MAM, JJA,

and SON) soil moisture across various test sites of distinct soil and vegetation types in the continental US. We will look at both surface (0–5cm) and root zone (0–100cm) soil moisture to examine if there exists any difference in their parametric uncertainties.

The structure of the paper is as follows: In Section 2, we describe the datasets (COSMOS, SMAP soil moisture product, meteorological forcing) and the model (CESM2) used in this study. The method is given in Section 3, including the choice of study sites and uncertain model parameters, Gaussian process emulator, variance-based sensitivity analyses, and model experiment design. Section 4 describes the brief comparison of soil moisture simulated with the default CLM5 model parameters against the SMAP at the selected study sites, the validation of emulator, and variance-based sensitivity analysis of uncertain parameters using emulation. Summary and conclusions are provided in Section 5.

## 2. Datasets and Model

### 2.1 COSMOS

The COSMOS serves as an option, particularly in heterogeneous regions, to fill the scale gap between point in situ soil moisture measurements and low-resolution satellite products by providing area-average soil moisture within a 150–250 m radius footprint. The instrument, called “cosmic-ray moisture probe”, measures low-energy cosmic-ray neutrons above the ground, whose intensity is inversely correlated with soil moisture content and with water in any form above the ground surface (i.e. snow/vegetation water) (Zreda *et al.*, 2008). It is built on existing and tested technologies to ensure its continuous operation after deployment and provide long-term measurements. The data are available in near-real time, including neutron counts in two energy bands (fast, >1 keV; and thermal, <0.5 eV), ancillary data (atmospheric temperature, atmospheric pressure and relative humidity), and computed soil moisture. In this study, we focus on the network of probes installed at sites throughout the USA with most of these sites lying in existing facilities. We use level 3 data, which have gone through quality control and correction for various factors (i.e. temporal change in pressure, other sources of water) that could affect the accuracy of computed soil moisture from the measured neutron intensity (Zreda *et al.*, 2012). Level 3 data includes the hourly and 12-hour boxcar-filtered soil moisture as well as effective measurement depth.

### 2.2 SMAP

Currently, multiple global surface soil moisture products from passive microwave satellites are available to the community, including the AMSR-E/AMSR2, SMOS, and the latest launched SMAP. Extensive evaluation efforts have been conducted to assess the reliability and accuracy of these products across different regions. These results gen-

erally showed the SMAP soil moisture mostly outperforms its counterparts and well reproduces the anomalies and temporal variation of in situ measurements with favorable performance metric values (Chen *et al.*, 2017; Montzka *et al.*, 2017; Zhang *et al.*, 2017; Chan *et al.*, 2018; Cui *et al.*, 2018; El Hajj *et al.*, 2018). Therefore, we will focus on the SMAP soil moisture only in this study.

The SMAP is launched on 31 January 2015 and is dedicated to providing global top 5cm soil moisture with an accuracy of  $0.04 \text{ m}^3/\text{m}^3$  and freeze/thaw state through moderate vegetation cover. SMAP instruments observe the Earth's surface with a near-polar, Sun-synchronous 6:00 A.M. (descending)/ 6:00 P.M. (ascending) orbit. Despite the failure of the SMAP radar on 7 July 2015, its L band radiometer continues to operate as planned and provides routine data starting from March 31, 2015. Among five alternative algorithms developed to produce soil moisture retrievals, the Single Channel Algorithm V-pol (SCA-V), which uses the vertically polarized brightness temperature for surface soil moisture estimation, is selected as the operational baseline algorithm. In this study, the SMAP Level-3 (L3) radiometer global daily soil moisture product (SPL3SMP, version 5) is used. The L3 SMAP soil moisture products are resampled to a global, cylindrical 36 km Equal-Area Scalable Earth Grid, Version 2.0 (EASE-Grid 2.0). In the SCA-V algorithm, the soil temperature and vegetation canopy temperature are assumed equal to be the effective soil temperature. This assumption is more reliable in the early morning due to the increased thermal equilibrium of vegetation canopy and near-surface soil. Therefore, here we only consider the SMAP radiometer soil moisture products from the 6:00 A.M. (local solar time) descending passes, spanning from March 31, 2015 to present.

### 2.3 Near-surface Meteorological Forcing - CRUNCEP Version 7

CRUNCEP V7 is a combination of the CRU TS v3.2  $0.5^\circ \times 0.5^\circ$  monthly data from 1901 to 2002 (Mitchell and Jones, 2005) and the NCEP reanalysis  $2.5^\circ \times 2.5^\circ$  6-hourly data from 1948 to 2016 (Kalnay *et al.*, 1996). It combines NCEP's temporal advantage and CRU's better spatial resolution to achieve the 6-hourly  $0.5^\circ$  global forcing product spanning from 1901 to 2016. The NCEP reanalysis is only used to generate the diurnal and daily variability added to CRU TS monthly means. The NCEP is first bi-linearly interpolated to the  $0.5^\circ \times 0.5^\circ$  resolution of CRU for all fields except for precipitation which is linearly interpolated. Rainfall, cloudiness, relative humidity and temperature are taken from the CRU, while the other fields (pressure, longwave radiation, windspeed) are directly derived from NCEP. Cloudiness is converted to incoming solar radiation based on calculation of clear sky incoming solar radiation as a function of date and latitude, while relative humidity is converted to specific humidity as a function of temperature

and surface pressure. Prior to 1948, the procedure is the same except that the variability from 1948 is applied every year (there is no interannual variability for these fields), while after CRU period the fields are directly extrapolated from NCEP<sup>1</sup>. The missing data is filled with Qian *et al.* (2006) from 1948 that is interpolated to the  $0.5^\circ$  grid. The CRUNCEP dataset has been used for studies of vegetation growth (Mao *et al.*, 2013), evapotranspiration (Shi *et al.*, 2013) and trends in net land-atmosphere carbon exchange (Piao *et al.*, 2012), among many other use cases.

### 2.4 Community Earth System Model Version 2 (CESM2)

The Community Land Model (CLM) is a well-established mathematical model of land surface processes, developed for use as the land component in the fully coupled CESM. It is a fully prognostic model, and calculates the cycling of energy, water, C, and nitrogen and updates state variables at each 30 minutes for each grid cell. Compared to CLM 3.5, later versions (CLM4, CLM4.5, and CLM5) include many updates relevant to representations of soil hydrology. In particular, the new implementations include a revised solution to the Richard's equation which improves the accuracy and stability of the numerical soil water solution (Zeng and Decker, 2009), a dry surface layer-based evaporation resistance parameterization (Swenson and Lawrence, 2014), a spatially variable soil thickness (0.4 ~ 8.5 m depth, Brunke *et al.*, 2016), increased model soil layer resolution, revised treatments of soil column-groundwater interactions, a revised determination of hydraulic properties of frozen soils, corrections that increase the consistency between soil water state and water table position, and a representation of the thermal and hydraulic properties of organic soil in conjunction with the mineral soil properties (Lawrence and Slater, 2008). These augmentations to CLM3.5 result in improved soil moisture dynamics that lead to higher soil moisture variability and drier soils. Excessively wet and unvarying soil moisture was recognized as a deficiency in CLM3.5 (Decker and Zeng, 2009).

### 2.5 Data Processing

We derive the CLM surface (0–5cm) and root zone (0–100cm) volumetric soil moisture as the weighted averages of those from the first two and twelve soil layers, respectively. We assume that soil density remains the same in various layers and thus weights are proportional to the thickness of each layer. The analyses of various soil moisture data are conducted on a monthly time scale. For COSMOS, the hourly soil moisture measurement is first averaged into daily values and further averaged into monthly values. We compared the monthly values from the directly-measured hourly and 12-hour boxcar-filtered hourly data and find

<sup>1</sup> See more details at [ftp://nacp.ornl.gov/synthesis/2009/frescati/model\\_driver/cru\\_ncep/analysis/readme.htm](ftp://nacp.ornl.gov/synthesis/2009/frescati/model_driver/cru_ncep/analysis/readme.htm).

that the resulting differences were fairly small. In this study, we show the results from the 12-hour smoothed hourly COSMOS data. The daily SMAP data is extracted at an individual 36km EASE grid in which each COSMOS site is located and then averaged into monthly value for comparison with COSMOS. The comparison between COSMOS and SMAP is performed between their greatest overlap period—April 2015 to May 2019. The analyses involved with CLM are focused on year 2016 only.

### 3. Methods

An understanding of model sensitivity to uncertain parameters can not only help attribute the uncertainty in modelled variable of interest (i.e. soil moisture) to various processes and the poorly constrained parameters that describe these processes, but also guide future model development. However, the computational demand involved in sensitivity analysis of complex global models often prevents the source of uncertainty at the process level from being rigorously quantified. In addition, the most commonly used single parameter perturbation or one-at-a-time (OAT) approach, is inadequate for sensitivity analysis in two aspects: 1) It severely undersamples the parameter uncertainty space when the number of parameters is large.; and 2) It does not account for interactions among parameters. Here we apply well-established Gaussian process emulator (O’Hagan, 2006) to the CLM5 embedded in CESM model to quantify the effect of parametric uncertainty on model-simulated soil moisture. The overall approach has been described in detail by Lee *et al.* (2011, 2013), we briefly elaborate the procedure here. First, we choose uncertain model parameters associated with the simulation of soil moisture in CLM5 and determine their uncertainty ranges and probability distributions to represent the uncertainty in these parameters. Second, we use a maximin Latin Hypercube to sample an appropriate number of sets of parameter values covering the multidimensional parameter uncertainty space for CLM5 simulations. Third, we use Gaussian process emulation conditioned on the performed CLM5 simulations to generate continuous soil moisture distribution throughout the entire parameter uncertainty space. Lastly, a full variance-based sensitivity analysis is carried out using the emulator for a Monte Carlo-type sampling of soil moisture to quantify its sensitivity to the parameters and their interactions. We exemplify the entire procedure at the selected COSMOS sites in the continental United States.

It is worth noting that Gaussian process emulation and variance-based sensitivity analysis do not involve the use of any observational data, while model parameter calibration is attained by optimizing (or iteratively changing) the parameter values to minimize the errors between model simulations and observations (the subject of a subsequent paper). Our criteria for site selection is to include differ-

ent major soil types and also ensure the availability of high-quality soil moisture observations. In addition, we will examine how CLM-simulated soil moisture with default model parameters performs again the observations at the selected sites. This could provide us a useful insight into various extents to which parameter calibration improves model simulated soil moisture across different soil types.

#### 3.1 Choice of Study Sites

Since neither COSMOS nor SMAP soil moisture could be considered as “pure” truth, we evaluate the consistency between two measurements over all the USA COSMOS sites. Through such cross-validation exercise, the consistency between the two may increase the confidence in both data sets. Among the total of 78 USA COSMOS sites, we excluded all the sites where the measurements do not overlap with the SMAP data period or overlap with the SMAP for less than a year, as well as all the sites uncalibrated or with gravimetric calibration. The soil moisture values at the uncalibrated sites are obtained with an assumed calibration parameter and thus may be unreliable, while the gravimetric soil moisture value (in unit g/g) may lead to a bias when converted to the volumetric unit ( $\text{cm}^3/\text{cm}^3$ ) due to unknown bulk density (M. Zreda, personal communication, July 30, 2019). This results in a total of 30 sites remaining (**Table 1**). One potential limitation in the cosmic-ray method is the presence in the footprint of any surface water (i.e. snow, runoff, or intercepted precipitation) other than that in the soil. Water at the surface can depress the neutron intensity and lead to overestimated soil moisture. Zreda *et al.* (2012) suggested not to determine soil moisture in the presence of snow because the correction may be substantial and the signal due to soil moisture too weak to produce a well-defined calibration function. The SMAP retrievals during the frozen season are not available as well. In consideration of these facts, the evaluation of the consistency between SMAP and COSMOS soil moisture measurements will be conducted in unfrozen and snow-free seasons for any site.

Four standard statistic metrics were employed to provide a more comprehensive description of their consistency performance (Entekhabi *et al.*, 2010a), including the bias, the Root Mean Square Error (RMSE), the unbiased RMSE (ubRMSE), and the Pearson correlation coefficient ( $r$ ). Soil moisture time series usually show a strong seasonal pattern, which may artificially increase the perceived agreement in term of  $r$  between satellite retrievals and in situ observation. To avoid seasonal effects, time series of anomalies after removing the mean seasonal cycle are also computed. To summarize, bias is computed using the original time series, while RMSE and  $r$  are computed using both the original and anomaly time series. The RMSE of soil moisture anomalies are assumed to reflect ubRMSE by accounting for slowly varying seasonal bias. These statistics are computed for each site using the entire overlapping

**Table 1.** Characteristics of COSMOS sites used in this study. Soil organic C and vegetation are from COSMOS data. The soil type is determined from CLM surface data for the corresponding grid cell at which each COSMOS site is located. The shading is placed to differentiate various CLM soil type groups across the sites.

COSMOS Site	Lat (N)	Lon (E)	Soil organic C (wt %)	Vegetation	CLM Soil
ARM-1	36.61	262.51	0.59	grass	Loam
CC_Pasture	41.27	262.05	1	pasture	Loam
Daniel Forest	41.87	248.49	1.72	mixed conifer, aspen forest	Loam
Fort Peck	48.31	254.9	1.18	grassland	Loam
Howland	45.2	291.26	8.15	mixed forest	Loam
Manitou Forest	39.1	254.9	0.5	forest	Loam
Metolius	44.45	238.44	1.36	ponderosa pine	Loam
P301	37.07	240.81	2.36	mixed conifer forest	Loam
Park Falls	45.95	269.73	1.42	forest, wetland	Loam
Reynolds Creek	43.12	243.28	2.62	sagebush, aspen	Loam
Rosemount2	44.69	266.94	1.49	corn-soybean	Loam
Shale Hills	40.66	282.09	1.26	forest	Loam
SMAP-OK	36.06	262.78	0.65	pasture/sparse tree	Loam
Tenderfoot Creek	46.95	249.11	1.11	pine	Loam
York Irrigated Soybean	40.93	262.54	1.26	soybean	Loam
Bondville	40.01	271.71	1.57	corn-soy	Clay loam
Coastal Sage UCI	33.73	242.3	1.21	open shrublands	Clay loam
Flag Wildfire	35.45	248.23	1.96	grass, forbes	Clay loam
Iowa Validation Site	41.98	266.32	1.59	corn-soybean	Clay loam
Mozark	38.74	267.8	1.45	deciduous forest	Clay loam
Neb Field 3	41.16	263.53	1.26	soy-corn cropland	Clay loam
NJ meadowlands	40.77	285.91	1.6	grassland	Clay loam
Silver Sword	19.76	204.58	1.81	silver sword	Clay loam
Desert Chaparral UCI	33.61	243.55	0.17	open shrublands	Sandy loam
JERC	31.24	275.54	0.66	evergreen forest, grass understory	Sandy loam
Santa Rita Creosote	31.91	249.16	0.3	shrubland	Sandy loam
Freeman Ranch	29.95	262	1.98	woody savanna	Clay
Goodwin Creek	34.25	270.13	1.6	pasture	Sandy clay loam
Tonzi Ranch	38.43	239.03	0.55	oak savanna	Sandy clay loam
Harvard Forest	42.54	287.83	6.05	mixed forest	Silt loam

monthly time series between SMAP and COSMOS from April 2015 to May 2019. Mean seasonal cycle is computed with maximum available years of data after excluding any missing or flagged values from SMAP or COSMOS (i.e. those under frozen conditions or outside the valid range).

In comparison with COSMOS, SMAP soil moisture may underestimate or overestimate the soil moisture to various extents across the sites with a bias from  $-0.3 \text{ cm}^3/\text{cm}^3$  (NJ Meadowlands) to  $0.39 \text{ cm}^3/\text{cm}^3$  (Park Falls) (Table 2). The RMSE is fairly large as well, ranging from  $0.03 \text{ cm}^3/\text{cm}^3$  to  $0.4 \text{ cm}^3/\text{cm}^3$  and with about 2/3 of the sites larger than  $0.1 \text{ cm}^3/\text{cm}^3$ . However, after the bias (temporal dynamic

variability) is removed, the SMAP product can achieve modest accuracy with the ubRMSE of half the sites smaller than  $0.04 \text{ cm}^3/\text{cm}^3$  and of only one site larger than  $0.08 \text{ cm}^3/\text{cm}^3$ , which are overall comparable to the SMAP mission requirement of  $0.04 \text{ cm}^3/\text{cm}^3$ . The time series correlations based on both the original ( $r$ ) and anomaly ( $r_A$ ) time series are largely consistent and statistically significant at the 99% confidence level for most of the sites, which suggests that the SMAP generally captures well the temporal variation of the COSMOS soil moisture.

We will build an emulator and carry out variance-based sensitivity analysis for each soil type. The USA COSMOS

**Table 2.** Statistics of performance metrics for the consistency between SMAP and COSMOS soil moisture measurements across the sites listed in Table 1. The unit for bias, RMSE, ubRMSE is  $\text{cm}^3/\text{cm}^3$ ; ubRMSE and  $r_A$  is the RMSE and  $r$  calculated from the anomaly time series, respectively.  $N$  is the number of samples. Italic and bold numbers indicate that correlation coefficient is statistically significant at the 95% and 99% confidence levels, respectively. \* indicates the sites selected for the emulator and variance-based sensitivity analyses (See the text for details).

Soil Type	Site	Bias	RMSE	ubRMSE	$r$	$r_A$	$N$
Loam	* ARM-1	0.042	0.053	0.026	<b>0.83</b>	<b>0.83</b>	43
	CC_Pasture	-0.023	0.054	0.027	0.27	0.28	23
	Daniel Forest	-0.152	0.260	0.061	<b>0.67</b>	<b>0.57</b>	19
	Fort Peck	0.006	0.048	0.011	<b>0.43</b>	<b>0.90</b>	30
	Howland	0.035	0.188	0.076	-0.10	<i>0.37</i>	37
	Manitou Forest	-0.022	0.063	0.050	<b>0.52</b>	<b>0.62</b>	46
	Metolius	-0.094	0.167	0.078	<b>0.69</b>	<b>0.48</b>	49
	* P301	0.037	0.079	0.047	<b>0.88</b>	<b>0.71</b>	50
	Park Falls	0.385	0.401	0.068	-0.60	-0.81	25
	Reynolds Creek	-0.048	0.092	0.037	<b>0.80</b>	<b>0.46</b>	43
	Rosemount2	0.053	0.141	0.110	-0.54	-0.58	43
	Shale Hills	0.094	0.100	0.025	<b>0.50</b>	<b>0.59</b>	49
	* SMAP-OK	0.060	0.062	0.014	<b>0.92</b>	<b>0.92</b>	40
	Tenderfoot Creek	-0.100	0.182	0.060	<b>0.44</b>	0.06	32
York Irrigated Soybean	-0.118	0.129	0.030	<b>0.53</b>	<b>0.76</b>	48	
Clay Loam	Bondville	-0.282	0.293	0.054	<b>0.42</b>	<b>0.64</b>	50
	Coastal Sage UCI	0.244	0.251	0.039	<b>0.63</b>	<b>0.57</b>	25
	Flag Wildfire	-0.098	0.138	0.062	<b>0.56</b>	<b>0.76</b>	39
	Iowa Validation Site	-0.074	0.122	0.051	0.02	<b>0.37</b>	49
	* Mozark	0.042	0.064	0.027	<b>0.63</b>	<b>0.82</b>	50
	Neb Field 3	-0.151	0.166	0.050	<b>0.35</b>	<b>0.45</b>	48
	NJ Meadowlands	-0.303	0.330	0.052	-0.01	<i>0.25</i>	49
	* Silver Sword	-0.211	0.217	0.047	<b>0.88</b>	<b>0.87</b>	50
Sandy Loam	* Desert Chaparral UCI	0.010	0.030	0.015	<b>0.85</b>	<b>0.80</b>	50
	* JERC	0.145	0.150	0.007	<b>0.80</b>	<b>0.73</b>	20
	* Santa Rita Creosote	0.038	0.040	0.012	<b>0.87</b>	<b>0.65</b>	30
Clay	* Freeman Ranch	-0.096	0.107	0.040	<b>0.64</b>	<b>0.66</b>	44
Sandy Clay Loam	* Goodwin Creek	0.191	0.194	0.014	<b>0.71</b>	<b>0.71</b>	25
	* Tonzi Ranch	0.010	0.026	0.017	<b>0.97</b>	<b>0.81</b>	46
Silt Loam	* Harvard Forest	0.139	0.183	0.069	-0.49	-0.14	48

sites are characterized by six major soil types. We selected all the available sites for the soil types of Sandy Loam (SL, 3), Clay (C, 1), Sandy Clay Loam (SCL, 2), and Silt Loam (SiL, 1). In consideration of computational demand, we chosen a few sites for the Loam (L, 3) and Clay Loam (CL, 2), which have relatively higher and statistically significant (99% level) anomaly correlations ( $r_A$ ), lower ubRMSE, and larger sample size. This selection is not meant to be exclusive but to be illustrative, so other sites can be used as well. This resulted in a total of 12 sites (marked with “\*” in Table 2). Inclusion of multiple sites for the same soil type helps us understand the impact on the emulator

performance of the factors other than soil texture (i.e. vegetation, soil organic C content, etc.).

### 3.2 Choice of Uncertain Model Parameters

The experiment design for building emulator depends primarily on the range of uncertainty given to individual parameter rather than the shape of the uncertainty distribution (assumed as Gaussian in this study). Therefore, it is crucial to get the uncertainty range within which each parameter normally falls. We identify five uncertain parameters associated with the hydraulic properties of the soil (Table 3). In CLM5, the bulk hydraulic properties of each soil layer are computed as weighted averages of the

**Table 3.** Range of the chosen uncertain parameters associated with hydraulic properties of mineral soils in CLM5.

	Sand%	Saturated Hydraulic Conductivity (mm/s)	Porosity (vol/vol)	Saturated Soil Matric Potential (mm)	Shape Parameter	Soil Organic Matter Fraction
<b>Sandy Clay Loam</b>	45 - 100	0.00223 - 0.18286	0.243 - 0.597	1.0 - 143.3	3.0 - 11.98	0.055-0.490
<b>Sandy Loam</b>	43 - 85	0.00208 - 0.09165	0.177 - 0.693	1.0 - 114.8	3.0 - 8.40	0.000-0.553
<b>Loam</b>	23 - 52	0.00103 - 0.02005	0.217 - 0.685	1.0 - 201.4	3.0 - 9.13	0.000-0.598
<b>Silt Loam</b>	0 - 50	0.00046 - 0.01829	0.308 - 0.662	1.0 - 232.2	3.0 - 9.22	0.000-0.950
<b>Clay Loam</b>	20 - 45	0.00092 - 0.01453	0.317 - 0.635	1.0 - 216.0	3.0 - 13.00	0.077-0.556
<b>Clay</b>	0 - 45	0.00046 - 0.01453	0.332 - 0.632	1.0 - 159.6	4.0 - 18.00	0.151-0.352

organic and mineral components with weights equal to soil organic matter fraction ( $f_{om}$ ) and  $(1-f_{om})$ , respectively. The properties of the mineral soil (saturated hydraulic conductivity  $K_{sat,min}$ , porosity  $\Theta_{sat,min}$ , saturated soil matric potential  $\Psi_{sat,min}$ , and shape parameter  $b_{min}$ ) are determined by the soil texture (fractions of sand and clay), while those of organic matter remain constant. In this paper, we thus focus on the properties of the mineral soil as uncertain parameters. The uncertainty ranges of mineral soil specified in this study are derived from Clapp and Hornberger (1978), which provided representative values ( $\mu$ ) and standard deviations ( $\sigma$ ) for porosity  $\Theta_{sat}$ , saturated soil matric potential  $\Psi_{sat}$ , and shape parameter  $b$  of different soil textures based on analysis of 1845 soils. The range of  $\Theta_{sat,min}$  is specified as  $\mu \pm 3\sigma$ .  $\Psi_{sat,min}$  is given as ranging from the greater of 1 and  $-3\sigma$  to  $3\sigma$ , while shape parameter  $b_{min}$  is given as ranging from the greater of 3 and  $-2\sigma$  to  $2\sigma$ . The statistics ( $\mu$  and  $\sigma$ ) of  $f_{om}$  for each soil texture are determined from all the gridpoints of specific soil texture over the USA based on the CLM global surface data of  $0.5^\circ$  resolution.  $f_{om}$  is given as ranging from the greater of 0 and  $-2\sigma$  to the lesser of 0.95 and  $2\sigma$ . These additional constraints for the minimum and maximum values of parameters are imposed to maintain the required water balance. The mean  $K_{sat}$  values are estimated from USDA<sup>2</sup>, where we assume the corresponding fractions of sand are 0%, 25%, 50%, 75%, and 100% for coarse, moderate coarse, medium, moderate fine, fine and very fine textural classes, respectively. We fit two new exponential functions across the range of sand fraction for the low and high limits of  $K_{sat}$ , by adjusting the coefficients (the initial value of the function or the y-intercept) of  $K_{sat}$  exponential function used in CLM (the base and exponent are unchanged). The coefficients are determined so that the mean  $K_{sat}$  values lie in the middle of two  $K_{sat}$  limits. The range of  $K_{sat}$  for each soil texture is then estimated by feeding the low and high sand fractions of specific soil texture into the corresponding new exponential functions, respectively.

2 [https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/office/ssr10/tr/?cid=nrcs144p2\\_074846](https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/office/ssr10/tr/?cid=nrcs144p2_074846)

### 3.3 Gaussian Process Emulator

Emulation is the process by which a complex model is replaced by a statistical surrogate model that can be run more efficiently to estimate the model output at a large number of unsampled parameter combinations so that variance-based sensitivity analysis becomes feasible. Various emulation methods are available and have been applied to climate and ocean models (Goldstein and Rougier, 2006; Sanderson *et al.*, 2008; Lee *et al.*, 2011; Lee *et al.*, 2013) as well as dynamic vegetation model (Kennedy *et al.*, 2008). In this study, we demonstrate that Gaussian process emulator (O’Hangan, 2006) can be used to study the sensitivity of land surface model (i.e. CLM5) simulated soil moisture across the uncertainty space of multiple parameters.

Gaussian process emulation combines prior beliefs on parameter uncertainty and model behavior (step 1) with soil moisture simulated from a designated number of CLM5 runs (training data, step 2) to produce a posterior probability distribution of soil moisture across the same parameter uncertainty space. Prior belief on model behavior is represented by a prior probability distribution which is assumed to be the Gaussian process in this study. This indicates that the posterior probability distribution conditioned on the training data is also a Gaussian process and both distributions can be specified completely by a mean function and a covariance function. The mean of the posterior distribution is used as an approximation for the CLM5 and the covariance provides the information on uncertainty resulting from using emulation rather than direct CLM5 simulation. Sampling from the posterior distribution provides the necessary data for sensitivity analysis. Any realization from the Gaussian process emulator passes exactly through all the training points. It is worth noting that there are two assumptions behind the use of the Gaussian process emulator for sensitivity analysis: 1) The response of the CLM5-simulated monthly soil moisture to the parameter uncertainty space is smooth and continuous. Such assumed smoothness denotes that each designated CLM5 simulation provides information about the soil moisture at neighboring parameter values and thus ensures a lower uncertainty in predictions far from the

training data. If there is discontinuity in the model with respect to its uncertain parameters, many runs would be required to build the emulator and its increased efficiency is lost. The smoothness assumption can be tested in the emulator validation (Section 4.2). 2) The uncertain model parameters should be separately identifiable (Section 3.2).

### 3.4 Variance-based Sensitivity Analyses

Variance-based sensitivity analysis decomposes the uncertainty in the model simulations into its parametric sources. Here we quantify the sensitivity of CLM5-simulated soil moisture to each of the five parameters in Table 3 and their interactions. In this study, parameter sensitivity is referred to as “contribution to the uncertainty”. The uncertainty of soil moisture is presented here as the standard deviation of its probability distribution around the mean and is estimated by sampling from the emulator mean function. Two measures of sensitivity are calculated. The main effect measures the percentage of the total variance (uncertainty) that will be reduced if a parameter is known precisely. The total effect measures all variance components involving a parameter, specifically both the individual effect and the interaction effect of each parameter with all others, as a percentage of the total variance. The two sensitivity measures are compared to assess the sensitivity of the soil moisture to interactions between uncertain parameters. If there is no interaction with a parameter, two measures are equivalent.

We use readily available software, the Gaussian Emulation Machine for Sensitivity Analysis (GEM-SA)<sup>3</sup>. GEM-SA will provide 1) 200 realizations from a five-dimensional emulator based on 50 points. The mean of the realizations is used to estimate the soil moisture and the spread of the realizations gives the emulator uncertainty; and 2) The main effect and total effect sensitivity measures for each of the uncertain parameters.

### 3.5 Model Experiments

We used a maximin Latin hypercube (McKay *et al.*, 1979) to sample parameter values across the uncertainty ranges of 5 parameters in Table 3. Latin hypercube sampling splits the range in every dimension into  $N$  equal intervals and then makes sure that each interval is sampled exactly once. As recommended by Loepky *et al.*, (2009), the number of points sampled in the Latin hypercube should be ten times the number of chosen uncertain parameters. Therefore, we configured 50 initial CESM model runs. To ensure that the designed emulator is adequate to describe the model behavior at the unsampled points, additional 15 model runs, equal to three times the number of uncertain parameters, were employed to validate the emulator. Parameter values for a third of these runs (5) were chosen deliberately close to those in the original 50 runs and for

the remaining two-thirds (10) were placed further away, determined by a separate Latin hypercube design. The validation process helped identify potential failures with the statistical assumptions made to build the emulator. We build an emulator for each month over the year 2016 and each of the 12 selected sites in Table 2.

In this study, the CESM2 simulations were conducted in land-only mode in which the default land component CLM5 was forced with historical forcing CRUNCEPv7 and with satellite phenology (SP)—prescribed LAI and 20<sup>th</sup> century aerosol deposition rates (described by an I2000Clm50SpGs component set). The CLM5 with default model parameters was first run from 1951 to 2015. The parameter perturbations were then applied with all model runs having identical initial conditions. Additional 4-year simulations driven with 2016 CRUNCEPv7 forcing were carried out for each model run to ensure that the perturbations take effect. The first year is treated as a spin-up period and the analysis is performed on the monthly soil moisture averaged for the last three years.

## 4. Results

### 4.1 CLM5-simulated Soil Moisture

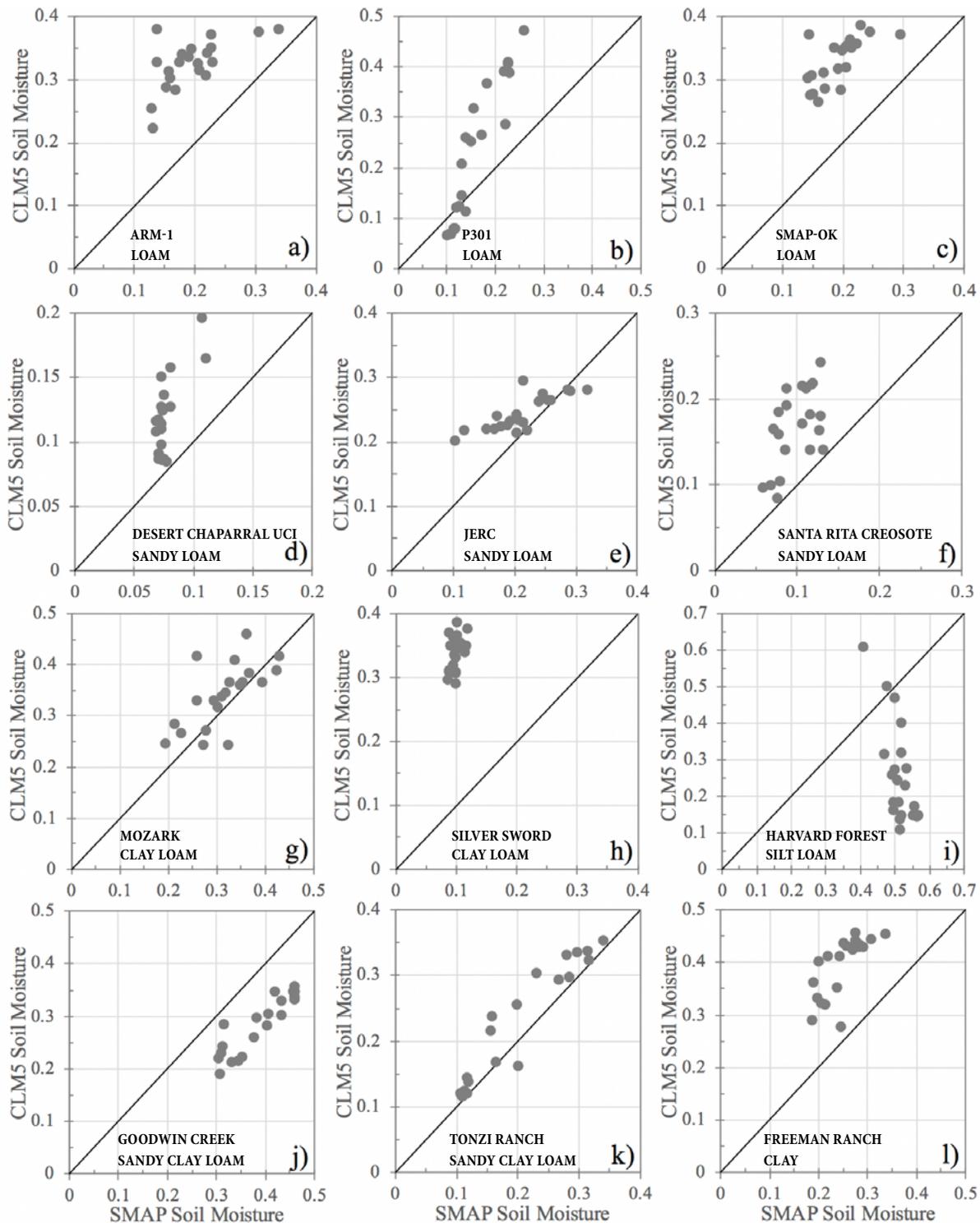
We first examined the performance of surface soil moisture (0–5cm) simulated from CLM5 with default model parameters against its SMAP counterpart from April 2015 to December 2016 at the selected 12 COSMOS sites listed in Table 2 (**Figure 1**). This exercise serves as the main motivation to use the emulator for understanding the relative contribution of various model parameters to the uncertainty in model-simulated soil moisture and further to calibrate/optimize these parameters for improved soil moisture prediction. All the performance metrics are calculated from the original time series because such a short period does not justify the derivation of anomalies. ubRMSE is calculated by removing the mean bias ( $\text{ubRMSE} = \sqrt{\text{RMSE}^2 - \text{Bias}^2}$ ) (Entekhabi *et al.*, 2010a). We did not include the comparison of CLM5 soil moisture with COSMOS due to the missing data at several sites, particularly Goodwin Creek.

The CLM5 simulations tend to overestimate the soil moisture with positive biases in all the sites, except for Goodwin Creek and Harvard Forest (**Table 4**). The CLM5-simulated soil moisture at Silver Sword and Harvard Forest have the worst performances with the largest positive (0.238  $\text{cm}^3/\text{cm}^3$ ) and negative (-0.253  $\text{cm}^3/\text{cm}^3$ ) biases, respectively. Their RMSE values are also larger than those at other sites, with 0.24  $\text{cm}^3/\text{cm}^3$  and 0.3  $\text{cm}^3/\text{cm}^3$  at Silver Sword and Harvard Forest, respectively. It is worth noting that Silver Sword attains one of the lowest ubRMSE values (0.025  $\text{cm}^3/\text{cm}^3$ ) after the bias is removed, while Harvard Forest still retains the highest ubRMSE (0.16  $\text{cm}^3/\text{cm}^3$ ). This indicates that the mismatch between CLM5-simulated and

3 <http://www.tonyohagan.co.uk/academic/GEM/>

SMAP soil moisture at Silver Sword is mostly attributed to its mean bias. In fact, all the sites, except for Harvard Forest, P301 and Mozark, meet the accuracy requirement of the SMAP mission with the uBRMSE values no greater than 0.04 cm<sup>3</sup>/cm<sup>3</sup>. Except for Silver Sword and Harvard

Forest, the CLM5-simulated soil moisture at all the other sites well captures the temporal variation of SMAP soil moisture with the time series correlation coefficients ranging from 0.56 to 0.95 and statistically significant at 99% level. Although the correlation coefficient of Silver Sword (0.4)



**Figure 1.** Comparison of monthly soil moisture from CLM5 with default model parameters and SMAP retrieval from Apr 2015 to December 2016 over the selected 12 COSMOS sites (Table 2) of different soil types (see text for details).

**Table 4.** Performance metrics of CLM5-simulated soil moisture against its SMAP counterpart across the 12 sites selected in Table 2. The unit for bias, RMSE, ubRMSE is  $\text{cm}^3/\text{cm}^3$ ; All the metrics are calculated from the original time series. N is the number of samples. Italic and bold numbers indicate that correlation coefficient is statistically significant at the 95% and 99% confidence levels, respectively.

	Site	Bias	RMSE	ubRMSE	r	N
Loam	ARM-1	0.128	0.134	0.042	<b>0.627</b>	21
	P301	0.071	0.114	0.090	<b>0.935</b>	21
	SMAP-OK	0.136	0.140	0.030	<b>0.678</b>	21
Clay Loam	Mozark	0.025	0.056	0.050	<b>0.671</b>	21
	Silver Sword	0.238	0.239	0.025	<i>0.395</i>	21
Sandy Loam	Desert Chaparral UCI	0.040	0.047	0.024	<b>0.697</b>	21
	JERC	0.031	0.047	0.035	<b>0.811</b>	21
	Santa Rita Creosote	0.067	0.077	0.037	<b>0.558</b>	21
Clay	Freeman Ranch	0.144	0.149	0.038	<b>0.730</b>	21
Sandy Clay Loam	Goodwin_Creek	-0.105	0.108	0.025	<b>0.905</b>	21
	Tonzi Ranch	0.023	0.035	0.027	<b>0.954</b>	21
Silt Loam	Harvard Forest	-0.253	0.299	0.159	-0.694	21

is statistically significant at 95% level, the dynamic range of SMAP soil moisture is much constrained and damped (temporal standard deviation of  $0.009 \text{ cm}^3/\text{cm}^3$  in contrast to that of  $0.028 \text{ cm}^3/\text{cm}^3$  for CLM5), which makes the resulting correlation coefficient severely compromised. Harvard forest is the only site showing a strong negative correlation (-0.7), indicating that CLM5 simulation fails to capture the temporal variation of SMAP soil moisture. The reason behind this feature needs further examination but is beyond the scope of this study.

It is known that land surface models are strongly limited in their ability to reproduce observed soil moisture by a lack of critical information on soil hydraulic properties and, more importantly, by the need to represent complex, nonlinear, and nonresolvable processes across large distances in a very simple way. There are often considerable mean biases or biases in the dynamic range and time variability of the modeled soil moisture (Reichle *et al.*, 2004). The poor performances of CLM5-simulated soil moisture, particularly at such sites as Harvard Forest and Silver Sword, confirm these biases and further inspires us to explore the potential for an improved prediction. And understanding the effect of parameter uncertainty in the modeled soil moisture is the first step to achieve this goal.

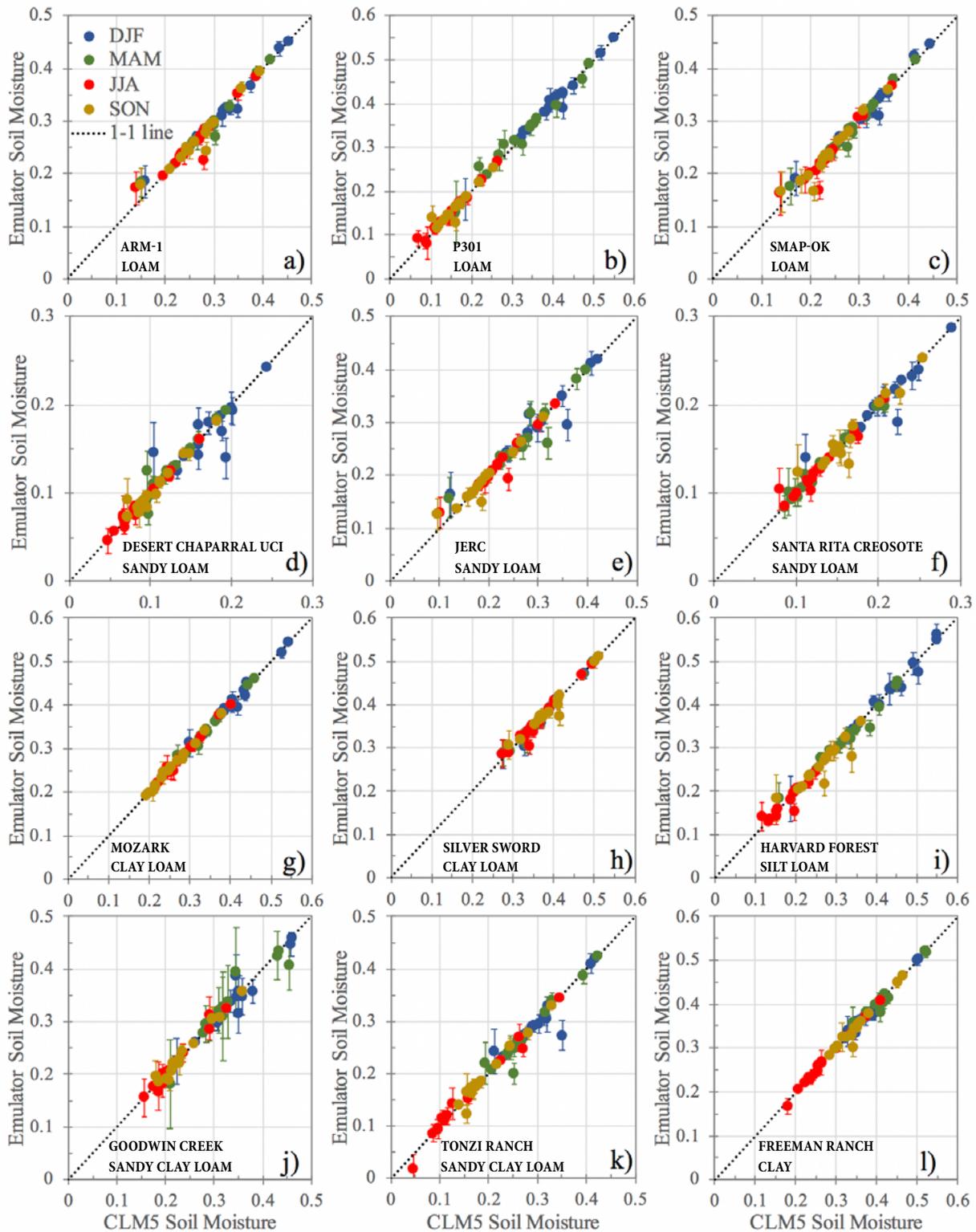
#### 4.2 Validation of the Emulator

The evaluation of the emulator is conducted by comparing the CLM5-simulated and emulator-predicted surface (0–5cm) soil moisture from 15 validation simulations at the 12 USA COSMOS sites of different soil types for the winter (DJF), spring (MAM), summer (JJA), and autumn (SON) seasons (Figure 2). Also shown is the 95% probability bound as a measure of the “confidence” of the emulator prediction based on 200 realizations. It is desirable to have CLM-simulated soil moisture from 95% of the validation

runs contained within the corresponding 95% probability bound of the emulator. Based on this criterion, we see that emulation of surface soil moisture performs best in JJA across various sites, followed by SON and MAM, and performs worst in DJF (Table 5). Two sites with clay loam performs best across different seasons, while three sites with sandy loam performs worst. At Harvard forest (silt loam), only 80% of the CLM simulations lie within the 95% confidence interval of the emulator for all the seasons, except for JJA. However, it is unfeasible to generalize the performance on this soil texture based on just one site. The five validation points placed close to the training data are generally characterized by very small 95% probability bounds that cover the CLM simulations for all the sites and seasons (indiscernible in Figure 2), indicating that the

**Table 5.** Number of the validation runs in which CLM-simulated soil moisture is contained within the corresponding 95% probability bound of the emulator. Bold number indicates approximately 93% of validation runs that meet the criterion.

	DJF	MAM	JJA	SON
ARM-1	13	12	13	14
P301	13	14	14	13
SMAP-OK	14	13	14	14
Mozark	12	14	15	14
Silver Sword	14	14	14	14
Desert Chaparral UCI	10	13	14	13
JERC	13	12	14	13
Santa Rita Creosote	13	13	13	12
Freeman Ranch	12	12	15	14
Goodwin Creek	13	13	15	14
Tonzi Ranch	14	14	13	13
Harvard Forest	12	12	14	12



**Figure 2.** Comparison between CLM5-simulated and emulator predicted seasonal surface (0-5cm) soil moisture ( $\text{cm}^3/\text{cm}^3$ ) of year 2016 across various selected USA COSMOS sites. The uncertainty of emulator prediction is presented as 95% probability ( $2\sigma$ ) limits. The points with indiscernible uncertainty bars are those deliberately placed close to the training data.

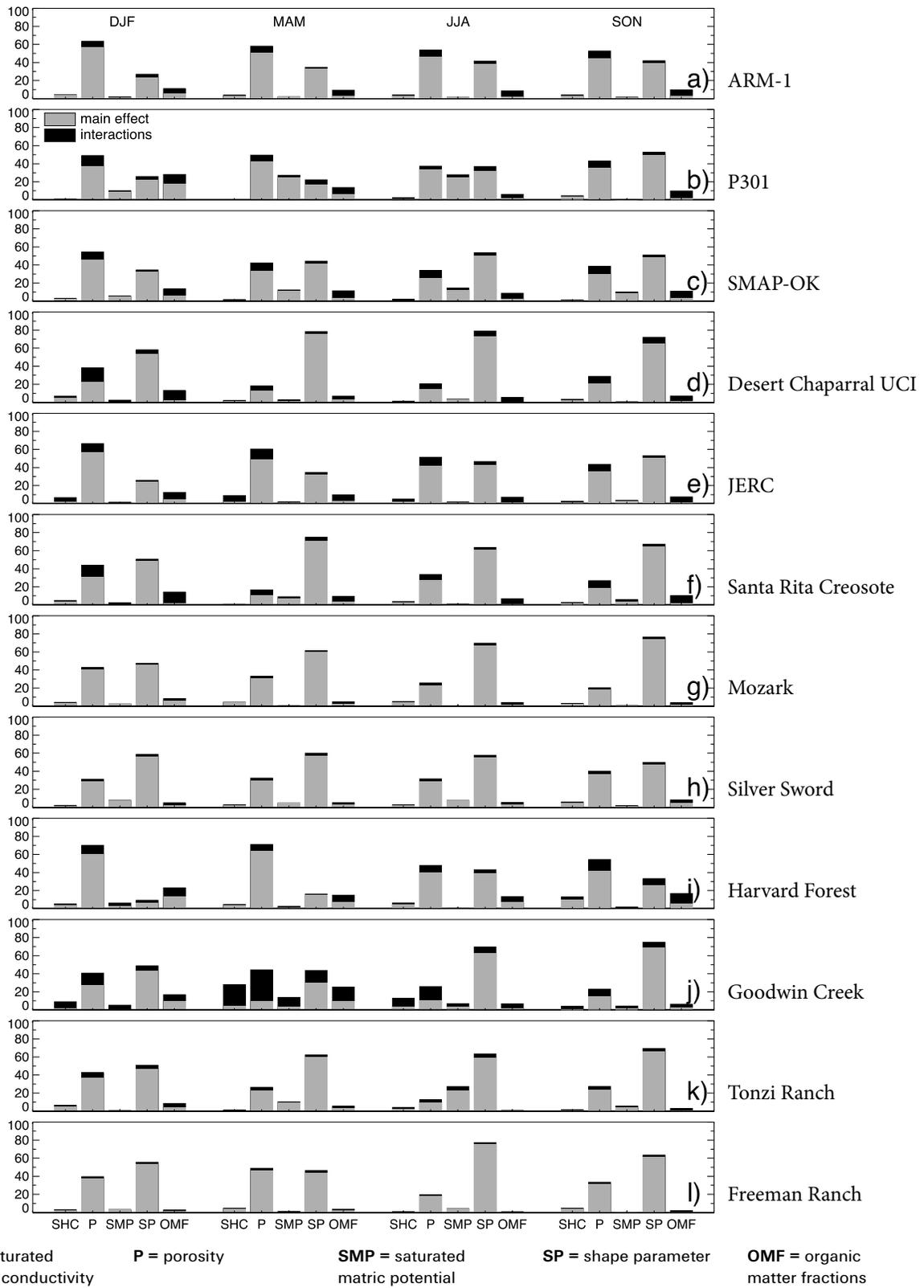
emulator estimates well close to the training data. With the exception of DJF at Desert Chaparral UCI, the emulator also estimates fairly well at points far away from the training data, with at least 7 out of these 10 points having 95% confidence intervals that cover the CLM simulations across various sites and seasons. In terms of the comparison between the emulator-predicted mean and CLM-simulated surface soil moisture, the correlations are all above 0.90 (ranging from 0.91–1.0) with the exception of the DJF at Desert Chaparral UCI (0.82) and the RMSEs range from 0.01 to 0.02 cm<sup>3</sup>/cm<sup>3</sup> which is much lower than the SMAP mission requirement of 0.04 cm<sup>3</sup>/cm<sup>3</sup> (Appendix B). The emulation of root zone soil moisture exhibits similar performances and statistics (Appendix C). Overall, the emulator captures well the model behavior across the entire parameter uncertainty space for different soil textures and seasons and is considered to be accurate and robust. The small emulator standard deviation indicates that its mean is representative of the CLM simulation and there will be a very small effect on the accuracy to estimate the parameter sensitivities using the emulator.

### 4.3 Sensitivity Analysis

The sensitivity analysis partitions the variance due to the five parameters into the variance due to each parameter and their interactions. The results for surface and root zone soil moisture are summarized in **Figures 3 and 4**, respectively, as well as in Appendix. The main effect variance is the percentage of the total variance due to the perturbation of each parameter individually. The interaction variance is the percentage of the total variance due to the interaction between each parameter and other parameters. Note the total variance contributions will typically not add to 100% because the variance contributions are shared between parameters in the presence of interactions (Appendix D). The further it is away from 100% gives an indication of the strength of interaction effects. The main effects for surface soil moisture are much more important than the interactions, with the portion of the variance contributed by the individual parameters ranging from 81% to 98% across almost all the sites and seasons (Appendix D). The only exception is MAM at Goodwin Creek where parameter interactions become strong with the summed total effect variance being 155% and main effect accounting for only 57% of the total variance. The main effects are generally stronger in soil textures with lower fraction than higher fraction of sand. For example, the summed total effect variances are generally smaller than 105% (standard deviation  $\sigma = 0.9\%$ , number of samples  $N = 12$ ) and the main effects account for more than 95% ( $\sigma = 0.8\%$ ) of the variance consistently across all the seasons for clay loam and clay, whose fractions of sand are 20–45% and 0–45% (Table 3), respectively. For sandy clam loam and sandy loam whose respective fractions of sand are 45–100%

and 43–85%, the main effects largely account for around 85–90% of the variance (excluding MAM at Goodwin Creek), variably across the seasons and sites ( $\sigma = 3.9\%$ ,  $N=20$ ). Among the main effects of five parameters, the majority of surface soil moisture variance is described by the uncertainties in porosity and shape parameter. The combination of porosity and shape parameters explains about 60–94% of the total variance across different sites and seasons, with the exception of MAM at Goodwin Creek which has only 40% of the total variance explained. Similarly, for soil textures with lower sand fractions (clay loam and clay), we see higher percentage (> 85%) of the variance explained by the combined porosity and shape parameter across different seasons, in comparison with soil textures of higher sand fractions. The other three parameters are shown to have relatively small individual effects with the explained variances by  $K_{\text{sat,min}}$ ,  $f_{\text{om}}$ , and  $\Psi_{\text{sat,min}}$  less than 10%, 18%, and 25%, respectively. However, the median and 75<sup>th</sup> percentile of their explained variances across all the sites and seasons ( $N = 48$ ) are only 2% and 3.7% for  $K_{\text{sat,min}}$ , 2.6% and 5% for  $f_{\text{om}}$ , as well as 1.8% and 4.5% for  $\Psi_{\text{sat,min}}$ , respectively.

Nevertheless, the relative importance of porosity versus shape parameter varies strongly with sites and seasons. The majority of the sites, including Desert Chaparral UCI, Santa Rita Creosote, Mozark, Silver Sword, Goodwin Creek, Tonzi Ranch, and Freeman Ranch, show that the variance in surface soil moisture is attributed distinctly more to the uncertainty in shape parameter than to that in porosity across different seasons. Over these sites, the explained variance by shape parameter ranges from about 45% to 75% (excluding MAM at Goodwin Creek), while those by porosity ranges from about 9% to 45%. The relative dominance by shape parameter is generally the weakest in DJF. Different pattern of parameter sensitivity is observed at the sites of ARM-1 and Harvard forest, where porosity contributes markedly more to the total variance than shape parameter across the seasons. The explained variance by porosity ranges from about 40% to 65%, while that by shape parameter ranges from 7% to 40%. Such difference in parametric variance contribution is more evident in DJF and MAM, but becomes smaller in JJA and SON. The sites of P301, SMAP-OK and JERC show mixed responses. The porosity at P301 and JERC presents their dominances in variance contribution in all the seasons except for SON, while shape parameter at SMAP-OK exhibit its dominance in all the seasons except for DJF. Overall, there is no simple rule of thumb for the parametric sensitivity across different sites and seasons. The same soil texture can exhibit different relative importance of uncertain parameters, while different soil textures can present the similar pattern of variance dominance. Examining the parametric sensitivity at more sites of the same soil texture would be useful to generalize these conclusions. These results suggest that some factors



**Figure 3.** The parameter sensitivities of surface soil moisture at the selected USA COSMOS sites. The gray and black bars show the main effect sensitivities and how much the interaction of each parameter with the others contributes to the surface soil moisture variance.

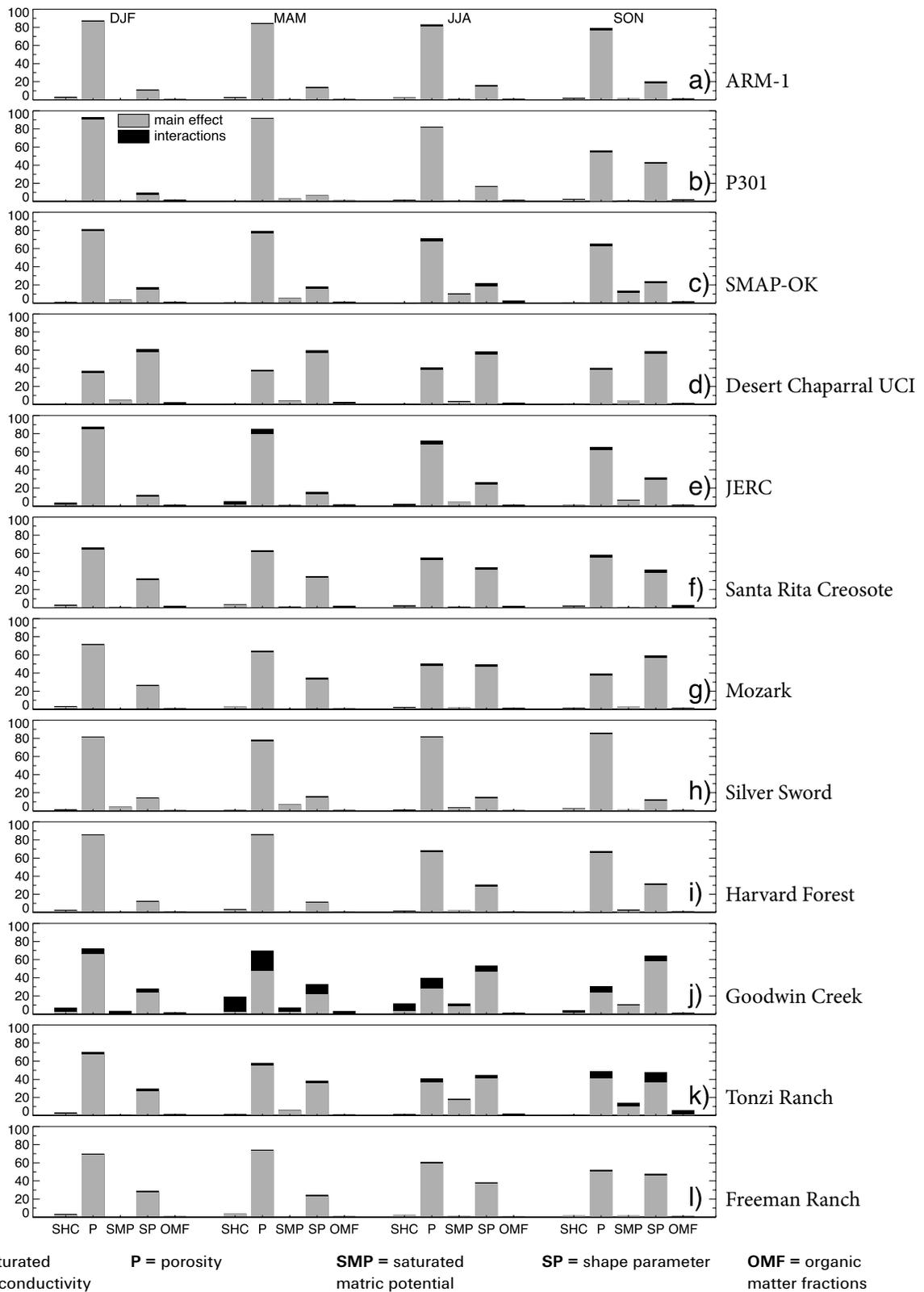


Figure 4. Same as Figure 3, but for root zone soil moisture.

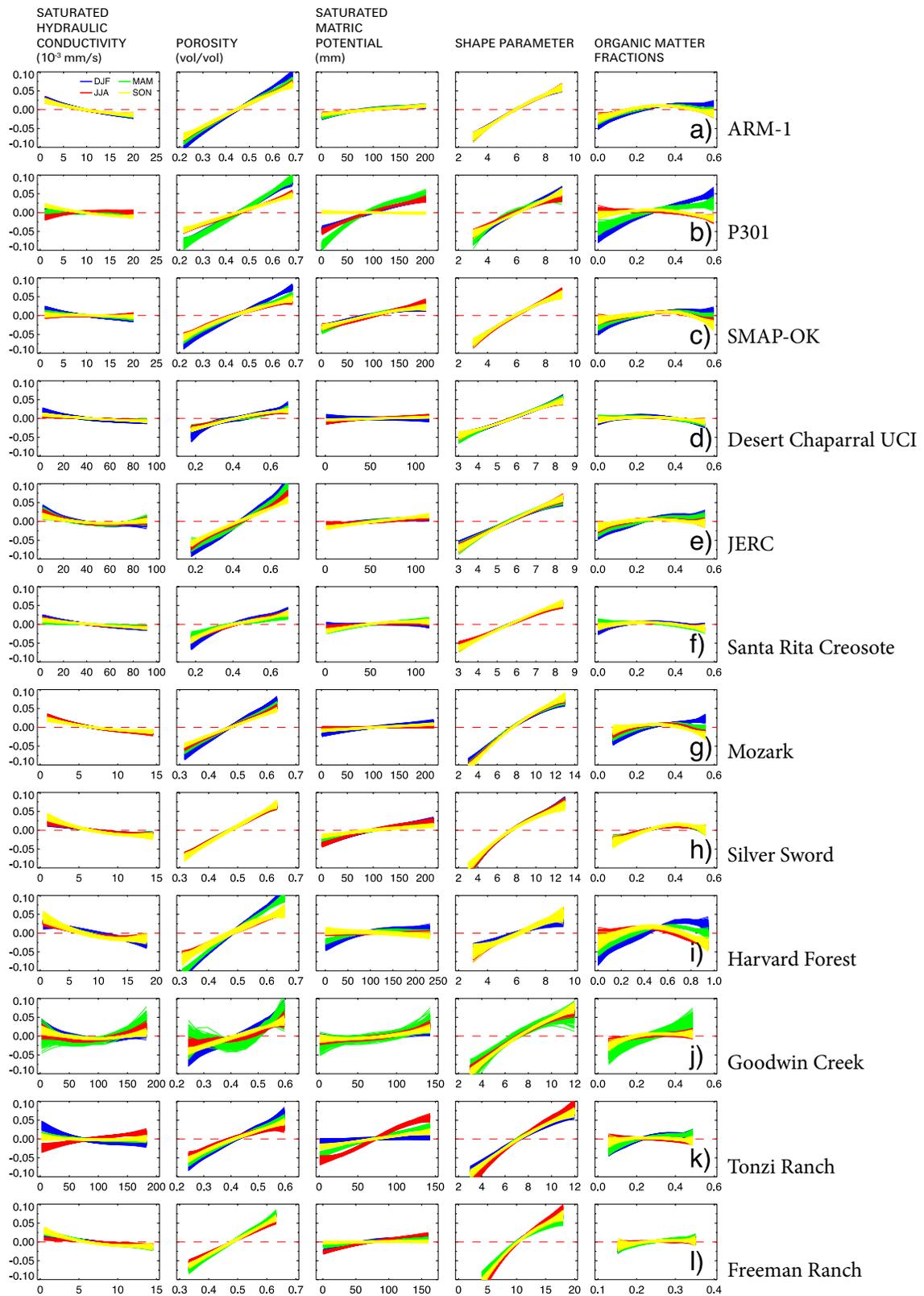
other than soil texture, such as overlaid land cover type, may also contribute to the process of determining the parametric sensitivity of soil moisture.

The parametric sensitivity of root zone soil moisture shares similar features to that of surface soil moisture but with several differences. The main effects prominently dominate the total variances in all the sites and seasons with the interactions between the model parameters much weaker for root zone than for surface soil moisture. With two sites (Goodwin Creek and Tonzi Branch) of sandy clay loam excluded, the total variances and the explained variances by individual parameters range from 101% to 107% ( $\sigma = 1.2\%$ ,  $N = 40$ ) and 93% to 99% ( $\sigma = 1.1\%$ ), respectively (Appendix E). The sandy clay loam shows stronger parametric interactions than other soil textures, with the total variances and the variances explained by the main effects ranging from 104% to 130% ( $\sigma = 8.8\%$ ,  $N=8$ ) and from 74% to 97% ( $\sigma = 7.4\%$ ), respectively. For surface soil moisture, the corresponding total variances and the explained variances by individual parameters are 102–119% ( $\sigma = 4.2\%$ ,  $N=40$ ) and 83–98% ( $\sigma = 3.8\%$ ) for soil textures other than sandy clay loam as well as 106–155% ( $\sigma = 16.4\%$ ,  $N=8$ ) and 57–95% ( $\sigma = 12.8\%$ ) for sandy clay loam. Also, the maximum interaction effect occurs at MAM of Goodwin Creek. The larger standard deviations of surface soil moisture indicate that both main and interaction effects are more variable across different soil textures and seasons than root zone soil moisture.

The main effects for root zone soil moisture are differentiated by parameter as well. However, the majority of the sites (8) show that uncertainty of porosity is more important than that of shape parameter in the variance of root zone soil moisture in all the seasons, with the explained variances by porosity ranging from 50% to 90% (median 77%, 75<sup>th</sup> percentile 83%,  $N = 32$ ) in contrast to 5–45% by shape parameter (median 17%, 75<sup>th</sup> percentile 29%). Desert Chaparral UCI is the only site in which shape parameter contributes slightly more to the variance of root zone soil moisture than porosity across the seasons, with the explained variances by shape parameter being 55–58% in contrast to 34–38% by porosity. Several sites, including Mozark, Goodwin Creek, and Tonzi Branch, give mixed responses with relative dominance of two parameters divided across the seasons. All three sites exhibit dominance of porosity in DJF and MAM, while Mozark and Goodwin Creek show dominance of shape parameter in SON. Overall,  $f_{om}$  and  $K_{sat,min}$  are shown to have negligible individual effects for root zone soil moisture with the explained variances less than 1% and 3% for all the sites and seasons, respectively.  $\Psi_{sat,min}$  exhibits a small individual effect with the explained variance less than 17% but its median and 75<sup>th</sup> percentile ( $N=48$ ) being only 1.3% and 3.8%, respectively. The individual effects of all three parameters ( $f_{om}$ ,  $K_{sat,min}$ , and  $\Psi_{sat,min}$ ) are smaller for root

zone than for surface soil moisture. However, the largest reduction in the individual effect occurs to  $f_{om}$ , which is likely attributed to the specification of depth as 0.5 meter where organic matter is assumed to act like peat.

We also examine the relationships between the anomaly of emulator-estimated surface soil moisture and each parameter (all other parameters fixed) for all the sites and seasons (Figure 5). The anomaly is calculated relative to the mean of surface soil moisture across the range of each parameter and 200 realizations. The dominance of  $\Theta_{sat,min}$  (2<sup>nd</sup> column) and  $b_{min}$  (4<sup>th</sup> column) in the main effect relationships is immediately evident. There are clear positive linear relationships of surface soil moisture to  $\Theta_{sat,min}$  and  $b_{min}$  in all the sites and seasons, with the exception of MAM at Goodwin Creek (row j) which exhibits a distinct curvilinear relationship between surface soil moisture and  $\Theta_{sat,min}$ . MAM at Goodwin Creek is also characterized by the larger emulator uncertainties (the spread of the green lines) than other seasons and sites for all the parameters, which are likely attributed to the stronger interaction effects. As expected, emulator uncertainty increases as parameter approaches the edges of parameter range in all the sites and seasons. There are also negative linear relationships of surface soil moisture to  $K_{sat,min}$  (1<sup>st</sup> column) and positive linear relationships to  $\Psi_{sat,min}$  (3<sup>rd</sup> column) in all the sites and seasons, except for curvilinear relationships to  $K_{sat,min}$  at Goodwin Creek caused by interaction effects. However, these relationships are not strong with soil moisture anomalies generally in the range of  $\pm 0.025 \text{ cm}^3/\text{cm}^3$ , except for relatively stronger relationships to  $\Psi_{sat,min}$  at P301 and Tonzi Ranch as well as to  $K_{sat,min}$  at Goodwin Creek. The soil moisture anomalies due to the changes of  $\Theta_{sat,min}$  or  $b_{min}$  generally range from  $-0.1$  to  $0.1 \text{ cm}^3/\text{cm}^3$ , with the relative strength of relationships to  $\Theta_{sat,min}$  versus  $b_{min}$  consistent with what were shown in Figure 3. Surface soil moisture presents a mixed response to  $f_{om}$ , with weak inverted U-shaped curvilinear relationships in most sites and seasons but strong linear relationships for DJF and MAM at P301. The resulting soil moisture anomalies largely vary between  $-0.025$  and  $0.025 \text{ cm}^3/\text{cm}^3$ , but could reach  $-0.075 \sim 0.05 \text{ cm}^3/\text{cm}^3$  for Harvard Forest as well as DJF and MAM of P301. The main effects relationships also demonstrate seasonal differences which are reflected by the changes in the direction (positive or negative), the shape (linear or curvilinear), and the strength of the relationship. Most sites present little seasonal differences with the realizations of all the seasons clustered into a single thick line, while P301, Harvard Forest, Goodwin Creek and Tonzi Ranch exhibit all kinds of changes in relationship characteristics across the seasons for different parameters. The high seasonal sensitivity of the main effect relationships is likely attributed to the strong parametric interactions at two sites of high sand fraction, but to the poor model performances at two forest sites. Figure 5 also reveals the values



**Figure 5.** The relationship between surface soil moisture and each of the parameters listed in Table 3 at the selected USA COSMOS sites. In each panel, the thickness of the line for each season (each color) represents the emulator uncertainty from 200 realizations. The y-axis is plotted as the anomaly of surface soil moisture relative to the mean across the range of each parameter and 200 realizations. The spread on the y-axis represents the effects of the parametric uncertainty on soil moisture.

of each parameter around which soil moisture anomalies change the signs at various sites. When the parameters are smaller than these values, the main effects will lead to drier (wetter) conditions relative to their mean states for  $\Theta_{\text{sat,min}}$ ,  $\Psi_{\text{sat,min}}$  and  $b_{\text{min}}(K_{\text{sat,min}})$  and vice versa. These values range from 0.005 to 0.1 mm/s for  $K_{\text{sat,min}}$ , 0.4 to 0.5 for  $\Theta_{\text{sat,min}}$ , 50 to 100 mm for  $\Psi_{\text{sat,min}}$ , and 5 to 10 for  $b_{\text{min}}$ , respectively. The higher fraction of sand the soil contains, the lower (higher) these values are for  $\Theta_{\text{sat,min}}$ ,  $\Psi_{\text{sat,min}}$  and  $b_{\text{min}}(K_{\text{sat,min}})$ . The main effect relationships of root zone soil moisture to each parameter have a lot in common with surface soil moisture, except that they consistently exhibit smaller emulator uncertainties and seasonal differences across all the sites (Figure 6). Regardless of the shape (linear or curvilinear), the strength of these relationships is also weaker, particularly with  $f_{\text{om}}$ . These results are consistent with what were shown in Figure 4.

## 5. Summary and Conclusions

Sensitivity analysis is essential for attributing the uncertainty in the output of a mathematical model to different sources of uncertainty in its inputs (parameters), identifying the parameters that have the most impact on model output, and for guiding further model development. The most commonly used sensitivity analysis method is one-at-a-time (OAT), which is straightforward to implement but presents two serious downsides: 1) It is unable to study the interactions between parameters; and 2) It severely undersamples the parameter uncertainty space when the number of parameters is large. An alternative that overcomes these downsides is variance-based methods, which decompose the variance of the output(s) into terms corresponding to the different parameters and their interactions. A full variance-based sensitivity analysis often requires information of the model output throughout the entire parameter uncertainty space, which is achieved by a large number of model runs in a Monte Carlo style. For complex global models with many parameters, the computational expense thus prevents the source of uncertainty from being rigorously quantified. In this study, we employed Gaussian process emulation, in which a statistical surrogate model is first trained with considerably fewer model runs than Monte Carlo, then used to estimate the model output at a large number of unsampled parameter combinations. Such statistical model can be run efficiently to generate the same level of information required by a full variance-based sensitivity analysis. The emulation approach has been applied to climate and ocean models as well as dynamic vegetation model. The primary aim of this study was to assess the method for carrying out a variance-based sensitivity analysis of Land Surface Model (LSM).

LSM-simulated soil moisture has been widely used to understand many complex processes in the Earth system, but it often presents considerable biases in its mean, dy-

namic range, and time variability. A realistic estimate of the parametric uncertainty of model-simulated soil moisture provides more useful information for improving its prediction. We quantified the relative contributions of different parameters and their interactions to the overall uncertainty in the CLM5-simulated surface and root zone soil moisture for four seasons in USA. The study sites were chosen to have high-quality soil moisture observations (will be used for parameter calibration in a subsequent paper) and also represent different major soil types. We focus on four hydraulic property parameters of mineral soil whose values are determined by soil texture, including saturated hydraulic conductivity, porosity, saturated matric potential, and shape parameter as well as organic matter fraction that constitutes the weight to calculate the bulk hydraulic properties from the weighted average of the organic and mineral components. The ranges of each parameter for different soil types were specified based on a widely-cited literature. A maximin Latin Hypercube was used to sample an appropriate number of combinations of parameter values covering the five-dimensional parameter uncertainty space for conducting CLM5 simulations to train the emulator.

The validation simulations have shown that the emulator captures well the model behavior across the entire parameter uncertainty space for different soil textures and seasons. The comparison between the emulator-predicted mean and CLM-simulated surface soil moisture indicates that the correlations are all above 0.90 (ranging from 0.91–1.0) with the exception of the DJF at Desert Chaparral UCI (0.82) and the RMSEs range from 0.01 to 0.02  $\text{cm}^3/\text{cm}^3$  which is much lower than the SMAP mission requirement of 0.04  $\text{cm}^3/\text{cm}^3$ . The emulation of root zone soil moisture exhibits even better performance statistics with the correlations all above 0.95 and RMSEs below 0.01  $\text{cm}^3/\text{cm}^3$ . The uncertainty in using the emulator instead of the model simulation is very small for the validation points placed both close to and far away from the training data, indicating that the emulator mean is representative of the CLM simulation and there will be a very small effect on the accuracy to estimate the parametric uncertainties using the emulator. These results suggest that the emulator is accurate and robust across different variables (surface versus root zone soil moisture), seasons, and soil textures. In particular, the number of model simulations required to train the emulator increases linearly with the number of parameters ( $10 \times$  number of parameters). The approach is therefore much more efficient than widely used factorial methods for a comprehensive coverage of parameter space of a large number of uncertain parameters.

The sensitivity analysis has shown that main effects for surface soil moisture are much more important than the interactions with the large portion of the variance contributed by individual parameters for almost all the sites and seasons. The majority of surface soil moisture vari-

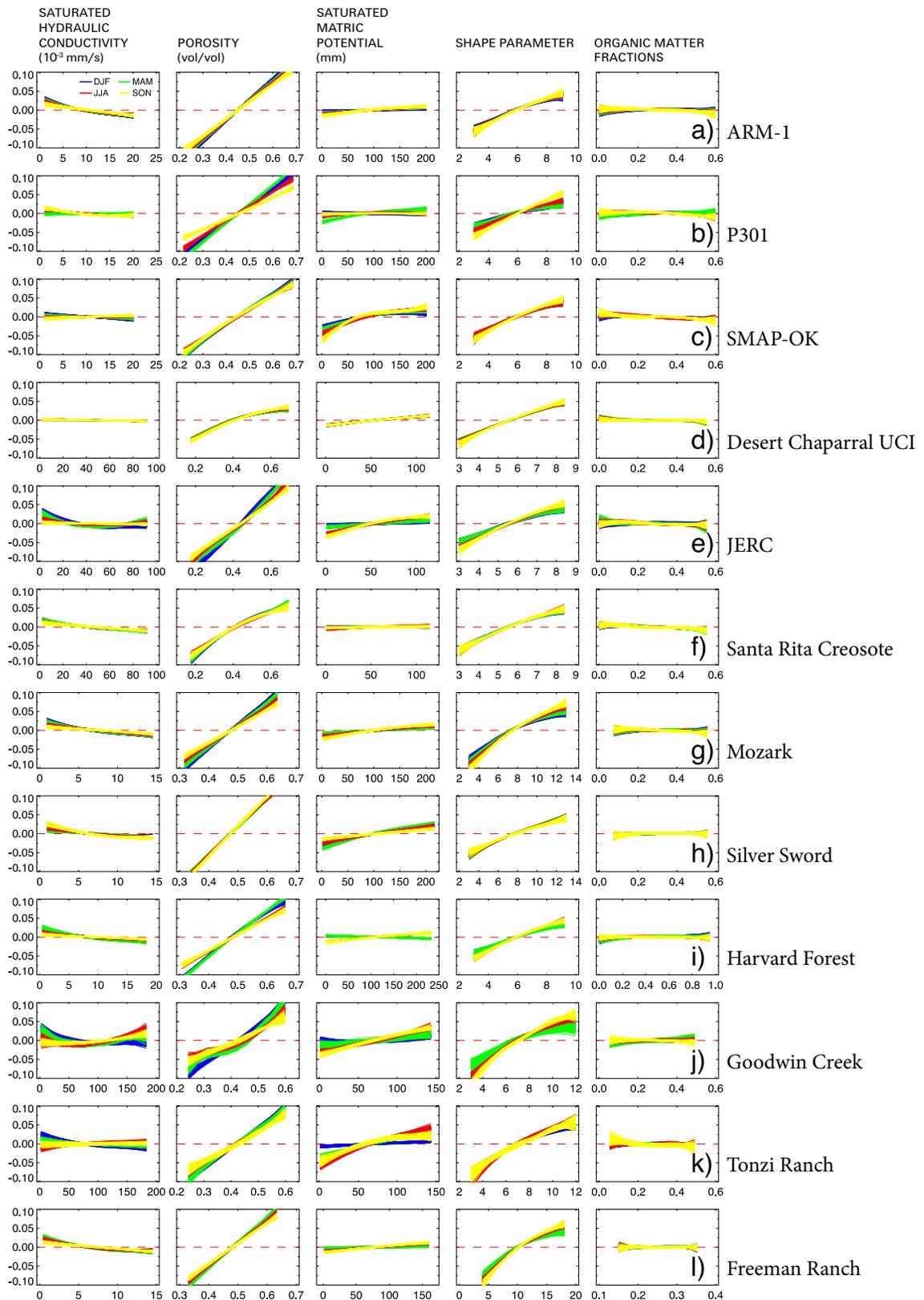


Figure 6. Same as Figure 5, but for root zone soil moisture.

ance is dominated by the uncertainties in porosity and shape parameter, while the other three parameters exhibit small individual effects with the 75<sup>th</sup> percentile of explained variances by  $K_{\text{sat,min}}$ ,  $f_{\text{om}}$ , and  $\Psi_{\text{sat,min}}$  across all the sites and seasons being 3.7%, 5%, and 4.5%, respectively. However, the relative importance of porosity versus shape parameter varies strongly with sites and seasons. The majority of the sites (7 out of 12) show that the variance in surface soil moisture is attributed distinctly more to the uncertainty in shape parameter than to that in porosity across different seasons, while two sites (ARM-1 and Harvard Forest) present the opposite uncertainty pattern of the two parameters and the other remaining sites give mixed responses. There seems no simple rule of thumb for the parametric uncertainty across different sites. The same soil texture can exhibit different relative importance of uncertain parameters, while different soil textures can present similar pattern of variance dominance. These results suggest that some factors other than soil texture, such as overlaid land cover type, may also contribute to the process of determining the parametric sensitivity of soil moisture. For root zone soil moisture, the main effects prominently dominate the total variances in all the sites and seasons with the interactions between the model parameters much weaker than for surface soil moisture. Also, both main and interaction effects are less variable for root zone soil moisture across different soil textures and seasons with smaller standard deviations of the explained variances by two effects. However, the majority of the sites show that uncertainty of porosity is more important than that of shape parameter in the variance of root zone soil moisture in all the seasons. The individual effects of the other three parameters ( $f_{\text{om}}$ ,  $K_{\text{sat,min}}$ , and  $\Psi_{\text{sat,min}}$ ) are smaller for root zone than for surface soil moisture, with the explained variances less than 1% and 3% for  $f_{\text{om}}$  and  $K_{\text{sat,min}}$  as well as the 75<sup>th</sup> percentile of the explained variances for  $\Psi_{\text{sat,min}}$  being 3.8% across all the sites and seasons. One common feature for both surface and root zone soil moisture is that the main effects are generally stronger in soil textures with lower fraction of sand (or parametric interactions are stronger in soil textures with higher fraction of sand).

We found that there are clear positive linear relationships of surface soil moisture to  $\Theta_{\text{sat,min}}$  and  $b_{\text{min}}$ , weak negative linear relationship to  $K_{\text{sat,min}}$ , as well as weak positive linear relationships to  $\Psi_{\text{sat,min}}$  in all the sites and seasons. Surface soil moisture presents a mixed response to  $f_{\text{om}}$ , with weak inverted U-shaped curvilinear relationships in most sites and seasons. These main effects relationships present little seasonal differences at most sites, except for two sites of mixed Forest and two sites of high sand fraction whose

relationships exhibit the changes in the direction, the shape, and the strength across the seasons for different parameters. In comparison with surface soil moisture, the main effects relationships of root zone soil moisture to each parameter consistently exhibit smaller emulator uncertainties and seasonal differences for all the sites. Regardless of the shape (linear or curvilinear), the strength of these relationships is also weaker, particularly with  $f_{\text{om}}$ .

Although our study focuses on soil moisture, the approach presented here could be applied to other key quantities of land surface model, such as trace gas flux ( $\text{N}_2\text{O}$ ,  $\text{CH}_4$ ), whose estimates exhibit considerable uncertainty and are affected by many poorly constrained parameters whose values are not unambiguously known or even knowable. The approach can also be carried out to quantify the uncertainty of the variable and its sources from different land surface models. The resulting sensitivity analyses can then be compared to better understand the uncertainty in model simulations attributed to model structure differences, and perhaps reduce such uncertainty. Our immediate future work will focus on reducing the uncertainty and biases in the model-simulated soil moisture, which can be achieved by constraining the results with the observation data, i.e. calibration. Calibration is the process to tune parameter values for optimal model simulations (against the observations), which simultaneously reduces the uncertainty in the model parameters and improve knowledge about them. Our study shows which parameters should be focused on to improve the model simulations of surface versus root zone soil moisture for different soil textures and seasons, which serves as a useful guidance for model development and improvement. The availability of global high-resolution soil moisture from SMAP, including the Level-3 surface soil moisture product (SPL3SMP) as well as the model-derived, value-added Level 4 surface and root zone soil moisture product (SPL4SMGP), will certainly provide an unprecedented potential in support of this exercise.

### Acknowledgements

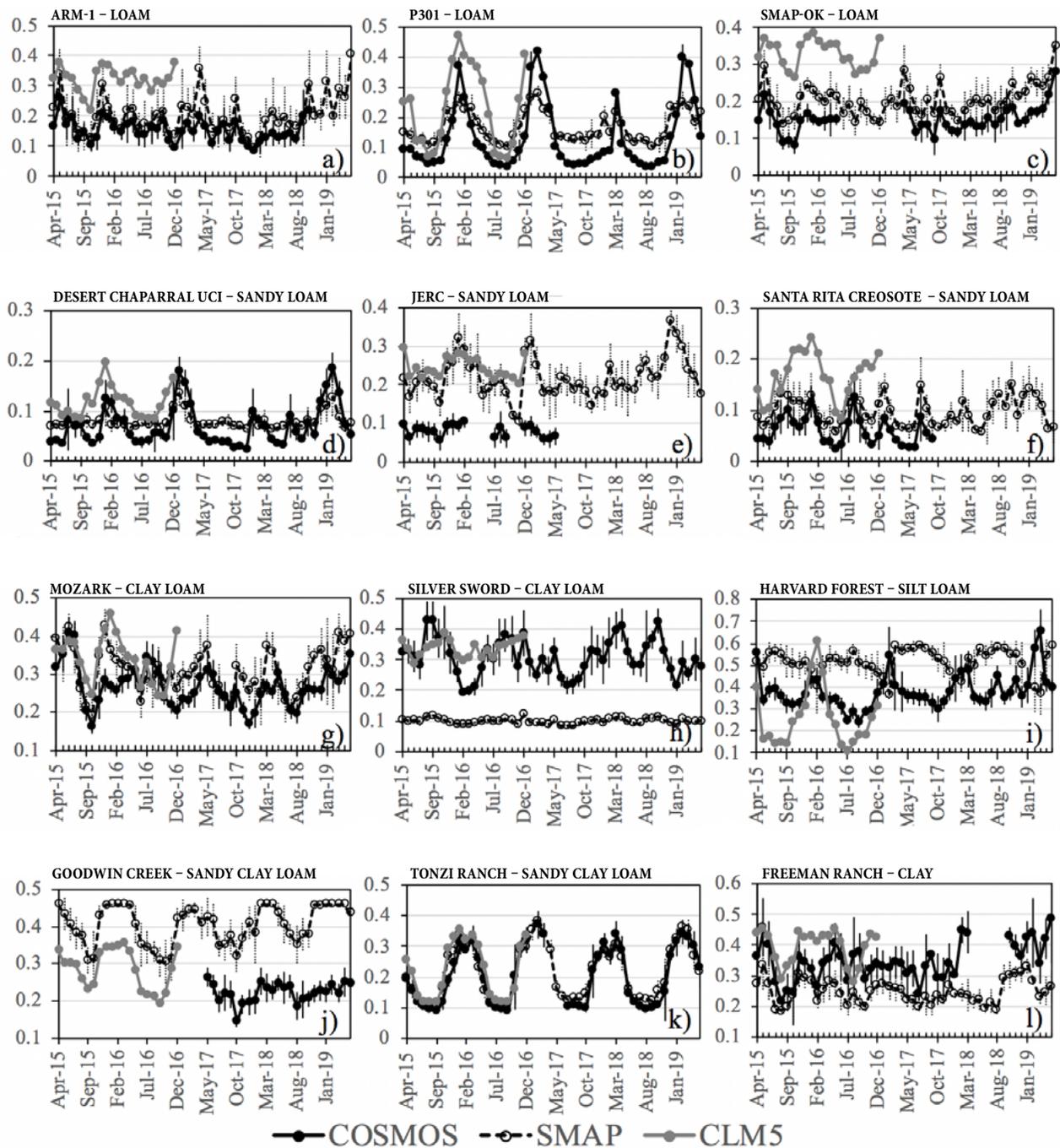
This work is funded by NASA Science Utilization of the Soil Moisture Active-Passive Mission (SUSMAP) (NNX16AN15G) and other government, industry and foundation sponsors of the MIT Joint Program on the Science and Policy of Global Change. For a complete list of sponsors, see <http://globalchange.mit.edu/sponsors>. We acknowledge the Soil Moisture Active-Passive Mission (SMAP) science team, Cosmic-ray Soil Moisture Observing System (COSMOS) project team, and the Climate and Global Dynamics Laboratory (CGD) at the National Center for Atmospheric Research (NCAR) for their roles in making available the relevant datasets and Community Land Model version 5. We thank Tony O'Hagan for making the Gaussian Emulation Machine for Sensitivity Analysis (GEM-SA, <http://www.tonyohagan.co.uk/academic/GEM/>) software available.

## 6. References

- Brunke, M.A., P. Broxton, J. Pelletier, D. Gochis, P. Hazenberg, D.M. Lawrence, L.R. Leung, G.-Y. Niu, P.A. Troch and X. Zeng (2016): Implementing and Evaluating Variable Soil Thickness in the Community Land Model, Version 4.5 (CLM4.5). *J Clim* 29: 3441–3461.
- Chan, S.K. *et al.*, (2018): Development and assessment of the SMAP enhanced passive soil moisture product. *Remote Sens Environ* 204, 931–941.
- Chen, Y., K. Yang, J. Qin, Q. Cui, H. Lu, Z. La, M. Han and W. Tang (2017): Evaluation of SMAP, SMOS, and AMSR2 soil moisture retrievals against observations from two networks on the Tibetan Plateau. *J Geophys Res Atmos* 122, 5780–5792, doi:10.1002/2016JD026388.
- Clapp, B.R. and G.M. Hornberger (1978): Empirical Equations for some soil hydraulic properties. *Water Resour Res* 14, 601–604.
- Cui, C., J. Xu, J. Zeng, K.-S. Chen, X. Bai, H. Lu, Q. Chen and T. Zhao (2018): Soil Moisture Mapping from Satellites: An Intercomparison of SMAP, SMOS, FY3B, AMSR2, and ESA CCI over Two Dense Network Regions at Different Spatial Scales. *Remote Sens* 10, 33.
- Dirmeyer, P.A., Z. Guo and X. Gao (2004): Comparison, Validation, and Transferability of Eight Multiyear Global Soil Wetness Products. *J Hydrometeorol* 5, 1011–1033.
- El Hajj, M., N. Baghdadi, M. Zribi, N. Rodríguez-Fernández, J.P. Wigneron, A. Al-Yaari, A. Al Bitar, C. Albergel and J.-C. Calvet (2018): Evaluation of SMOS, SMAP, ASCAT and Sentinel-1 Soil Moisture Products at Sites in Southwestern France. *Remote Sens* 10, 569.
- Entekhabi, D., E. Njoku, P. Houser, M. Spencer, T. Doiron, J. Smith, R. Girard, S. Belair, W. Crow, T. Jackson, Y. Kerr, J. Kimball, R. Koster, K. McDonald, P. O'Neill, T. Pultz, S. Running, J.C. Shi, E. Wood and J. van Zyl (2004): The Hydrosphere State (HYDROS) mission concept: An Earth system pathfinder for global mapping of soil moisture and land freeze/thaw. *IEEE Trans Geosci Remote Sens* 42, 2184–2195.
- Entekhabi, D., R.H. Reichle, R.D. Koster and W.T. Crow (2010a): Performance Metrics for Soil Moisture Retrievals and Application Requirements. *J Hydrometeorol* 11, 832–840, doi: 10.1175/2010JHM1223.1
- Entekhabi, D. *et al.* (2010b): The Soil Moisture Active Passive (SMAP) Mission. In *Proceedings of the IEEE* 98(5): 704–716. doi: 10.1109/JPROC.2010.2043918
- Famiglietti, J.S., D. Ryu, A.A. Berg, M. Rodell and T.J. Jackson (2008): Field observations of soil moisture variability across scales. *Water Resour Res* 44, W01423, doi:10.1029/2006WR005804.
- Goldstein, M. and Rougier, J. (2006): Bayes linear calibrated prediction for complex systems. *J Am Stat Assoc* 101, 1132–1143.
- Guo, Z., P.A. Dirmeyer, Z.-Z. Hu, X. Gao and M. Zhao (2006): Evaluation of the Second Global Soil Wetness Project soil moisture simulations: 2. sensitivity to external meteorological forcing. *J Geophys Res* 111, D22S03, doi:10.1029/2006JD007845
- Hornbuckle, B., S. Irvin and J. Patton (2011): The impact of rapidly growing vegetation on COSMOS soil moisture measurements. Abstract presented at 2011 Soil Science Society of America Annual Meeting, San Antonio, Texas.
- Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K.C. Mo, C. Ropelewski, J. Wang, R. Jenne and D. Joseph (1996): The NCEP/NCAR 40-year reanalysis project. *Bull Am Meteorol Soc* 77, 437–471.
- Kerr, Y.H., P. Waldteufel, J.P. Wigneron, S. Delwart, F. Cabot, J. Boutin, M.J. Escorihuela, J. Font, N. Reul, C. Gruhier, *et al.* (2010): The smos mission: New tool for monitoring key elements of the global water cycle. *Proceedings of the IEEE* 98, 666–687.
- Lawrence, D.M. and A.G. Slater (2008): Incorporating organic soil into a global climate model. *Clim Dyn* 30. DOI:10.1007/s00382-007-0278-1.
- Loeppky, J., J. Sack and W. Welch (2009): Choosing the Sample Size of a Computer Experiment: A Practical Guide. *Technometrics* 51(4), 366–376.
- Mao, J., X. Shi, P.E. Thornton, F.M. Hoffman, Z. Zhu and R.B. Myneni (2013): Global latitudinal-asymmetric vegetation growth trends and their driving mechanisms: 1982–2009. *Remote Sensing* 5: 1484–1497.
- McKay, M., W. Conover and R. Beckman (1979): A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21, 239–245, .
- Mitchell, T.D. and P.D. Jones (2005): An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int J Climatol* 25: 693–712.
- Montzka, C., H.R. Bogen, M. Zreda, A. Monerris, R. Morrison, S. Muddu and H. Vereecken (2017): Validation of Spaceborne and Modelled Surface Soil Moisture Products with Cosmic-Ray Neutron Probes. *Remote Sens* 9, 103.
- Njoku, E.G., T.J. Jackson, V. Lakshmi, T.K. Chan and S.V. Nghiem (2003): Soil moisture retrieval from AMSR-E. *IEEE Trans Geosci Remote Sens* 41, 215–229.
- O'Hagan, A. (2006): Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering and System Safety* 91, 1290–1300.
- Parinussa, R.M., T.R.H. Holmes, N. Wanders, W.A. Dorigo and R.A.M. de Jeu (2015): A preliminary study toward consistent soil moisture from AMSR2. *J Hydrometeorol* 16, 932–947.
- Piao, S.L., *et al.* (2012): The carbon budget of terrestrial ecosystems in East Asia over the last two decades. *Biogeosciences* 9: 3571–3586.
- Qian, T. *et al.* (2006): Simulation of global land surface conditions from 1948 to 2004: Part I: Forcing data and evaluations. *J Hydrometeorology* 7: 953–975.
- Sanderson, B., R. Knutti, T. Aina, C. Christensen, N. Faull, D. Frame, W. Ingram, C. Piani, D. Stainforth, D. Stone and M. Allen (2008): Constraints on model response to greenhouse gas forcing and the role of subgrid-scale processes. *J Climate* 21, 2384–2400.
- Seneviratne, S.I., T. Corti, E.L. Davin, M. Hirschi, E.B. Jaeger, I. Lehner, B. Orlowsky and A.J. Teuling (2010): Investigating soil moisture-climate interactions in a changing climate: A review. *Earth-Sci Rev* 99, 125–161
- Sheffield, J., E.F. Wood, N. Chaney, K.Y. Guan, S. Sadri, X. Yuan, L. Olang, A. Amani, A. Ali, S. Demuth, *et al.* (2014): A drought monitoring and forecasting system for Sub-Saharan African water resources and food security. *Bull Am Meteorol Soc* 95, 861–882.
- Shi, X., J. Mao, P.E. Thornton and M. Huang (2013): Spatiotemporal patterns of evapotranspiration in response to multiple environmental factors simulated by the Community Land Model. *Environ Res Lett* 8:024012.

- Sospedra-Alfonso, R. and W.J. Merryfield (2018): Initialization and Potential Predictability of Soil Moisture in the Canadian Seasonal to Interannual Prediction System. *J Climate* 31, 5205–5224, doi: 10.1175/JCLI-D-17-0707.1
- Swenson, S.C. and D.M. Lawrence (2014): Assessing a dry surface layer-based soil resistance parameterization for the Community Land Model using GRACE and FLUXNET-MTE data. *J Geophys Res* 119, 10, 299–10,312, DOI:10.1002/2014JD022314.
- van den Hurk, B., F. Doblas-Reyes, G. Balsamo *et al.* (2012): Soil moisture effects on seasonal temperature and precipitation forecast scores in Europe. *Clim Dyn* 38: 349, doi: 10.1007/s00382-010-0956-2
- Wagner, W., S. Hahn, R. Kidd, T. Melzer, Z. Bartalis, S. Hasenauer, J. Figa-Saldana, P. de Rosnay, A. Jann, S. Schneider, *et al.* (2013): The ASCAT soil moisture product: A review of its specifications, validation results, and emerging applications. *Meteorol Z* 22, 5–33.
- Wanders, N., D. Karssenbergh, A. de Roo, S.M. de Jong and M.F.P. Bierkens (2014): The suitability of remotely sensed soil moisture for improving operational flood forecasting. *Hydrol Earth Syst Sci* 18, 2343–2357.
- Zeng, X. and M. Decker (2009): Improving the numerical solution of soil moisture-based Richards equation for land models with a deep or shallow water table. *J Hydrometeorol* 10: 308–319.
- Zhang, X., T. Zhang, P. Zhou, Y. Shao and S. Gao (2017): Validation Analysis of SMAP and AMSR2 Soil Moisture Products over the United States Using Ground-Based Measurements. *Remote Sens* 9, 104.
- Zreda, M., D. Desilets, T.P.A. Ferré and R.L. Scott (2008): Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons. *Geophys Res Lett* 35, L21402, doi:10.1029/2008GL035655.
- Zreda, M., W.J. Shuttleworth, X. Zeng, C. Zweck, D. Desilets, T. Franz and R. Rosolem (2012): COSMOS: the COsmic-ray Soil Moisture Observing System. *Hydrol Earth Syst Sci* 16, 4079–4099, doi: 10.5194/hess-16-4079-2012.
-

**Appendix A**



**Figure A1.** Comparison of soil moisture from CLM5 with default model parameters, COSMOS measurement, and SMAP retrieval over the selected 12 COSMOS sites (Table 2) of different soil types (see text for details).

## Appendix B

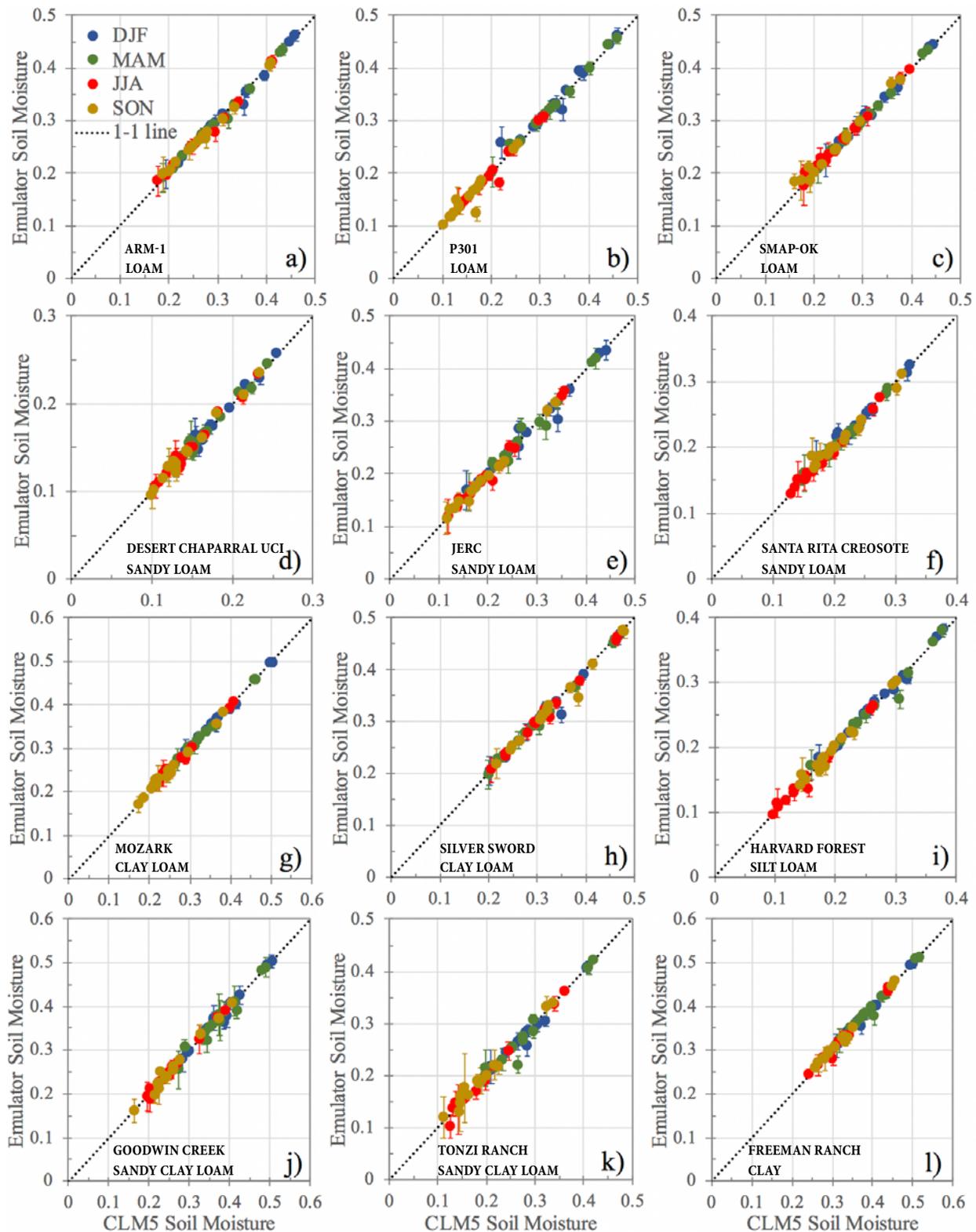
**Table B1.** Correlations and RMSEs ( $\text{cm}^3/\text{cm}^3$ ) between CLM5-simulated and emulator-predicted surface soil moisture for 15 validation runs.

		Correlation				RMSE			
		DJF	MAM	JJA	SON	DJF	MAM	JJA	SON
Loam	ARM-1	0.99	0.98	0.96	0.97	0.01	0.01	0.02	0.01
	P301	0.99	0.99	0.99	0.94	0.01	0.01	0.01	0.01
	SMAP-OK	0.99	0.99	0.96	0.97	0.01	0.01	0.02	0.01
Clay Loam	Mozark	0.99	0.99	0.99	1.00	0.01	0.01	0.01	0.00
	Silver Sword	0.99	0.99	0.98	0.97	0.01	0.01	0.01	0.01
Sandy Loam	Desert Chaparral UCI	0.82	0.96	0.99	0.97	0.02	0.01	0.00	0.01
	JERC	0.96	0.95	0.96	0.97	0.02	0.02	0.01	0.01
	Santa Rita Creosote	0.94	0.99	0.97	0.96	0.01	0.01	0.01	0.01
Clay	Freeman Ranch	0.99	0.98	1.00	0.98	0.01	0.01	0.00	0.01
Sandy Clay Loam	Goodwin Creek	0.97	0.95	0.99	0.98	0.02	0.02	0.01	0.01
	Tonzi Ranch	0.92	0.97	0.99	0.98	0.02	0.02	0.01	0.01
Silt Loam	Harvard Forest	0.99	0.99	0.96	0.91	0.01	0.01	0.01	0.02

## Appendix C

**Table C1.** Correlations and RMSEs ( $\text{cm}^3/\text{cm}^3$ ) between CLM5-simulated and emulator-predicted root zone soil moisture for 15 validation runs.

		Correlation				RMSE			
		DJF	MAM	JJA	SON	DJF	MAM	JJA	SON
Loam	ARM-1	1.00	1.00	1.00	1.00	0.01	0.01	0.01	0.01
	P301	0.98	1.00	0.98	0.95	0.01	0.00	0.01	0.01
	SMAP-OK	1.00	1.00	1.00	0.99	0.00	0.00	0.01	0.01
Clay Loam	Mozark	1.00	1.00	0.99	1.00	0.01	0.00	0.01	0.01
	Silver Sword	0.99	1.00	1.00	0.99	0.01	0.01	0.01	0.01
Sandy Loam	Desert Chaparral UCI	0.99	0.99	0.99	0.99	0.01	0.01	0.01	0.00
	JERC	0.99	0.99	0.99	1.00	0.01	0.01	0.01	0.01
	Santa Rita Creosote	0.99	0.99	1.00	0.99	0.01	0.00	0.00	0.01
Clay	Freeman Ranch	1.00	0.99	0.99	1.00	0.01	0.01	0.01	0.01
Sandy Clay Loam	Goodwin Creek	0.99	0.98	1.00	0.99	0.01	0.01	0.01	0.01
	Tonzi Ranch	0.99	0.98	0.99	0.99	0.01	0.01	0.01	0.01
Silt Loam	Harvard Forest	1.00	0.99	0.99	0.99	0.01	0.01	0.01	0.01



**Figure C1.** Comparison between CLM5-simulated and emulator predicted seasonal root zone (0-100cm) soil moisture ( $\text{cm}^3/\text{cm}^3$ ) of year 2016 across various selected USA COSMOS sites. The uncertainty of emulator prediction is presented as 95% probability ( $2\sigma$ ) limits. The points with indiscernible uncertainty bars are those deliberately placed close to the training data.

## Appendix D

**Table D1.** Model parameters and their effects on the variance of surface soil moisture at the ARM-1 site (Loam). “Main” and “Total” represents main and total effect variance contribution (%), respectively.

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{sat,min}$	3.65	4.19	3.42	3.99	3.41	4.07	3.34	4.03
$\theta_{sat,min}$	57.03	63.45	51.07	57.96	46.10	53.80	44.24	52.62
$\Psi_{sat,min}$	1.50	1.81	1.78	2.02	1.00	1.23	1.22	1.47
$b_{min}$	25.49	26.91	33.14	34.70	38.80	41.37	39.42	41.92
$f_{om}$	5.36	11.09	3.18	9.27	2.31	8.52	2.68	9.64
All	93.02	107.45	92.60	107.94	91.63	108.99	90.90	109.68

**Table D2.** Same as Table D1 but at the P301 site (Loam).

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{sat,min}$	0.21	1.03	0.03	0.39	0.91	2.44	3.72	4.48
$\theta_{sat,min}$	37.54	48.97	42.64	49.50	33.20	37.25	34.94	43.06
$\Psi_{sat,min}$	9.08	9.99	24.80	27.00	24.71	27.68	0.07	0.32
$b_{min}$	22.28	25.80	16.87	22.13	31.56	36.64	49.86	52.60
$f_{om}$	18.02	28.00	6.24	13.53	1.67	6.10	2.04	9.78
All	87.12	113.79	90.57	112.55	92.05	110.11	90.63	110.24

**Table D3.** Same as Table D1 but at the SMAP-OK site (Loam).

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{sat,min}$	0.21	1.03	0.03	0.39	0.91	2.44	3.72	4.48
$\theta_{sat,min}$	37.54	48.97	42.64	49.50	33.20	37.25	34.94	43.06
$\Psi_{sat,min}$	9.08	9.99	24.80	27.00	24.71	27.68	0.07	0.32
$b_{min}$	22.28	25.80	16.87	22.13	31.56	36.64	49.86	52.60
$f_{om}$	18.02	28.00	6.24	13.53	1.67	6.10	2.04	9.78
All	87.12	113.79	90.57	112.55	92.05	110.11	90.63	110.24

**Table D4.** Same as Table D1 but at the Desert Chaparral UCI site (Sandy Loam).

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{sat,min}$	4.36	6.80	1.10	2.15	0.85	1.34	2.40	3.36
$\theta_{sat,min}$	22.76	38.25	12.44	18.26	14.08	20.36	20.62	28.80
$\Psi_{sat,min}$	0.34	2.36	1.27	2.74	3.04	3.50	0.40	0.71
$b_{min}$	53.59	58.19	75.42	78.35	73.25	78.88	65.30	71.99
$f_{om}$	2.13	13.09	2.58	6.88	0.71	5.50	1.07	6.95
All	83.19	118.69	92.81	108.38	91.93	109.58	89.80	111.81

**Table D5.** Same as Table D1 but at the JERC site (Sandy Loam).

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{\text{sat,min}}$	2.49	6.51	1.68	8.90	2.09	5.06	1.03	2.54
$\theta_{\text{sat,min}}$	57.11	66.42	48.47	60.36	41.99	51.37	35.18	43.52
$\Psi_{\text{sat,min}}$	0.70	1.47	1.23	2.01	0.79	1.86	2.80	3.69
$b_{\text{min}}$	23.75	25.76	32.04	34.49	42.87	46.47	50.41	52.98
$f_{\text{om}}$	5.10	12.29	3.03	9.77	1.31	7.09	1.10	7.45
All	89.15	112.45	86.44	115.53	89.05	111.85	90.52	110.18

**Table D6.** Same as Table D1 but at the Santa Rita Creosote site (Sandy Loam).

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{\text{sat,min}}$	3.70	4.77	0.51	0.89	2.97	3.71	1.96	2.75
$\theta_{\text{sat,min}}$	31.05	43.93	10.41	16.58	27.36	33.62	17.99	26.78
$\Psi_{\text{sat,min}}$	0.09	2.26	6.56	8.99	0.30	0.99	3.06	5.83
$b_{\text{min}}$	48.69	50.55	70.87	74.65	60.53	63.50	64.20	67.16
$f_{\text{om}}$	1.84	14.20	3.04	9.24	1.04	6.77	1.47	10.25
All	85.38	115.71	91.39	110.35	92.20	108.59	88.67	112.77

**Table D7.** Same as Table D1 but at the Mozark site (Clay Loam).

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{\text{sat,min}}$	3.05	4.06	3.74	4.40	4.19	5.13	2.39	3.05
$\theta_{\text{sat,min}}$	39.92	42.58	30.14	32.98	22.89	25.69	18.03	20.27
$\Psi_{\text{sat,min}}$	2.08	2.41	0.46	0.56	0.02	0.21	0.56	0.79
$b_{\text{min}}$	45.43	47.45	59.34	61.47	67.08	69.70	73.80	76.16
$f_{\text{om}}$	5.66	8.16	2.42	4.84	1.57	3.98	1.59	3.75
All	96.14	104.66	96.10	104.25	95.74	104.71	96.37	104.02

**Table D8.** Same as Table D1 but at the Silver Sword site (Clay Loam).

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{\text{sat,min}}$	3.05	4.06	3.74	4.40	4.19	5.13	2.39	3.05
$\theta_{\text{sat,min}}$	39.92	42.58	30.14	32.98	22.89	25.69	18.03	20.27
$\Psi_{\text{sat,min}}$	2.08	2.41	0.46	0.56	0.02	0.21	0.56	0.79
$b_{\text{min}}$	45.43	47.45	59.34	61.47	67.08	69.70	73.80	76.16
$f_{\text{om}}$	5.66	8.16	2.42	4.84	1.57	3.98	1.59	3.75
All	96.14	104.66	96.10	104.25	95.74	104.71	96.37	104.02

**Table D9.** Same as Table D1 but at the Harvard Forest site (Silt Loam).

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{\text{sat,min}}$	3.76	5.31	3.93	4.67	4.81	6.38	9.58	12.95
$\Theta_{\text{sat,min}}$	60.41	70.04	64.01	71.05	39.78	47.77	41.44	54.27
$\Psi_{\text{sat,min}}$	3.19	6.20	1.29	2.42	0.06	0.31	0.44	1.91
$b_{\text{min}}$	6.79	9.16	15.09	16.14	38.69	42.95	26.05	33.23
$f_{\text{om}}$	13.74	23.08	7.18	14.79	6.92	13.30	5.93	16.72
All	87.89	113.79	91.51	109.07	90.28	110.71	83.45	119.08

**Table D10.** Same as Table D1 but at the Goodwin Creek site (Sandy Clay Loam).

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{\text{sat,min}}$	1.92	8.98	4.01	27.97	3.29	12.85	1.09	4.01
$\Theta_{\text{sat,min}}$	27.35	40.67	9.03	44.25	10.48	25.78	15.09	22.92
$\Psi_{\text{sat,min}}$	1.12	5.14	3.59	13.83	3.59	6.92	1.91	4.25
$b_{\text{min}}$	43.36	48.74	30.13	43.53	62.61	69.64	68.82	74.76
$f_{\text{om}}$	9.59	16.79	9.90	25.24	1.44	6.74	2.40	6.27
All	83.34	120.32	56.67	154.82	81.41	121.93	89.31	112.21

**Table D11.** Same as Table D1 but at the Tonzi Ranch site (Sandy Clay Loam).

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{\text{sat,min}}$	4.61	6.24	0.38	1.26	2.20	4.16	0.15	1.49
$\Theta_{\text{sat,min}}$	36.85	42.46	22.89	26.31	9.18	12.51	23.45	27.33
$\Psi_{\text{sat,min}}$	0.19	0.73	9.44	10.23	22.69	27.13	4.03	5.63
$b_{\text{min}}$	46.50	50.68	59.56	62.24	58.76	63.30	66.32	69.25
$f_{\text{om}}$	4.40	8.44	2.91	5.52	0.23	0.75	0.35	2.92
All	92.56	108.55	95.17	105.56	93.06	107.85	94.30	106.62

**Table D12.** Same as Table D1 but at the Freeman Ranch site (Clay).

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{\text{sat,min}}$	2.21	2.97	3.60	4.64	0.43	0.78	3.65	4.61
$\Theta_{\text{sat,min}}$	37.74	39.53	46.42	48.74	18.13	19.60	31.50	33.34
$\Psi_{\text{sat,min}}$	2.56	3.11	0.82	1.18	3.78	4.03	0.03	0.23
$b_{\text{min}}$	53.20	55.49	43.82	46.41	75.47	77.25	61.25	63.46
$f_{\text{om}}$	1.32	2.53	1.81	3.49	0.12	0.71	0.53	1.96
All	97.03	103.63	96.48	104.46	97.92	102.37	96.96	103.60

## Appendix E

**Table E1.** Model parameters and their effects on the variance of root zone soil moisture at the ARM-1 site (Loam). “Main” and “Total” represents main and total effect variance contribution (%), respectively.

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{\text{sat,min}}$	2.57	2.96	2.22	2.62	1.81	2.25	1.46	1.86
$\theta_{\text{sat,min}}$	85.96	87.09	83.12	84.41	81.18	82.74	76.87	78.91
$\Psi_{\text{sat,min}}$	0.05	0.09	0.27	0.35	0.50	0.63	1.01	1.23
$b_{\text{min}}$	9.79	10.98	12.63	13.94	14.40	15.97	17.85	19.88
$f_{\text{om}}$	0.12	0.64	0.09	0.62	0.14	0.75	0.29	1.14
All	98.49	101.76	98.33	101.94	98.03	102.34	97.48	103.02

**Table E2.** Same as Table E1 but at the P301 site (Loam).

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{\text{sat,min}}$	0.01	0.03	0.02	0.02	0.89	1.24	1.39	2.01
$\theta_{\text{sat,min}}$	89.80	92.34	90.83	91.57	81.23	82.08	54.10	55.75
$\Psi_{\text{sat,min}}$	0.05	0.37	2.05	2.69	0.24	0.56	0.02	0.65
$b_{\text{min}}$	6.93	9.11	5.58	6.36	15.80	16.53	41.25	42.95
$f_{\text{om}}$	0.19	1.59	0.34	0.89	0.53	1.25	0.56	1.93
All	96.98	103.44	98.81	101.53	98.69	101.66	97.32	103.29

**Table E3.** Same as Table E1 but at the SMAP-OK site (Loam).

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{\text{sat,min}}$	0.54	0.77	0.19	0.43	0.03	0.05	0.14	0.29
$\theta_{\text{sat,min}}$	78.85	81.02	76.52	79.17	67.81	70.87	62.77	65.03
$\Psi_{\text{sat,min}}$	2.67	3.23	4.65	5.05	9.35	10.20	11.57	13.27
$b_{\text{min}}$	15.07	17.15	15.30	17.77	18.19	21.60	21.88	23.59
$f_{\text{om}}$	0.18	1.04	0.31	1.10	0.54	2.26	0.52	1.45
All	97.31	103.21	96.97	103.52	95.91	104.98	96.88	103.63

**Table E4.** Same as Table E1 but at the Desert Chaparral UCI site (Sandy Loam).

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{\text{sat,min}}$	0.08	0.17	0.13	0.27	0.16	0.28	0.11	0.29
$\theta_{\text{sat,min}}$	34.30	36.72	35.90	38.02	38.13	40.34	37.71	40.06
$\Psi_{\text{sat,min}}$	4.02	4.69	3.17	3.82	2.83	3.45	3.10	3.67
$b_{\text{min}}$	57.63	60.68	56.79	59.50	55.30	58.13	55.58	58.55
$f_{\text{om}}$	0.30	2.17	0.51	2.56	0.21	1.73	0.10	1.27
All	96.33	104.43	96.50	104.17	96.63	103.93	96.61	103.84

**Table E5.** Same as Table E1 but at the JERC site (Sandy Loam).

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{\text{sat,min}}$	1.70	3.40	1.14	5.04	0.30	1.98	0.04	0.76
$\Theta_{\text{sat,min}}$	84.45	87.35	79.05	84.89	67.81	71.88	61.80	64.86
$\Psi_{\text{sat,min}}$	0.11	0.19	0.51	0.75	3.77	4.14	5.87	6.32
$b_{\text{min}}$	10.15	12.02	12.53	15.05	23.33	26.08	28.45	31.18
$f_{\text{om}}$	0.10	1.10	0.24	1.59	0.23	1.01	0.19	0.98
All	96.52	104.06	93.48	107.32	95.45	105.09	96.35	104.10

**Table E6.** Same as Table E1 but at the Santa Rita Creosote site (Sandy Loam).

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{\text{sat,min}}$	2.40	3.03	2.57	3.31	1.67	2.29	1.56	2.06
$\Theta_{\text{sat,min}}$	64.24	66.12	60.96	62.99	52.47	54.95	55.28	58.02
$\Psi_{\text{sat,min}}$	0.01	0.27	0.04	0.87	0.29	0.75	0.04	0.17
$b_{\text{min}}$	30.37	32.01	32.82	34.45	41.73	44.19	38.64	41.62
$f_{\text{om}}$	0.43	1.73	0.74	1.79	0.57	1.65	0.61	2.61
All	97.45	103.16	97.14	103.41	96.73	103.83	96.12	104.48

**Table E7.** Same as Table E1 but at the Mozark site (Clay Loam).

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{\text{sat,min}}$	2.60	3.06	2.16	2.60	1.37	1.82	0.71	1.00
$\Theta_{\text{sat,min}}$	70.10	71.62	82.69	64.19	47.69	50.06	36.83	39.02
$\Psi_{\text{sat,min}}$	0.35	0.42	0.33	0.42	1.42	1.66	2.84	3.10
$b_{\text{min}}$	24.81	26.50	32.72	34.45	46.34	48.96	56.61	59.04
$f_{\text{om}}$	0.12	0.84	0.08	0.73	0.14	1.11	0.25	1.02
All	97.98	102.44	97.98	102.39	96.96	103.61	97.25	103.18

**Table E8.** Same as Table E1 but at the Silver Sword site (Clay Loam).

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{\text{sat,min}}$	0.85	1.31	0.36	0.58	0.83	1.15	1.95	2.40
$\Theta_{\text{sat,min}}$	80.54	81.53	76.90	78.27	80.53	81.72	84.49	85.69
$\Psi_{\text{sat,min}}$	3.80	4.14	6.26	6.76	3.29	3.52	1.00	1.11
$b_{\text{min}}$	13.28	14.24	14.65	15.87	13.74	14.94	10.80	12.13
$f_{\text{om}}$	0.08	0.49	0.10	0.62	0.09	0.44	0.10	0.60
All	98.55	101.71	98.26	102.10	98.48	101.77	98.34	101.93

**Table E9.** Same as Table E1 but for at the Harvard Forest site (Silt Loam).

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{\text{sat,min}}$	2.15	2.43	2.76	3.10	1.08	1.59	0.48	0.83
$\theta_{\text{sat,min}}$	84.53	85.73	85.04	85.92	66.57	68.18	65.01	67.26
$\Psi_{\text{sat,min}}$	0.10	0.35	0.11	0.21	1.46	1.73	1.81	2.69
$b_{\text{min}}$	11.34	12.42	10.60	11.51	28.57	30.29	29.70	31.68
$f_{\text{om}}$	0.38	0.79	0.24	0.72	0.15	0.68	0.08	0.93
All	98.50	101.72	98.75	101.46	97.83	102.47	97.08	103.39

**Table E10.** Same as Table E1 but for at the Goodwin Creek site (Sandy Clay Loam).

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{\text{sat,min}}$	2.34	6.72	2.27	18.64	3.07	11.33	1.21	3.66
$\theta_{\text{sat,min}}$	65.53	71.61	47.53	69.55	27.89	39.47	23.45	30.27
$\Psi_{\text{sat,min}}$	0.48	3.06	2.22	6.81	8.69	11.24	9.26	10.39
$b_{\text{min}}$	23.00	27.67	21.91	32.64	46.75	52.97	57.86	63.91
$f_{\text{om}}$	0.43	1.53	0.34	3.01	0.06	1.08	0.14	1.02
All	91.79	110.59	74.26	130.65	86.46	116.09	91.93	109.25

**Table E11.** Same as Table E1 but at the Tonzi Ranch site (Sandy Clay Loam).

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{\text{sat,min}}$	2.26	3.08	0.53	1.31	0.53	1.42	0.04	0.16
$\theta_{\text{sat,min}}$	67.32	69.75	55.37	57.76	36.63	40.52	40.57	48.51
$\Psi_{\text{sat,min}}$	0.33	0.50	5.15	5.69	16.68	18.15	9.87	13.64
$b_{\text{min}}$	26.62	29.36	35.55	38.07	40.74	44.45	36.68	47.48
$f_{\text{om}}$	0.11	1.33	0.08	1.00	0.11	1.82	0.70	5.59
All	96.63	104.02	96.69	103.83	94.69	106.36	87.86	115.38

**Table E12.** Same as Table E1 but at the Freeman Ranch site (Clay).

Parameter	DJF		MAM		JJA		SON	
	Main	Total	Main	Total	Main	Total	Main	Total
$K_{\text{sat,min}}$	2.47	3.01	2.78	3.32	1.79	2.29	1.10	1.47
$\theta_{\text{sat,min}}$	68.41	69.71	72.83	73.98	59.15	60.38	49.73	51.82
$\Psi_{\text{sat,min}}$	0.05	0.10	0.03	0.05	0.70	0.79	1.27	1.43
$b_{\text{min}}$	27.09	28.71	22.57	24.01	36.53	38.02	45.20	47.48
$f_{\text{om}}$	0.07	0.74	0.06	0.72	0.05	0.63	0.06	0.85
All	98.08	102.27	98.26	102.08	98.21	102.11	97.36	103.05

# Joint Program Report Series - Recent Articles

For limited quantities, Joint Program Reports are available free of charge. Contact the Joint Program Office to order.

Complete list: <http://globalchange.mit.edu/publications>

- 341. Emulation of Community Land Model Version 5 (CLM5) to Quantify Sensitivity of Soil Moisture to Uncertain Parameters.** *Gao et al., Feb 2020*
- 340. Can a growing world be fed when the climate is changing?** *Dietz and Lanz, Feb 2020*
- 339. MIT Scenarios for Assessing Climate-Related Financial Risk.** *Landry et al., Dec 2019*
- 338. Deep Decarbonization of the U.S. Electricity Sector: Is There a Role for Nuclear Power?** *Tapia-Ahumada et al., Sep 2019*
- 337. Health Co-Benefits of Sub-National Renewable Energy Policy in the U.S.** *Dimanchev et al., Jun 2019*
- 336. Did the shale gas boom reduce US CO<sub>2</sub> emissions?** *Chen et al., Apr 2019*
- 335. Designing Successful Greenhouse Gas Emission Reduction Policies: A Primer for Policymakers – The Perfect or the Good?** *Phillips & Reilly, Feb 2019*
- 334. Implications of Updating the Input-output Database of a Computable General Equilibrium Model on Emissions Mitigation Policy Analyses.** *Hong et al., Feb 2019*
- 333. Statistical Emulators of Irrigated Crop Yields and Irrigation Water Requirements.** *Blanc, Aug 2018*
- 332. Turkish Energy Sector Development and the Paris Agreement Goals: A CGE Model Assessment.** *Kat et al., Jul 2018*
- 331. The economic and emissions benefits of engineered wood products in a low-carbon future.** *Winchester & Reilly, Jun 2018*
- 330. Meeting the Goals of the Paris Agreement: Temperature Implications of the Shell Sky Scenario.** *Paltsev et al., Mar 2018*
- 329. Next Steps in Tax Reform.** *Jacoby et al., Mar 2018*
- 328. The Economic, Energy, and Emissions Impacts of Climate Policy in South Korea.** *Winchester & Reilly, Mar 2018*
- 327. Evaluating India's climate targets: the implications of economy-wide and sector specific policies.** *Singh et al., Mar 2018*
- 326. MIT Climate Resilience Planning: Flood Vulnerability Study.** *Strzepek et al., Mar 2018*
- 325. Description and Evaluation of the MIT Earth System Model (MESM).** *Sokolov et al., Feb 2018*
- 324. Finding Itself in the Post-Paris World: Russia in the New Global Energy Landscape.** *Makarov et al., Dec 2017*
- 323. The Economic Projection and Policy Analysis Model for Taiwan: A Global Computable General Equilibrium Analysis.** *Chai et al., Nov 2017*
- 322. Mid-Western U.S. Heavy Summer-Precipitation in Regional and Global Climate Models: The Impact on Model Skill and Consensus Through an Analogue Lens.** *Gao & Schlosser, Oct 2017*
- 321. New data for representing irrigated agriculture in economy-wide models.** *Ledvina et al., Oct 2017*
- 320. Probabilistic projections of the future climate for the world and the continental USA.** *Sokolov et al., Sep 2017*
- 319. Estimating the potential of U.S. urban infrastructure albedo enhancement as climate mitigation in the face of climate variability.** *Xu et al., Sep 2017*
- 318. A Win-Win Solution to Abate Aviation CO<sub>2</sub> emissions.** *Winchester, Aug 2017*
- 317. Application of the Analogue Method to Modeling Heat Waves: A Case Study With Power Transformers.** *Gao et al., Aug 2017*
- 316. The Revenue Implications of a Carbon Tax.** *Yuan et al., Jul 2017*
- 315. The Future Water Risks Under Global Change in Southern and Eastern Asia: Implications of Mitigation.** *Gao et al., Jul 2017*
- 314. Modeling the Income Dependence of Household Energy Consumption and its Implications for Climate Policy in China.** *Caron et al., Jul 2017*
- 313. Global economic growth and agricultural land conversion under uncertain productivity improvements in agriculture.** *Lanz et al., Jun 2017*
- 312. Can Tariffs be Used to Enforce Paris Climate Commitments?** *Winchester, Jun 2017*
- 311. A Review of and Perspectives on Global Change Modeling for Northern Eurasia.** *Monier et al., May 2017*
- 310. The Future of Coal in China.** *Zhang et al., Apr 2017*
- 309. Climate Stabilization at 2°C and Net Zero Carbon Emissions.** *Sokolov et al., Mar 2017*
- 308. Transparency in the Paris Agreement.** *Jacoby et al., Feb 2017*
- 307. Economic Projection with Non-homothetic Preferences: The Performance and Application of a CDE Demand System.** *Chen, Dec 2016*
- 306. A Drought Indicator based on Ecosystem Responses to Water Availability: The Normalized Ecosystem Drought Index.** *Chang et al., Nov 2016*