DISSERTATION

ENHANCING ROOTZONE SOIL MOISTURE ESTIMATION USING REMOTE SENSING, REGIONAL CHARACTERISTICS, AND MACHINE LEARNING

Submitted by

Ahmad Shukran Sahaar

Department of Civil and Environmental Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2023

Doctoral Committee:

Advisor: Jeffrey D. Niemann

Jose Luis Chavez
Timothy R. Green
Gregory Butters

ABSTRACT

ENHANCING ROOTZONE SOIL MOISTURE ESTIMATION USING REMOTE SENSING, REGIONAL CHARACTERISTICS, AND MACHINE LEARNING

Accurate estimation of root-zone soil moisture ($\bar{\theta}$) is essential for various agricultural applications, including crop yield estimation, precision irrigation, and groundwater management. This dissertation encompasses three interconnected studies that collectively investigate different approaches for improving soil moisture estimation. The first study delves into the utilization of remote sensing methods, particularly optical and thermal satellite imagery, to estimate fine-resolution (30 m) root-zone soil moisture across diverse regions. Traditionally, these methods relied on empirical relationships with evaporative fraction Λ_{SEB} or evaporative index Λ_{PET}. However, it has been shown that a single relationship does not universally apply to all regions. This study evaluates the influence of regional soil, vegetation, and climatic conditions on the shape and strength of these relationships using global sensitivity analysis. The results highlight that soil characteristics, such as clay and silt content, and vegetation properties, like leaf area index and rooting depth, play pivotal roles in determining these relationships. Moreover, the impact of annual precipitation in defining climatic regions is crucial. Consequently, region-specific relationships are proposed, adapting to local conditions and potentially enhancing soil moisture estimates. The second study extends this investigation by applying the regionally adapted relationships for the Λ_{SEB} vs. $\bar{\theta}$ and Λ_{PET} vs. $\bar{\theta}$ to estimate rootzone soil moisture ($\bar{\theta}$) from remote sensing data across four study regions. The results consistently demonstrate the

superior performance of the regionally adapted relationships over a single empirical relationship, with a substantial reduction in root mean squared error. These adapted relationships are particularly effective in arid and semiarid regions. The third study explores the application of machine learning models, including XGBoost, CatBoost, RF, LightGBM, and artificial neural networks, to predict soil moisture levels across various climates and depths in the contiguous United States. The findings emphasize the high accuracy and effectiveness of machine learning models, especially XGBoost, in predicting soil moisture across diverse climate regions. XGBoost outperforms other models, making it a potentially valuable tool for soil moisture prediction in environmental monitoring and management. The study also highlights the influence of climate and soil depth on prediction accuracy, with deeper layers having improved forecasts. Additionally, feature importance analysis identifies key predictors for predicting soil moisture, such as elevation, aridity index, soil composition, and depth. These findings contribute to the advancement of soil moisture monitoring and management, with practical applications in agriculture and environmental sciences.

ACKNOWLEDGEMENTS

This project was funded by the USDA National Institute for Food and Agriculture Hatch grants 1000065-COL00693A and 1000065-COL00797. It was also funded in part by the National Science Foundation (2312319).

DEDICATION

To Maryam, Hamza, and Omar

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

1.1 Background

Soil moisture (volumetric water content) is a key hydrological variable that controls various land surface, soil, and atmospheric processes. It is important for the partitioning of incoming energy into latent and sensible heat fluxes, the dynamics of the atmospheric boundary layer, and weather and climate (Entekhabi et al., 1996; Patel et al., 2009). It also controls hydrological processes such as infiltration, runoff, and soil evaporation, and provides the source of water for root-water uptake and transpiration (Bronstert and Plate, 1997; Sadeghi et al., 2017). Soil moisture is also important for crop health and productivity, and it helps regulate the terrestrial carbon cycle (Gardner et al., 1981; Melillo et al., 2002; Ntukamazina et al., 2017). Accurate measurement and understanding of the spatiotemporal dynamics of soil moisture are vital for agricultural water management, water resources sustainability, runoff generation, numerical weather forecasting, climate modeling, and monitoring of extreme events such as floods and droughts (Hu et al., 2020; Petropoulos et al., 2018).

Over the past three decades, substantial progress has been made in measuring soil moisture at various scales. Ground-based methods such as time domain reflectometry (TDR) (Černý, 2009), frequency domain reflectometers (Xu et al., 2012), and capacitance probes (Mohamed et al., 1997), neutron probes (Evett et al., 2003) have been used successfully. However, these methods

only provide point-scale measurements, which limits their ability to capture soil moisture heterogeneity over large areas. This heterogeneity is caused by complex interactions between meteorological, soil, land cover, and topographic factors (Peng et al., 2017). Additionally, establishing an in-situ soil moisture network is costly, labor-intensive, and often fails to provide adequate spatial coverage (Wang et al., 2021).

Recent advances in satellite remote sensing techniques have enabled the monitoring of soil moisture at regional to global scales. (SMAP) (Entekhabi et al., 2010a), Advanced Scatterometer (ASCAT) (Bartalis et al., 2007), Soil Moisture and Ocean Salinity (SMOS) (Kerr et al., 2001), and Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) (Njoku et al., 2003) have been dedicated to soil moisture measurements in the microwave frequency range. Microwave sensors, both passive and active, have been used to retrieve surface soil moisture by leveraging the relationship between soil dielectric constant and water content (Cho et al., 2015; Wang and Qu, 2009). Passive sensors, such as radiometers, measure natural emissions from the land surface, whereas active sensors, such as synthetic aperture radars (SAR), provide their own energy source for illumination (Chan and Koo, 2008). Passive sensors are more sensitive to surface soil moisture (~5 cm) but have coarse spatial resolutions (~9-60 km), making them suitable for continental and global scale applications (Peng et al., 2017). The main limitations of these methods include the coarse spatial resolution (~9-60 km grid cells), the shallow depth of soil moisture measurement (approximately the top 5 cm of the soil column), and lower accuracy for densely vegetated and rougher surfaces (Peng et al., 2017).

Downscaling approaches are alternatives for providing soil moisture data with fine spatial resolution. Significant efforts have been made to downscale microwave soil moisture products (Peng et al., 2017; Sabaghy et al., 2018). For example, Piles et al., (2009) used a change detection method for downscaling SMAP soil moisture data from 36 km to 10 km resolution using 3-km radar backscatter coefficients and 36-km radiometer brightness temperature. Wu et al. (2017) evaluated three methods to downscale SMAP soil moisture data from 36 km to 9 km resolution using 1-km synthetic aperture radar backscatter data, and the optional method that was based on a background soil moisture value is recommended for global implantation by SMAP due to its higher correlation between the downscaled soil moisture and the reference soil moisture. Tagesson et al. (2018) utilized land surface temperature and vegetation dryness index from a geostationary satellite to disaggregate SMOS soil moisture from ~40 km resolution soil moisture data to ~5 km resolution soil moisture in West Africa. Similarly, Das et al. (2019) used the Sentinel-1A and Sentinel-1B SAR data to disaggregate SMAP L-band radiometer measurements from ~40 km radiometer measurement to 3 km and 1 km gridded products. Nuñez-Olivieri et al. (2017) used MODIS products, sand fraction, and elevation to downscale AMSR2 soil moisture estimates from 25 km to 1 km in Puerto Rico. Vergopolan et al. (2020) utilized a hyper-resolution land surface model, a radiative transfer model, and a Bayesian scheme to merge and downscale SMAP 36 km coarse resolution soil moisture data to a 30 m spatial resolution and validated it across four watersheds in the United States.

Soil moisture can alternatively be estimated from optical and thermal remote sensing data. This approach uses the visible and near infrared bands to characterize soil moisture's relationship with water stressed vegetation and uses the thermal infrared band to derive soil moisture's

relationship with the thermal properties of soil (Zhang and Zhou, 2016). This approach provides much finer spatial resolutions (30 m to 1 km grid cells) and estimates the average soil moisture throughout the rootzone ($\overline{\theta}$). Several different methods have been employed for optical and thermal remote sensing, including the triangle and trapezoid method (Carlson and Petropoulos, 2019; Krishnan and Indu, 2023; Rawat et al., 2022; Wang et al., 2022; Yuting Yang et al., 2015), drought index methods (Amani et al., 2017; Feng et al., 2013; Nie et al., 2020; Sims et al., 2002), thermal inertia methods (Hao et al., 2022; Matsushima et al., 2012; Soliman et al., 2013; Van doninck et al., 2011), optical methods (Babaeian et al., 2019; Gao et al., 2017; Ngo et al., 2019; Qiu et al., 2019; Sadeghi et al., 2017), and energy balance methods (Akuraju et al., 2021; Bastiaanssen et al., 2000; Hain et al., 2009; Scott et al., 2003). However, the approach is limited to clear sky conditions because optical and thermal sensors do not penetrate clouds (Zhao and Li, 2013).

Energy balance methods use optical and thermal remote sensing data to calculate all the components of the surface energy balance (SEB). Once the energy balance is calculated, the evaporative fraction Λ_{SEB} can be estimated. Λ_{SEB} is the ratio of the latent heat flux to the available energy (net radiation minus ground heat flux). The Bastiaanssen et al. (2000) method then uses an empirical logarithmic function to convert Λ_{SEB} to average rootzone soil moisture $\overline{\theta}$, while the Scott et al. (2003) method uses a similar function to obtain an average root-zone degree of saturation $\overline{s}$. Other studies have suggested relationships between $\overline{\theta}$ or $\overline{s}$ and the evaporative index Λ_{PET}, which is the ratio of the actual evapotranspiration to the potential evapotranspiration (PET) (Baier, 1969; Crago, 1996; Fleming et al., 2005; Thornthwaite and

Mather, 1955). The empirical logarithmic function proposed in these studies have been tested in several regions (Ahmad and Bastiaanssen, 2003; Hendrickx et al., 2016; Rahimzadeh et al., 2013). However, field research has shown that a single empirical relationship proposed by these studies is not applicable for all regions (Alburn et al., 2015). Calibration to in situ soil moisture observations is often needed to obtain accurate soil moisture estimates.

Several studies have shown that the Λ_{SEB} and Λ_{PET} relationships with $\overline{\theta}$ or $\overline{s}$ depend on regional soil, vegetation, and climatic characteristics. For example, the $\Lambda_{SEB} - \overline{\theta}$ relationship is strongly influenced by vegetation cover, with thicker vegetation cover producing higher Λ_{SEB} values for a given $\overline{\theta}$ (Gentine et al., 2007). Other studies have shown that the $\Lambda_{SEB} - \overline{\theta}$ relationship has strong dependency on soil hydraulic properties, net incoming radiation, meteorological conditions, and the total depth of the rootzone layer (Basara and Crawford, 2002; Dirmeyer et al., 2000; Ford et al., 2014). The $\Lambda_{PET} - \overline{s}$ relationship can have different forms depending on the soil texture (Wilson et al., 1997). The $\Lambda_{PET} - \overline{\theta}$ relationship is also influenced by the magnitude of the PET, with low atmospheric demands producing higher Λ_{PET} values and high atmospheric demands producing lower Λ_{PET} values for the same $\overline{\theta}$ value (Denmead and Shaw, 1962).

Multi-sensor data fusion is an alternative soil moisture estimation method that combines the strengths of coarse-resolution microwave remote sensing and high-resolution optical and thermal remote sensing products. In recent years, machine learning methods have gained popularity in

the areas of data fusion and downscaling/disaggregation of soil moisture (Abbaszadeh et al., 2019; Abowarda et al., 2021; Liu et al., 2020; Peng et al., 2021). Machine learning methods are data-driven approaches that learn patterns and relationships directly from data, making them well-suited for complex and nonlinear relationships between soil moisture and land surface processes without making assumptions about the processes governing soil water dynamics (Ali et al., 2015; Fang et al., 2019). Unlike process-based models, which often require data from particular sources, machine learning can effectively merge and process large volumes of data from various sources, including in-situ measurements, meteorological variables, and remote sensing datasets (Ali et al., 2015; Fang et al., 2019). Additionally, machine learning models have shown a strong correlation between in-situ measured soil moisture and predicted soil moisture, indicating high predictive accuracy. Several recent studies have reported successful applications of machine learning methods in this area (Ali et al., 2015; Rani et al., 2022).

1.2 Objectives

The overall goal of this research is to develop and evaluate methods for estimating rootzone soil moisture $\overline{\theta}$ using optical and thermal remote sensing data, considering regional soil, vegetation, topography, weather, and climatic conditions.

To achieve this goal, the one-dimensional variably saturated flow model HYDRUS -1D, Version 4.17 (Šimůnek et al., 2013) is used to simulate $\overline{\theta}$ and $\overline{s}$ through time for various soil, vegetation, and climatic conditions. Then the most important regional characteristics that affect the relationship constants are identified and a new set of empirical equations are proposed to

estimate these constants. The aim is to adapt optical and thermal methods to local conditions and potentially improve soil moisture estimates. In addition, Remote Sensing of ET (ReSET) (Elhaddad and Garcia, 2008) is used to estimate the components of the energy balance using optical/thermal satellite imagery. Then the rootzone soil moisture is estimated using the relationships developed in the first study. The rootzone soil moisture estimates are evaluated against in-situ soil moisture at four application regions with different climate conditions. Furthermore, five machine learning models including artificial neural network (ANN), random forest (RF), extreme gradient boosting (XGBoost), categorical boosting (CatBoost), and light gradient boosting machines (LightGBM) are used to fuse information from various satellite sources and datasets to generate soil moisture estimates at different depth ranges. The soil moisture predictions from each model are evaluated against in situ soil moisture stations across the Continental United States.

This dissertation consists of three individual studies that accomplish the three main tasks of this research. These studies are briefly described below:

(1) Assess the impact of regional soil, vegetation, and climatic conditions on the form and strength of the Λ_{SEB}vs. $\overline{\theta}$, Λ_{PET}vs. $\overline{\theta}$, Λ_{SEB}vs. $\overline{s}$, and Λ_{PET}vs. $\overline{s}$ relationships.

A global sensitivity analysis is performed by first generating combinations of soil, vegetation, and climate characteristics. These combinations are then supplied to the one-dimensional HYDRUS-1D model. A new set of Λ_{SEB}vs. $\overline{\theta}$, Λ_{PET}vs. $\overline{\theta}$, Λ_{SEB}vs. $\overline{s}$, and Λ_{PET}vs. $\overline{s}$ relationships are then fitted to the datasets, and a sensitivity analysis is performed to identify the soil, vegetation, and climate characteristics that have the

greatest control on the relationship constants. Finally, empirical equations are proposed to estimate the relationship constants based on the most important regional characteristics. Estimating the relationship parameters in this way allow the optical and thermal methods to be adapted to local conditions and has the potential to improve the $\overline{\theta}$ and $\overline{s}$ estimates.

(2) Evaluate the performance of proposed methods for estimating rootzone soil moisture from optical/thermal remote sensing using regional characteristics. ReSET is used to estimate the components of the surface energy balance as well as Λ_{SEB} and Λ_{PET}. Then $\overline{\theta}$ is estimated from the Λ_{SEB} vs. $\overline{\theta}$ and Λ_{PET} vs. $\overline{\theta}$ relationships proposed in the previous study. The evaluation of the proposed methods is conducted in four application regions that span arid to humid climates and have extensive in situ soil moisture observations. The methods are evaluated by comparing their estimated $\overline{\theta}$ to in situ observations and to $\overline{\theta}$ estimates from the empirical relationship of Bastiaanssen et al. (2000) and Scott et al. (2003).

(3) Apply machine learning to estimate soil moisture at multiple depths across the contiguous United States using multi-sensor satellite data fusion with regional characteristics. A soil moisture retrieval framework is developed based on five machine learning algorithms. Multi-sensor satellite information is fused, including coarse-resolution microwave satellite products, fine-resolution optical and thermal ECOSTRESS products, MODIS vegetation products, aridity index, topography, soil, and National Land Cover Database (NLCD) datasets. Multi-depth (0-5 cm, 0-10 cm, 0-20 cm, 0-50 cm, and 0-100 cm) soil moisture at 70 m spatial resolution is generated using the machine learning methods. The performance of the retrieval models is evaluated against 730 in situ soil moisture stations obtained from the International Soil Moisture Network (ISMN) across the continental

United States. The performance of the best-performing model is tested across different climates, depths, and networks to assess its prediction ability. The relative importance of predictors across machine learning models is analyzed.

CHAPTER 2. IMPACT OF REGIONAL CHARACTERISTICS ON THE ESTIMATION OF ROOT-ZONE SOIL MOISTURE FROM THE EVAPORATIVE INDEX OR EVAPORATIVE FRACTION

2.1 Introduction

Soil moisture is an important factor in the transfer of water and energy at the land-atmosphere boundary and governs the partitioning of incoming energy into latent and sensible heat fluxes through soil evaporation and plant transpiration (Entekhabi et al., 1996; Sun et al., 2011). Soil moisture also has a strong influence on the dynamics of the atmospheric boundary layer and has a direct influence on weather and climate (Patel et al., 2009). Thus, accurate estimation of soil moisture is useful for planning and management of water resources as well as sustainable agricultural production, particularly in semiarid and arid ecosystems where water is a limited resource.

In recent decades, spaceborne remote-sensing methods have been developed to monitor soil moisture nearly globally at regular time intervals. These methods include passive and active microwave instruments such as the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), Soil Moisture and Ocean Salinity (SMOS), and Soil Moisture Active and Passive (SMAP) missions (Entekhabi et al., 2010a; Kerr et al., 2012; Njoku et al., 2003). These satellites estimate soil moisture at very coarse spatial resolutions (9-60 km grid

cells) and are only sensitive to moisture in the upper 5-10 cm of the soil layer (Entekhabi et al., 2004; Kerr et al., 2012).

Soil moisture can also be estimated indirectly using optical (Liu et al., 2003; Lobell and Asner, 2002; Weidong et al., 2002) and thermal infrared (TIR) remote sensing (Lu et al., 2009; Verstraeten et al., 2006). These methods provide much finer spatial resolutions (30 m to 1 km) and estimate the average soil moisture throughout the root-zone. However, they often have poor temporal resolutions and, like the active/passive microwave methods, have lower performance as the vegetation cover increases (Muller and Décamps, 2001). Soil moisture can also be estimated from the combination of optical and TIR data using the triangle or trapezoidal method (Gillies et al., 1997; Stisen et al., 2008), the perpendicular drought index (PDI) method (Ghulam et al., 2007), or energy-balance methods (Bastiaanssen et al., 2000; Scott et al., 2003).

The methods proposed by Bastiaanssen et al. (2000) and Scott et al. (2003) use multispectral imagery to calculate all the components of the surface energy balance (SEB). Several models are available for this purpose including the Two-Source Model (TSM) (Norman et al., 1995), Surface Energy Balance System (SEBS) (Su, 2002), Two-Source Energy Balance (TSEB) (French et al., 2002), Mapping evapotranspiration (ET) at High Resolution with Internalized Calibration (METRIC) (Allen et al., 2007a), and Remote Sensing of ET (ReSET) (Elhaddad and Garcia, 2008). Once the energy balance is calculated, the Bastiaanssen et al. (2000) and Scott et al. (2003) methods estimate the evaporative fraction Λ_{SEB}, which is the ratio of the latent heat flux to the available energy (net radiation minus ground heat flux) (Shuttleworth et al., 1989). The Bastiaanssen et al. (2000) method then uses an empirical logarithmic function to convert

Λ_{SEB} to average root-zone volumetric water content $\overline{\theta}$, while the Scott et al. (2003) method uses a similar function to obtain an average root-zone degree of saturation $\overline{s}$.

The reliability of these approaches has been evaluated in several studies. Smith et al. (1992) plotted Λ_{SEB} values from surface stations against the soil moisture by mass in the top 2.5 cm of the soil for grazed and ungrazed grasslands on alluvial soils and loess in Kansas. Most of these gravimetric soil moisture measurements were above 25% by mass, and little correlation was observed between Λ_{SEB} and soil moisture. However, when Bastiaanssen et al. (2000) fitted their logarithmic function using remotely-sensed Λ_{SEB} values and field measurements of $\overline{\theta}$ at 23 locations, they obtained a coefficient of determination r^2 of 0.855. Ahmad and Bastiaanssen (2003) compared SEBAL-derived estimates of $\overline{s}$ to observations from theta probes (top 100 cm of the soil) for wheat, cotton, and rice fields in Pakistan. They found a root mean square error (RMSE) of 0.05 m^3m^{-3}, but they only had 26 observations in this comparison. Scott et al. (2003) also compared SEBAL-derived estimates of $\overline{s}$ to observations from irrigated fields of rice, cotton, wheat, alfalfa, and sorghum in Pakistan and Mexico (depths ranging from 45 cm to 100 cm). They also found an RMSE of 0.05 m^3m^{-3}, but this dataset included only 52 observations.

Extensive evaluations of the $\Lambda_{SEB} - \overline{\theta}$ and $\Lambda_{SEB} - \overline{s}$ relationships were performed by Alburn et al. (2015). ReSET was used to estimate Λ_{SEB} from Landsat 5 and Landsat 7 images across a semiarid region in southeastern Colorado for the summer months of three years. The Bastiaanssen et al. (2000) and Scott et al. (2003) relationships were then applied to estimate $\overline{\theta}$ and $\overline{s}$. These estimates were compared to 280 $\overline{\theta}$ and $\overline{s}$ observations from time-domain

reflectometry (TDR), but the estimates failed to explain any of the space-time variations in the TDR measurements. The Nash-Sutcliffe Coefficient of Efficiency (NSCE) (Nash and Sutcliffe, 1970) of the $\overline{\theta}$ and $\overline{s}$ estimates was -1.12 and -1.02, respectively (+1.00 would indicate a perfect estimate, while a negative NSCE value indicates that the mean of the observed soil moisture is a more accurate estimate than the estimated soil moisture). The remote-sensing estimates were much more successful when the relationships were calibrated using the TDR observations, (NSCE was 0.32 for $\overline{\theta}$ and 0.36 for $\overline{s}$), but calibration is only possible when in-situ soil moisture observations are available for the region of interest. Overall, these results suggest that the empirical relationships proposed by Bastiaanssen et al. (2000) and Scott et al. (2003) are not applicable for all regions, and region-specific relationships are required to achieve adequate performance in some cases.

Although previous studies have used Λ_{SEB} to estimate $\overline{\theta}$ or $\overline{s}$, numerous other studies have suggested similar relationships between $\overline{\theta}$ or $\overline{s}$ and the evaporative index Λ_{PET}, which is the ratio of actual to potential ET. $\Lambda_{PET} - \overline{\theta}$ and $\Lambda_{PET} - \overline{s}$ relationships have been used to evaluate water budgets (Thornthwaite and Mather, 1955), estimate the availability of soil moisture to plants under different meteorological conditions (Baier, 1969; Denmead and Shaw, 1962), assess agricultural potential for different land classifications (Yao, 1974), improve estimates of actual ET from soil moisture (Crago, 1996; Crago and Brutsaert, 1994), estimate root-zone soil moisture in the Middle Rio Grande Basin in New Mexico (Fleming et al., 2005), downscale soil moisture (Ranney et al., 2015), and insert real-time ET fluxes into hydrologic models (Hendrickx et al., 2016).

The shapes of the $\Lambda_{SEB} - \overline{\theta}$, $\Lambda_{SEB} - \overline{s}$, $\Lambda_{PET} - \overline{\theta}$, and $\Lambda_{PET} - \overline{s}$ relationships are expected to depend on regional soil, vegetation, and climatic characteristics (Davies and Allen, 1973; Jobson, 1982; Rouse et al., 1977). For example, Wilson et al. (1997) analyzed the $\Lambda_{PET} - \overline{s}$ relationship for thin soil specimens in the laboratory and observed different relationship forms depending on the specimen's texture (sand, silt, and clay). (Gentine et al., 2007) used a calibrated model to show that the shape of the $\Lambda_{SEB} - \overline{\theta}$ relationship depends on the vegetation cover with thicker vegetation covers producing higher Λ_{SEB} values for a given value $\overline{\theta}$. Denmead and Shaw (1962) used field experiments to examine the $\Lambda_{PET} - \overline{\theta}$ relationship for corn when the ET is dominated by transpiration. They found that low atmospheric demands (associated with low potential ET values) can be supplied at low $\overline{\theta}$ values, which produces high Λ_{PET} values. However, high atmospheric demands cannot be fully supplied, which produces lower Λ_{PET} values for the same $\overline{\theta}$. These results suggest that the shape of the $\Lambda_{PET} - \overline{\theta}$ relationship also depends on the magnitude of the potential ET.

The main objectives of the present study are to (1) quantify the impacts of soil, vegetation, and climatic characteristics on the strength and form of the $\Lambda_{SEB} - \overline{\theta}$, $\Lambda_{SEB} - \overline{s}$, $\Lambda_{PET} - \overline{\theta}$, and $\Lambda_{PET} - \overline{s}$ relationships and (2) propose a set of empirical relationships that account for the relevant regional characteristics. To achieve these goals, a global sensitivity analysis is performed using Extended Fourier Amplitude Sensitivity Test (eFAST) (Saltelli et al., 1999). Combinations of soil, vegetation, and climate characteristics are generated with eFAST and supplied to the physically-based vadose-zone model HYDRUS-1D (Šimůnek et al., 2013). For each set of characteristics, HYDRUS-1D is used to simulate the soil moisture dynamics over a

period of nine months, which produces paired datasets for $\left(\Lambda_{SEB},\overline{\theta}\right)$, $\left(\Lambda_{SEB},\overline{s}\right)$, $\left(\Lambda_{PET},\overline{\theta}\right)$ and $\left(\Lambda_{PET},\overline{s}\right)$. The $\Lambda_{SEB}-\overline{\theta}$, $\Lambda_{SEB}-\overline{s}$, $\Lambda_{PET}-\overline{\theta}$, and $\Lambda_{PET}-\overline{s}$ relationships are then fitted to the datasets, and eFAST is used to identify the soil, vegetation, and climate characteristics that have the greatest control on the relationship constants and r^2 values. Finally, empirical equations are determined that estimate the relationship constants based on the most important regional characteristics.

2.2 Methodology

2.2.1 HYDRUS-1D Model

The one-dimensional variably saturated flow model HYDRUS-1D, Version 4.17 (Šimůnek et al., 2013) is used to simulate $\overline{\theta}$ and $\overline{s}$ through time for various soil, vegetation, and climatic conditions. HYDRUS-1D has been used in many studies including studies on solute transport (Boudreau et al., 2009), groundwater recharge (Jiménez-Martínez et al., 2009), soil CO_2 fluxes (Buchner et al., 2008), surface energy balance (Saito and Šimůnek, 2009), and soil moisture (Chen et al., 2014). HYDRUS-1D simulates one-dimensional water movement in a partially saturated rigid medium using Richard's equation:

$$\frac{\partial\theta(\psi)}{\partial t}=\frac{\partial}{\partial z}\left[K(\psi)\left(\frac{\partial\psi}{\partial z}+1\right)\right]-S_v(\psi) \tag{1}$$

where $\theta(\psi)$ is the volumetric water content at a point in the soil profile (cm^3cm^{-3}), t is time (h), z is the vertical coordinate in the soil (positive upward) (cm), $K(\psi)$ is the hydraulic conductivity (cm h^{-1}), ψ is the pressure head (cm), and $S_v(\psi)$ is a sink term that accounts for root-water uptake (cm^3 cm^{-3} h^{-1}) (Šimůnek et al., 2013).

The van Genuchten (1980) functions are used for $\theta(\psi)$ and $K(\psi)$. Specifically:

$$\theta(\psi) = \begin{cases} \theta_r + \dfrac{\theta_s - \theta_r}{\left[1 + |\kappa\psi|^n\right]^m} & \psi < 0 \\ \theta_s & \psi \geq 0 \end{cases} \tag{2}$$

where θ_r is the residual water content ($cm^3\ cm^{-3}$), θ_s is the saturated water content ($cm^3\ cm^{-3}$), κ is the inverse of the air entry suction, n describes the pore size distribution, and $m = 1 - 1/n$. Furthermore:

$$K(\psi) = K_s s_e^l \left[1 - \left(1 - s_e^{1/m}\right)^m\right]^2 \tag{3}$$

where K_s is the saturated hydraulic conductivity and l is the pore tortuosity/connectivity index. The effective saturation s_e is:

$$s_e = \frac{\theta(\psi) - \theta_r}{\theta_s - \theta_r} \tag{4}$$

To implement the van Genuchten (1980) equations, several soil hydraulic characteristics are required (θ_r, θ_s, κ, n, K_s, and l). In the present study, the ROSETTA model, Version 1 (Schaap et al., 2001) is used to obtain the required soil parameters. ROSETTA was selected because it can estimate soil hydraulic properties using different pedotransfer functions depending on the types of available input data. Here, the soil hydraulic properties are estimated using the percent sand Sa, silt Si, and clay Cl, which will be varied in the sensitivity analysis to simulate the soil properties of different regions. To reduce the number of required simulations, the soil profile is always 100 cm deep, vertical, and homogeneous with an initial volumetric water content of 0.15. For simplicity, soil hysteresis is also neglected, which is consistent with the

level of simplifications made elsewhere in the analysis (e.g., homogeneous soil texture with depth) (Sutanto et al., 2012).

The sink term in Richard's equation is calculated as:

$$S_v(\psi) = \Omega(\psi) S_p \tag{5}$$

where S_p is the potential water uptake rate (h^{-1}) and $\Omega(\psi)$ describes the uptake as a dimensionless function of the soil water pressure head ($0 \leq \Omega(\psi) \leq 1$). The term $\Omega(\psi)$ is defined as an S-shaped function following Van Genuchten (1987).

To calculate S_p, a provided potential ET (ET_p) is partitioned into potential soil evaporation E_S and potential plant transpiration E_T using a provided leaf area index LAI Ritchie (1972). In particular:

$$E_S = ET_p \exp(-k_{ext} LAI) \tag{6}$$

where k_{ext} is an extinction coefficient, which depends on the vegetation type. The parameters k_{ext} and LAI will be varied in the sensitivity analysis to simulate the vegetation of different regions. E_T is then found from:

$$E_T = ET_p - E_S \tag{7}$$

The potential root-water uptake is derived from potential transpiration rate E_T (cm h^{-1}):

$$S_p = \beta(z) E_T \tag{8}$$

where $\beta(z)$ is the normalized root-water uptake distribution (cm^{-1}), which in part characterizes the root density (Šimůnek et al., 2013). $\beta(z)$ is defined so that its integral up to the maximum

rooting depth R_M (cm) is equal to one. $\beta(z)$ is expected to decay exponentially with depth. R_M will also be varied in the sensitivity analysis to simulate different types of vegetation in different regions.

Boundary fluxes such as actual evaporation, infiltration, and runoff at the soil surface and drainage from the bottom of soil profile are all calculated using Richards' equation (Šimůnek et al., 2013). For simplicity, the bottom boundary is selected to be freely draining in all simulations, which implies a deep water table. The top boundary condition is determined by temporally-varying precipitation and potential ET forcing data. Actual evaporation is constrained to be less than E_S. For each region (i.e. a set of soil, vegetation, and climate characteristics), HYDRUS-1D is used to simulate March through November at an hourly time step. This period is used to avoid possible freezing and thawing cycles during the winter.

2.2.2 Precipitation Forcing

The precipitation forcing for HYDRUS-1D is modeled as a series of random storm events using the rectangular pulse model (Gopalakrishnan and Biegler, 2013; Hawk and Eagleson, 1992). The storms are generated using three independent random variables (I_R', t_R', and t_B'). I_R' is the constant rainfall rate during a given storm, t_R' is the duration of a given storm, and t_B' is the time between consecutive storms. The variables I_R', t_R', and t_B' are assumed to be exponentially distributed with probability density functions (PDFs) as follows:

$$f\left(I_R'\right) = \frac{1}{I_R}\exp\left(-\frac{I_R'}{I_R}\right) \tag{9}$$

$$f\left(t_R'\right)=\frac{1}{t_R}\exp\left(-\frac{t_R'}{t_R}\right) \tag{10}$$

$$f\left(t_B'\right)=\frac{1}{t_B}\exp\left(-\frac{t_B'}{t_B}\right) \tag{11}$$

where $f\left(I_R'\right)$, $f\left(t_R'\right)$, and $f\left(t_B'\right)$ are the probability densities for I_R', t_R', and t_B', respectively, and I_R, t_R, and t_B are the mean rainfall intensity (cm h^{-1}), mean rainfall duration (h), and mean inter-storm period (h), respectively. These three means will be varied in the sensitivity analysis to simulate different climatic conditions.

2.2.3 Potential ET Forcing

For this study, ET_p is estimated as the ASCE hourly standardized reference ET for short crops (ASCE-EWRI, 2005; Walter et al., 2001). Using this equation:

$$ET_p=\frac{0.408\,\Delta\left(R_n-G\right)+\gamma\frac{C_n}{T+273}u_2\left(e_s-e_a\right)}{\Delta+\gamma\left(1+C_d\,u_2\right)} \tag{12}$$

where ET_p has units of mm h^{-1}, R_n is net radiation (MJ m^{-2} h^{-1}), G is the ground heat flux at the soil surface (MJ m^{-2} h^{-1}), T is hourly air temperature (ºC), u_2 is hourly wind speed at 2 m height (m s^{-1}), e_s is saturation vapor pressure (kPa), e_a is actual vapor pressure (kPa), Δ is the slope of the relationship between saturation vapor pressure and temperature (kPa ºC^{-1}), γ is the psychrometric constant (kPa ºC^{-1}), C_n is a constant that changes with reference crop type and calculation time step (K mm s^3 Mg^{-1} h^{-1}), and C_d is a constant that changes with the time step (s m^{-1}) (Allen et al., 1998).

R_n is obtained using an approach provided by ASCE-EWRI (2005). Specifically:

$$R_n = R_{ns} - R_{nl} \tag{13}$$

where R_{ns} is net shortwave radiation (MJ m^{-2} h^{-1}) and R_{nl} is net outgoing longwave radiation (MJ m^{-2} h^{-1}). Furthermore:

$$R_{ns} = (1-\alpha)R_s \tag{14}$$

where R_s is the incident solar radiation and α is the surface albedo. The albedo α is estimated using (Zeng et al., 1999):

$$\alpha = 0.38 - 0.3(1 - \exp(-k_{ext} \cdot LAI)) \tag{15}$$

which produces an albedo of 0.38 when LAI is zero and 0.08 as LAI becomes very large. The incident solar radiation R_s can be estimated:

$$R_s = \frac{f_{cd} + 0.35}{1.35} R_{SO} \tag{16}$$

where f_{cd} is the cloudiness and R_{SO} is the clear-sky solar radiation. f_{cd} is allowed to vary randomly in each time step and is generated from a beta distribution with shape parameters p_1 and p_2 following Falls (1974). The selection of the shape parameters is discussed later. The clear sky radiation R_{SO} can be estimated:

$$R_{SO} = \left(0.75 + 2 \times 10^{-5}\ Z\right) R_a \tag{17}$$

where Z is the elevation above sea level (m) and R_a is the extraterrestrial radiation (MJ m^{-2} h^{-1}), which is a function of the day of the year, the time of day, latitude ϕ, and longitude λ (see Duffie and Beckman (2013)). Z, ϕ, and λ will be varied as part of the sensitivity analysis. R_{nl} in Equation (13) is estimated:

$$R_{nl} = \sigma f_{cd} \left(0.34 - 0.14\sqrt{e_a}\right) T^4 \quad (18)$$

where σ is the Stefan-Boltzmann constant (2.042 x 10^{-10} MJ K^{-4} m^{-2} h^{-1}) and T is the air temperature (K) (Brunt, 1952).

Both G and T are required to calculate ET_p with Equation (12). For simplicity, G is estimated from R_n using equations from Allen et al., (1998), and T (ºC) is modeled as the sum of two sine waves, which describe the seasonal and diurnal variations of air temperature. Specifically:

$$T = A_{ST} \sin(F_{ST} N + C_{ST}) + A_{DT} \sin(F_{DT} N + C_{DT}) + T_m \quad (19)$$

where A_{ST} (ºC), F_{ST} (day^{-1}), and C_{ST} (-) are the amplitude, frequency, and phase angle, respectively, of the seasonal temperature variation. A_{DT} (ºC), F_{DT} (day^{-1}), C_{ST} (-) are the amplitude, frequency, and phase angle, respectively, of the daily temperature variation. N is the day of the year (day), and T_m is the mean annual air temperature (ºC). The variables F_{ST}, F_{DT} C_{ST}, and C_{DT} are specified to provide realistic timing and wavelengths for the seasonal and daily temperature variations, and the same values are used for all simulations for simplicity. The parameters A_{ST}, A_{DT}, and T_m will be varied in the sensitivity analysis to describe different climates. The air temperature T is also used to estimate Δ and e_s using expressions from Murray (1967) and Jensen et al. (1990). The actual vapor pressure e_a is calculated by assuming that the dew point temperature is the daily minimum T and using the method from ASCE-EWRI (2005). The psychrometric constant γ is calculated using the equation given in Allen et al. (1998) where atmospheric pressure is calculated from elevation Z (Burman et al., 1987). The C_n and C_d values for daytime and nighttime hours are from ASCE-EWRI (2005). The wind

speed u_2 in each time step is randomly generated according to the Weibull distribution with a specified scale parameter u_a that is similar to the mean wind speed and a constant shape parameter of 1.7 following Garcia et al. (1998). u_a will be varied in the sensitivity analysis and is called the wind speed magnitude. In the end, calculation of ET_p through time for a given region requires specification of A_{ST}, A_{DT}, and T_m (to simulate air temperature) and ϕ, λ, Z, u_a, k_{ext}, and LAI to calculate the associated ET_p.

2.2.4 Use of Model Outputs

When HYDRUS-1D is used to simulate a given hypothetical region, that simulation produces datasets for $\left(\Lambda_{SEB}, \overline{\theta}\right)$, $\left(\Lambda_{SEB}, \overline{s}\right)$, $\left(\Lambda_{PET}, \overline{\theta}\right)$, and $\left(\Lambda_{PET}, \overline{s}\right)$, where each pair of data is associated with a particular time during a simulation. Recall that Λ_{SEB} can be found from the components of the energy balance:

$$\Lambda_{SEB} = \frac{L \cdot ET_a}{R_n - G} \tag{20}$$

where L is the latent heat of vaporization of water (J kg^{-1}) and ET_a is the actual ET (kg m^{-2} s^{-1}). The variables R_n and G are model inputs, which are used to calculate the ET_p forcing, while ET_a is a model output. Similarly, Λ_{PET} can be calculated as:

$$\Lambda_{PET} = \frac{ET_a}{ET_p} \tag{21}$$

Both Λ_{SEB} and Λ_{PET} are unstable during the nighttime (Crago, 1996; Peng et al., 2013), so only daytime hours (excluding the hours of sunrise and sunset) are included in the paired datasets.

The average root-zone soil moisture $\overline{\theta}$ and degree of saturation $\overline{s}$ can also be found from the model outputs at the selected times steps. $\overline{\theta}$ is obtained from an arithmetic average of the individual θ values within the root-zone. All θ values in the root-zone are given equal weight regardless of the root-distribution to be consistent past evaluations of the remote sensing methods (Ahmad and Bastiaanssen, 2003; Scott et al., 2003). The degree of saturation $\overline{s}$ is calculated as $\overline{\theta} / \theta_s$.

These datasets are then used to estimate the parameters of the $\Lambda_{SEB} - \overline{\theta}$, $\Lambda_{SEB} - \overline{s}$, $\Lambda_{PET} - \overline{\theta}$, and $\Lambda_{PET} - \overline{s}$ relationships. Using the logarithmic functions proposed by Bastiaanssen et al. (2000) and Scott et al. (2003), one can write:

$$\Lambda_{SEB} = a + b\ln(\overline{\theta}) \quad \text{or} \quad \overline{\theta} = \exp\left(\frac{\Lambda_{SEB} - a}{b}\right) \tag{22}$$

$$\Lambda_{SEB} = c + d\ln(\overline{s}) \quad \text{or} \quad \overline{s} = \frac{\overline{\theta}}{\overline{\theta}_{sat}} = \exp\left(\frac{\Lambda_{SEB} - c}{d}\right) \tag{23}$$

$$\Lambda_{PET} = e + f\ln(\overline{\theta}) \quad \text{or} \quad \overline{\theta} = \exp\left(\frac{\Lambda_{PET} - e}{f}\right) \tag{24}$$

$$\Lambda_{PET} = g + h\ln(\overline{s}) \quad \text{or} \quad \overline{s} = \frac{\overline{\theta}}{\overline{\theta}_{sat}} = \exp\left(\frac{\Lambda_{PET} - g}{h}\right) \tag{25}$$

where a, b, c, d, e, f, g, and h are empirical constants. Bastiaanssen et al. (2000) and Scott et al. (2003) both estimated nearly identical values: $a = 1.284$ and $b = 0.421$. Scott et al. (2003) also estimated $c = 1$ and $d = 0.421$, where $c = 1$ was imposed from the assumption that $\Lambda_{SEB} = 1$ when the soil is fully saturated ($\overline{s} = 1$).

Figure 1a-c show the effects of a and b on the $\Lambda_{SEB} - \bar{\theta}$ relationship. As a increases, thee relationship shifts to the left and becomes steeper (Figure 1a). As b increases, the relationship moves to the right and becomes steeper (Figure 1b). Increasing both a and b simultaneously holds the position of the relationship approximately constant but makes the relationship steeper (Figure 1c). The constants e and f have similar effects on the $\Lambda_{PET} - \bar{\theta}$ relationship as shown in Figure 1d-f.

2.2.5 eFAST

eFAST is used to estimate the sensitivity of the relationship constants (a through h) and the relationship r^2 values to the climate, soil, and vegetation parameters that are provided to the model. eFast was selected in part because it performs a global sensitivity analysis (GSA) (Saltelli et al., 2007). GSA describes the sensitivity over the entire feasible ranges of the parameters rather than at a specified base point in the parameter space. GSA is better suited for current purposes because no well-defined base point exists. Among the available GSA methods (Saltelli et al., 2007) , eFAST was selected because it is computationally efficient (Saltelli et al., 1999) and has been used with other hydrological models (Jefferson et al., 2015; Reusser et al., 2011; Sanadhya et al., 2014).

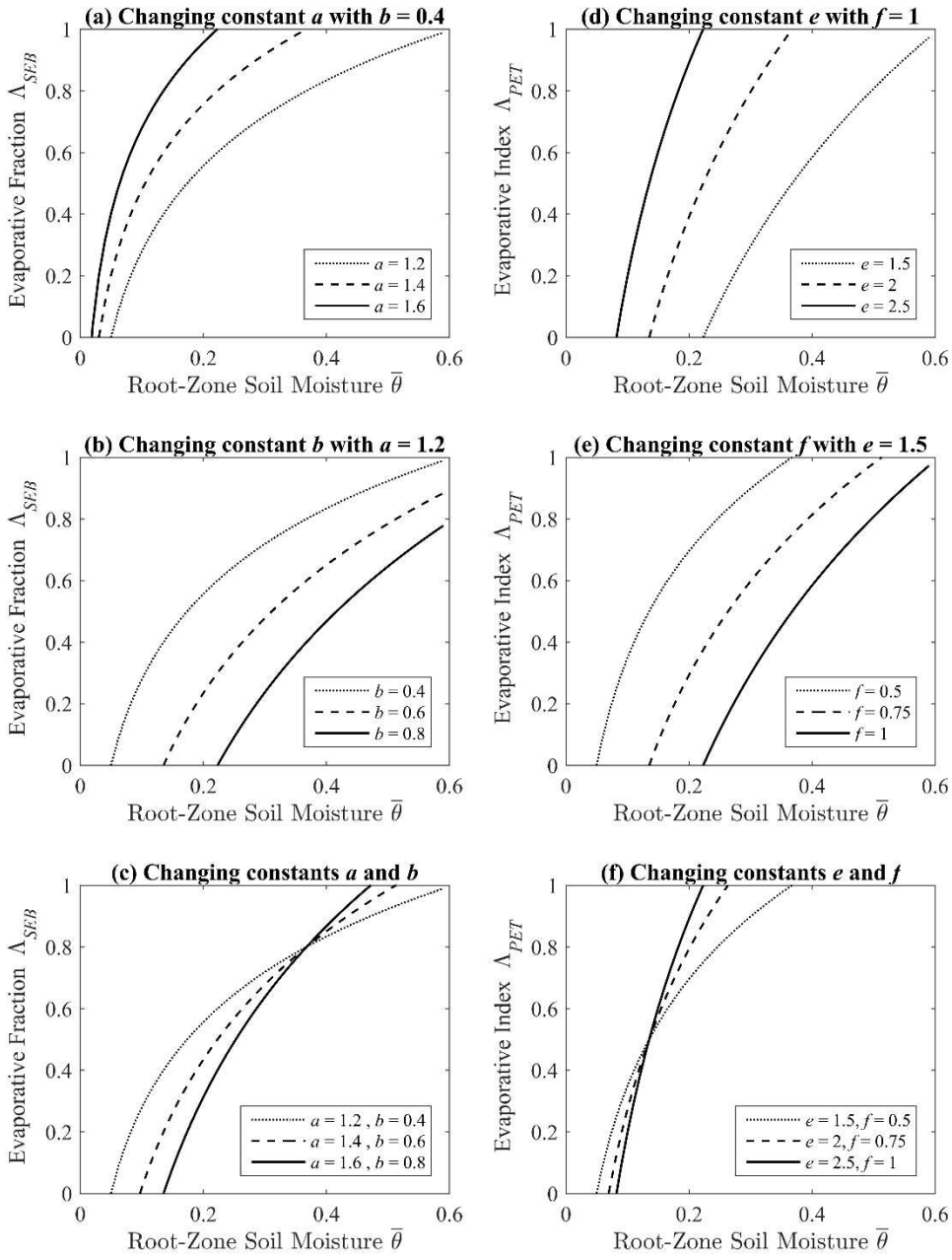

Figure 1. The form of the $\Lambda_{SEB} - \overline{\theta}$ relationship as (a) a changes while b is held constant, (b) b changes while a is held constant, (c) both a and b change, and the form of the $\Lambda_{PET} - \overline{\theta}$ relationship as (d) e changes while f is held constant, (e) f changes while e is constant, and (f) both e and f change.

eFAST uses two variance-based measures of sensitivity (Saltelli et al., 2000). One is the first-order sensitivity index to parameter j (denoted S_j), which can be written:

$$S_j = \frac{\mathrm{var}\left[E\left(Y\middle|X_j\right)\right]}{\mathrm{var}(Y)} \tag{26}$$

where $\mathrm{var}(Y)$ is the total variance of model output Y when all parameters are varied over their feasible ranges, $E\left(Y\middle|X_j\right)$ is the expected value of output Y for a given value of the parameter X_j, and $\mathrm{var}\left[E\left(Y\middle|X_j\right)\right]$ is the variance of $E\left(Y\middle|X_j\right)$ as X_j varies over its feasible range. The first order sensitivity evaluates the direct contribution that a parameter makes to the variability of the output. The second sensitivity is the total order index, which evaluates the total contribution of a parameter to the output variability when all parameter interactions within the model structure are included. Both indices were calculated in this study and found to exhibit very similar behaviors, which suggests relatively small parameter interactions in the model. Only the first order index is shown for brevity.

eFAST was applied using the SIMLAB sensitivity and uncertainty package (Giglioli and Saltelli, 2003) in conjunction with MATLAB. SIMLAB determines the minimum required number of simulations (sample size) for a given set of parameters. For the 17 parameters in this analysis (Table 1), eFAST requires at least 1,105 model simulations (65 runs per parameter). To ensure the accuracy of the results, 5049 model simulations (or 297 runs per parameter) were performed for each scenario. In addition, the sensitivity indices were generated twice for selected scenarios to confirm that they remain stable between repeated analyses.

Table 1. Parameters included in the sensitivity analysis.

	Parameter	Symbol
Soil	1. Sand (%)	Sa
	2. Silt (%)	Si
	3. Clay (%)	Cl
Vegetation	4. Root distribution index (-)	β
	5. Maximum rooting depth (cm)	R_M
	6. Leaf area index (-)	LAI
	7. Radiation extinction by the canopy (-)	k_{ext}
Climate	8. Amplitude of seasonal variation in temperature (°C)	A_{ST}
	9. Amplitude of daily variation in temperature (°C)	A_{DT}
	10. Mean annual temperature (°C)	T_m
	11. Wind speed magnitude (m/s)	u_a
	12. Mean rainfall pulse duration (h)	t_R
	13. Mean rainfall pulse intensity (cm h^{-1})	I_R
	14. Mean inter-storm period (h)	t_B
Location	15. Latitude (°N)	ϕ
	16. Longitude (°W)	λ
	17. Elevation above sea level (m)	Z

To perform the GSA, all the parameters in Table 1 were varied independently within feasible ranges according to specified continuous PDFs. In reality, several of these parameters could have interdependence because they are partially determined by the region's climate. To partially account for such dependence, eFAST was applied separately for four different climatic regions that occur in the conterminous United States (USA48), and different parameter distributions were allowed for each climatic region. The four climatic regions were defined using the aridity index (AI), which is the ratio of annual precipitation and potential ET, and the generalized

climate classifications of UNEP (1997): arid (AI below 0.20), semiarid (AI from 0.20 to 0.50), sub-humid (AI from 0.50 to 0.65), and humid (AI above 0.65). An AI map with a resolution of 30 arc-seconds was obtained from Trabucco and Zomer (2009). While some parameters may still be dependent within a climatic region, such dependence will be weaker and may be overwhelmed by other factors (e.g., human activities). Neglecting weak dependence introduces errors in the sensitivity index values but is unlikely to affect the determination of the most important parameters.

The distributions of some parameters were assumed to be independent of the climatic region (Table 2). In eFAST, all these parameters were assumed to vary according to a uniform distribution due to lack of information about their true variability. Sa, Si, and Cl were varied such that their sum was always 100%. Sa and Si were varied from 0 to 100%, but Cl was varied from 0 to 90% to avoid numerical instabilities that occur for very high clay contents. The range for β was obtained from Jackson et al. (1996), who performed a global analysis of root distributions for various terrestrial biomes. The maximum R_M value was set to 100 cm because 100 cm is the thickness of the modeled layer. The range for k_{ext} was obtained from the HYDRUS-1D manual (Šimůnek et al., 2013). In reality, k_{ext} depends on the sun angle, the distribution of plants, and the arrangement of leaves (Šimůnek et al., 2013). The range for u_a was obtained from Archer and Jacobson (2003). The range for ϕ was determined from the approximate extent of the USA48.

Table 3 describes the parameters whose distributions change between the four climatic regions. LAI was calculated for the USA48 for June 28, 2014, using MODIS-MOD15A2 images and the

procedure of (Wang and Zlotnik, 2012). Beta distributions were found to approximate the LAI variability within each region best among 15 distributions considered, so LAI is assumed to vary according to a beta distribution in eFAST. The distribution's shape parameters p and q were found by fitting the beta distribution to the data within each region.

Table 2. Probability density functions (PDFs) and limits of the feasible ranges used in eFAST for parameters that do not vary between climatic regions.

Parameter	PDF	Lower Limit	Upper Limit
Sa	Uniform	0	100
Si	Uniform	0	100
Cl	Uniform	0	90
β	Uniform	0.914	0.976
R_M	Uniform	0	100
k_{ext}	Uniform	0.5	0.75
u_a	Uniform	1	10
ϕ	Uniform	27	49

Table 3. Probability density functions (PDFs) and limits of the feasible ranges used in eFAST for parameters that vary between climatic regions. p and q are the parameters for the beta distribution, and μ and σ are the mean and the standard deviation of the natural log of the variable.

Region	Parameter	PDF	PDF Parameters	Lower Limit	Upper Limit
Arid	LAI	Beta	$p = 1.365,\ q = 15.73$	0	6.8
	A_{ST}	Uniform	-	11.38	12.61
	A_{DT}	Uniform	-	5.14	7.10
	T_m	Uniform	-	10.03	23.92
	t_R	LogNormal	$\mu = 2.084, \sigma = 0.687$	0.005	0.99

	I_R	LogNormal	$\mu = -2.475, \sigma = 0.427$	0.005	0.99
	t_B	LogNormal	$\mu = 5.434, \sigma = 0.503$	0.005	0.99
	λ	Uniform	-	-122	-100
	Z	Beta	$p = 1.299,\ q = 1.862$	0	2488
Semiarid	LAI	Beta	$p = 1.434,\ q = 12.177$	0	6.8
	A_{ST}	Uniform	-	10.56	15.91
	A_{DT}	Uniform	-	4.67	4.67
	T_m	Uniform	-	6.28	18.39
	t_R	LogNormal	$\mu = 2.138, \sigma = 0.678$	0.005	0.99
	I_R	LogNormal	$\mu = -2.615, \sigma = 0.610$	0.005	0.99
	t_B	LogNormal	$\mu = 5.065,\ \sigma = 0.695$	0.005	0.99
	λ	Uniform	-	-122	-100
	Z	Beta	$p = 2.814,\ q = 4.300$	0	3250
Sub-humid	LAI	Beta	$p = 2.598,\ q = 5$	0	6.8
	A_{ST}	Uniform	-	9.61	17.38
	A_{DT}	Uniform	-	3.95	4.83
	T_m	Uniform	-	4.90	20.04
	t_R	LogNormal	$\mu = 2.069,\ \sigma = 0.704$	0.005	0.99
	I_R	LogNormal	$\mu = -2.326, \sigma = 0.652$	0.005	0.99
	t_B	LogNormal	$\mu = 4.933,\ \sigma = 0.572$	0.005	0.99
	λ	Uniform	-	-100	-97
	Z	Beta	$p = 1.444,\ q = 6.951$	71	3575
Humid	LAI	Beta	$p = 1.369,\ q = 1.311$	0	6.8
	A_{ST}	Uniform	-	9.77	13.35
	A_{DT}	Uniform	-	4.51	4.60
	T_m	Uniform	-	8.24	18.37
	t_R	LogNormal	$\mu = 1.937,\ \sigma = 0.547$	0.005	0.99
	I_R	LogNormal	$\mu = -2.01,\ \sigma = 0.609$	0.005	0.99
	t_B	LogNormal	$\mu = 4.29,\ \sigma = 0.400$	0.005	0.99
	λ	Uniform	-	-97	-67
	Z	Beta	$p = 0.663,\ q = 3.878$	0	4371

The temperature parameters A_{ST}, A_{DT}, and T_m were determined from hourly U.S. Climate Normals from the National Oceanic and Atmospheric Administration (Arguez et al., 2012). Each climatic region was divided into three subregions, and a representative station was identified towards the middle of each subregion. The sinusoidal temperature model was then fitted to the data from each station. The ranges for A_{ST}, A_{DT}, and T_m were determined from the range of calibrated parameters within each region. Uniform distributions were selected for these parameters for simplicity (the number of stations is not sufficient for detailed characterization of the variability in each region). The variables F_{ST}, F_{DT}, C_{ST}, C_{DT} in the temperature model exhibited little difference between the stations, so they were set to consistent values of 0.01713 day^{-1}, 6.284 day^{-1}, -1.844, and -2.388, respectively.

The ranges of the precipitation model parameters I_R, t_R, and t_B were obtained from Hawk and Eagleson (1992). They estimated these parameters at more than 70 stations in the USA48. Those stations were plotted on an AI map to obtain I_R, t_R, and t_B values for each climatic region. A log-normal distribution was found to approximate the variability of each parameter in each region, so this distribution is used in eFAST. The distribution parameters (the mean of the log of the variable μ and the standard deviation of the log of the variable σ) were determined by fitting the log-normal distribution to the data in each region. Each distribution was also truncated at an upper bound to prevent very rare outlier rainfall values from occurring.

The ranges for λ were determined from the approximate extents of the climatic regions within the USA48. Similarly, the distributions and ranges for Z were obtained from the elevation

values in each region based on the United States Geological Survey National Elevation Dataset (1 arc-second resolution).

The cloudiness distribution parameters p_1 and p_2 were determined as follows. First, hourly clear sky solar radiation and incident solar radiation data were obtained from the National Solar Radiation Database for 2000-2010 for twelve stations within the USA48 (three per climatic region). For each station, the cloudiness was calculated by a rearrangement of Eq. (16), and beta distributions were fitted to the data. The average parameters in each region are used in the model: arid ($p_1 = 0.2434$, $p_2 = 0.0713$), semiarid ($p_1 = 0.1847$, $p_2 = 0.0532$), sub-humid ($p_1 = 0.1890$, $p_2 = 0.0664$), and humid ($p_1 = 0.2950$, $p_2 = 0.1052$).

2.3 Results and Discussion

2.3.1 Average Strength and Form of Relationships

We first examine the strength of the $\Lambda_{SEB} - \overline{\theta}$, $\Lambda_{PET} - \overline{\theta}$, $\Lambda_{SEB} - \overline{s}$, and $\Lambda_{PET} - \overline{s}$ relationships under the assumption that a single $\Lambda_{SEB} - \overline{\theta}$, $\Lambda_{PET} - \overline{\theta}$, $\Lambda_{SEB} - \overline{s}$, or $\Lambda_{PET} - \overline{s}$ relationship applies to all locations in the USA48. Specifically, Figure 2 shows the r^2 values for the four relationships when they are calculated using data from all times in all simulations from all regions. Overall, low r^2 values are observed for all four relationships. These low values occur because a single relationship cannot adapt to the local soil, vegetation, and climatic characteristics. Higher r^2 values are observed for the relationships that use $\overline{s}$ than the equivalent relationships that use $\overline{\theta}$. By using $\overline{s}$, the relationships account the local saturation water content θ_s and thus partially capture the local soil characteristics. This modeling result is

consistent with the hypothesis of Scott et al. (2003) and the empirical results from Alburn et al. (2015), who observed better performance using $\overline{s}$ than $\overline{\theta}$. Higher r^2 values are also observed for the relationships that use Λ_{PET} instead of Λ_{SEB}. Λ_{SEB} only considers the available energy in its denominator (i.e. $R_n - G$), but ET depends on not only the available energy but also the turbulent transport (i.e. wind speed and vapor pressure deficit). Λ_{PET} accounts for all these components by using ET_p in its denominator. Thus, Λ_{PET} is more directly related to the actual ET and the associated $\overline{\theta}$ and $\overline{s}$ values.

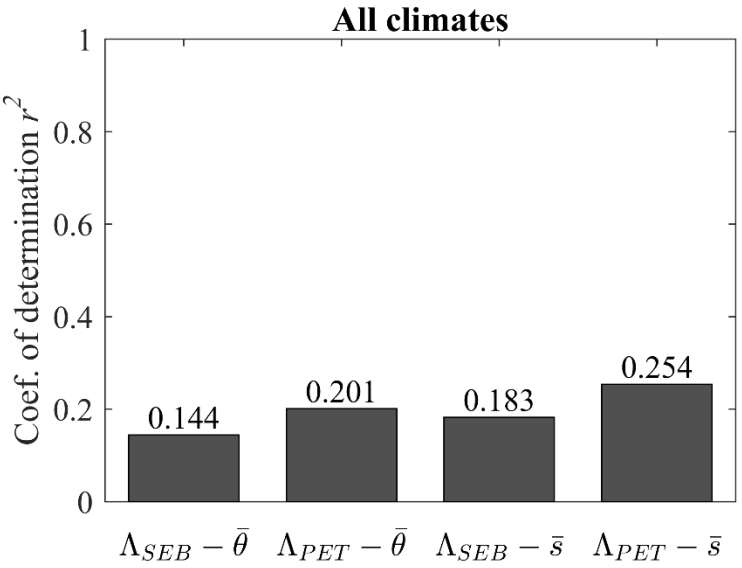

Figure 2. Coefficient of determination r^2 for the $\Lambda_{SEB} - \overline{\theta}$, $\Lambda_{PET} - \overline{\theta}$, $\Lambda_{SEB} - \overline{s}$, and $\Lambda_{PET} - \overline{s}$ relationships based on all data from all climatic regions

We next examine the strength of the $\Lambda_{SEB} - \overline{\theta}$, $\Lambda_{PET} - \overline{\theta}$, $\Lambda_{SEB} - \overline{s}$, and $\Lambda_{PET} - \overline{s}$ relationships under the assumption that the relationships change for each combination of soil, vegetation, and climate characteristics. Specifically, Figure 3 shows the r^2 values for the four relationships

when the relationships are calculated separately from the data in each simulation (and the r^2 values are then averaged for each climatic region). The r^2 values in Figure 3 are roughly twice the values in Figure 2, which suggests that the performance of the remote-sensing methods might be improved greatly if the local characteristics are accommodated in the relationships. This modeling result is consistent with the empirical results of Alburn et al. (2015), who observed that calibrated relationships substantially outperform the universal relationships. In Figure 3, the relationships that use Λ_{PET} again outperform those that use Λ_{SEB} because Λ_{PET} is more directly connected with the actual ET. However, the relationships that use $\bar{\theta}$ and $\bar{s}$ exhibit nearly identical performance in Figure 3 because different relationships are fitted to the data from each simulation. Thus, the constants that define the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships accommodate the local θ_s values. Figure 3 also shows that the r^2 values are higher for semiarid and sub-humid regions than for arid and humid regions. The simulations for arid regions include significant periods where $\bar{\theta}$ is less than θ_r, in which case the ET has almost no relationship with $\bar{\theta}$. Similarly, the simulations for humid regions have long periods where $\bar{\theta}$ approaches θ_s and ET changes little as $\bar{\theta}$ changes. Because the relationships for $\bar{\theta}$ and $\bar{s}$ exhibit nearly identical performance (Figure 3) and $\bar{\theta}$ is expected to be more useful for practical applications than $\bar{s}$, only the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships are considered for further analysis.

Figure 4 shows the average forms of the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships for the four climatic regions. These relationships were determined by first obtaining the empirical constants for each simulation (a and b for the $\Lambda_{SEB} - \bar{\theta}$ relationship and e and f for the $\Lambda_{PET} - \bar{\theta}$ relationship) and then averaging those constants within each climate. Arid and semiarid regions

have steeper $\Lambda_{SEB} - \overline{\theta}$ relationships than sub-humid and humid regions. Storms in arid and semiarid climates tend to be shorter and have lower intensity, so the soil moisture tends to be in the upper portion of the soil profile. Also, lower vegetation cover causes surface evaporation to dominate the ET. As a result, high Λ_{SEB} values can occur in these regions even though the average root-zone soil moisture is small. In addition, arid and semiarid regions tend to have lower $\overline{\theta}$ values than sub-humid and humid regions. A single logarithmic relationship does not always fit the data for the full range of $\overline{\theta}$ values, so differences in the distribution of $\overline{\theta}$ values lead to differences in the fitted relationships.

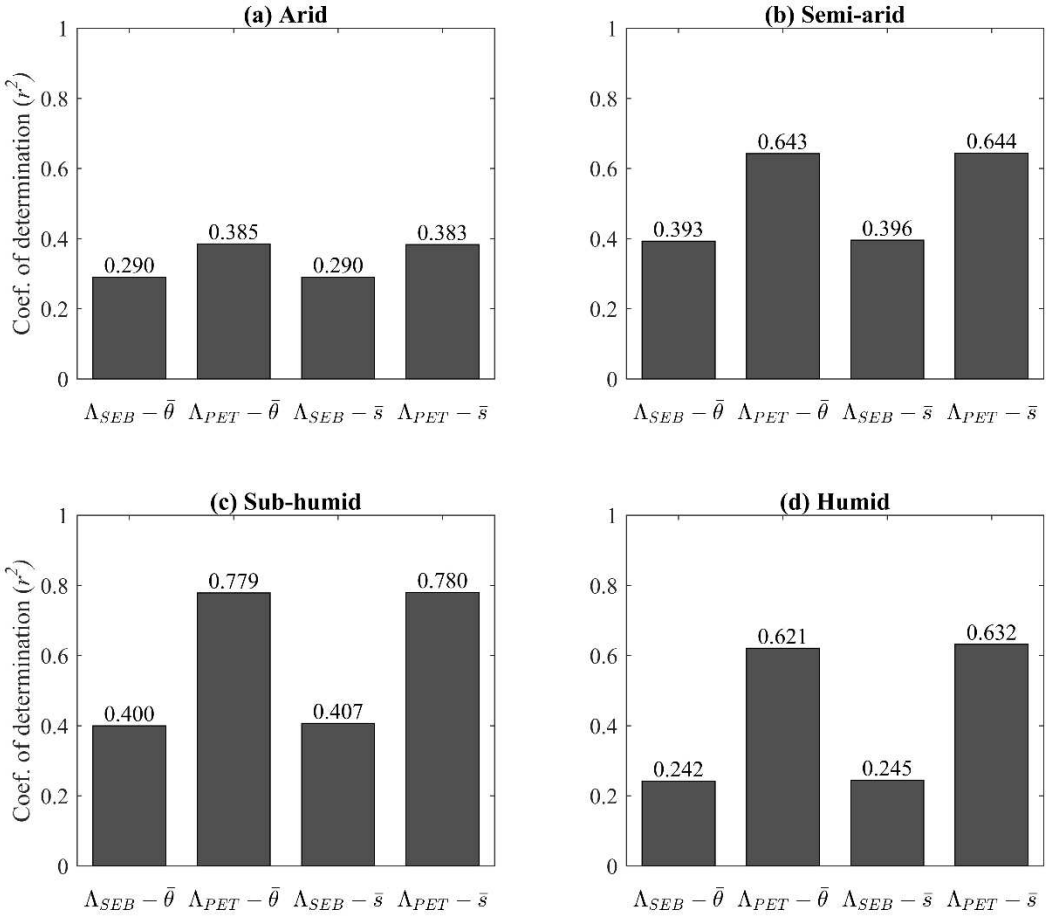

Figure 3. Coefficient of determination r^2 for the $\Lambda_{SEB}-\bar{\theta}$, $\Lambda_{PET}-\bar{\theta}$, $\Lambda_{SEB}-\bar{s}$, and $\Lambda_{PET}-\bar{s}$ relationships when the relationships are determined separately for each simulation for (a) arid, (b) semiarid, (c) sub-humid, and (d) humid climates

Figure 4b shows that the $\Lambda_{PET}-\bar{\theta}$ relationship is more linear than the $\Lambda_{SEB}-\bar{\theta}$ relationship and tends to fall in a narrower range of $\bar{\theta}$. The lower and upper limits of the linear section are implied by the S-shaped function that characterizes root-water uptake and can be thought of as an approximate average permanent wilting point (PWP) and average field capacity (FC), respectively. HYDRUS-1D implies a nearly linear relationship between the soil moisture at a

particular depth and Λ_{PET} between the PWP and FC if transpiration dominates. Similar linear $\Lambda_{PET} - \bar{\theta}$ relationships have also been seen in previous studies (Baier, 1969; Thornthwaite and Mather, 1955). The relationships change between climatic regions due to variations in the moisture with depth. In particular, due to the frequent rainfall events that occur throughout the year and associated wetting and drying fronts, humid regions tend to have the most variable Λ_{PET} and moisture changes with depth, which leads to increased scatter in the Λ_{PET} values for a given $\bar{\theta}$. Because Λ_{PET} always falls between 0 and 1, such scatter inevitably leads to a relationship with a lower slope as seen in Figure 4b.

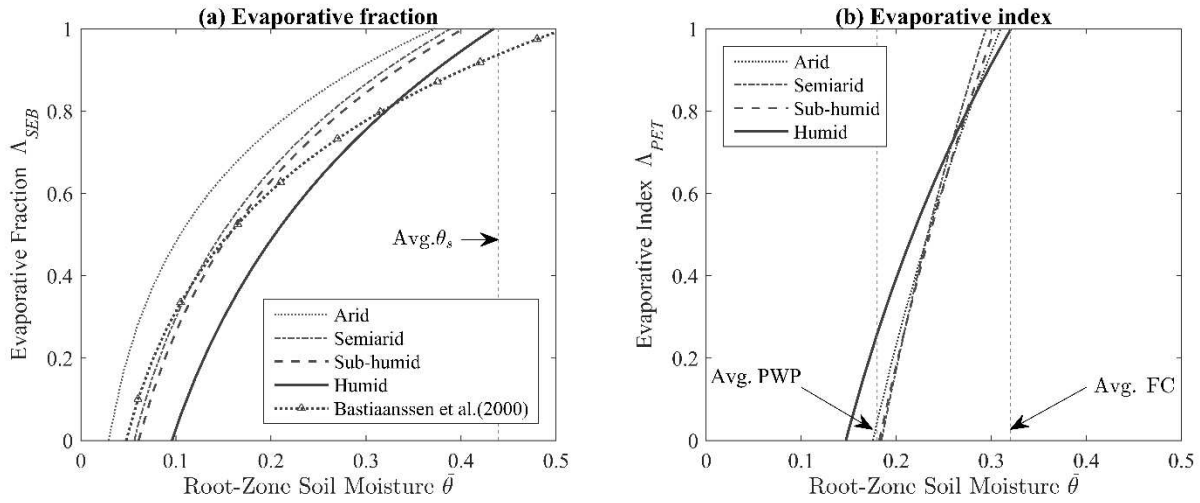

Figure 4. Comparison of the average (a) $\Lambda_{SEB} - \bar{\theta}$ and (b) $\Lambda_{PET} - \bar{\theta}$ relationships for the four climatic regions. The $\Lambda_{SEB} - \bar{\theta}$ relationship proposed by Bastiaanssen et al. (2000) is also shown for reference. Approximate average $\bar{\theta}_s$, permanent wilting point (PWP), and field capacity (FC) are also indicated.

2.3.2 Sensitivity Results

The eFAST results for the $\Lambda_{PET} - \overline{\theta}$ relationship are shown in Figure 5, 6, and 7. Only this relationship is shown because Λ_{PET} was found to provide more reliable estimates of $\overline{\theta}$ than Λ_{SEB} in the previous section. eFAST results were obtained for the other three cases and typically have similar behaviors to the case that is shown.

Figure 5 shows the first-order sensitivity S of the $\Lambda_{PET} - \overline{\theta}$ relationship's r^2 to each climatic, soil, and vegetation parameter. In this figure, the parameters are ranked from the most important to least important based on their S values. To facilitate inter-comparisons, the S values are normalized so that their sum is 100% (Ruark et al., 2011). Overall, vegetation characteristics primarily control the strength of the $\Lambda_{PET} - \overline{\theta}$ relationship. LAI is the most important parameter for all four regions, and R_M is the second most important for three out of four regions. The average dependence of r^2 on LAI and R_M is shown in Figure 5e and 5f, respectively. When LAI is very low, r^2 is also low because soil evaporation dominates the ET (see Equation (6)). Soil evaporation is insensitive to deep soil moisture because it draws water from near the ground surface. Thus, Λ_{PET} becomes less connected to $\overline{\theta}$ when LAI is small. When LAI is above about 2, its value has little impact on r^2 because transpiration more consistently dominants the ET. For arid regions, the dependence on LAI extends over a wider range of LAI because the limited precipitated water sometimes does not reach deep into the soil, which emphasizes the importance of evaporation and weakens the $\Lambda_{PET} - \overline{\theta}$ relationship. r^2 exhibits a similar dependence on LAI (and other parameters) for the sub-humid and humid regions

because the distributions from which LAI and other parameters are generated are similar for those two regions (Table 3).

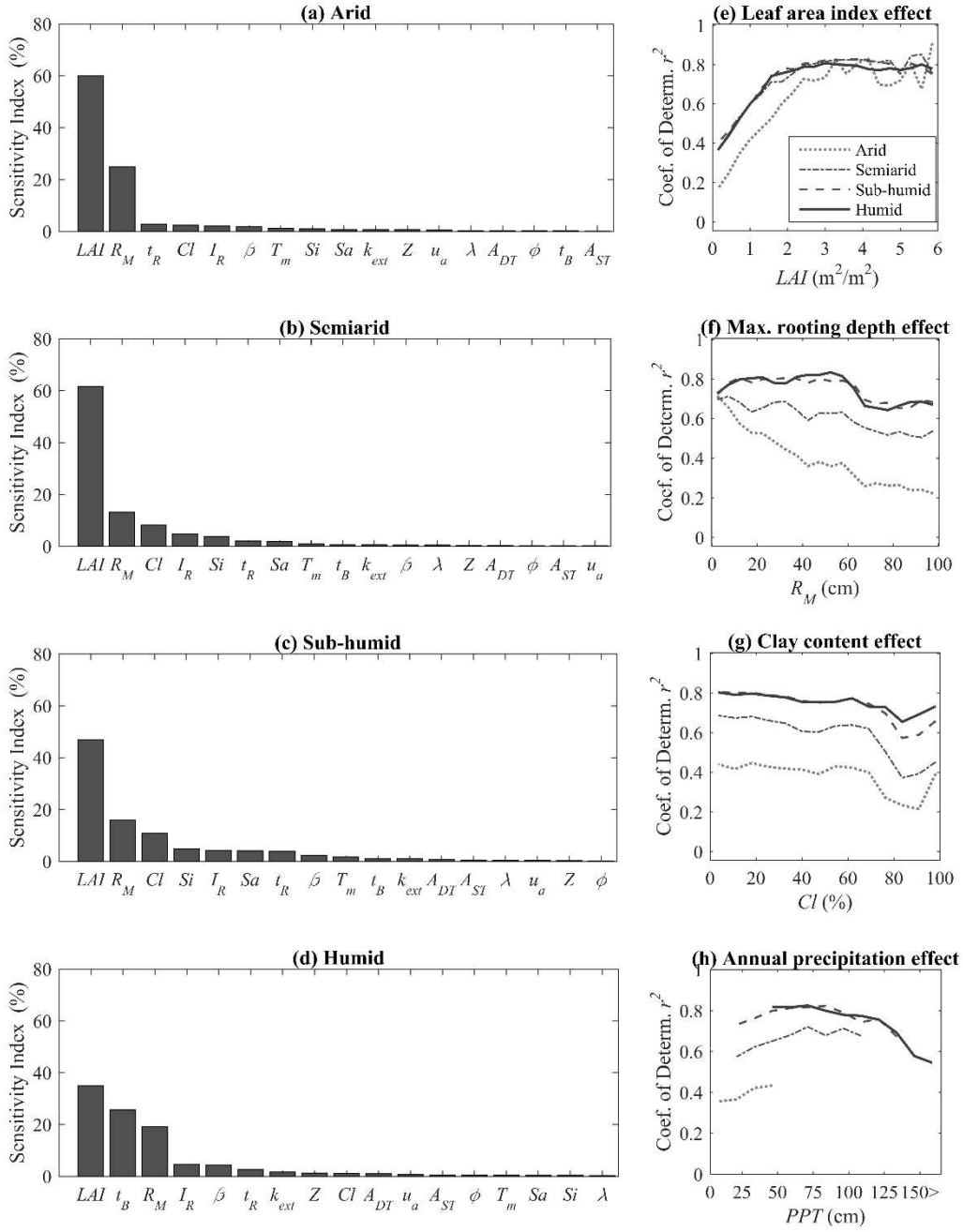

Figure 5. Sensitivity of the coefficient of determination r^2 of the $\Lambda_{PET} - \bar{\theta}$ relationship to the model parameters for (a) arid, (b) semiarid, (c) sub-humid, and (d) humid climates, and the average dependence of r^2 on (e) leaf area index, (f) maximum rooting depth, (g) clay content, and (h) annual precipitation.

Figure 5f shows that smaller R_M values improve the strength of the $\Lambda_{PET} - \bar{\theta}$ relationship, particularly for arid regions. As noted earlier, precipitated water sometimes does not reach deep into the soil profile for arid regions. If R_M is large, the importance of transpiration relative to evaporation changes through time depending on whether the precipitated water has reached deep into the soil. Evaporation and transpiration have different relationships with soil moisture, so this changing importance tends to weaken the overall $\Lambda_{PET} - \bar{\theta}$ relationship. In humid regions, precipitation is sufficient to keep the soil profile wet, so transpiration tends to be more consistently important irrespective of R_M.

Figure 5h plots the average dependence of r^2 on the precipitation that occurred in each simulation. To facilitate interpretation, the annual precipitation PPT is estimated by multiplying the precipitation received in each 9-month simulation by 12/9. PPT was not included in eFAST because it is not a parameter in the model simulations. Instead, it is a byproduct of the parameters of the stochastic precipitation model (I_R , t_R and t_B). For all climates, very large values of PPT weaken the $\Lambda_{PET} - \bar{\theta}$ relationship. When PPT is very large, $\bar{\theta}$ is above the FC for substantial portions of each simulation. Thus, Λ_{PET} approaches one irrespective of the $\bar{\theta}$ value and the relationship between these two variables weakens. Low PPT values also tend to weaken the $\Lambda_{PET} - \bar{\theta}$ relationship for arid regions. Arid regions tend to have shorter and less intense storms than other regions. Thus, when precipitation is low for arid regions, the precipitated water is more likely to stay near the ground surface. In such a case, soil evaporation tends to play a larger role. Soil evaporation does not depend on deeper soil moisture, which also weakens the $\Lambda_{PET} - \bar{\theta}$ relationship.

Figure 6 shows the sensitivity of the $\Lambda_{PET} - \bar{\theta}$ relationship constant f to the model parameters. Overall, soil characteristics are most important for determining the value of f. Cl is the single most important parameter for all four climatic regions, and Si is the second most important for three of the four regions. R_M plays a secondary role for semiarid and sub-humid regions. Figure 6e-h show the average dependence of f on Cl, Si, R_M, and PPT for each climate. Increasing Cl causes a nearly linear increase in f (Figure 6e), while increasing Si causes a nearly linear decrease in f (Figure 6f). When R_M is small, the constant f also increases with R_M (Figure 6g). In addition, f increases with PPT until it reaches a peak around 50 cm, at which point f decreases with PPT.

Figure 7 shows the sensitivity of the relationship constant e to the model parameters. The highest sensivity values in Figure 7 are smaller than the highest sensitivity values in Figure 6, which suggests that more parameters have a meaningful influence on e than f. Similar to f, e depends primarily on soil characteristics. Cl is the most important characteristic in all four climates, while Si is the second most important in three out of the four climates. Vegetation characteristics (LAI and R_M) again have a secondary role for semiarid and sub-humid climates. The nature of e's dependence on Cl, Si, R_M, and PPT is shown in Figure 7e-h. Overall, e shows a similar dependence on these parameters as f.

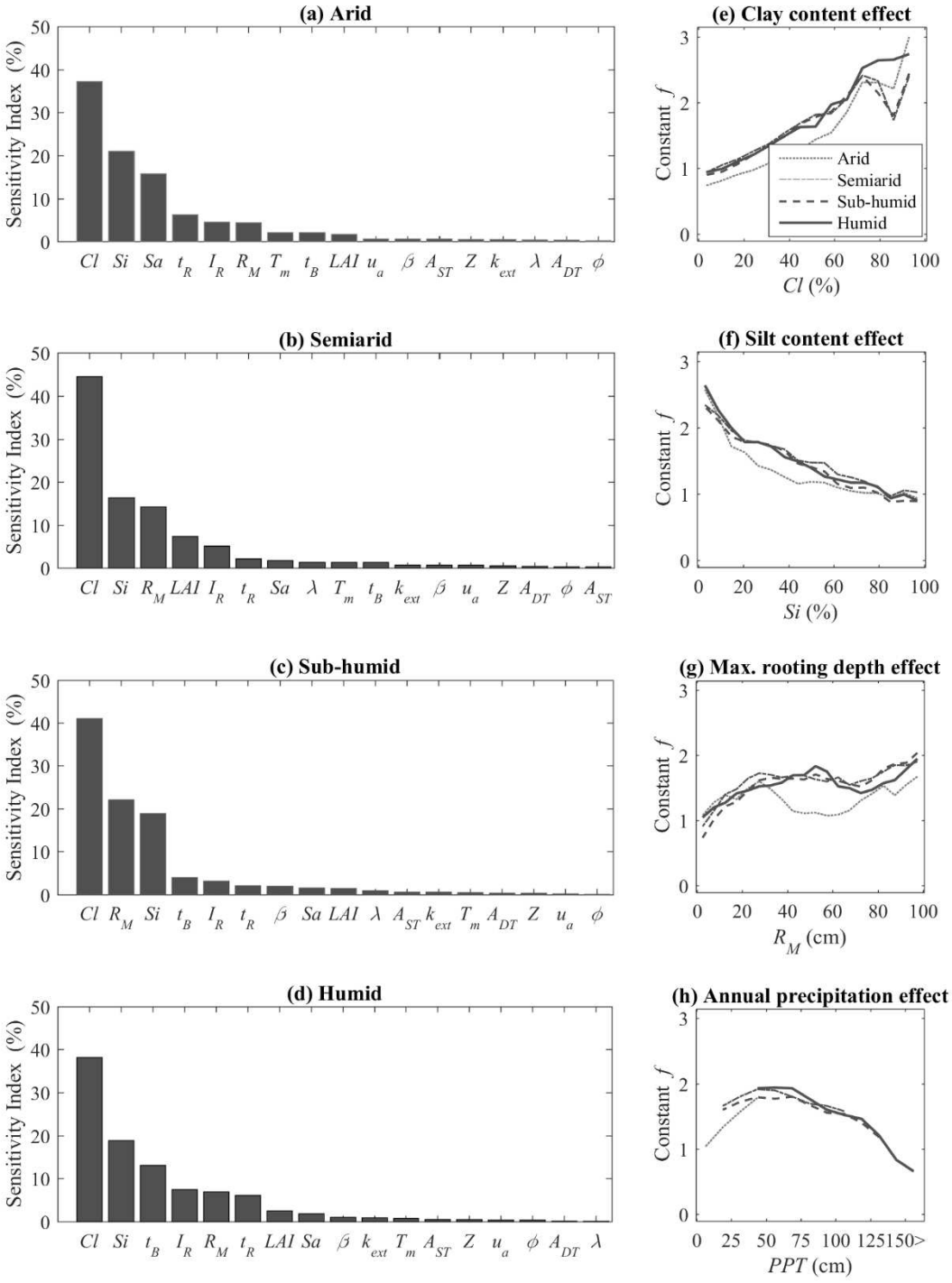

Figure 6. Sensitivity of relationship constant f to the model parameters for (a) arid, (b) semiarid, (c) sub-humid, and (d) humid climates and the average dependence of f on (e) clay content, (f) silt content, (g) maximum rooting depth, and (h) annual precipitation for the different climates

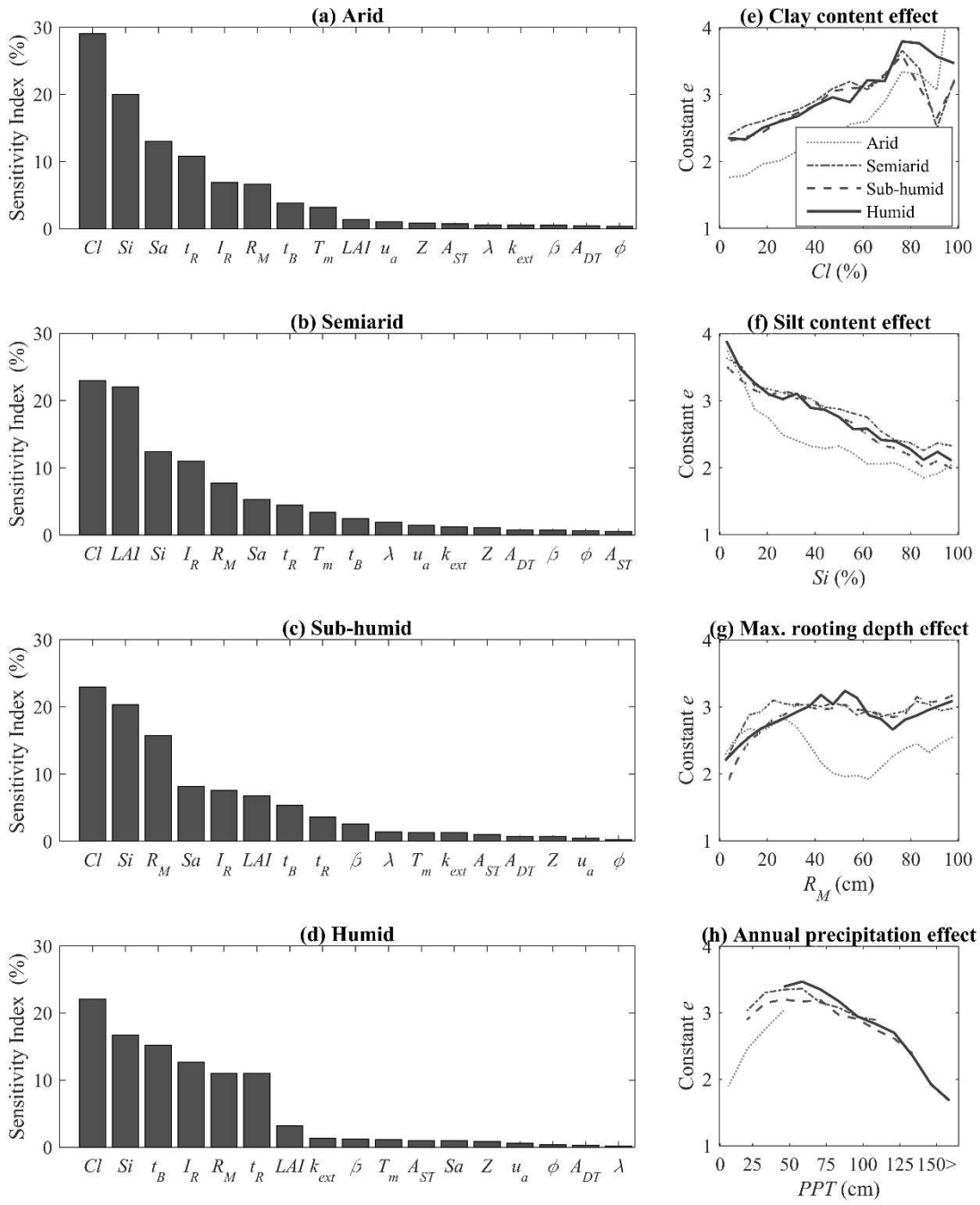

Figure 7. Sensitivity of relationship constant e to the model parameters for (a) arid, (b) semiarid, (c) sub-humid, and (d) humid climates and the average dependence of e on (e) clay content, (f) silt content, (g) maximum rooting depth, and (h) annual precipitation for the different climates

The tendency of e and f to increase simultaneously suggests that the most important model parameters (i.e. regional characteristics) affect the $\Lambda_{PET} - \overline{\theta}$ relationship as shown in Figure 1f. Specifically, when e and f both increase, the $\Lambda_{PET} - \overline{\theta}$ relationship becomes steeper. An increase in Cl or a decrease in Si reduces the range between the PWP and FC (i.e. the available water for root-water uptake). This increases e and f and produces a steeper $\Lambda_{PET} - \overline{\theta}$ relationship. Low values of R_M (below about 30 cm) tend to restrict transpiration's access to soil moisture. Thus, Λ_{PET} tends to be lower for a given $\overline{\theta}$, which also reduces e and f (see example curves in Figure 1f). As discussed earlier, very large PPT values tend to produce many soil moisture values in each simulation that exceed the FC, where Λ_{PET} does not depend strongly on the particular value of $\overline{\theta}$. Similarly, very low PPT values tend to produce soil moisture values that are below the PWP, where Λ_{PET} also does not depend strongly on the particular value of $\overline{\theta}$. Thus, both extremes introduce more scatter into the relationship, which reduces the estimates of e and f.

2.4 Proposed Equations

Regression equations were developed to estimate the constants of the $\Lambda_{SEB} - \overline{\theta}$ and $\Lambda_{PET} - \overline{\theta}$ relationships from the most important regional characteristics. To develop these equations, four cases were considered. Case 1 assumes no regional information is available, so the relationship constants are simply numbers. These numbers can be compared to those suggested by Bastiaanssen et al. (2000), but they are derived from all the combinations of soil, vegetation, and climate characteristics supplied to HYDRUS-1D. Case 2 assumes that only the climate

classification is available. Thus, the relationship constants differ for arid, semiarid, sub-humid, and humid regions. Case 3 assumes that the climate classification and total annual precipitation are known. Thus, the relationship constants are estimated as a function of PPT, where the function differs for each climate. Finally, Case 4 assumes that the climate classification, PPT, Cl, Si, and LAI are all available. Although R_M was observed to be important in the previous section, it is unlikely to be known for a particular region with any certainty. Thus, it is not used in the equations. Linear regression equations are used to estimate the relationship constants. As noted earlier, 50 cm appears to separate two distinct dependencies on PPT, so separate relationships are used for these two ranges of PPT.

For the $\Lambda_{SEB} - \bar{\theta}$ relationship, the following equations are proposed to estimate a and b for Case 3:

$$a = C_1 + C_2 \cdot PPT \tag{27}$$

$$b = D_1 + D_2 \cdot PPT \tag{28}$$

where PPT is the annual precipitation (cm) and C_1 (-), C_2 (cm^{-1}), D_1 (-), and D_2 (cm^{-1}) are the regression constants. For Case 4, the following equations are proposed to estimate a and b:

$$a = C_1 + C_2 \cdot PPT + C_3 \cdot Cl + C_4 \cdot Si + C_5 \cdot LAI \tag{29}$$

$$b = D_1 + D_2 \cdot PPT + D_3 \cdot Cl + D_4 \cdot Si + D_5 \cdot LAI \tag{30}$$

where Cl is a percentage, Si is a percentage, and C_3 (-), C_4 (-), C_5 (-), D_3 (-), D_4 (-), and D_5 (-) are regression constants.

For the $\Lambda_{PET} - \bar{\theta}$ relationship, the following equations are proposed to estimate e and f for Case 3:

$$e = E_1 + E_2 \cdot PPT \tag{31}$$

$$f = F_1 + F_2 \cdot PPT \tag{32}$$

where E_1 (-), E_2 (cm^{-1}), F_1 (-), and F_2 (cm^{-1}) are regression constants. For Case 4, the following equations are proposed to estimate e and f:

$$e = E_1 + E_2 \cdot PPT + E_3 \cdot Cl + E_4 \cdot Si \tag{33}$$

$$f = F_1 + F_2 \cdot PPT + F_3 \cdot Cl + F_4 \cdot Si \tag{34}$$

where E_3 (-), E_4 (-), F_3 (-), and F_4 (-) are regression constants.

Table 4 shows the regression constants for the $\Lambda_{SEB} - \bar{\theta}$ relationship, and Table 5 shows the constants for the $\Lambda_{PET} - \bar{\theta}$ relationship. In all cases, the regression constants were calibrated to maximize the NSCE between the *a* and *b* values (or *e* and *f* values) obtained from the HYDRUS-1D simulations and the regression equations. The constants for the $\Lambda_{SEB} - \bar{\theta}$ relationship in Case 1 (Table 4) are similar to the values proposed by Bastiaanssen et al. (2000) and Scott et al. (2003). Based on their local datasets, they suggested $a = 1.284$ and $b = 0.421$, whereas Table 4 reports $a = 1.484$ and $b = 0.522$, respectively. The NSCE values for Case 1 and Case 2 in Tables 4 and 5 are all zero because these cases explain no variation in the relationship constants within a given climatic region. The NSCE increases when the precipitation effects are considered (Case 3). For the $\Lambda_{SEB} - \bar{\theta}$ relationship, the average NCSE is 0.17 for a and 0.12 for b. For the $\Lambda_{PET} - \bar{\theta}$ relationship, the average NSCE is 0.24 for e and 0.15 for f. The NSCE values

further improve when the effects of PPT, Cl, Si, and LAI are included (Case 4). For the $\Lambda_{SEB} - \bar{\theta}$ relationship, the average NCSE is 0.40 for a and 0.53 for b. For the $\Lambda_{PET} - \bar{\theta}$ relationship, the average NSCE is 0.44 for e and 0.55 for f.

To evaluate the overall performance, Figure 8 shows the NSCE values for $\bar{\theta}$ when the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships are estimated using Bastiaanssen et al. (2000) and Cases 1-4. Using Λ_{PET} typically provides better NSCE values than Λ_{SEB}. This improvement also occurs when using the empirical relationship, even though Bastiaanssen et al. (2000) did not suggest using the empirical relationship with Λ_{PET}. However, the $\bar{\theta}$ estimates from the Bastiaanssen et al. (2000) relationship typically have negative NSCE values. The $\bar{\theta}$ estimates from all four cases have higher NSCE's than the estimates from the Bastiaanssen et al. (2000) relationship. Furthermore, as one uses more information to estimate the relationship (i.e. one moves from Case 1 to Case 4), the estimates of $\bar{\theta}$ typically improve. Like the empirical relationship, Case 1 uses no regional characteristics, but it was derived by considering a much wider range of climate, soil, and vegetation conditions than the empirical relationship. Case 2 uses distinct $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships for each climatic region, so it adapts to the typical conditions for each climate. Better performance is observed for the semiarid and subhumid regions than the arid and humid regions. Case 3 includes the annual precipitation (PPT) when determining the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships, but it shows only slight improvements compared to Case 2. Case 4 provides the best $\bar{\theta}$ estimates for all regions by including variations in precipitation (PPT), soil texture (Si and Cl), and vegetation (LAI). Although these evaluations are made using

modeled soil moisture, the substantial improvement in NSCE between using the Bastiaanssen et al. (2000) relationship and Case 4 indicates that further testing of the proposed methods are warranted.

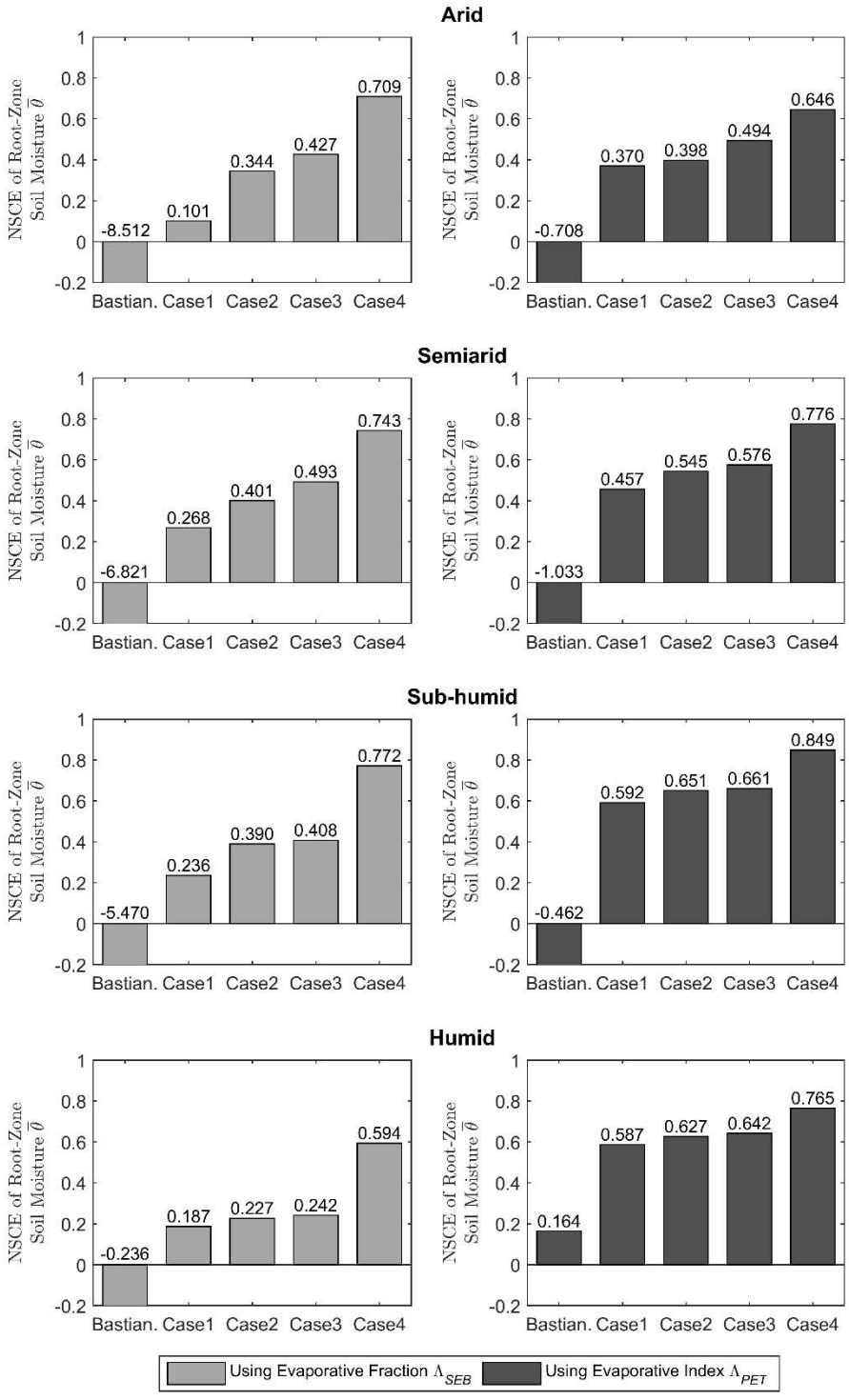

Figure 8. Nash-Sutcliffe Coefficient of Efficiency (NSCE) of root-zone soil moisture $\bar{\theta}$ estimates from the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships proposed by Bastiaanssen et al. (2000) and Cases 1 to 4 for arid, semiarid, sub-humid, and humid regions.

Table 4. Proposed regression equation constants for estimating a and b of the $\Lambda_{SEB} - \bar{\theta}$ relationship.

Climatic Region	Constant a						Constant b					
	C_1	C_2	C_3	C_4	C_5	NSCE	D_1	D_2	D_3	D_4	D_5	NSCE
Case 1												
All Climates	1.4844	--	--	--	--	0.00	0.5222	--	--	--	--	0.00
Case 2												
Arid	1.3884	--	--	--	--	0.00	0.3932	--	--	--	--	0.00
Semi-arid	1.4873	--	--	--	--	0.00	0.5158	--	--	--	--	0.00
Sub-humid	1.4814	--	--	--	--	0.00	0.5286	--	--	--	--	0.00
Humid	1.5517	--	--	--	--	0.00	0.6612	--	--	--	--	0.00
Case 3												
Arid	1.3669	0.0057	--	--	--	0.07	0.4160	0.0045	--	--	--	0.05
Semi-arid*	1.3709	0.0024	--	--	--	0.17	0.3968	0.0011	--	--	--	0.10
Semi-arid**	1.5634	-0.0021	--	--	--	0.19	0.5128	-0.0014	--	--	--	0.14
Sub-humid*	1.3967	0.0025	--	--	--	0.12	0.4803	0.0013	--	--	--	0.07
Sub-humid **	1.3545	0.0019	--	--	--	0.28	0.4665	0.0009	--	--	--	0.22
Humid	3.4866	-0.0082	--	--	--	0.17	2.9917	-0.0096	--	--	--	0.13
Case 4												
Arid	1.4457	0.0084	0.0042	-0.0031	-0.059	0.36	0.3195	0.0061	0.0073	-0.0012	-0.0369	0.40
Semi-arid*	1.2327	0.0065	0.006	-0.0023	-0.0542	0.30	0.1086	0.0046	0.0085	-0.0011	-0.02	0.54
Semi-arid**	1.7498	-0.0026	0.0017	-0.0032	-0.0321	0.39	0.5127	-0.0017	0.0047	-0.0019	-0.0055	0.56
Sub-humid*	1.7462	0.0054	0.0061	-0.0051	-0.08	0.40	0.5550	0.0037	0.0109	-0.0028	-0.0417	0.51
Sub-humid **	1.7578	0.0013	0.0041	-0.0055	-0.0281	0.51	0.4651	0.0025	0.0095	-0.0032	-0.0071	0.60
Humid	5.6182	-0.0181	0.0286	-0.0309	-0.0651	0.41	4.4269	-0.0197	0.0452	-0.0286	-0.0249	0.58

* $PPT \leq 50$ cm and ** $PPT > 50$ cm

Table 5. Proposed regression equation constants for estimating e and f of the $\Lambda_{PET} - \overline{\theta}$ relationship.

Climatic Region	Constant e					Constant f				
	E_1	E_2	E_3	E_4	NSCE	F_1	F_2	F_3	F_4	NSCE
Case 1										
All Climates	1.8597	--	--	--	0.00	0.7423	--	--	--	0.00
Case 2										
Arid	1.6292	--	--	--	0.00	0.5314	--	--	--	0.00
Semi-arid	1.6895	--	--	--	0.00	0.5953	--	--	--	0.00
Sub-humid	2.0299	--	--	--	0.00	0.8893	--	--	--	0.00
Humid	3.0385	--	--	--	0.00	1.8528	--	--	--	0.00
Case 3										
Arid	1.4484	0.0102	--	--	0.08	0.4809	0.0041	--	--	0.05
Semi-arid*	1.6180	0.0007	--	--	0.18	0.5102	0.0008	--	--	0.1
Semi-arid**	1.8433	-0.0031	--	--	0.28	0.6440	-0.0015	--	--	0.17
Sub-humid*	1.7358	0.0101	--	--	0.17	0.7179	0.0054	--	--	0.1
Sub-humid **	2.3901	-0.0048	--	--	0.33	1.0573	-0.0021	--	--	0.2
Humid	3.8706	-0.0099	--	--	0.39	3.3920	-0.0113	--	--	0.28
Case 4										
Arid	1.1161	0.0167	0.0122	-0.0014	0.34	0.1714	0.0089	0.0123	0.0000	0.38
Semi-arid*	1.3567	0.0032	0.0091	-0.0003	0.33	0.1955	0.0030	0.0100	0.0003	0.49
Semi-arid**	1.8118	-0.0041	0.0056	-0.0001	0.5	0.5428	-0.0021	0.0072	-0.0006	0.61
Sub-humid*	1.4607	0.0231	0.0219	0.0008	0.32	0.2354	0.0154	0.0259	0.0020	0.48
Sub-humid **	2.7372	-0.0114	0.0294	-0.0003	0.54	1.0697	-0.0064	0.0317	-0.0004	0.64
Humid	4.3430	-0.0093	0.0101	-0.0247	0.62	3.3385	-0.0115	0.0254	-0.0211	0.67

* $PPT \leq 50$ cm and ** $PPT > 50$ cm

2.5 Conclusions

The purposes of this study were to (1) quantify the impacts of soil, vegetation, and climatic characteristics on the strength and form of the $\Lambda_{SEB} - \overline{\theta}$, $\Lambda_{SEB} - \overline{s}$, $\Lambda_{PET} - \overline{\theta}$, and $\Lambda_{PET} - \overline{s}$ relationships and (2) propose a set of specific relationships that account for the most important regional characteristics. Based on this modeling study, the following conclusions can be reached:

1. The relationships that use Λ_{PET} are better predictors of $\overline{\theta}$ and $\overline{s}$ than the relationships that use Λ_{SEB}. Λ_{PET} is preferred not only when a single relationship shape is assumed to apply for all locations but also when the constants that determine the relationship shape are calibrated for each location. Λ_{PET} includes the effects of both available energy and turbulent transport (wind speed and vapor pressure deficit), while Λ_{SEB} includes only the effects of available energy. Therefore, Λ_{PET} is more directly related to ET and thus $\overline{\theta}$ and $\overline{s}$. This result suggests that the optical/thermal remote sensing methods might be improved by using Λ_{PET} instead of Λ_{SEB} if Λ_{PET} can be reliably estimated from remote sensing and/or weather station data.
2. When a single relationship is assumed to apply to all locations, the relationships are more reliable for estimating $\overline{s}$ than $\overline{\theta}$. By estimating $\overline{s}$, local variations in the saturation water content are implicitly accommodated. However, when the relationship is calibrated separately for each location, both $\overline{\theta}$ and $\overline{s}$ can be estimated with similar accuracy. In this case, the calibration accounts for the local saturation water content.
3. The $\Lambda_{SEB}-\overline{\theta}$, $\Lambda_{SEB}-\overline{s}$, $\Lambda_{PET}-\overline{\theta}$, and $\Lambda_{PET}-\overline{s}$ relationships are stronger for semiarid and sub-humid regions than for arid and humid regions. For arid and humid regions, $\overline{\theta}$ and $\overline{s}$ are more commonly below the PWP or above the FC, where Λ_{SEB} and Λ_{PET} depend less on the specific values of $\overline{\theta}$ and $\overline{s}$.
4. The strength of the $\Lambda_{PET}-\overline{\theta}$ relationship primarily depends on vegetation characteristics such as the leaf area index LAI and maximum rooting depth R_M. LAI is important

because it impacts the relative importance of soil evaporation and plant transpiration. R_M is important because it controls the soil depths where transpiration obtains water.

5. The shape of the $\Lambda_{PET} - \bar{\theta}$ relationship primarily depends on soil characteristics such as the percent clay Cl and percent silt Si. Cl and Si determine the range of soil moisture between the PWP and FC and thus affect the steepness of the $\Lambda_{PET} - \bar{\theta}$ relationship.
6. The annual precipitation PPT affects both the strength and the shape of the $\Lambda_{PET} - \bar{\theta}$ relationship. Λ_{PET} already accounts for one of the two main forcing variables for soil moisture dynamics in its denominator (potential ET), but it does not account for the other main forcing variable (precipitation). PPT helps determine how frequently the soil moisture reaches extreme values where the $\Lambda_{PET} - \bar{\theta}$ relationship weakens.
7. Four methods were proposed to estimate the forms of the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships. Case 1 uses no regional information, Case 2 uses only the climate classification, Case 3 uses the climate classification and PPT, and Case 4 uses the climate classification, PPT, Cl, Si, and LAI. The NSCE values from these cases suggest that the accuracy of the estimated $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships substantially improve as more regional characteristics are considered.

The relationships proposed in this study are based on Richards' equation for variably-saturated flow and soil characteristic curves from van Genuchten (1980) as implemented in the one dimentional physically-based model HYDRUS-1D. The proposed relationships were not tested by application in a remote-sensing method or comparisons to in-situ observations of $\bar{\theta}$ or $\bar{s}$. Such testing is a critical avenue of future research. When used in a practical context, it is

possible that errors in the estimation of Λ_{SEB} and Λ_{PET} or regional characteristics such as PPT, Cl, Si, and LAI affects the relative performance of the various methods. Such testing could also consider how heterogeneity of soil, vegetation, and climatic conditions within a region affects the performance of the methods. For example, if soil characteristics vary within a region, can separate relationships be used for different locations with adequate reliability? Future research could also repeat this analysis using other vadose-zone model structures. Although HYDRUS-1D is physically-based, the results likely depend on assumptions embedded in the model structure. Future modeling studies could include thinner soils, shallow water tables, compensation in the root-water uptake from different soil depths, more detailed descriptions of the surface energy balance, interdependence between regional characteristics, etc.

CHAPTER 3. USING REGIONAL CHARACTERISTICS TO IMPROVE UNCALIBRATED ESTIMATION OF ROOTZONE SOIL MOISTURE FROM OPTICAL/THERMAL REMOTE-SENSING

3.1 Introduction

Soil moisture is an important factor in the exchange of energy, water, and carbon between land surface and atmosphere. It helps partition incoming solar radiation into sensible and latent heat fluxes at the land surface, which in turn affects air temperature and humidity in the lower boundary layer (Eltahir, 1998; Entekhabi et al., 1996; Haghighi et al., 2018). Soil moisture controls several hydrological processes such as infiltration, runoff, and evaporation (Bronstert and Plate, 1997; Ochsner et al., 2013; Penman, 1948; Penna et al., 2011; Sadeghi et al., 2017). It provides the source of water for root-water uptake and transpiration, and maintaining soil moisture within a controlled range is critical for crop health and productivity (Çakir, 2004; Campbell and Campbell, 1982; Lyons et al., 2021; Ntukamazina et al., 2017; Osakabe et al., 2014). Soil moisture also helps regulate the terrestrial carbon cycle by impacting the stomatal control of photosynthesis and transpiration (Bastiaanssen et al., 2012; Gardner et al., 1981; Manabe and Wetherald, 1986; Melillo et al., 2002; Monteith, 1965; Pastor and Post, 1986; Scholze et al., 2016; Wu et al., 2020).

Several satellite remote-sensing methods have used active and passive microwave sensors to monitor soil moisture. These methods include the Advanced Microwave Scanning Radiometer–EOS (AMSR-E) (Njoku et al., 2003), Advanced Scatterometer (ASCAT) (Bartalis et al., 2007), Soil Moisture and Ocean Salinity (SMOS) (Kerr et al., 2001), Soil Moisture Active Passive (SMAP) (Entekhabi et al., 2010a), and the Sentinel-1 satellite from the European Space Agency (Torres et al., 2012). The main limitations of these methods include the coarse spatial resolution (~9-60 km grid cells), the shallow depth of soil moisture measurement (approximately the top 5 cm of the soil column), and lower accuracy for densely vegetated and rougher surfaces (Peng et al., 2017).

Soil moisture can alternatively be estimated from optical and thermal remote sensing data. This approach uses the visible and near infrared bands to characterize soil moisture's relationship with water stressed vegetation and uses the thermal infrared band to derive soil moisture's relationship with the thermal properties of soil (Zhang and Zhou, 2016). This approach provides much finer spatial resolutions (30 m to 1 km grid cells) and estimates the average soil moisture throughout the rootzone. However, the approach is limited to clear sky conditions because optical and thermal sensors do not penetrate clouds (Zhao and Li, 2013). Furthermore, no single physics-based model has been established for reliable calculation of soil moisture from optical and thermal data (Sadeghi et al., 2017). Optical and thermal remote sensing methods include the triangle and trapezoid (Carlson, 2007; Carlson and Petropoulos, 2019; Petropoulos et al., 2009; Sadeghi et al., 2017; Wang et al., 2011), drought index (Amani et al., 2017; Anderson et al., 2012; Feng et al., 2013; Ghulam et al., 2007; Hu et al., 2020), thermal inertia (Van doninck et al., 2011), combined passive/optical/thermal (Chauhan et al., 2003; Piles et al., 2016), single optical

(Amazirh et al., 2018; Gao et al., 2017), machine learning (Greifeneder et al., 2021), and energy balance methods (Bastiaanssen et al., 2000; Chen and Liu, 2020; Hain et al., 2009; Scott et al., 2003).

Energy balance methods use optical and thermal data to calculate all the components of the surface energy balance (SEB). Several approaches are available including: Surface Energy Balance Algorithm (SEBAL) (Bastiaanssen et al., 1998), Surface Energy Balance System (SEBS) (Su, 2002), Mapping evapotranspiration (ET) at High Resolution with Internalized Calibration (METRIC) (Allen et al., 2007a), Remote Sensing of ET (ReSET) (Elhaddad and Garcia, 2008), Operational Simplified Surface Energy Balance (SSEBop) (Senay et al., 2013), and Enhanced Two-Source Evapotranspiration Model For Land (ETEML) (Yongmin Yang et al., 2015). Once the components of the energy balance are estimated, these methods are most frequently used to estimate ET. However, one can instead calculate the evaporative fraction Λ_{SEB}, which is the ratio of the latent heat flux to the available energy (i.e., net radiation minus ground heat flux) (Shuttleworth et al., 1989). Then, an empirical logarithmic function proposed by Bastiaanssen et al. (2000) and Scott et al. (2003) is used to convert Λ_{SEB} to average rootzone soil moisture (or volumetric water content) $\bar{\theta}$. Other studies (Baier, 1969; Crago, 1996; Denmead and Shaw, 1962; Fleming et al., 2005; Sahaar and Niemann, 2020; Thornthwaite and Mather, 1955; Yao, 1974) have suggested relationships between $\bar{\theta}$ and the evaporative index Λ_{PET}, which is the ratio of the actual ET to the potential ET (PET), where potential ET is the atmospheric demand for evapotranspiration over a region (Thornthwaite, 1948; Xiang et al., 2020).

Soil moisture estimates from the empirical $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships have been tested in several regions. For example, Ahmad and Bastiaanssen (2003) used the $\Lambda_{SEB} - \bar{\theta}$ relationship to estimate rootzone soil moisture in agricultural fields in Indus River Basin in Pakistan. Rahimzadeh et al. (2013) used the $\Lambda_{SEB} - \bar{\theta}$ relationship to estimate soil moisture from MODIS data for the Canadian prairies. Hendrickx et al. (2016) used the $\Lambda_{SEB} - \bar{\theta}$ relationship to insert real-time ET fluxes into hydrologic models. Fleming et al. (2005) used the $\Lambda_{PET} - \bar{\theta}$ relationship to conduct a qualitative analysis of soil moisture in the Middle Rio Grande Basin.

Previous research has shown that the empirical $\Lambda_{SEB} - \bar{\theta}$ relationship proposed by Bastiaanssen et al. (2000) and Scott et al. (2003) is not applicable for all regions. Alburn et al. (2015) found that the relationship provides poor estimates of $\bar{\theta}$ for a semiarid region in southeastern Colorado. Calibration to in situ soil moisture observations was needed to obtain soil moisture estimates with a root mean squared error (RMSE) of 0.042 m^3/m^3. However, in situ soil moisture observations are not available for method calibration in many regions.

Several studies have shown that the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships depend on regional soil, vegetation, and climatic characteristics (Davies and Allen, 1973; Jobson, 1982; Rouse et al., 1977). For example, Kustas et al. (1993) showed that the $\Lambda_{SEB} - \bar{\theta}$ relationship strongly depends on vegetation cover in semiarid regions. Gentine et al. (2007) also showed that the $\Lambda_{SEB} - \bar{\theta}$ relationship depends on vegetation cover (specifically, the leaf area index) with thicker vegetation cover producing higher Λ_{SEB} values for a given $\bar{\theta}$. Ford et al. (2014) showed that the

$\Lambda_{SEB} - \bar{\theta}$ relationship not only depends on soil hydraulic properties such as wilting point and field capacity but also depends on net incoming radiation and meteorological conditions such as cloudiness and presence/absence of precipitation during preceding days. Basara and Crawford (2002) showed that the $\Lambda_{SEB} - \bar{\theta}$ relationship strongly depends on the total depth of the rootzone layer with shallow layers showing more nonlinear relationships. For forested areas, Dirmeyer et al. (2000) showed that Λ_{SEB} has a high sensitivity to $\bar{\theta}$ when the soil is dry and the sensitivity decreases for moderate to high soil moisture.

Sahaar and Niemann (2020) considered the impacts of local climate, soil, and vegetation characteristics on the strength and form of the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships for naturally vegetated sites. They generated more than 5000 hypothetical regions based on combinations of seventeen soil, climate, and vegetation characteristics and calculated time series of Λ_{SEB}, Λ_{PET}, and $\bar{\theta}$ with a physically based model that simulates Richard's Equation (HYDRUS-1D) (Šimůnek et al., 2008). They performed a global sensitivity analysis using Extended Fourier Amplitude Sensitivity Test (Saltelli et al., 1999) to determine the characteristics that most strongly impact the strength and form of the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships. They found that within a given climatic region (i.e., arid, semiarid, sub-humid, and humid) soil characteristics such as percent clay and percent silt are most important in determining the shape of the relationships, while vegetation characteristics such as leaf area index and maximum rooting depth have the greatest effect on the strength of the relationships. They showed that total annual precipitation, which helps determine the climatic region, also has a strong effect on both the form and strength of the relationships.

Based on those results, Sahaar and Niemann (2020) proposed four methods by which the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships can be estimated to account for the relevant regional characteristics. The methods make use of the most important regional characteristics as determined by the global sensitivity analysis. Four cases were considered to accommodate situations where different types of data are available. Case 1 uses no regional information but provides an alternative to the empirical relationship proposed by Bastiaanssen et al. (2000) and Scott et al. (2003). Case 2 uses only the climate classification to estimate the relationship forms. Case 3 uses the climate classification and annual precipitation, and Case 4 uses the climate classification, annual precipitation, soil texture, and leaf area index. The four methods were developed by maximizing the coefficient of determination r^2 between the estimated relationships and the relationships generated by the HYDRUS-1D simulations. The methods were then tested by how well they reproduced the model generated soil moisture time series. The proposed methods were not applied using remote-sensing data nor tested by comparisons to real observations of $\bar{\theta}$ for any study regions.

The main objective of this study is to evaluate the $\bar{\theta}$ estimates produced from optical/thermal remote-sensing when the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships are inferred from regional characteristics using the methods proposed by Sahaar and Niemann (2020). To achieve this goal, ReSET (Elhaddad and Garcia, 2008) is used to calculate Λ_{SEB}, and ReSET is combined with weather station data to estimate Λ_{PET}. Then, $\bar{\theta}$ is estimated from the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships using Cases 1 to 4. The methods are applied to four regions that span arid to humid

climates and have extensive in situ soil moisture observations. The $\bar{\theta}$ estimates are evaluated by comparing to in situ soil moisture observations and $\bar{\theta}$ estimates from the empirical relationship of Bastiaanssen et al. (2000) and Scott et al. (2003). The performance of each method is also evaluated by separately analyzing the reproduction of observed spatial and temporal variations of $\bar{\theta}$.

3.2 Methodology

3.2.1 ReSET

Detailed presentations of ReSET are provided by (Eldeiry et al., 2016; Elhaddad et al., 2011; Elhaddad and Garcia, 2011, 2008), so only a summary is included here. ReSET uses the same approach as SEBAL and METRIC, but it interpolates between multiple weather stations to account for spatial variations within the region. The components of the energy balance are estimated using optical/thermal satellite imagery, a digital elevation model (DEM), and weather station data. The energy balance can be written:

$$L_v E_a = R_n - G - H \tag{35}$$

where L_v is the latent heat of vaporization of water (J kg^{-1}), E_a is the actual ET (kg m^{-2} s^{-1}), R_n is the net radiation (W m^{-2}), G is the ground heat flux at the soil surface (W m^{-2}), and H is the sensible heat flux (W m^{-2}). For each pixel in an image, the components of the energy balance are calculated using the visible and thermal bands. For example, using Landsat 7 imagery, the albedo is derived from the visible bands (1, 2, and 3) and infrared bands (4, 5, and 7). The normalized vegetation difference index (NDVI) is calculated from the visible band 3 and the near-infrared band 4. The surface temperature is derived from the thermal infrared band (band 6). Surface

characteristics from a digital elevation model are combined with the imagery to calculate R_n. G is calculated empirically based on albedo, NDVI, and surface temperature using the Bastiaanssen et al. (2000) method. To calculate the latent heat flux $L_v E_a$, two points in the region are selected to represent the extremes of the thermal continuum. These points are usually referred to as the hot and cold (or dry and wet) pixels. At the cold pixel, ET is expected to occur at a rate that fully meets atmospheric demand. At the hot pixel, the ET is assumed to be zero. The sensible heat is calculated using the Monin-Obhukov (Monin and Obukhov, 1954) functions by iterating the air density and aerodynamic heat transport resistance. For more information on estimating E_a from optical/thermal satellite imagery see Bastiaanssen et al. (1998), Bastiaanssen et al. (2000), Allen et al. (2007), Tasumi et al. (2005), and Elhaddad and Garcia (2011).

Weather station data are also required for the energy balance calculations. Required data include daily wind run as well as hourly and daily PET E_p. E_p is estimated as reference crop ET (Xiang et al., 2020) using the ASCE standardized equation (Giteau, 1965), which is a version of the Penman-Monteith equation (Allen et al., 1998; Itenfisu et al., 2003), assuming a short grass reference crop. A crop coefficient is not included in the estimation of E_p because the crop coefficient is difficult to estimate for natural vegetation without in situ observations (see, for example, the wide range of crop coefficients produced by different methods for similar conditions in Corbari et al. (2017), Liu et al., (2017), and Wight and Hanson (1990)). To calculate E_p, hourly air temperature, relative humidity, wind speed, and solar radiation from weather stations within the image domain are used. Hourly and daily E_p as well as daily wind run are estimated first at the weather stations, then ordinary kriging is performed to interpolate

between the stations (Elhaddad and Garcia, 2008). ReSET can be used in calibrated or uncalibrated forms. The uncalibrated form is similar to SEBAL except ReSET interpolates wind run to better capture spatial wind variability in the area. The calibrated form is similar to METRIC, but it also interpolates hourly and daily ET as well as the daily wind run data to better estimate H. The calibrated form of RESET model provides better estimates of the energy balance components unless the weather stations are far apart or elevation varies significantly within a region (Elhaddad and Garcia, 2011).

3.2.2 Proposed Equations

Once the components of the energy balance are determined from ReSET, Λ_{SEB} is calculated as:

$$\Lambda_{SEB} = \frac{L_v E_a}{R_n - G} \tag{36}$$

and Λ_{PET} is calculated:

$$\Lambda_{PET} = \frac{E_a}{E_p} \tag{37}$$

The difference between Λ_{SEB} and Λ_{PET} is that Λ_{SEB} only considers the available energy in its denominator (i.e., $R_n - G$). ET depends on not only the available energy but also the turbulent transport (wind speed and vapor pressure deficit), and Λ_{PET} accounts for all these components by using E_p in its denominator. The accuracies of Λ_{SEB} and Λ_{PET} have been evaluated for several remote sensing methods that are similar to ReSET (Allen et al., 2007b; Bastiaanssen, 1998; Bastiaanssen et al., 1997; Chávez et al., 2008; Roerink et al., 2000; Zhu et al., 2021). For example, Bastiaanssen et al. (1998) reported RMSE values between 0.10 and 0.20. Finally, $\bar{\theta}$ is calculated using either:

$$\bar{\theta} = \exp\left(\frac{\Lambda_{SEB} - a}{b}\right) \tag{38}$$

or:

$$\bar{\theta} = \exp\left(\frac{\Lambda_{PET} - e}{f}\right) \tag{39}$$

where the parameters a, b, e, and f determine the relationship shapes. Bastiaanssen et al. (2000) and Scott et al. (2003) estimated $a = 1.284$ and $b = 0.421$ based on comparisons to in situ soil moisture observations from experimental fields in Kansas and Spain (they did not estimate e or f). It is worth noting that there are other factors, aside from soil moisture, that may affect the Λ_{SEB} and Λ_{PET} that are not considered in this method. These include soil salinity (through osmotic potential) as well as vegetation nutrition and diseases.

Sahaar and Niemann (2020) proposed four methods to estimate a, b, e, and f based on their sensitivity analysis of HYDRUS-1D simulations. Case 1 assumes that no regional information is available, so the relationship parameters are merely numbers, which can be compared to the values proposed by Bastiaanssen et al. (2000) and Scott et al. (2003). Case 2 assumes the climate classification is known, so different values are available for arid, semiarid, sub-humid, and humid regions (Table 6). Case 3 assumes the climate classification and annual precipitation for the preceding year are known, so the Λ_{SEB} vs. $\bar{\theta}$ relationship constants are estimated:

$$a = A_1 + A_2 P \tag{40}$$

$$b = B_1 + B_2 P \tag{41}$$

where P (cm) is the annual precipitation and A_1 (-), A_2 (cm^{-1}), B_1 (-), and B_2 (cm^{-1}) are regression constants that vary by climatic region (Table 6). Similarly, the $\Lambda_{PET} - \bar{\theta}$ relationship constants can be estimated:

$$e = E_1 + E_2 P \tag{42}$$

$$f = F_1 + F_2 P \tag{43}$$

where E_1 (-), E_2 (cm^{-1}), F_1 (-), and F_2 (cm^{-1}) are regression constants that vary by climatic region (Table 7). Including precipitation in the estimation method partially accounts for transitions between climatic regions. Precipitation controls the depth that water reaches in the soil profile. If the water is concentrated near the surface, soil evaporation is more important. If water reaches deeper, then transpiration dominates. The $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships differ for these two processes (Sahaar and Niemann, 2020).

For Case 4, the climate classification, annual precipitation, soil texture, and leaf area index are all used. These variables were determined to impact most strongly the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships by Sahaar and Niemann (2020). The $\Lambda_{SEB} - \bar{\theta}$ relationship constants are estimated:

$$a = A_1 + A_2 P + A_3 C_l + A_4 S_i + A_5 L_{AI} \tag{44}$$

$$b = B_1 + B_2 P + B_3 C_l + B_4 S_i + B_5 L_{AI} \tag{45}$$

where C_l is the percent clay, S_i the percent sand, L_{AI} is the leaf area index, and A_3 (-), A_4 (-), A_5 (-), B_3 (-), B_4 (-), and B_5 (-) are regional regression constants (Table 6). Similarly:

$$e = E_1 + E_2 P + E_3 C_l + E_4 S_i \tag{46}$$

$$f = F_1 + F_2 P + F_3 C_l + F_4 S_i \tag{47}$$

where E_3 (-), E_4 (-), F_3 (-), and F_4 (-) are regional regression constants (Table 7). Sahaar and Niemann (2020) showed that increasing C_l or decreasing S_i reduces the range between the permanent wilting point and field capacity, which affects the form of the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships. L_{AI} affects the relationships because it helps determine whether soil evaporation or transpiration dominates the ET. Soil evaporation and transpiration have different $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships. See Sahaar and Niemann (2020) for additional information about the method development.

3.3 Study Regions

The four study regions are: (1) the Walnut Gulch Experimental Watershed in Arizona, (2) the Piñon Canyon Maneuver Site and Lower Arkansas River Valley in Colorado, (3) the Little Washita and Fort Cobb Experimental Watersheds in Oklahoma, and (4) the Mississippi Delta region in Mississippi (Figure 9). The study regions focus on naturally vegetated conditions because Sahaar and Niemann (2020) developed the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships based on such conditions. None of the selected study regions were considered by Sahaar and Niemann (2020). For simplicity, they are referred to as the Arizona, Colorado, Oklahoma, and Mississippi regions. The climate classification for each region is determined using the UNEP (1997) system. That system is based on the aridity index (AI), which is the ratio of the average annual precipitation and PET. An arid region has an AI below 0.20, a semiarid region has an AI from 0.20 to 0.50, a sub-humid region has an AI from 0.50 to 0.65, and a humid region has an AI above 0.65.

Table 6. Proposed regression constants for estimating a and b in the $\Lambda_{SEB} - \overline{\theta}$ relationship

Climatic Region	Constant *a*					Constant *b*				
	A_1	A_2	A_3	A_4	A_5	B_1	B_2	B_3	B_4	B_5
Case 1										
All Climates	1.4844	--	--	--	--	0.5222	--	--	--	--
Case 2										
Arid	1.3884	--	--	--	--	0.3932	--	--	--	--
Semiarid	1.4873	--	--	--	--	0.5158	--	--	--	--
Sub-humid	1.4814	--	--	--	--	0.5286	--	--	--	--
Humid	1.5517	--	--	--	--	0.6612	--	--	--	--
Case 3										
Arid	1.3669	0.0057	--	--	--	0.4160	0.0045	--	--	--
Semiarid*	1.3709	0.0024	--	--	--	0.3968	0.0011	--	--	--
Semiarid**	1.5634	-0.0021	--	--	--	0.5128	-0.0014	--	--	--
Sub-humid*	1.3967	0.0025	--	--	--	0.4803	0.0013	--	--	--
Sub-humid **	1.3545	0.0019	--	--	--	0.4665	0.0009	--	--	--
Humid	3.4866	-0.0082	--	--	--	2.9917	-0.0096	--	--	--
Case 4										
Arid	1.4457	0.0084	0.0042	-0.0031	-0.059	0.3195	0.0061	0.0073	-0.0012	-0.0369
Semiarid*	1.2327	0.0065	0.006	-0.0023	-0.0542	0.1086	0.0046	0.0085	-0.0011	-0.02
Semiarid**	1.7498	-0.0026	0.0017	-0.0032	-0.0321	0.5127	-0.0017	0.0047	-0.0019	-0.0055
Sub-humid*	1.7462	0.0054	0.0061	-0.0051	-0.08	0.5550	0.0037	0.0109	-0.0028	-0.0417
Sub-humid **	1.7578	0.0013	0.0041	-0.0055	-0.0281	0.4651	0.0025	0.0095	-0.0032	-0.0071
Humid	5.6182	-0.0181	0.0286	-0.0309	-0.0651	4.4269	-0.0197	0.0452	-0.0286	-0.0249

* $P \leq 50$ cm and ** $P > 50$ cm

Table 7. Proposed regression constants for estimating e and f in the $\Lambda_{PET} - \bar{\theta}$ relationship.

Climatic Region	Constant *e*				Constant *f*			
	E_1	E_2	E_3	E_4	F_1	F_2	F_3	F_4
Case 1								
All Climates	1.8597	--	--	--	0.7423	--	--	--
Case 2								
Arid	1.6292	--	--	--	0.5314	--	--	--
Semiarid	1.6895	--	--	--	0.5953	--	--	--
Sub-humid	2.0299	--	--	--	0.8893	--	--	--
Humid	3.0385	--	--	--	1.8528	--	--	--
Case 3								
Arid	1.4484	0.0102	--	--	0.4809	0.0041	--	--
Semiarid*	1.6180	0.0007	--	--	0.5102	0.0008	--	--
Semiarid**	1.8433	-0.0031	--	--	0.6440	-0.0015	--	--
Sub-humid*	1.7358	0.0101	--	--	0.7179	0.0054	--	--
Sub-humid **	2.3901	-0.0048	--	--	1.0573	-0.0021	--	--
Humid	3.8706	-0.0099	--	--	3.3920	-0.0113	--	--
Case 4								
Arid	1.1161	0.0167	0.0122	-0.0014	0.1714	0.0089	0.0123	0.0000
Semiarid*	1.3567	0.0032	0.0091	-0.0003	0.1955	0.0030	0.0100	0.0003
Semiarid**	1.8118	-0.0041	0.0056	-0.0001	0.5428	-0.0021	0.0072	-0.0006
Sub-humid*	1.4607	0.0231	0.0219	0.0008	0.2354	0.0154	0.0259	0.0020
Sub-humid **	2.7372	-0.0114	0.0294	-0.0003	1.0697	-0.0064	0.0317	-0.0004
Humid	4.3430	-0.0093	0.0101	-0.0247	3.3385	-0.0115	0.0254	-0.0211

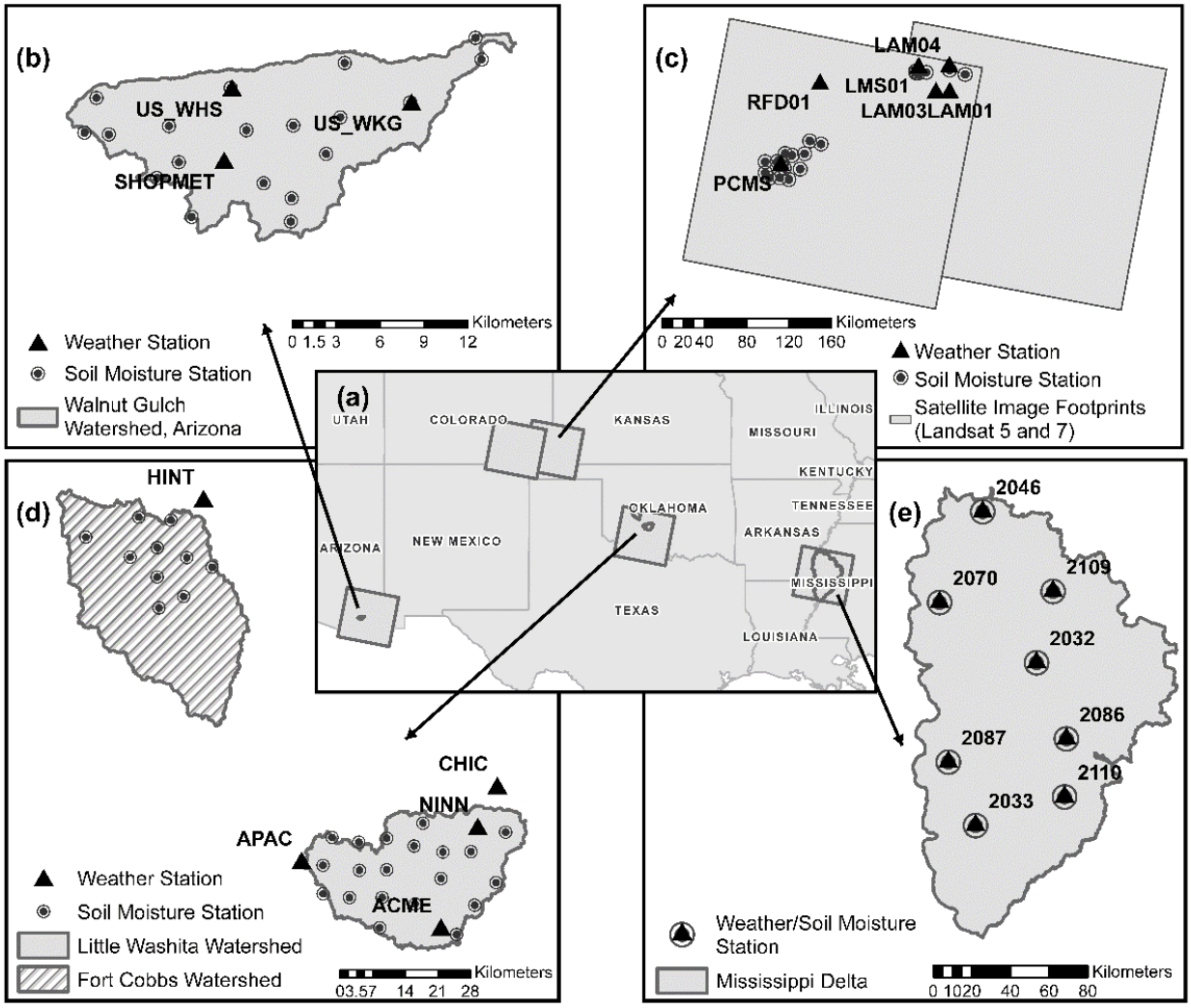

Figure 9. (a) Locations of the four study regions in the U.S. with approximate Landsat image footprints and watershed boundaries. Locations of weather and soil moisture stations in the (b) Arizona, (c) Colorado, (d) Oklahoma, and (e) Mississippi study regions.

3.3.1 Arizona

The Walnut Gulch Experimental Watershed is in the San Pedro Valley of southeastern Arizona (Figure 9b). The region is classified as arid with an approximate AI of 0.13 using the UNEP (1997) classification (Trabucco and Zomer, 2019). Based on the National Land Cover Database

(NLCD) 2011, the typical vegetation cover includes desert shrubs at lower elevations and desert grasses at upper elevations (Table 8) (King et al., 2008).

In situ soil moisture data have been collected at 5 cm depth using Stevens Hydra Probes (Stevens et al., 2008; Keefer et al., 2008) and are available from the USDA-Agricultural Research Service (USDA-ARS) (https://www.tucson.ars.ag.gov/dap/). The data are available at 19 locations, and each location falls within a different Landsat pixel (Figure 9b). Cox et al. (1986) assessed the root distribution in shrublands and grasslands in Arizona and found that approximately 70% of the total fine root biomass is within 15 cm of the surface. Kurc and Small (2004) considered shrublands and grasslands in New Mexico and showed that the evaporative fraction is strongly correlated with 0-5 cm soil moisture and poorly correlated with deeper soil moisture because infiltrated water rarely reaches deeper layers. Thus, the 5 cm soil moisture measurements are expected to describe the most active portion of the rootzone. Soil moisture observations were used for nine dates between February 18, 2015 and October 18, 2016 that align with cloud-free Landsat images. Soil moisture data are available at a 30 min resolution and were interpolated to match the Landsat acquisition time of 10:52 a.m. (local time). A total of 171 in situ soil moisture measurements (ranging from 0.01 to 0.30) were used to evaluate the remote-sensing estimates (Table 8).

The soil moisture estimates from remote sensing were developed using Landsat 8 images acquired from the USGS Earth Explorer website (Table A.2). The 30 m Shuttle Radar Topography Mission (SRTM) DEM was used for all analysis regions (available from USGS Earth Explorer website). Hourly E_p was estimated as the ASCE standardized reference crop ET

without use of a crop coefficient (consistent with ReSET). The required inputs (air temperature, relative humidity, wind speed, and solar radiation) were determined from three AmeriFlux weather stations (Figure 9b and Table A.1). The resulting E_p values were spatially interpolated throughout the analysis region using ordinary kriging method. P was also spatially interpolated with ordinary kriging using 19 rain gauges that are collocated with the soil moisture probes (Fig 1b). The soil texture (specifically C_l and S_i) at each monitoring site was obtained from the Soil Survey Geographic (SSURGO) database of the Natural Resource Conservation Service (NRCS, 2018). The selected texture from SSURGO corresponds to the depth of the soil moisture sensor (5 cm). Sandy loam is abundant at the sensor locations (Table 8), but the soils include an abundance of gravel, which deviates from the soil types originally used to estimate the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships (Sahaar and Niemann, 2020). L_{AI} was obtained from ReSET for all regions. ReSET estimates L_{AI} using the method proposed by Bastiaanssen et al. (2012), which is based on the NDVI estimates.

Table 8. Typical soil characteristics, land cover, and soil moisture ranges at the soil moisture monitoring sites in each study region

Study Region	**Typical Soil Types**	**Typical Land Cover**	**Soil Moisture Sensor Depths (cm)**	**Soil Moisture Range (cm^3 cm^{-3})**
Arizona	Very gravelly sandy loam	Shrub, scrub	5	0.01-0.30
Colorado	Loam and silt loam	Grassland, herbaceous	0-10	0.02-0.30
Oklahoma	Sandy loam and silt loam	Grassland, herbaceous	5, 25, 45	0.01-0.37
Mississippi	Silt loam and silt clay	Grassland, cropland	5, 10, 20, 50, 100	0.12 -0.62

3.3.2 Colorado

The Piñon Canyon Maneuver Site (PCMS) constitutes a large area of non-grazed and relatively undisturbed grasslands (Figure 9c). The Lower Arkansas River Valley (LARV) is an alluvial valley that includes both grasslands and irrigated farmland. The overall region is semiarid with an AI of 0.29 (Burkhalter and Gates, 2005). Based on NLCD 2011, the typical vegetation at the soil moisture monitoring sites includes shallow-root prairie grasses and herbaceous vegetation (Homer et al., 2015). The active root layer is typically in the top 50 cm of the soil (Coupland and Johnson, 1965; West, 1983).

In situ soil moisture observations were obtained by combining three datasets. The first dataset (CSU-PCMS) consists of 3 sites within PCMS with hourly measurements at 0–10 cm depth. The data were collected using time-domain reflectometry (TDR) probes that were installed perpendicular to the soil surface (Melliger and Niemann, 2010). The second dataset (USGS-PCMS) was obtained from the USGS and has 13 sites within PCMS. It includes hourly measurements from TDR probes that were buried horizontally at depths ranging from 3.8 to 10.2 cm. Although both the CSU-PCMS and USGS-PCMS datasets describe the soil moisture at relatively shallow depths, these observations are assumed to characterize the entire rootzone due to limited data availability. The third dataset (CSU-LARV) includes 12 non-cultivated fields in the LARV. In each field, the average soil moisture from two layers (0-11 cm and 25-36 cm) was used as the rootzone soil moisture. Overall, in situ soil moisture data are available for 28 Landsat grid cells in the Colorado region. Soil moisture measurements were used for 24 cloud-free Landsat dates between June 2009 and August 2012. The in situ soil moisture values range from

0.02 to 0.30 (Table 8). In total, 281 observations were used to evaluate the remote-sensing estimates.

To produce the remote-sensing estimates of soil moisture, 24 Landsat 5 and 7 images were obtained from the USGS Earth Explorer website (Table A.2). Because the LARV region lies in the overlap between the 32/34 and 31/34 Landsat image footprints, satellite images could be accessed every 8 days instead of every 16 days as for other sites. E_p and P were calculated similar to the Arizona region using 6 weather stations located in the LARV and PCMS regions (Figure 9c and Table A.1). Five of those stations are part of the Colorado Agricultural Meteorological Network (CoAgMet) (Andales et al., 2009), and 1 station at PCMS was operated by Colorado State University (Melliger and Niemann, 2010). C_l and S_i were obtained from SSURGO at the soil moisture measurement depth for each site (0-10 cm for PCMS and USGS fields, and 0 – 36 cm for LARV fields). Silt loam and loam are most common soil types for this region (Table 8), but some sites are sandy loam and clay loam.

3.3.3 Oklahoma

The Little Washita and Fort Cobb experimental watersheds are in southwest Oklahoma (Figure 9d). The climate is sub-humid with an AI of 0.64 (Steiner et al., 2008; Starks et al., 2014a). The main vegetation covers in Little Washita are grasslands and rangelands, and for Fort Cobb, the dominant vegetation covers are cropland and rangeland (P. J. Starks et al., 2014; Patrick J. Starks et al., 2014).

USDA-ARS operates an extensive network of soil moisture stations referred to as ARS Micronet (http:/ars.mesonet.org/). In situ soil moisture measurements are available at 20 Little Washita and 10 Fort Cobb sites (Figure 9d and Table A.1). The sites are often near county roads, which might introduce errors in the remote-sensing estimates of soil moisture due to heterogeneity in reflectance within the Landsat pixels. The in situ measurements are available every 30 min from Hydra Probes at three depths (5, 25, and 45 cm) (Table 8). The average rooting depth ranges from 1.5 to 1.8 m in sub-humid regions (Schenk and Jackson, 2002), so the average from all three measurement depths is used to represent the rootzone layer. Nine cloud-free Landsat dates were selected for analysis between April 21, 2009 and October 1, 2010. The monitoring period includes rainfall events of 26 mm to 40 mm in the days before three Landsat images, and the in situ soil moisture values range from 0.01 to 0.37 (Table 8). Overall, the in situ soil moisture dataset contains 270 soil moisture observations.

The remote-sensing estimates of soil moisture were generated using Landsat 8 images from the USGS Earth Explorer website (Table A.2). E_p and P were determined from hourly data at 5 Mesonet weather stations (http://www.mesonet.org and McPherson et al., 2007). C_l and S_i in the top 50 cm for each soil moisture monitoring site were obtained from USDA-ARS (Moriasi et al., 2014). The soil texture varies from sand to clay (Cosh et al., 2006), but sandy loam and silt loam are most abundant in the Oklahoma region (Table 8).

3.3.4 Mississippi

The Mississippi Delta region is bounded to the west by the Mississippi River and the east by the Yazoo River in northwestern Mississippi (Zachos et al., 2016). The climate is humid with an AI

ranging from 1.1 in the north to 1.5 in the south (Mostovoy and Anantharaj, 2008; Bengtson et al., 1995). The typical land cover at the soil moisture monitoring sites is natural grasslands (Table 8), but some Landsat pixels that contain the monitoring sites also contain croplands. The crops in this area include soybeans, cotton, corn, and rice (Kebede et al., 2014).

The Mississippi Delta has 8 Soil Climate Analysis Network (SCAN) sites that provide soil moisture observations from Hydra Probes at five depths from 5 cm to 100 cm (Table 8, Figure 9e) (Keefer et al., 2008). The data are available from the USDA-Natural Resources Conservation Service (https:/www.wcc.nrcs.usda.gov/scan/). Due to the variation in cultivated crops in this region, the rootzone is assumed to be 100 cm, and the average soil moisture from all five soil moisture sensors (Table 8) is used to represent the rootzone soil moisture. Eleven clear Landsat dates were selected for analysis between April 2016 and October 2017. The soil moisture ranges from 0.12 to 0.62. Overall, 88 location/date combinations are used in this region.

The remote sensing estimates of soil moisture were generated using Landsat 8 images (Table A.2). E_p and P values were determined from the SCAN dataset (Schaefer et al., 2007) (Figure 9e). C_l and S_i at each probe location and depth were also obtained from the SCAN dataset. The C_l and S_i data were each depth averaged to match the average of the rootzone layer. Silt loam and silt clay are the dominant soil types at the SCAN sites (Table 8).

3.4 Results

Figure 10 shows example results from the remote sensing methods for a small sub-region in the LARV in Colorado on August 07, 2009. In the false color composite image (Figure 10a), areas

with denser vegetation (such as center-pivot irrigation fields) are bright red, while grasslands and barren lands are gray. The agricultural fields in this region are irrigated to ensure crop productivity, so they are expected to have higher Λ_{SEB} and Λ_{PET} values and higher $\overline{\theta}$ values than the surrounding areas. The associated Λ_{SEB} and Λ_{PET} images are shown in Figure 10b and Figure 10c, respectively. Both Λ_{SEB} and Λ_{PET} can vary from 0 to 1, but Λ_{SEB} typically has higher values than Λ_{PET} in this sub-region. Λ_{SEB} divides the latent heat flux by the available energy (Eq. 2), which produces Λ_{SEB} values near one in the cultivated areas. In contrast, Λ_{PET} divides the actual ET (which is very similar to the latent heat flux) by E_p (Eq. 3). E_p considers not only the available energy but also the turbulent transport (i.e., wind speed and vapor pressure deficit). The low humidity in this semiarid region increases E_p, which reduces Λ_{PET}. Both Λ_{SEB} and Λ_{PET} display complex spatial variation within the sub-region, but Λ_{PET} exhibits more variability within the center pivot fields and is more similar to the false-color image in Figure 10a.

The soil moisture estimated using the empirical $\Lambda_{SEB} - \overline{\theta}$ relationship from Bastiaanssen et al. (2000) is shown in Figure 10d. Because the Λ_{SEB} values are near one in the cultivated areas of the sub-region, $\overline{\theta}$ is also large and often exceeds realistic porosity values for this region (e.g., 0.45). If the Bastiaanssen et al. (2000) relationship is used instead with Λ_{PET} (Figure 12e), then the $\overline{\theta}$ estimates are lower and more realistic. Nonetheless, $\overline{\theta}$ still has unrealistically high values for much of the cultivated area. Figure 10f and Figure 10g show the $\overline{\theta}$ estimates if Case 4 is used (instead of the Bastiaanssen et al. (2000) relationship) with Λ_{SEB} and Λ_{PET}, respectively. The

$\Lambda_{SEB} - \bar{\theta}$ relationship from Case 4 reduces the $\bar{\theta}$ estimates throughout the sub-region, and the $\Lambda_{PET} - \bar{\theta}$ relationship from Case 4 produces the lowest $\bar{\theta}$ estimates among all cases shown in the figure. Overall, Case 4 produces fewer supersaturated locations, and thus appears to generate more realistic $\bar{\theta}$ results for this sub-region. Because Case 4 includes spatial variations in soil texture within the region, it also shows a clear boundary in the $\bar{\theta}$ values in the southern part of the figure. This boundary is associated with a change in soil texture at the edge of the alluvial valley.

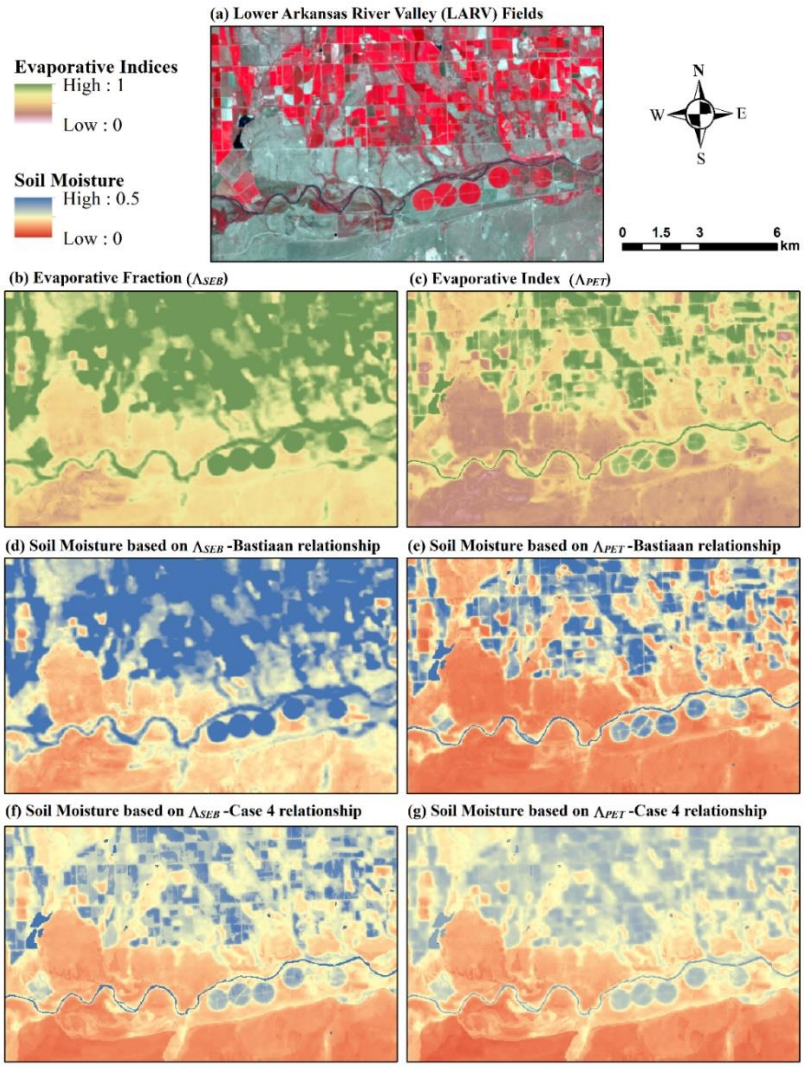

Figure 10. Example sub-region of the Lower Arkansas River Valley (LARV) within the Colorado study region including: (a) false color Landsat image, (b) evaporative fraction Λ_{SEB}, (c) evaporative index Λ_{PET}, and rootzone soil moisture $\overline{\theta}$ from (d) Λ_{SEB} and Bastiaanssen et al. (2000), (e) Λ_{PET} and Bastiaanssen et al. (2000), (f) Λ_{SEB} and Case 4, and (g) Λ_{PET} and Case 4.

Figure 11 shows the RMSE of the $\overline{\theta}$ estimates for all four study regions when $\overline{\theta}$ is estimated using Λ_{SEB} and the Bastiaanssen et al. (2000) relationship or Cases 1-4. All the methods perform better for the arid (Arizona) and semiarid (Colorado) regions than for the sub-humid (Oklahoma) and humid (Mississippi) regions. One reason could be that Arizona and Colorado have moisture-

limited ET, while Oklahoma and Mississippi have energy-limited ET. Several studies have suggested three distinct regimes for the relationship between soil moisture and Λ_{SEB} or Λ_{PET} (Akbar et al., 2018; Gentine et al., 2007; Gopalakrishnan and Biegler, 2013; Niemann and Eltahir, 2004; Rodriguez-Iturbe et al., 1999; Seneviratne et al., 2010). Below the permanent wilting point θ_{PWP}, little to no ET occurs, so Λ_{SEB} and Λ_{PET} are independent of soil moisture. When the soil moisture is above θ_{PWP} but below the field capacity θ_{FC}, soil moisture strongly affects Λ_{SEB} and Λ_{PET} because the ET is moisture limited (Seneviratne et al., 2006). Above θ_{FC}, Λ_{SEB} and Λ_{PET} are independent of soil moisture because ET is energy limited (Seneviratne et al., 2006). θ_{FC} was calculated from the soil texture data at each soil moisture monitoring sites using the pedotransfer functions from Saxton and Rawls, (2006). $\bar{\theta}$ briefly exceeds field capacity at only 2 of 19 sites in Arizona and 2 of 28 sites in Colorado due to extended rain-free periods. However, soil moisture frequently exceeds field capacity for 19 of 30 sites in Oklahoma and all sites in Mississippi. Another reason for the poorer performance in Oklahoma could be the proximity of some soil moisture stations to county roads. Similarly, for the Mississippi region, some soil moisture monitoring sites are located within mixed vegetation pixels (the grasslands that include the soil moisture monitoring site also includes the cropland within the same Landsat pixel). The mixed vegetation produces heterogeneous surface temperature and vegetation reflectance within the Landsat pixels, which might cause the remote sensing results to deviate from the conditions at the measurement locations. Also, in Oklahoma and Mississippi, vegetation rooting depth is more important because plants in these environments have access to deeper soil moisture levels than in arid and semiarid climates (Schenk and Jackson, 2002).

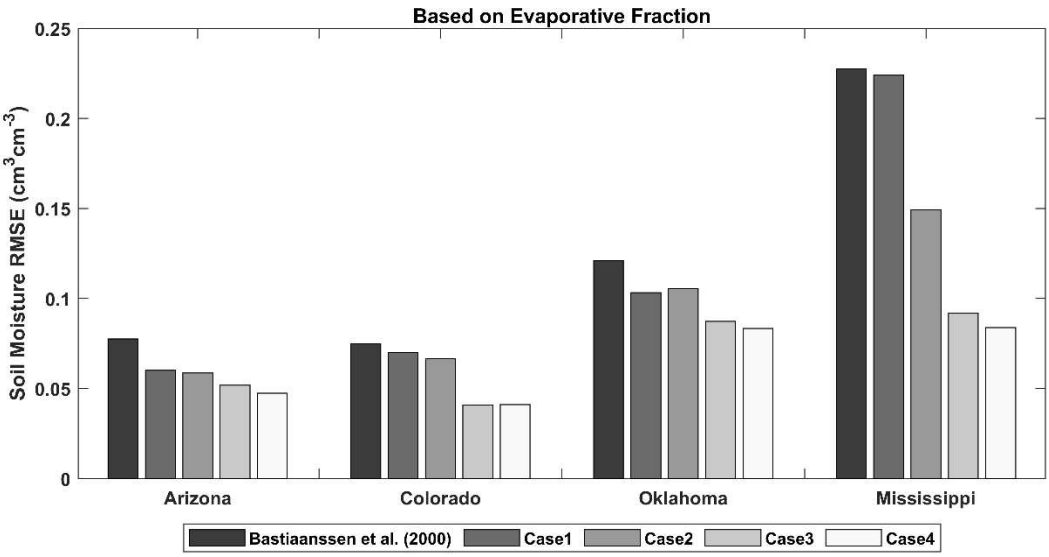

Figure 11. Root mean squared error (RMSE) of rootzone soil moisture $\bar{\theta}$ when $\bar{\theta}$ is estimated from evaporative fraction Λ_{SEB} using the Bastiaanssen et al. (2000) relationship or Cases 1-4 for the four study regions

Figure 12 shows the RMSE's of the $\bar{\theta}$ estimates when Λ_{PET} is used. Comparing Figure 11 and Figure 12 suggests using Λ_{PET} in place of Λ_{SEB} typically reduces the RMSE's of the $\bar{\theta}$ estimates. This improvement is also seen for the empirical relationship, even though Bastiaanssen et al. (2000) did not suggest using the empirical relationship with Λ_{PET}. As previously mentioned, Λ_{PET} considers both the available energy and the turbulent transport, so it is expected to provide a better normalization than Λ_{SEB}. The relative performance of the various cases in Figure 12 is similar to that in Figure 11. The $\bar{\theta}$ estimates from all four cases (except Case 1 in Colorado) have lower RMSE's than the estimates from the empirical relationship. Furthermore, as one includes more information in the estimation of the relationship (i.e., one moves from Case 1 to Case 4), the $\bar{\theta}$ estimates typically improve. The improvement in

performance for the Mississippi region is especially noteworthy. When the empirical $\Lambda_{PET} - \bar{\theta}$ relationship or Case 1 is used, the RMSE exceeds 0.20 $cm^3\ cm^{-3}$. Thus, the $\bar{\theta}$ estimates from these approaches likely have too much error for any practical use. However, when Case 4 is used, the RMSE drops below 0.10 $cm^3\ cm^{-3}$. While this error is still substantial, Case 4 can provide rough estimates of $\bar{\theta}$ for this humid region. The reasons for the improvement in the Mississippi region are discussed later.

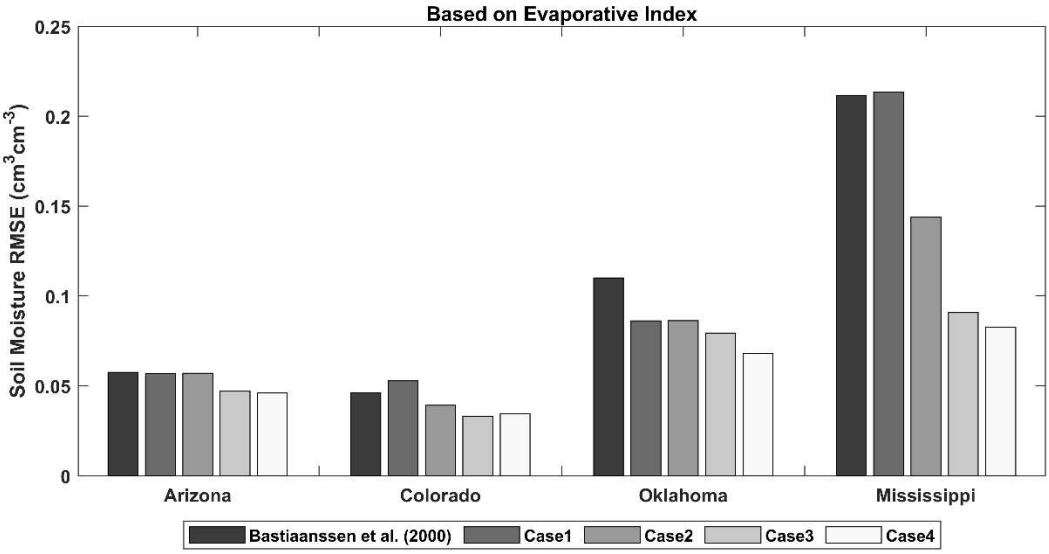

Figure 12. Root mean squared error (RMSE) of rootzone soil moisture $\bar{\theta}$ when $\bar{\theta}$ is estimated from evaporative index Λ_{PET} using the Bastiaanssen et al. (2000) relationship or Cases 1-4 for the four study regions

To test whether Cases 1-4 improve the reproduction of spatial variations in $\bar{\theta}$, the RMSE was calculated separately for each date using all monitoring locations within each region. Then, the RMSE values for all dates were characterized using a box-and-whisker plot (Figure 13). For Arizona and Oklahoma, all cases have relatively similar spatial RMSE values. However, the

median spatial RMSE usually decreases when Λ_{SEB} is replaced with Λ_{PET} or when more regional information is included (i.e., as one shifts from the empirical relationship to Case 4). Furthermore, the range of the spatial RMSE values decreases as more information is included, which indicates more consistent reproduction of spatial variability. For Colorado, the reproduction of spatial variability improves most (among the regions) when Λ_{SEB} is replaced with Λ_{PET}, but this improvement reduces as one moves from the empirical relationship to Case 4. The use of Λ_{PET} likely improves the soil moisture estimates in the Colorado region due to the importance of turbulent transport in this region as discussed earlier. For Mississippi, using Λ_{PET} provides only a small improvement in performance over Λ_{SEB}. The much larger improvement occurs as one includes more regional information. This large improvement likely occurs for two reasons. First, the annual precipitation varies substantially within this region, which is considered in Cases 3-4. Second, soil texture was obtained from the probe's actual position (not SSURGO), so spatial variations in soil texture and vegetation are likely well represented in Case 4.

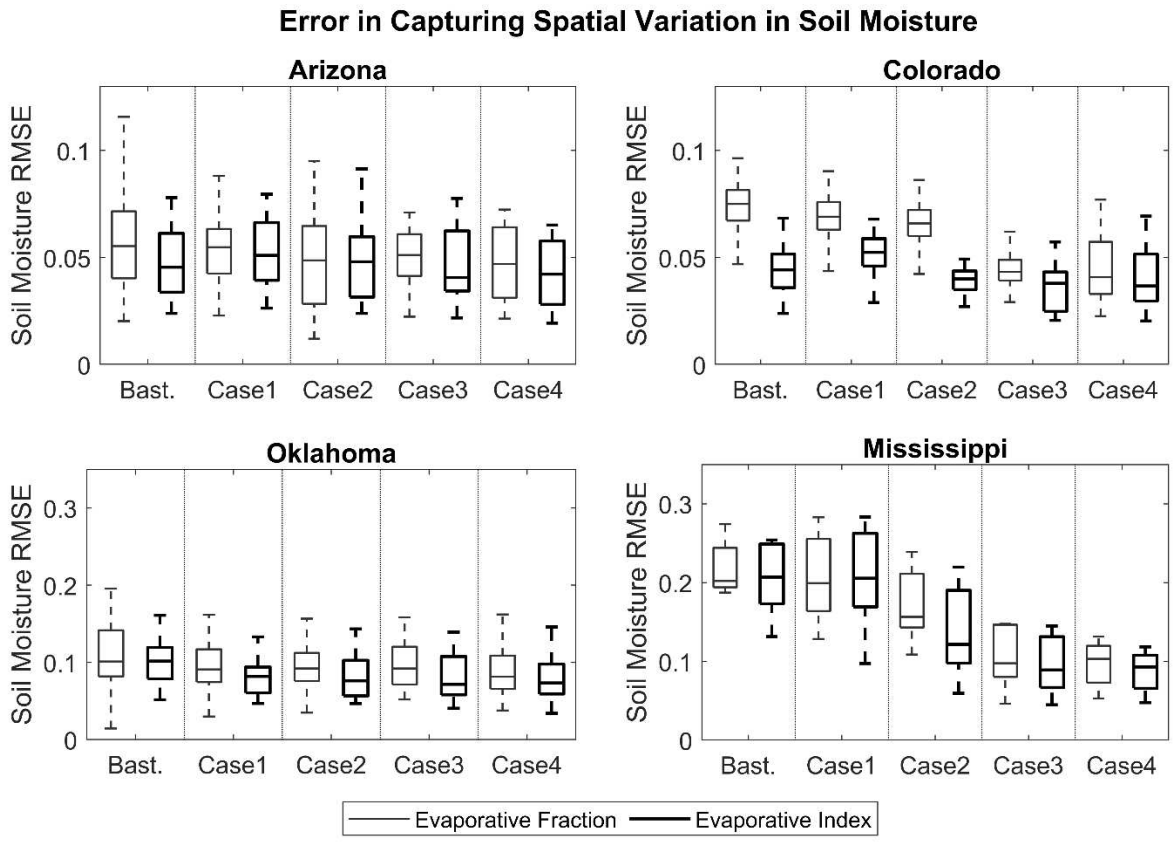

Figure 13. Root mean squared error (RMSE) of rootzone soil moisture when calculated using all locations on a given date. The horizontal line inside each box marks the median RMSE, the upper and lower edges of each box correspond to upper and lower quartiles, and the upper and lower limits of the whiskers indicate maximum and minimum values among all dates. The label "Bast." indicates the use of the Bastiaanssen et al. (2000) relationship.

To test how well the methods reproduce temporal variations in $\overline{\theta}$, the RMSE was calculated separately for each monitoring site using all available dates. Then, the RMSE values for all sites were characterized using a box-and-whisker plot for each case and region (Figure 14). Overall, the patterns observed in Figure 14 resemble those in Figure 13. Thus, better reproduction of spatial variability in $\overline{\theta}$ tends to be associated with better reproduction of temporal variability. Yet, two differences are apparent between Figs. 5 and 6. For Colorado, the median spatial RMSE (Figure 13) is consistently higher than the median temporal RMSE (Figure 14) for any given

case. Thus, for this region, the methods have more error determining where it is wet than when it is wet. The Colorado region has diverse Λ_{SEB} and Λ_{PET} values in close proximity due to the presence of irrigated fields within the semiarid rangeland. That heterogeneity likely induces advection between cells, which may not be well captured by ReSET due to the limited number of the available meteorological stations in this region. The second difference between the figures is observed for the Oklahoma region. Specifically, the boxes and whiskers are wider for the spatial RMSE in Figure 13 than the temporal RMSE in Figure 14. This difference suggests that the reproduction of spatial variations is more variable between dates, while the reproduction of temporal variability is more consistent between locations. Alternatively stated, certain dates have spatial patterns of soil moisture that are poorly reproduced for the Oklahoma region.

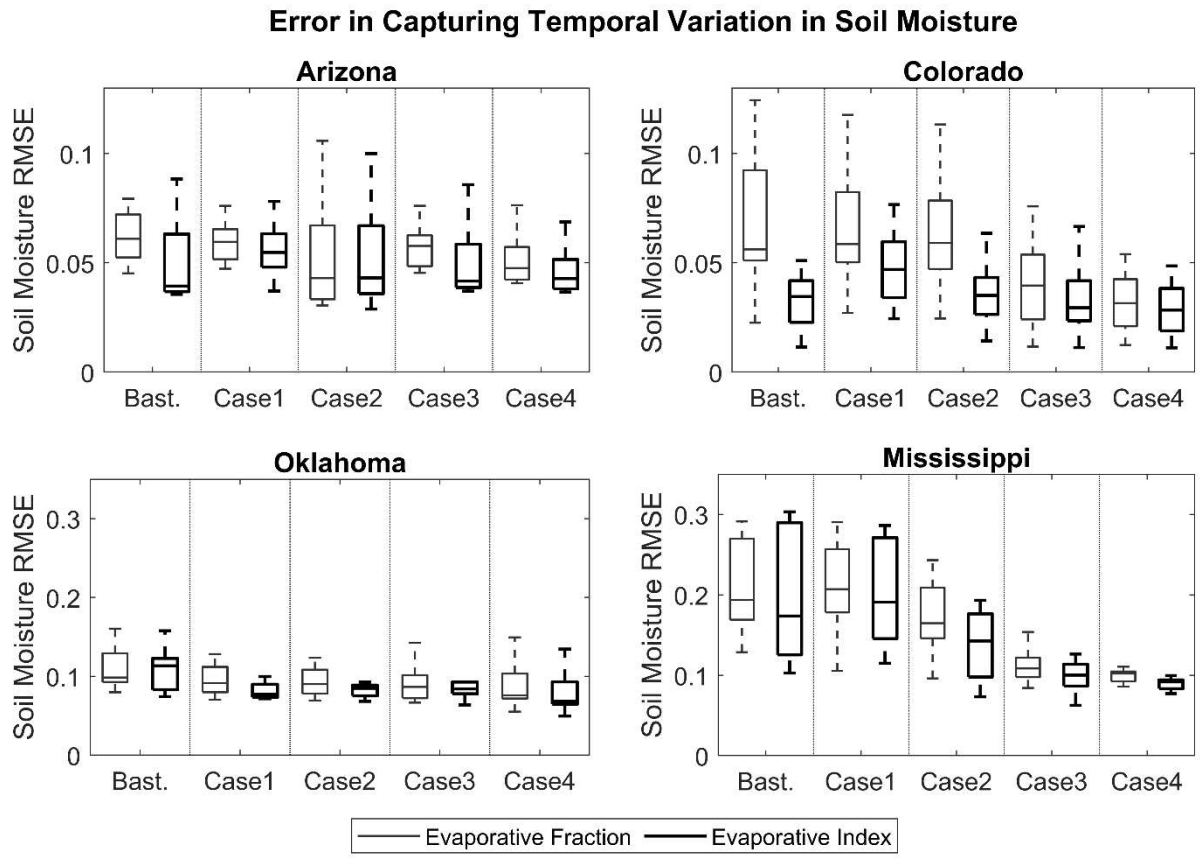

Figure 14. Root mean squared error (RMSE) of rootzone soil moisture when calculated using all dates at a given location. The horizontal line inside each box marks the median, the upper and lower edges of each box corresponds to upper and lower quartiles, and the upper and lower limits of the whiskers indicate maximum and minimum values among different locations. The label "Bast." indicates the use of the Bastiaanssen et al. (2000) relationship.

Table 9 shows the RMSE, mean absolute error (MAE), mean bias error (MBE), and correlation coefficient (*r*) for $\bar{\theta}$ using each method and study region. Overall, using Λ_{PET} instead of Λ_{SEB} improves the accuracy of the $\bar{\theta}$ estimates with the largest improvements occurring for Bastiaanssen et al. (2000) in Colorado. Including all cases and regions, using Λ_{PET} rather than Λ_{SEB} reduces RMSE, MAE, and MBE by 19%, 21%, and 39%, respectively, and increases *r* by 70%. Table 9 also shows that, overall, Case 4 performs better than all the other cases. When

Λ_{SEB} is used, Case 4 reduces the RMSE, MAE, and MBE by 64%, 74%, and 5%, respectively, and increases *r* by 73% compared to using the Bastiaanssen et al. (2000) relationship. Similarly, when Λ_{PET} is used, Case 4 reduces RMSE, MAE, MBE by 52%, 55%, and 12%, respectively, and increases *r* by 29% compared to using the Bastiaanssen et al. (2000) relationship. Comparing between regions, Colorado exhibits the best overall performance. Including all cases for Λ_{PET} relationships, Colorado has an average RMSE, MAE, MBE, and *r* of 0.043, 0.034, 0.012, and 0.695, respectively. In contrast, Mississippi has the worst overall performance with an average RMSE, MAE, MBE, and *r* of 0.152, 0.129, -0.095, and 0.396, respectively.

Table 9. Performance measures of the soil moisture estimates $\overline{\theta}$ when either Λ_{SEB} or Λ_{PET} is used with the Bastiaanssen et al. (2000) relationship (labeled "Bast.") or Cases 1-4 for the four study regions. A positive mean bias error indicates the remote sensing method overestimates the in situ soil moisture.

Study Regions	**Relationship**	**Root Mean Squared Error (RMSE)**		**Mean Absolute Error (MAE)**		**Mean Bias Error (MBE)**		**Correlation Coefficient (*r*)**	
		Λ_{SEB}	Λ_{PET}	Λ_{SEB}	Λ_{PET}	Λ_{SEB}	Λ_{PET}	Λ_{SEB}	Λ_{PET}
Arizona	Bast.	0.078	0.058	0.063	0.045	0.045	0.016	0.429	0.456
	Case 1	0.060	0.057	0.049	0.046	0.023	0.020	0.434	0.465
	Case 2	0.059	0.057	0.041	0.042	-0.020	-0.020	0.427	0.461
	Case 3	0.058	0.052	0.046	0.039	0.021	-0.008	0.467	0.493
	Case 4	0.053	0.051	0.041	0.039	0.007	-0.004	0.506	0.521
Colorado	Bast.	0.075	0.046	0.064	0.036	0.058	0.019	0.566	0.697
	Case 1	0.070	0.053	0.061	0.046	0.054	0.038	0.561	0.694
	Case 2	0.067	0.039	0.058	0.033	0.050	0.013	0.561	0.696
	Case 3	0.045	0.037	0.036	0.028	0.015	-0.008	0.565	0.701
	Case 4	0.046	0.038	0.035	0.029	0.008	-0.004	0.576	0.686
Oklahoma	Bast.	0.121	0.111	0.097	0.090	0.054	0.028	-0.116	0.108
	Case 1	0.103	0.086	0.084	0.072	0.036	0.016	-0.116	0.102
	Case 2	0.101	0.086	0.083	0.071	0.031	-0.006	-0.116	0.104
	Case 3	0.097	0.088	0.078	0.071	-0.003	-0.015	-0.133	0.090
	Case 4	0.093	0.076	0.075	0.063	0.040	0.026	0.128	0.397

Mississippi	Bast.	0.221	0.211	0.189	0.178	-0.131	-0.159	0.060	0.379
	Case 1	0.213	0.214	0.185	0.190	-0.160	-0.188	0.071	0.367
	Case 2	0.177	0.144	0.152	0.122	-0.111	-0.106	0.081	0.355
	Case 3	0.111	0.101	0.091	0.081	-0.010	-0.018	0.098	0.343
	Case 4	0.101	0.092	0.082	0.074	-0.012	-0.005	0.412	0.537

To gain a better understanding of how the improvements in performance occur, Figure 15 plots $\overline{\theta}$ from the empirical relationship (Bastiaanssen et al. 2000) and Case 4 (both using Λ_{SEB}) against $\overline{\theta}$ from the in situ observations. For Arizona, the graph shows that the empirical relationship typically overestimates $\overline{\theta}$, and this overestimation is confirmed by a positive MBE (MBE = 0.045). One reason could be that the Bastiaanssen et al. (2000) relationship was initially developed using agricultural fields in a semiarid region and tall prairie grasslands in a sub-humid region. Both the soil texture and vegetation in Arizona reduce the soil moisture compared to the estimate from the Bastiaanssen et al. (2000) relationship. Unlike the empirical relationship, Case 4 exhibits no clear overestimation of $\overline{\theta}$, and its MBE reduces to 0.007. Only a small improvement is seen in the scatter of the data when Case 4 is used in place of the empirical relationship (r^2 increases from 0.184 to 0.256). A similar improvement is observed for Colorado. Using Case 4 produces a visible improvement in the overestimation of $\overline{\theta}$, but it provides relatively little improvement in the scatter of the data. For Oklahoma, Case 4 provides a small improvement in the overestimation of $\overline{\theta}$, and some improvement in the scatter. For the Mississippi region, both the MBE and r^2 greatly improve, which explains why such large gains in RMSE were observed earlier.

Figure 16 plots $\bar{\theta}$ from the empirical relationship and Case 4 (both using Λ_{PET}) against $\bar{\theta}$ from the in situ observations. Comparing Figure 15 and Figure 16, the $\bar{\theta}$ estimates from the empirical relationship have much less bias when Λ_{PET} is used. As a result, changing from the empirical relationship to Case 4 in Figure 16 tends to produce a smaller improvement in MBE, but improvement in the scatter does occur. Case 4 reduces the scatter in Figure 16 for two reasons. First, as indicated by Figure 13, Case 4 better captures the local form of the $\Lambda_{PET} - \bar{\theta}$ relationship because it accounts for differences in annual precipitation, soil texture, and vegetation cover between monitoring locations. Second, as indicated by Figure 14, the $\Lambda_{PET} - \bar{\theta}$ relationship from Case 4 better captures the temporal variations in $\bar{\theta}$ as Λ_{PET} changes through time.

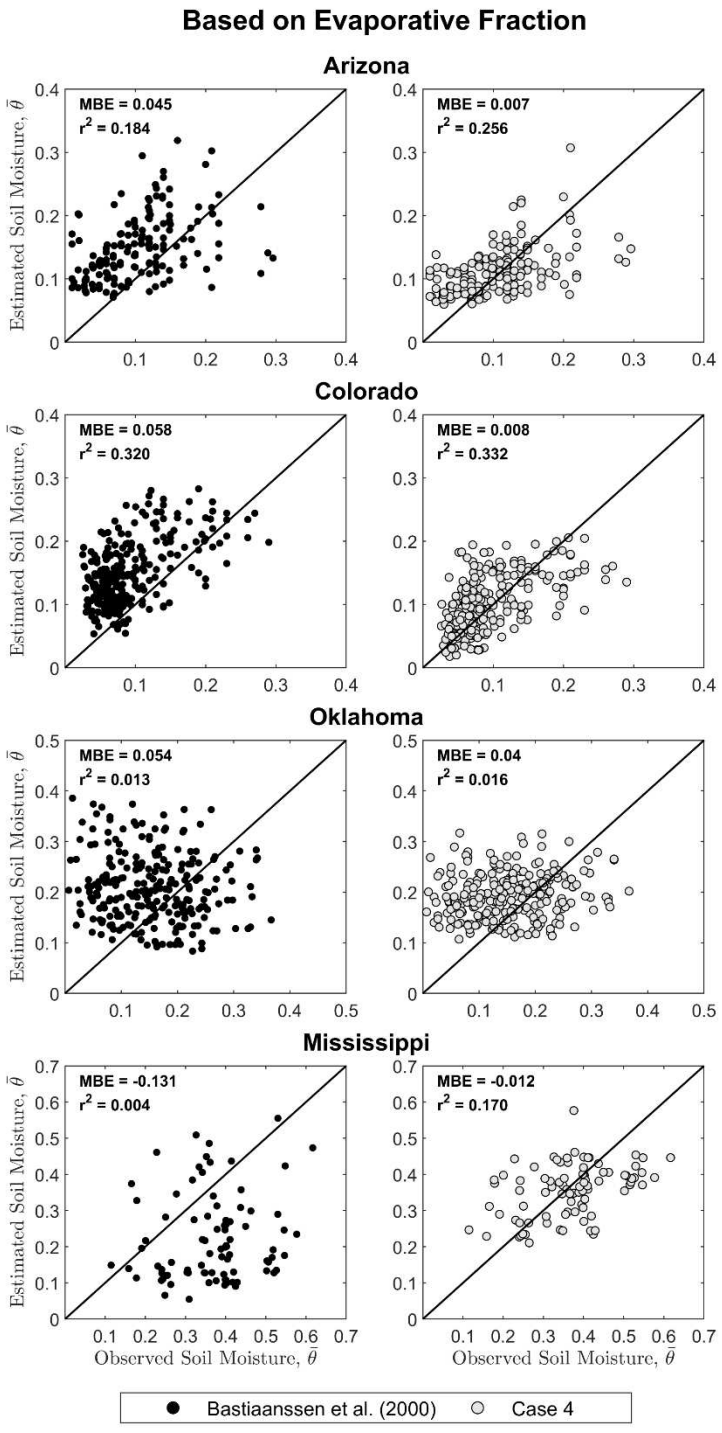

Figure 15. Comparison of observed rootzone soil moisture and estimated $\bar{\theta}$ from the Bastiaanssen et al. (2000) relationship (left side of figure) and Case 4 (right side of figure) using evaporative fraction Λ_{SEB} for the four study regions

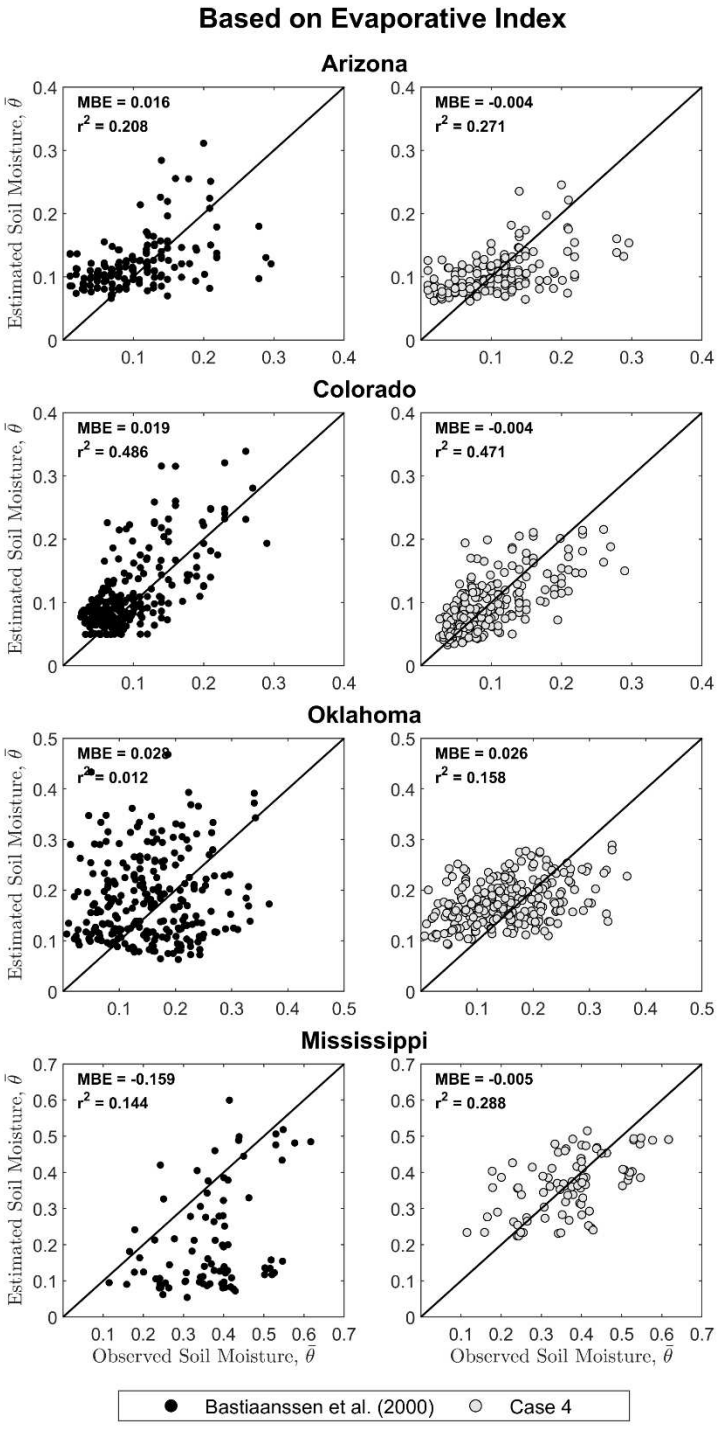

Figure 16. Comparison of observed rootzone soil moisture and estimated $\bar{\theta}$ from the Bastiaanssen et al. (2000) relationship (left side of figure) and Case 4 (right side of figure) using evaporative index Λ_{PET} for the four study regions

3.5 Discussion

Overall, the methods proposed by Sahaar and Niemann (2020) show promise for estimating rootzone soil moisture without calibration. Previously, Scott et al. (2003) considered using Λ_{SEB} to estimate rootzone soil moisture and showed that normalizing by porosity (i.e. estimating degree of soil saturation instead of soil moisture) improved the performance. The results of the present study show that moving from Λ_{SEB} to Λ_{PET} and using additional regional characteristics to determine the $\Lambda_{SEB} - \overline{\theta}$ and $\Lambda_{PET} - \overline{\theta}$ relationships further improves the performance. For our test regions, the methods that Sahaar and Niemann (2020) proposed consistently outperform pre-existing methods. The gains in performance are mostly achieved by reducing biases; the methods have difficulty capturing spatial and temporal variations. The RMSE's of the proposed methods range between 0.037-0.221, which are similar to errors in active/passive microwave methods (Colliander et al., 2017; Karthikeyan et al., 2017; Ma et al., 2019; Zeng et al., 2015). The proposed methods produce soil moisture estimates at a much finer spatial resolution but with more intermittent temporal frequency than active/passive microwave satellites. The proposed methods rely on empirical relationships, but those relationships are based on a global sensitivity analysis of diverse hypothetical regions (Sahaar and Niemann, 2020). Microwave methods also involve some empiricism (Chan et al., 2018; Fung et al., 1992; Jackson et al., 1996b; Mo and Schmugge, 1987; Zribi, 2003).

Further testing of the proposed methods is warranted. Sahaar and Niemann (2020) developed the methods by considering naturally vegetated areas, and the testing in the present study considered similar conditions. However, the methods should be tested in irrigated fields because the

methods have practical utility in such circumstances (e.g., for irrigation scheduling). Variations in the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships are expected within a given climate classification. While considering mean annual precipitation partially accounts for such variations, worse performance is expected for regions near the boundaries of the classifications. The performance of the methods should also be examined at different resolutions and using different remote sensing products. The methods were applied at 30 m resolution using manual application of ReSET. Coarser resolution data will produce more heterogeneous characteristics, which will likely cause greater errors. Recently, methods became available to estimate evapotranspiration automatically from Landsat satellites using the Google Earth Engine (Laipelt et al., 2021; Mhawej and Faour, 2020). Automatic application of the remote sensing methods is expected to produce higher errors (Zakeri and Mariethoz, 2021). Other factors can affect the relationships such as soil salinity, plant nutrition, and depth to groundwater (Nagler et al., 2005; Wang et al., 2020). Performance is expected to be worse when such factors impact the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships.

3.6 Conclusions

The purpose of this study was to evaluate the $\bar{\theta}$ estimates from optical/thermal remote sensing when the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships are inferred from regional characteristics using the methods proposed by Sahaar and Niemann (2020). The estimates were evaluated by comparing to in situ $\bar{\theta}$ observations and $\bar{\theta}$ estimates from the empirical relationship from Bastiaanssen et al. (2000) and Scott et al. (2003). The proposed methods were used in four study

regions (in Arizona, Colorado, Oklahoma, and Mississippi). Based on this study, the following conclusions can be made:

1. All the cases proposed by Sahaar and Niemann (2020) outperform the empirical relationship from Bastiaanssen et al. (2000) and Scott et al. (2003). The regionalized $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships better account for differences in precipitation, soil texture, and vegetation cover between monitoring locations, which allows better reproduction of spatial and temporal variability of $\bar{\theta}$ within the region.
2. Estimating soil moisture based on Λ_{PET} provides better results than estimating soil moisture based on Λ_{SEB}. Combining the results from all study regions, using Λ_{PET} rather than Λ_{SEB} reduces the RMSE, MAE, and MBE by 19%, 21%, 39%, and increases the correlation coefficient by 70%, respectively. These improvements occur because Λ_{PET} accounts for both the available energy and turbulent transport whereas Λ_{SEB} considers only the available energy.
3. Estimates of soil moisture improve as more regional information is used to estimate the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships. Specifically, the performance improves as one moves from Case 1 to Case 4. The constants that determine the shape of the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships depend on regional characteristics such as the soil texture and the leaf area index. Thus, the relationships are better estimated when more regional information is used. Overall, Case 4 performs the best. Using Λ_{PET} and Case 4, the soil moisture estimates range from a RMSE of 0.038 for Colorado to 0.092 for Mississippi.
4. The performance of the methods is better for the arid and semiarid regions than the sub-humid and humid regions. For arid and semiarid regions, the soil moisture is typically

below the field capacity, which suggests that ET is moisture limited and depends on $\bar{\theta}$. In contrast, the sub-humid and humid regions more frequently exceed the field capacity, which tends to produce energy limited ET that is independent of changes in $\bar{\theta}$.

CHAPTER 4. ENHANCING ROOTZONE SOIL MOISTURE ESTIMATION USING MULTI-SENSOR DATA FUSION OF OPTICAL AND THERMAL REMOTE SENSING AND MACHINE LEARNING

4.1 Introduction

Many applications can benefit from accurate knowledge of soil moisture at a fine resolution (10-100 m) over large spatial extents (30-300 km). For example, agricultural water management (Foster et al., 2020; Martínez-Fernández et al., 2016), water resources sustainability (Petropoulos et al., 2018), runoff generation (Sumargo et al., 2021), numerical weather forecasting (Koster and Suarez, 2003; Zhan et al., 2016), climate modeling (Brocca et al., 2017; Koster and Suarez, 2001), wildfire prediction (Bartsch et al., 2009; Hou and Orth, 2020), and monitoring of floods and droughts (Kim et al., 2019; Xu et al., 2020) can all benefit from reliable soil moisture information.

Despite its wide range of applications, measuring soil moisture at a fine spatial resolution over a large spatial extent can be quite challenging. Soil moisture measurements from fields are often obtained through gravimetric, electromagnetic, tensiometric, and hygrometric methods (Schmugge et al., 1980). However, these methods provide essentially point measurements, which cannot capture the heterogeneity of soil moisture over large spatial scales (Karthikeyan and Kumar, 2016). This heterogeneity is due to complex interactions among meteorologic,

vegetation, soil, and topographic variables (Peng et al., 2017). Furthermore, establishing and maintaining an in-situ soil moisture network can be costly and time consuming (Peng et al., 2017).

Satellite remote sensing techniques in the microwave frequency range can provide soil moisture estimates with the spatial extent necessary for many applications, but the spatial resolutions of these estimates are very coarse. For example, Soil Moisture Active Passive (SMAP) (Entekhabi et al., 2010), Advanced Scatterometer (ASCAT) (Wagner et al., 2013), Soil Moisture and Ocean Salinity (SMOS) (Kerr et al., 2001), and Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) (Njoku et al., 2003) provide soil moisture estimates at ~9-60 km spatial resolutions. The coarse spatial resolutions and shallow depth of measurement (~5 cm) limit the applicability of such data for many applications (Peng et al., 2017).

Several methods have been developed to downscale microwave soil moisture products. For example, Tagesson et al. (2018) used land surface temperature and vegetation dryness index from a geostationary satellite to disaggregate SMOS soil moisture from ~40 km resolution to ~5 km resolution in West Africa. They reported an average RMSE values of 0.063 cm^3/cm^3. Das et al. (2019) used the Sentinel-1A and Sentinel-1B synthetic aperture radar data to disaggregate SMAP L-band radiometer measurements from ~40 km to 3 km and 1 km gridded products. They reported RMSE values of 0.063 cm^3/cm^3 for 3 km and 0.064 cm^3/cm^3 for 1 km disaggregated soil moisture products, respectively. Wei et al. (2019) utilized Moderate Resolution Imaging Spectroradiometer (MODIS) and digital elevation data with a gradient boosting decision tree regression model to downscale SMAP soil moisture estimates from 36 km to 1 km resolution

across the Tibetan Plateau. They reported RMSE values ranging from 0.057 to 0.107 cm^3/cm^3. Nuñez et al. (2017) used MODIS products, sand fraction, and elevation to downscale AMSR2 soil moisture estimates from 25 km to 1 km in Puerto Rico. They reported RMSE values ranging from 0.075 to 0.095 cm^3/cm^3. Vergopolan et al. (2020) used a hyper-resolution land surface model, a radiative transfer model, and a Bayesian scheme to merge and downscale SMAP 36 km coarse resolution soil moisture data to a 30 m spatial resolution and evaluated the results using four watersheds in the United States. They reported an average Kling-Gupta efficiency KGE value of 0.56. The resulting product captured the spatial and temporal soil moisture dynamics better than the SMAP L3 and L4 products at field and watershed scales. Most soil moisture downscaling methods have focused on disaggregating surface soil moisture (top 5 cm); fewer methods have been developed to downscale rootzone soil moisture. Dumedah et al. (2015) used Disaggregation based on Physical And Theoretical scale Change (DisPATCh) to downscale satellite soil moisture data and estimate root-zone soil moisture at a spatial resolution of 1 km and depths of 0-30 cm, 30-60 cm, and 60-90 cm. They reported RMSE values of 0.071 cm^3/cm^3 for 0-30 cm and 0.058 cm^3/cm^3 for 30-60 cm, and 0.067 cm^3/cm^3 for 60-90 cm soil moisture products, respectively.

Soil moisture can also be estimated using optical and thermal remote sensing data obtained from satellite sensors such as Landsat and MODIS. These methods characterizes the relationship between soil moisture and water-stressed vegetation using the visible and near-infrared bands, and they use the thermal infrared band to derive the relationship between soil moisture and soil thermal properties (Zhang and Zhou, 2016). These methods provide soil moisture estimates at fine spatial resolutions (30 m to 1 km) over large spatial extent (185 km to 2330 km). The

methods include the triangle and trapezoid methods (Carlson, 2007; Carlson and Petropoulos, 2019; Petropoulos et al., 2009; Sadeghi et al., 2017; Wang et al., 2011), the drought index method(Amani et al., 2017; Anderson et al., 2012; Feng et al., 2013; Ghulam et al., 2007; Hu et al., 2020), the thermal inertia method (Van Doninck et al., 2011), single optical methods (Amazirh et al., 2018; Gao et al., 2017), energy balance methods (Bastiaanssen et al., 2000; Chen and Liu, 2020; Hain et al., 2009; Scott et al., 2003), and synergistic optical/thermal and microwave methods (Chauhan et al., 2003; Merlin et al., 2012; Piles et al., 2016).

Optical and thermal remote sensing methods have limitations. The triangle method, for instance, relies on a full range of vegetation and moisture conditions and assumes a linear relationship between vegetation index and surface temperature, which may not hold in non-flat terrain (Carlson, 2007). The drought index method calculates soil moisture retroactively and considers only water stress on vegetation indices, neglecting temperature and rainfall effects (Zhang and Zhou, 2016). The thermal inertia method assumes consistent soil properties in both horizontal and vertical directions, which may not be accurate in real-world conditions, especially in high-vegetation areas (Zhang and Zhou, 2016). Single optical methods can be precise for soil samples but are influenced by various factors like vegetation, atmospheric conditions, and topography, and rely on empirical relationships and lacks physical basis (Zhang and Zhou, 2016).

An alternative approach to soil moisture estimation uses multi-sensor data fusion of coarse-resolution microwave remote sensing and fine-resolution optical and thermal remote sensing products (Bai et al., 2019; Huang et al., 2022; Leng et al., 2019; Long et al., 2019; Mishra et al., 2018; Owe et al., 2001; Portal et al., 2018). Such approaches also use regional characteristics,

which can contribute to the variability of soil moisture (Vergopolan et al., 2022). Previous studies have shown that soil moisture can be influenced by antecedent soil moisture, which affects infiltration, percolation, and runoff (Wangemann et al., 2000; Wang et al., 2019). Landcover also plays a role because thickly vegetated regions (cropland and forest) tend to retain more soil moisture compared to bare soils (Ranney et al., 2015; Zhang and Schilling, 2006). Soil properties also impact soil water storage, hydraulic conductivity, infiltration, and drying rates (Crow et al., 2012; Takagi and Lin, 2011). Meteorological variables such as precipitation and land surface temperature regulate runoff and evapotranspiration rates (Evans et al., 2016; Rosenbaum et al., 2012), and topography drives surface and subsurface water flow and influences evapotranspiration rates through impacts on insolation (Coleman and Niemann, 2013; Cowley et al., 2017; Famiglietti et al., 2008; Geroy et al., 2011).

In recent years, machine learning methods have been explored for data fusion and downscaling/disaggregation of soil moisture (Abbaszadeh et al., 2019; Abowarda et al., 2021; Liu et al., 2020; Peng et al., 2021). Machine learning methods are data-driven approaches that learn patterns and relationships from data without making assumptions about the processes that govern soil water dynamics (Ali et al., 2015; Fang et al., 2019). Unlike process-based models that often have specific data requirements, machine learning can merge large volumes of data from various sources, including in-situ measurements, meteorological variables, and remote sensing datasets (Abbaszadeh et al., 2019; Greifeneder et al., 2021). Machine learning algorithms can also perform feature selection, automatically identifying the inputs that are most relevant for estimating soil moisture (Meyer et al., 2018). Machine learning models have shown a strong correlations between in-situ soil moisture observations and predicted soil moisture values (Ali et

al., 2015; Rani et al., 2022). Abowarda et al. (2021) employed a random forest model to produce surface soil moisture at a resolution of 30 m × 30 m for the Haihe Basin in northern China. SMAP Level 4 surface soil moisture was incorporated as the background field, and Landsat and MODIS data were used to determine fine resolution Normalized Difference Vegetation Index (NDVI), surface albedo, and land surface temperature. Precipitation and soil texture were also used as model inputs. The study reported RMSE values ranging from 0.031 to 0.050 cm^3/cm^3. Singh and Gaurav (2023) used a feed-forward artificial neural network (ANN) with nine input variables derived from the Sentinel-1 and Sentinel-2 satellites as well as digital elevation model products to estimate surface soil moisture at a spatial resolution of 60 m over the Kosi Fan in the Himalayan Foreland in the north Bihar plain, India. The study reported that the ANN model accurately predicts soil moisture with an RMSE value of 0.004 m^3/m^3. They also compared the ANN results to ten other machine learning models and found that the ANN outperformed the other machine learning methods. Karthikeyan and Mishra (2021) used extreme gradient boosting (XGBoost) to estimate soil moisture at 1 km resolution and multiple depths (5, 10, 20, 50, and 100 cm) over the CONUS. They included SMAP Level 4 soil moisture, soil texture, elevation, vegetation parameters (NDVI, EVI and Gross Primary Productivity), MODIS land surface temperature, and precipitation as their predictor variables. The study reported an unbiased root mean squared error ubRMSE of less than 0.040 m^3/m^3. Fuentes et al. (2022) used a deep learning approach to fuse SMAP, Sentinel-1, MODIS products (surface reflectance, land surface temperature, and land cover) and gridded soil properties to downscale soil moisture to 90 m resolution and multiple depths. They used a multilayer perceptron model for the surface (0–10 cm) soil moisture and recurrent neural networks model for the 0–30 cm and 30–60 cm soil moisture at Ozflux and Oznet soil moisture networks across Australia. The study reported RMSE

values of 0.073 cm^3/cm^3 for 0–10 cm and 0.070 cm^3/cm^3 for 0–30 cm and 30–60 cm, respectively.

In this study, an advanced soil moisture retrieval framework was developed based on machine learning algorithms (Feed Forward Artificial Neural Network (ANN), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Categorical Boosting (Catboost), and Light Gradient-Boosting Machine (LightGBM)) and a comprehensive suite of satellite-based datasets was integrated, including microwave soil moisture data from SMAP, precipitation data from the Global Precipitation Measurement (GPM) mission, optical and thermal evapotranspiration products from ECOSTRESS, vegetation data from MODIS, soil properties from gNATSGO, and land cover information from the National Land Cover Database (NLCD). The primary goal was to produce fine-scale, multi-depth (0-5 cm, 0-10 cm, 0-20 cm, 0-50 cm, and 0-100 cm) soil moisture estimates at a spatial resolution of 70 m × 70 m across the continental United States (CONUS). This research aimed to contribute to the field of remote sensing by advancing the accuracy and spatial resolution of soil moisture data, which is useful for applications in agriculture, hydrology, and environmental science.

4.2 Datasets

Most of the satellite data obtained in this study were collected using Google Earth Engine (GEE) or the Land Processes Distributed Active Archive Center (LPDAAC). GEE is a cloud-based platform for planetary-scale geospatial analysis. It provides users with access to a catalog of satellite imagery and geospatial datasets, as well as a suite of tools for analyzing and visualizing

this data (Gorelick et al., 2017). The LPDAAC) is an archive and distribution center for land-related satellite data and derived products (Davis et al., 2015).

4.2.1 In-Situ Soil Moisture Data

In this study, soil moisture estimation models were developed for the contiguous United States (CONUS) because soil and landcover data are only available for this region. The reference soil moisture data for training and evaluating the performance of our machine learning soil moisture estimates was obtained from the International Soil Moisture Network (ISMN). The ISMN is a centralized data hosting facility initiated by the European Space Agency in 2011. As of June 2023, the ISMN contained soil moisture data from 14 monitoring networks and 1430 stations in CONUS.

For this study, only soil moisture observations measured at the depth range of 0-102 cm were utilized, as most stations (approximately 99.9%) collected data up to this depth only a small portion of the stations measured soil moisture at greater depths (about 0.1%; see Table 10). Also, to simplify the prediction of soil moisture estimates, a maximum root-zone depth of 100 cm was chosen. In nature, the depths tapped by roots vary depending on vegetation type, soil type, and other environmental factors (Jackson et al., 1996a). All soil moisture measurements from the ISMN are available at a 1-hr time step.

Table 10. Percentage of soil moisture observations at ISMN networks within each depth range across the study area

Depth range	% observed data
0-5 cm	8.3
5-10 cm	25.0
10-20 cm	13.2
20-50 cm	24.5
50-102 cm	29.1
> 102cm	0.1

To estimate soil moisture for this study, the depth-weighted mean of soil moisture values were calculated for five depth ranges: 0-5 cm, 0-10 cm, 0-20 cm, 0-50 cm, and 0-100 cm, at each station, using the method described by Gao et al., (2017) and Liu et al. (2023). The following equation is used to determine the in-situ soil moisture for each depth range:

$$\theta_{0\_n} = \frac{2\theta_1 L_1 + (\theta_1 + \theta_2) L_2 + (\theta_2 + \theta_3) L_3 + ...(\theta_{n-1} + \theta_n) L_n}{2(L_1 + L_2 + L_3 + ...L_n)} \quad (48)$$

where θ_{0_n} is the volumetric soil moisture from the surface to the *n*th observation depth (cm^3/cm^3), θ_n is the volumetric soil moisture at *n*th observation depth (cm^3/cm^3), and L_n is the soil layer thickness between adjacent observation depths (cm).

The study period was selected based on the combined availability of all data used. Among the 1430 stations in CONUS, 801 stations have data coverage from January 2019 to December 2022 and soil moisture observation from 0 to 102 cm. All soil moisture values flagged by ISMN

(Dorigo et al., 2013), including those under frozen conditions and those outside the expected soil moisture range (<0.0 cm^3/cm^3or >0.6 cm^3/cm^3), were excluded from the analysis. The final in situ dataset includes soil moisture measurements from 731 stations. These in-situ soil moisture stations belong to eight different operational networks (Table 11), including the Snow Telemetry (SNOTEL), U.S. Department of Agriculture Soil Climate Analysis Network (SCAN), U.S. Climate Reference Network (USCRN), Texas Soil Observation Network (TxSON), Cosmic-ray Soil Moisture Observing System (COSMOS), Atmospheric Radiation Measurement Climate Research Facility (ARM), Roaring Fork Observation Network (iRON), and the FLUXNET network. The spatial distribution of the 731 selected stations is shown in Figure 17. Universal Time Coordinated (UTC) was used throughout this study because the hourly in situ observations from the ISMN were recorded in UTC time.

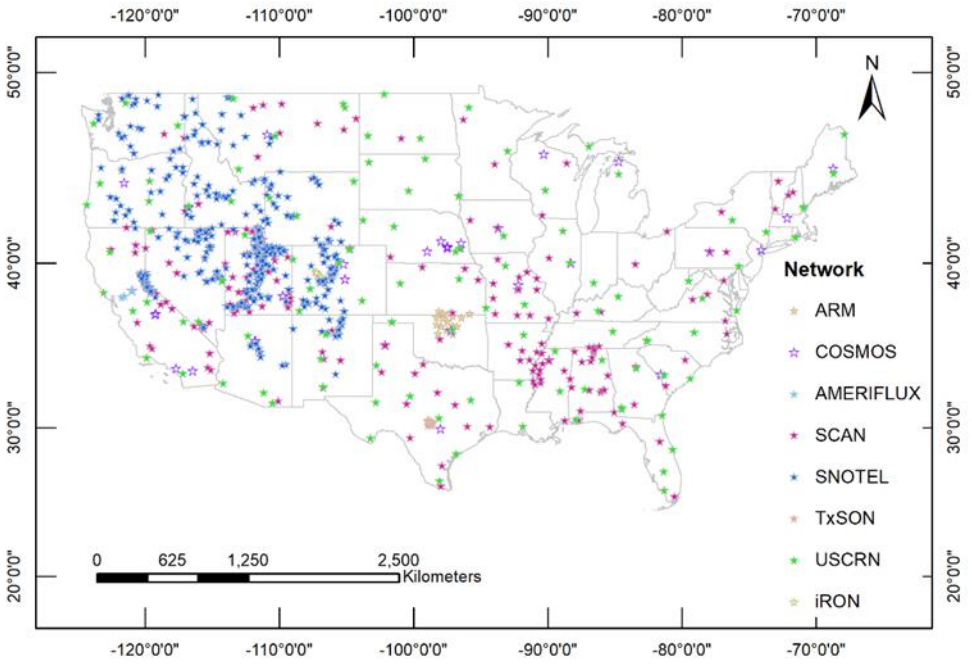

Figure 17. Locations of the 731 in-situ soil moisture stations with data available from January 2019 to December 2022 that were used in this study

Table 11. Soil moisture networks and number of stations used from each network

Network	No. of stations	Reference
ARM	17	Cook (2016)
COSMOS	29	Zreda et al. (2008)
FLUXNET-AMERIFLUX	3	Baldocchi et al. (2001)
iRON	8	Osenga et al. (2021)
SCAN	168	Schaefer et al. (2007)
SNOTEL	359	Leavesley et al. (2008)
TxSON	40	Caldwell et al. (2019)
USCRN	107	Bell et al. (2013)

4.2.2 Satellite Soil Moisture Data

The seventh version of the SMAP Level-4 (SPL4SMGP.007) (Reichle et al., 2022) surface soil moisture (SSM) (0-5 cm), rootzone soil moisture (RZSM) (0-100 cm), and profile soil moisture (PSM) (0-bedrock depth) products were used in this study (Table 12). The SMAP Level-4 SSM and RZSM products are generated using a land data assimilation system that combines satellite-based L-band brightness temperature measurements, precipitation observations, and land surface modeling (Reichle et al., 2017). The SMAP Level-4 products were used because they provide complete spatial and temporal coverage and include soil moisture values not only for the surface layer (0-5 cm) but also for deeper root-zone layers (Reichle et al., 2017). SMAP Level-4 products have a spatial resolution of 9 km and a temporal resolution of 3 hr. The SMAP Level-4 SSM, RZSM, and PSM data were downloaded from the LPDAAC (https://lpdaac.usgs.gov/).

The satellite soil moisture data were also used to obtain antecedent soil moisture values. Higher antecedent soil moisture leads to a slower initial infiltration rate and a more rapid decline in the infiltration rate through time (Wei et al., 2022). Soil Moisture Active Passive surface, root zone, and profile soil moisture data were used to calculate antecedent soil moisture values for each hour of soil moisture data using a rolling mean window of 1 day, 3 days, 7 days, and 14 days.

The parameters estimated were SMAP surface (SSM), root zone (RZSM), and profile (PSM) soil moisture, and antecedent 1-day (A1DAY), 3-day (A3DAY), 7-day (A7DAY), and 14-day (A14DAY) SMAP SSM, RZSM, and PSM (Table 14).

4.2.3 Land Cover and Vegetation Data

Land cover type affects rootzone soil moisture by influencing the surface albedo and the root water uptake (Wang et al., 2012; Zhang and Zhou, 2016). Land covers with deep-rooted vegetation, such as forests, can access water from deeper soil layers, reducing the soil moisture in those layers as transpiration occurs (James et al., 2003). Land cover with shallow-rooted vegetation, such as grasslands, relies more on near surface soil moisture. The National Land Cover Database 2019 (NLCD 2019) was used to investigate land cover effect on soil moisture. NLCD 2019 contains 16 landcover classes at a spatial resolution of 30 m (Dewitz and USGS, 2021) (Table 12). NLCD 2019 was downloaded from the U.S. Geological Survey National Land Cover Database (https://www.mrlc.gov/data/nlcd-2019-land-cover-conus).

The density of the vegetation cover is also expected to influence the rootzone soil moisture. Thicker vegetation cover promotes infiltration and reduces runoff by breaking up the soil aggregates (Bot and Benites, 2015; Cerdà, 1999; Peng and Wang, 2012). Thicker vegetation also shades the soil, reducing soil evaporation while simultaneously increasing transpiration rates (Ranney et al., 2015). To characterize the density and photosynthetic activity (greenness) of the vegetation cover, the following indices were used: Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI), Leaf Area Index (LAI) and Fraction of Photosynthetically Active Radiation (fPAR) information.

NDVI and EVI were obtained from the MODIS Vegetation Indices (MOD13Q1) Version 6.1 dataset. MOD13Q1 provides NDVI and EVI values every 16 days with a spatial resolution of 250 m (Table 12). MOD13Q1 data were downloaded from the LPDAAC (https://lpdaac.usgs.gov/). NDVI and EVI are vegetation indices that use different combinations of spectral bands to characterize vegetation (Matsushita et al., 2007). Higher NDVI and EVI values indicate thicker and/or healthier vegetation. NDVI uses the red and near infrared bands, while EVI also uses the blue band. EVI is less sensitive to atmospheric conditions than NDVI and is therefore preferred in situations where aerosol content is high or soil/background influences are significant (Matsushita et al., 2007).

The MODIS Vegetation Indices (MCD15A3H) Version 6.1 dataset provides LAI and FPAR values at regular intervals of 4 days with a spatial resolution of 500 m (Table 12). MCD15A3H data were downloaded from the LPDAAC (https://lpdaac.usgs.gov/). LAI is the total one-sided green leaf surface area per unit ground area, while fPAR is the fraction of photosynthetically active radiation that is absorbed by vegetation (Xu et al., 2018). FPAR is a key indicator of vegetative water, energy, and carbon balance, which plants require for the photosynthesis process (Churkina and Running, 1998). LAI typically ranges from 0 for no vegetation to >5 for dense forests. fPAR range between 0 for no vegetation and 1 for dense, healthy vegetation (Xu et al., 2018).

4.2.4 Soil Data

Soil properties can also affect rootzone soil moisture dynamics. For example, soil texture determines the water-holding capacity, with sandy soils drying out quickly and clay soils retaining more water (Nath, 2014). Soil depth impacts plant access to water, with shallow soils being more susceptible to drought stress (Barbeta et al., 2015). Soil salinity can hinder water uptake by plants, and soil acidity can affect nutrient availability (Machado and Serralheiro, 2017; Yadav et al., 2020).

The gridded National Soil Survey Geographic (gNATSGO) dataset was used to obtain the soil properties, including percent sand, silt, and clay, organic matter, bulk density, pH, electrical conductivity (EC), and depth to restrictive layer. All soil properties except the depth to restrictive layer were calculated for five depth ranges: 0-5 cm, 0-10 cm, 0-20 cm, 0-50 cm, and 0-100 cm. The gNATSGO dataset has a nominal 30-m resolution, but the resolution of the underlying data is far coarser. The gNATSGO dataset is a composite of the Soil Survey Geographic (SSURGO) dataset (mostly 1:24,000 scale), State Soil Geographic 2 (STATSGO2) dataset (1:250,000 scale), and the detailed Raster Soil Surveys (RSS) dataset (Rossiter et al., 2022). The gNATSGO dataset downloaded from the U.S. Department of Agriculture-Natural Resources Conservation Service website (https://www.nrcs.usda.gov/resources/data-and-reports/gridded-national-soil-survey-geographic-database-gnatsgo). EC characterizes soil salinity (Smith and Doran, 2015).

4.2.4.1 Weather and Climate Data

Weather and climate contribute significantly to rootzone soil moisture variations. Humid climates, characterized by increased precipitation, tend to have higher rootzone soil moisture levels, favoring water retention in the soil. In contrast, arid climates with limited rainfall experience lower rootzone soil moisture, increasing the risk of drought stress (Seleiman et al., 2021; Zohaib et al., 2017). Moreover, warm and sunny climates lead to higher rates of evapotranspiration, depleting soil moisture more rapidly compared to cold and cloudy climates (Evans et al., 2016). Key datasets are relied on to better understand these weather and climate factors, including land surface temperature (LST) to gauge the temperature's impact on soil moisture, evapotranspiration (ET) data to track moisture loss processes, and the evaporative stress index (ESI) to assess moisture-related stress. Potential evapotranspiration (PET) is also considered to quantify climate-driven water loss potential, and the global aridity index (AI) is considered to account for regional climate's influence on rootzone soil moisture. These datasets are essential for connecting weather and climate to rootzone soil moisture dynamics.

Precipitation was determined using the Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG) Final Precipitation Level 3 Half Hourly (GPM_3IMERGHH) precipitation product. GPM_3IMERGHH is a global, 0.5-hr, gridded precipitation product (Table 12). GPM_3IMERGHH precipitation estimates are produced using the 2017 version of the Goddard Profiling Algorithm (Huffman et al., 2019), adjusted using the GPM Combined Ku Radar-Radiometer Algorithm, and combined into 0.1° (roughly 10 km) resolution (Huffman et al., 2019). GPM-IMERG was chosen because it has high temporal resolution (Huffman et al., 2019) and has been applied in other soil moisture studies (Beck et al.,

2021; Farahani et al., 2022; Krishnan et al., 2022). Its higher temporal resolution helps quantify the short-term precipitation effects on soil moisture. GPM_3IMERGHH was obtained using the GEE platform (Gorelick et al., 2017). GPM precipitation data was used to calculate antecedent precipitation data for each hour of precipitation record using a rolling sum window of 6 hours, 1 day, 3 days, 7 days, and 14 days. The parameters estimated were antecedent 6-hour (P6H), 1-day (P1DAY), 3-day (P3DAY), 7-day (P7DAY), and 14-day (P14DAY) precipitation (Table 14).

The ECOsystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) was launched in June 2018. Compared to the Landsat and Sentinel-2 missions, ECOSTRESS offers improved temporal sampling with an average 4-day revisit time. Four ECOSTRESS products were used including Land Surface Temperature (LST), Potential Evapotranspiration (PET), actual Evapotranspiration (ET), and Evaporative Stress Index (ESI) (Table 12). These products are at a spatial resolution of 70 m. The LST product is derived from five thermal infrared bands. The actual ET product provides actual instantaneous ET using the Priestly-Taylor Jet Propulsion Laboratory algorithm (Hook and Fisher, 2019), which uses a series of ecophysiological scaling functions to reduce the potential ET to the actual ET, and partition total ET into three sources for canopy transpiration, soil evaporation, and interception evaporation. These functions are based on atmospheric vapor pressure deficit, relative humidity, and vegetation indices, including NDVI and Soil Adjusted Vegetation Index (Fisher et al., 2020). ESI is the ratio of Level 3 and PET. It serves as an indicator of potential drought and plant water stress (Fisher et al., 2020). ECOSTRESS data were obtained from the LPDAAC (https://lpdaac.usgs.gov/)

The Global Aridity Index and Potential Evapotranspiration Database v3 (Global-AI_PET_v3) was used to obtain the Aridity Index (AI) for the study region. The AI is the ratio of mean annual precipitation to mean annual PET, so AI is lower for more arid regions. The Global-AI_PET_v3 dataset was developed by Trabucco and Zomer (2019). It is based on the FAO Penman-Monteith reference crop evapotranspiration equation, which was implemented using the WorldClim global climate data, and mean annual precipitation from the WorldClim v2.1 dataset (Fick and Hijmans, 2017). The Global-AI_PET_v3 dataset is available at 30 arc-second (~1 km) resolution and was obtain from CGIAR-Consortium for Spatial Information (CGIAR-CSI , https://csidotinfo.wordpress.com) (Table 12).

4.2.5 Topographic Data

Topography has also been shown to impact soil moisture. Areas with higher elevations often experience varying soil moisture levels due to cooler temperatures, reduced atmospheric pressure, decreased soil depths, orographic precipitation effect, and steeply sloped regions that are more susceptible to erosion and runoff (Cowley et al., 2017; Perez, 1987). Conversely, gentle slopes reduce lateral hydraulic gradients, maintaining higher rootzone soil moisture (Famiglietti et al., 1998). Topographic aspect also contributes to rootzone soil moisture variability. For example, north-facing slopes can have higher soil moisture in the northern hemisphere due to reduced insolation and ET rates (Coleman and Niemann, 2013; Geroy et al., 2011; Western et al., 1999). Topographic position, whether on ridges or in valleys, affects soil moisture distribution. For example, valleys floor tend to have higher soil moisture levels than ridge locations due to a combination of factors, including reduced exposure to wind and solar radiation, and slower runoff rates (Zhao et al., 2012).

The Shuttle Radar Topography Mission Global 1-arc second (SRTMGL1) digital elevation model (DEM) was used to obtain elevation, slope, aspect, and hillshade values. The SRTMGL1 DEM has a resolution of 1-arc second (~30 m). Slope, aspect, and hillshade values were calculated using the GEE terrain products library. Hillshade is a variable that describes the amount of sunlight that a particular location on a landscape receives, based on its elevation and the sun's position. The calculation involves considering the differences in elevation across the landscape, as well as the azimuth and altitude of the sun. The constants of 270 and 45 proposed in the GEE terrain products library were used for the sun's azimuth and altitude angle for simplicity. Hillshade values range from 0 to 255, with lower values indicating less sunlight and higher values indicating more sunlight. Hillshade has practical applications, especially in creating depictions of topography, by mimicking the way sunlight interacts with hills and canyons, resulting in the creation of shadows and shading effects. This can be useful for visualizing the landscape and identifying potential areas of high or low sunlight exposure. Previous studies (Celik et al., 2022; Iverson et al., 2002; Luo et al., 2022) have explored the effect of hillshade on soil moisture. These studies have found that hillshade has impact on soil moisture levels, with areas of lower hillshade (i.e., more sunlight exposure) tending to have drier soils.

The Shuttle Radar Topography Mission multi-scale Topographic Position Index (SRTM_mTPI) was used to measure the elevation of a location relative to its surrounding area. The mTPI is calculated by subtracting the mean elevation within a 3x3 pixel neighborhood from the elevation at each location. It distinguishes valleys, ridges, peaks, and plains (Theobald et al., 2015). The spatial resolution of the SRTM_mTPI dataset is 270 m (Table 12).

Table 12. List of datasets used in this study (Glossary: CONUS - Contiguous United States, ISMN - International Soil Moisture Network, NLCD - National Land Cover Database, GEE - Google Earth Engine, SRTM - Shuttle Radar Topography Mission, SMAP - Soil Moisture Active Passive, ECOSTRESS - Ecosystem Spaceborne Thermal Radiometer Experiment on Space Station, LST - Land Surface Temperature, ET – Evapotranspiration, ESI - Evaporative Stress Index, PET – Potential Evapotranspiration, LAI - Leaf Area Index, fPAR - Fraction of Photosynthetically Active Radiation, NDVI - Normalized Difference Vegetation Index, EVI - Enhanced Vegetation Index)

Data Type	Variables	Product/Access	Spatial Resolution	Temporal Resolution	Reference
In-situ Soil Moisture	0-5 cm, 0-10 cm, 0-20 cm, 0-50 cm, 0-100 cm	Soil moisture/ ISMN	Point or filed measurements	Hourly	Dorigo et al. (2021)
Satellite Soil Moisture	SSM (0-5 cm), RZSM (0-100 cm), PSM (0-bedrock)	SMAP SPL4SMGP.007/GEE	9 km	3 Hour	Reichle et al. (2022)
Landcover and Vegetation	NLCD 2019	NLCD 2019/ GEE	30 m	Static	Dewitz (2021)
	NDVI, EVI	MOD13Q1.061/ LPDAAC	250 m	16 Days	Didan (2021)
	LAI, fPAR	MCD15A3H.061/ LPDAAC	500 m	4 Days	Myneni et al. (2021)
Soil	Sand, Silt, Clay, Organic Matter, Bulk Density, Electrical Conductivity, PH, Depth to Restrictive Layer	Soil layers/ USDA-NRCS (gNATSGO)	30 m	Static	Soil Survey Staff, (2020)

	Precipitation Measurement	GPM_3IMERGHH / GEE	11 km	Half-hourly	Huffman et al. (2019)
Weather and Climate	Instantaneous LST	ECO_L2_LSTE v002/ LPDAAC	70 m	Varies	Hook and Hulley (2022)
	Instantaneous ET	ECO3ETPTJPLv001/ LPDAAC	70 m	Varies	Hook and Fisher (2019)
	Instantaneous ESI, Instantaneous PET	ECO4ESIPTJPLv001/ LPDAAC	70 m	Varies	Hook and Fisher (2019)
	Aridity Index	Climate Database v3/ CGIAR-CSI	1 km	Static	Zomer et al. (2022)
Topography	Elevation, Slope, Aspect, Hillshade	SRTMGL1 v003/ GEE	30 m	Static	NASA-JPL (2013)
	mTPI	Global SRTM mTPI/ GEE	270 m	Static	Theobald et al. (2015)

4.3 Methodology

In this study, soil moisture levels at various depths (0-5 cm, 0-10 cm, 0-20 cm, 0-50 cm, and 0-100 cm) are assessed at a spatial resolution of 70 m by 70 m. The performance of five machine learning algorithms is evaluated using a dataset comprising 730 in-situ soil moisture stations from ISMN networks across CONUS. The best-performing model's predictive ability across

different climates, depths, and networks is determined, and predictor importance across the machine learning models is examined.

Instead of developing separate models for each depth range, a single model is implemented to predict moisture levels at any depth. The methodology includes meticulous processing and analysis of remote sensing data for soil moisture prediction. This process encompasses spatial and temporal resampling, data cleaning, preprocessing of categorical input variables, dataset splitting for training and testing, feature scaling, and the assessment of prediction accuracy using correlation coefficients and error metrics. Throughout the methodology section, detailed insights into each step of the approach are provided.

4.3.1 Spatial resampling

All remote sensing datasets were projected to the NAD_1983_2011 Contiguous USA Albers projection and resampled to 70 m to match the ECOSTRESS resolution. Then, the values from the resampled products were obtained using the ISMN latitude and longitude for each in-situ soil moisture location.

4.3.2 Temporal resampling

For the dynamic products such as vegetation, meteorological variables, and SMAP soil moisture the data were collected for the study period of Jan 2019-Dec 2022. After eliminating unrealistic values, datasets with intermittent temporal coverage were linearly interpolated to an hourly time

step. ECOSTRESS products had the most missing data due to clouds and shadows, so the timestamps of cleaned ECOSTRESS data were selected as the temporal reference for the resampling method. To ensure data alignment, the timestamps of the other remote sensing products were cross-referenced with the ECOSTRESS timestamps and the corresponding data was extracted.

After resampling the data based on the time stamps of ECOSTRESS data and location of in-situ soil moisture stations. The data was cleaned and resulted in 76 independent input variables with 5 in-situ soil moisture measurement depths as dependent data. A total of 72,233 observation rows were collected. Each row contains data observation on satellite soil moisture, landcover, vegetation, soil information at five depth ranges, weather, climate, topography and in situ soil moisture at five depth ranges, at a particular timestep for a particular station. Aligning input variables based on depth is important. This means that measurements taken at the same depth are grouped together in the dataset. As a result, the dataset is restructured so that each depth range has its set of associated input variables and soil moisture values. Due to this depth-associated feature alignment, the number of input variables in the dataset is reduced from 76 to 41. Additionally, the number of columns related to in-situ soil moisture measurements decreases from 5 to 1. The resulting dataset, with depth-aligned input variables and associated soil moisture values, is considered the final dataset. This dataset is then used for training and validating machine learning models to predict soil moisture levels.

4.3.3 Input variables preprocessing

The Python library sklearn.preprocessing was used to preprocess the data (Bisong, 2019). The categorical columns Depth (0-5 cm, 0-10 cm, 0-20 cm, 0-50 cm, and 0-100 cm) and NLCD were encoded using LabelEncoder and OneHotEncoder, respectively. LabelEncoder and OneHotEncoder are machine learning techniques used to convert categorical variables into numerical variables (Jiang et al., 2020). LabelEncoder assigns a unique numerical value to each category, while OneHotEncoder creates a new binary feature for each possible value of the categorical feature. This is necessary because the machine learning algorithms can only operate on numerical data. The dataset was then divided into 70% for training and validation and 30% for testing. Stratified splitting was performed based on the in-situ soil moisture network. Stratified splitting preserves the distribution of the selected variable in both sets (Sechidis et al., 2011). This ensures that all networks are represented in the training and testing sets, and that the model is trained on a representative sample of data. Finally, feature scaling was performed using sklearn StandardScaler on training and testing datasets. StandardScaler normalizes the input variables by subtracting the mean and dividing by the standard deviation (Thara et al., 2019). This normalization ensures that the scale or units of a feature does not influence the learning process.

4.3.4 Machine Learning algorithms

Five machine learning methods were used: Feed Forward Artificial Neural Network (ANN), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Categorical Boosting (Catboost), and Light Gradient-Boosting Machine (LightGBM). All the methods were utilized as regression

tools (they can also be used as classification tools). The machine learning methods were selected because they are well-suited for modeling complex nonlinear relationships, handling high-dimensional data, processing large datasets efficiently and capturing variable interactions (Abiy et al., 2022; Breiman, 2001; Chen and Guestrin, 2016). RF, XGBoost, Catboost, and LightGBM were selected because they provide feature importance, which is a useful tool for understanding which input variables are most important for modeling soil moisture. Coarse-resolution microwave satellite products (SMAP and GPM-precipitation), fine-resolution optical and thermal ECOSTRESS products (LST, ET, ESI, and PET), MODIS vegetation products (NDVI, EVI, LAI, and fPAR), aridity index, topography, soil and NLCD datasets were used as input predictors, and in situ soil moisture stations data was used as the training data.

4.3.4.1 Feed-Forward Artificial Neural Network (ANN)

ANNs have been used for a wide range of different applications in water resources and water management, for example, assessing the potential impact of climate change scenarios on salinity levels for Florida Bay (Abiy et al., 2022), forecasting monthly water consumption based on socio-economic and climatic factors (Firat et al., 2009), and estimating root-zone soil moisture using surface soil moisture observations, meteorological data, and soil information (Kornelsen and Coulibaly, 2014). ANNs have also been used to downscale soil moisture (Adab et al., 2020; Ahmad et al., 2010; Grillakis et al., 2021; Kornelsen and Coulibaly, 2014; Senyurek et al., 2020; Singh and Gaurav, 2023).

A feed-forward artificial neural network (ANN) (McCulloch and Pitts, 1943) is inspired by the structure and functioning of the human brain. It consists of interconnected nodes, or neurons, organized into layers. The network typically comprises an input layer that receives preprocessed data, one or more hidden layers that process this data through weighted connections, and an output layer that generates predictions or outputs. Feed-forward ANNs are characterized by the fact that information flows in one direction only, from the input layer to the output layer, without any feedback loops. During the training process of the ANN algorithm, the weights of the connections between neurons are adjusted iteratively through the backpropagation process to minimize the discrepancy between the predicted outputs and the actual target values (Thirumalaiah and Deo, 1998). The optimization is typically achieved using algorithms like gradient descent and its variants, which update the weights in a direction that reduces the prediction error as quantified by a loss function. Activation functions, applied to each neuron's output, introduce non-linearity to the network, enabling it to capture complex relationships in data. Tuning the parameters of the ANN improves its performance. Six parameters are selected for this purpose. They include, 1) hidden_layer_sizes – number of hidden layers and the number of neurons in each hidden layer, 2) activation – activation function used in the hidden layers, 3) solver – algorithm used to train the model, 4) alpha – L2 regularization term to prevent overfitting , 5) learning_rate – rate at which the model learns, 6) max_iter – maximum number of iterations of the training algorithm. L2 regularization is a technique that is used to prevent overfitting in machine learning models (Demir-Kavuk et al., 2011). For more information on model parameters, see Table 13.

4.3.4.2 Random Forest (RF)

RF is a widely used machine learning algorithm due to its accuracy, and resilience to overfitting(Breiman, 2001; He et al., 2016). RF has been used for various applications in hydrology and water resources, for example, predicting daily evapotranspiration using in-situ meteorological data and satellite-based information (Douna et al., 2021), improving the accuracy of rainfall rate estimation from optical satellite data (Kühnlein et al., 2014), and accurately classifying landcover using remotely sensed data (Rodriguez-Galiano et al., 2012). It has also been used in many studies for downscaling soil moisture (Abbaszadeh et al., 2019; Abowarda et al., 2021; Adab et al., 2020; Kisekka et al., 2022; Liu et al., 2020; Park et al., 2015; Senyurek et al., 2020; Tong et al., 2020).

RF (Breiman, 2001) uses decision trees to make predictions. The algorithm works by first creating a random sample of the training data with replacement. This means that some data points may be included multiple times in the sample, while others may not be included at all. A decision tree is then grown on each sample. When growing a decision tree, the algorithm randomly selects a subset of the predictor variables to consider at each node. This helps to ensure that the trees are diverse and not too correlated with each other. The predictions of the individual trees are then averaged to produce the final prediction. This averaging process helps improve the robustness and generalization of the regression model, as it reduces the impact of individual tree outliers or overfitting (Breiman, 2001). Six RF algorithm parameters were tuned to improve its performance. They include, 1) n_estimators – number of trees in the forest, 2) max_depth – maximum depth of each tree in the forest, 3) min_samples_split – minimum number of samples required to split a node, 4) min_samples_leaf – minimum number of samples required in a leaf

node, 5) max_features – maximum number of features to consider at each split, and 6) criterion – measure used to evaluate the quality of a split. For more information on model parameters, see Table 13.

4.3.4.3 Extreme Gradient Boosting (XGBoost)

XGBoost (Chen and Guestrin, 2016) has been used for variety of applications in remote sensing and water resources including forecasting streamflow (Ni et al., 2020), predicting droughts (Zhang et al., 2019), and downscaling groundwater storage data (Zhang et al., 2021). In recent years, it has been used for downscaling of surface soil moisture (He et al., 2021; Jamei et al., 2022; Karthikeyan and Mishra, 2021; Zhang et al., 2022).

Like RF, XGBoost is an ensemble learning algorithm that utilizes decision trees as base learners for predictive tasks. However, XGBoost employs a gradient boosting framework. Decision trees are sequentially trained with each tree aiming to correct the errors made by the preceding ones. This sequential training process, combined with regularization techniques, focuses on minimizing the residual errors and enhancing predictive performance. Thus, unlike RF, XGBoost trees are not constructed independently, and there is a strong interplay between the trees during the boosting process (Chen and Guestrin, 2016). Six XGBoost algorithm parameters were tuned, including: 1) eta – controls the learning rate, 2) max_depth – maximum depth of each tree, 3) n_estimators – number of trees to be included, 4) subsample – a fraction of observations to be randomly sampled for each tree, 5) colsample_bylevel – a fraction of input variables to be randomly sampled for each new depth level, in each tree, and 6) min_child_weight – a minimum

sum of instance weight required in a leaf node to proceed with further partitioning. For more information on model parameters, see Table 13.

4.3.4.4 Categorical Boosting (Catboost)

CatBoost has been used in various water resources applications including estimating daily reference crop evapotranspiration using limited meteorological data (Zhang et al., 2020), predicting flash flood susceptibility (Saber et al., 2022), optimizing wheat yield prediction by combining Sentinel-1 and Sentinel-2 data (Uribeetxebarria et al., 2023), and soil moisture estimation (Jamei et al., 2022; Nguyen et al., 2022; Tao et al., 2023).

CatBoost (Prokhorenkova et al., 2017) is a gradient boosting decision tree algorithm like XGBoost, but it differs from XGBoost in its handling of categorical features and its approach to training weak learners. CatBoost uses a greedy algorithm to effectively combine categorical features and their interactions, and it utilizes a prior value to reduce noise from infrequent categories (Hancock and Khoshgoftaar, 2020; Prokhorenkova et al., 2017). This allows CatBoost to learn more complex relationships between categorical inputs and the target variable (Prokhorenkova et al., 2017). CatBoost also employs ordered boosting, which trains a model for each sample in the training dataset to estimate the gradient of the loss function (Prokhorenkova et al., 2017). These gradient estimates are then aggregated to construct the final model. Ordered boosting improves gradient estimation precision and reduces the risk of overfitting (Hancock and Khoshgoftaar, 2020). Six algorithm parameters were tuned, including 1) iterations– number of trees to build in the model, 2) learning_rate– step size that CatBoost takes when updating the

model parameters, 3) depth– maximum depth of each tree, 4) subsample – proportion of the training data to use at each iteration, 5) colsample_bylevel – proportion of features to use at each level of each tree, and 6) l2_leaf_reg– L2 regularization parameter. For more information on model parameters, see Table 13.

4.3.4.5 Light Gradient-Boosting Machine (LightGBM)

LightGBM has been used for enhancing landcover mapping using Sentinel-2 imagery (Candido et al., 2021), quantifying monthly surface water dynamics (Li et al., 2023), forecasting seasonal changes of regional soil moisture (Chandra et al., 2023), and soil moisture downscaling (Du et al., 2022; Huang et al., 2023; Sun et al., 2022; Yang et al., 2023).

LightGBM (Ke et al., 2017) is a gradient boosting decision tree algorithm that is designed to be fast and memory-efficient. It is similar to XGBoost but incorporates techniques to enhance performance. One of the key features of LightGBM is gradient-based one-side sampling, which selectively excludes data instances with small gradients to reduce sample size and computational demands (Ke et al., 2017). LightGBM also integrates exclusive feature bundling, which groups highly correlated features together, reducing the number of input variables that need to be considered during decision tree growth. This further enhances computational efficiency (Ke et al., 2017). LightGBM is often applied to large datasets and scenarios where speed and efficiency are important. Six algorithm parameters were tuned, including1) num_iterations– number of boosting iterations (trees to build), 2) learning_rate – Controls how much each tree contributes to the final prediction, 3) num_leaves – number of leaves in each tree, 4) max_depth– maximum

depth of each tree, 5) min_data_in_leaf – minimum number of data points in a leaf, and 6) reg_alpha– regularization term that helps to prevent overfitting. For more information on model parameters, see Table 13.

Table 13. Optimal hyperparameter settings for five machine learning models used for training and testing

Model	**Tuning Parameters**	**Defaults**
ANN	hidden_layer_sizes=(100, 100, 100), activation='relu', solver='adam', alpha=0.01, learning_rate=0.001, max_iter=100	hidden_layer_sizes=(100,), activation='relu', solver='adam', alpha=0.0001, learning_rate='constant', max_iter=200
RF	n_estimators=200, max_depth=10, min_samples_split=5, min_samples_leaf=2, max_features='sqrt', criterion='squared_error'	n_estimators=100, max_depth=None, min_samples_split=2, min_samples_leaf=1, max_features=1, criterion='squared_error'
XGBoost	eta: 0.3, max_depth : 6, n_estimators : 500, subsample : 1, colsample_bylevel : 1, min_child_weight = 0.8	eta: 0.3, max_depth : 6, n_estimators : 100, subsample : 1, colsample_bylevel : 1, min_child_weight = 1
CatBoost	iterations: 1000, learning_rate: 0.05, depth: 10, subsample: 1, colsample_bylevel: 1, l2_leaf_reg: 3	iterations: 1000, learning_rate: 0.03, depth: 6, subsample: 1, colsample_bylevel: 1, l2_leaf_reg: 3
LightGBM	num_iterations: 1000, learning_rate: 0.05, num_leaves: 31, max_depth: 10, min_data_in_leaf: 20, reg_alpha: 0.1	num_iterations: 100, learning_rate: 0.01, num_leaves: 31, max_depth: -1 (unlimited), min_data_in_leaf: 20, reg_alpha: 0.0

4.4 Evaluation of Prediction Accuracy

The Pearson correlation coefficient (R), mean bias error ($Bias$), unbiased root mean squared error ($ubRMSE$), Nash-Sutcliffe efficiency (NSE), and Kling-Gupta efficiency (KGE) were used to evaluate the accuracy of the soil moisture predictions (θ_{pred}) from the ML methods in reproducing the in-situ (θ_{obs}) observations. These metrics are calculated:

$$R = \frac{\sum_{i=1}^{N}\left(\theta_{obs,i} - \overline{\theta_{obs}}\right)\left(\theta_{pred,i} - \overline{\theta_{pred}}\right)}{\sqrt{\sum_{i=1}^{N}\left(\theta_{obs,i} - \overline{\theta_{obs}}\right)^2 \sum_{i=1}^{N}\left(\theta_{pred,i} - \overline{\theta_{pred}}\right)^2}} \tag{49}$$

$$Bias = \frac{1}{N}\sum_{i=1}^{N}\left(\theta_{pred,i} - \theta_{obs,i}\right) \tag{50}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\theta_{pred,i} - \theta_{obs,i}\right)^2} \tag{51}$$

$$ubRMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\theta_{pred,i} - \theta_{obs,i}\right)^2} - Bias \tag{52}$$

$$NSE = 1 - \left[\frac{\sum_{i=1}^{N}\left(\theta_{obs,i} - \theta_{pred,i}\right)^2}{\sum_{i=1}^{N}\left(\theta_{obs,i} - \overline{\theta_{obs}}\right)^2}\right] \tag{53}$$

$$KGE = 1 - \sqrt{\left(R-1\right)^2 + \left(\alpha-1\right)^2 + \left(\beta-1\right)^2} \tag{54}$$

R describes the linear correlation between the observed and simulated values (Paulik et al., 2014). $Bias$ describes whether the model typically overestimates or underestimates its predictions (Morales et al., 2005). $RMSE$ measures the average squared difference between the prediction and the observations. $ubRMSE$ considers the error that remains if the bias is removed

from the model prediction (Entekhabi et al., 2010b). The NSE compares the squared error to the variance of the observations (Li et al., 2022; Nash and Sutcliffe, 1970) . KGE (Gupta et al., 2009) combines three measures of model performance: R, $\beta = \mu_{pred} / \mu_{obs}$, and $\alpha = \left(\sigma_{pred} / \sigma_{obs} \right)$, where μ_{pred} and μ_{obs} are the predicted and observed means and σ_{pred} and σ_{obs} are the predicted and observed standard deviations. NSE ranges from -∞ to +1, with a value of +1 indicating a perfect model. The KGE ranges from -∞ to 1, with values closer to 1 indicating that the model predictions are more accurate.

4.5 Results and discussion

Table 14 presents the correlations between all the input variables used in this study and the in-situ soil moisture at different depths. The correlations between the in-situ soil moisture values at different depths indicate that the soil moisture at a given depth is most related to nearby layers. It also shows that the soil moisture values for 0-100 cm are most different from other layers.

Antecedent soil moisture from SMAP is highly correlated with in-situ soil moisture. It is interesting that SMAP surface soil moisture exhibits higher correlations with deep in-situ soil moisture than the SMAP RZSM and PSM products. This behavior likely occurs because the surface soil moisture is inferred more directly from the satellite sensors while the RZSM and PSM products rely in part on models. Current SMAP soil moisture values and antecedent SMAP soil moisture values exhibit similar correlations with the in-situ soil moisture values.

Vegetation indices such as NDVI and EVI exhibit moderate correlations to the in-situ soil moisture values, with EVI displaying the highest correlations. EVI is less sensitive to variations in canopy structure as well as background factors such as soil conditions and atmospheric influences, and it is more sensitive to changes in canopy chlorophyll content, which make it better indicator of plant health and moisture stress than NDVI (Jamali et al., 2011; Méndez-Barroso et al., 2009; Santos et al., 2014). Among the vegetation indices, LAI usually exhibits the weakest correlations with soil moisture, perhaps because it is less indicative of the degree of soil surface shading than NDVI or EVI.

Soil moisture exhibits positive correlations with clay and silt content, and negative correlations with sand content. Higher sand content (and lower clay and silt content) is expected to increase drainage and reduce soil moisture. Wang et al. (2017) found that the spatial pattern of soil moisture is greatly influenced by soil factors, such as the sand and clay fractions. Bulk density, organic matter, and electrical conductivity exhibit low correlations with soil moisture. Higher bulk densities and lower organic matter are expected to reduce porosity, which would reduce the allowable range for soil moisture. However, errors in the estimated bulk density and organic matter values may cause the low correlations seen in Table 14. pH exhibits relatively strong negative correlations with in-situ soil moisture. Soil pH plays a role in determining microbial activity, nutrient availability, and soil structure, which indirectly influence the soil moisture (Zhang et al., 2019).

Positive correlations occur between in-situ soil moisture and precipitation, which is expected because precipitation replenishes soil moisture. The correlations are strongest for 14-day antecedent precipitation and weakens as the time period shortens to 6 hr. This result highlights soil moisture's memory (i.e., its ability to retain information about past precipitation events (Martínez-Fernández et al., 2021; Orth and Seneviratne, 2012; Song et al., 2019)). The evaporative stress index (ESI) and instantaneous actual evapotranspiration (ETins) are positively correlated with soil moisture, meaning that higher soil moisture allows higher evapotranspiration rates. Land Surface Temperature (LST) exhibits a weak negative correlation with soil moisture. Lower soil moisture reduces the latent heat flux and warms the land surface. Moreover, aridity index is positively correlated with soil moisture, reflecting greater availability of water in more humid climates.

Topographic variables also exhibit correlations with in-situ soil moisture. Elevation displays a strong negative correlation with soil moisture. Rong et al. (2017) reported that soil moisture at low elevation is often supplemented by surface runoff and subsurface flow from higher elevation points. This leads to a negative correlation between soil moisture and elevation. However, Cowley et al. (2017) found that soil moisture has a positive correlation with elevation at Reynolds Creek watershed in southern Idaho, with higher elevations having much more soil moisture than lower elevations. Slope exhibits a negative correlation, likely because it promotes lateral outflow of moisture (Jawson and Niemann, 2007; Ranney et al., 2015). Aspect, hillside, and mTPI exhibit only weak correlations.

Table 14. Correlation table between indicator variables and in-situ soil moisture layers. Non colored cells indicate a statistically insignificant correlation between two variables, while the positive correlation is shown in blue and negative correlation is shown in red color.

Factors	Parameters	Acronyms	Depth (cm)	Soil moisture 0-5	0-10	0-20	0-50	0-100
In-situ Soil moisture	Soil Moisture	SM	0-5	1	0.915	0.822	0.852	0.432
			0-10	0.915	1	0.908	0.858	0.460
			0-20	0.822	0.908	1	0.853	0.451
			0-50	0.852	0.858	0.853	1	0.537
			0-100	0.432	0.460	0.451	0.537	1
Antecedent Soil Moisture	SMAP surface soil moisture (0-5 cm)	SSM	0-5	0.492	0.527	0.531	0.476	0.390
	SMAP Root zone soil moisture (0-100 cm)	RZSM	0-100	0.400	0.438	0.455	0.406	0.357
	SMAP profile soil moisture (0-bedrock depth)	PSM	--	0.379	0.418	0.432	0.392	0.340
	Antecedent 1-Day SMAP SSM	A1DAY_SSM	0-5	0.498	0.534	0.536	0.482	0.394
	Antecedent 1-Day SMAP RZSM	A1DAY_RZSM	0-100	0.400	0.439	0.455	0.407	0.358
	Antecedent 1-Day SMAP PSM	A1DAY_PSM	--	0.379	0.418	0.432	0.392	0.341
	Antecedent 3-Day SMAP SSM	A3DAY_SSM	0-5	0.501	0.538	0.542	0.488	0.398
	Antecedent 3-Day SMAP RZSM	A3DAY_RZSM	0-100	0.399	0.437	0.454	0.406	0.358
	Antecedent 3-Day SMAP PSM	A3DAY_PSM	--	0.377	0.416	0.431	0.391	0.341
	Antecedent 7-Day SMAP SSM	A7DAY_SSM	0-5	0.498	0.536	0.543	0.490	0.400
	Antecedent 7-Day SMAP RZSM	A7DAY_RZSM	0-100	0.394	0.432	0.450	0.404	0.356
	Antecedent 7-Day SMAP PSM	A7DAY_PSM	--	0.373	0.412	0.427	0.388	0.339
	Antecedent 14-Day SMAP SSM	A14DAY_SSM	0-5	0.490	0.528	0.538	0.488	0.398
	Antecedent 14-Day SMAP RZSM	A14DAY_RZSM	0-100	0.387	0.425	0.444	0.398	0.352
	Antecedent 14-Day SMAP PSM	A14DAY_PSM	--	0.366	0.405	0.421	0.383	0.336
Vegetation	Normalized Difference Vegetation Index	NDVI	--	0.149	0.151	0.143	0.136	0.106
	Enhanced Vegetation Index	EVI	--	0.175	0.193	0.191	0.191	0.179
	Leaf Area Index	LAI	--	0.091	0.085	0.073	0.090	0.086
	fraction of Photosynthetically Active Radiation	fPAR	--	0.095	0.094	0.089	0.086	0.089
Soil Properties and Landcover	Landcover	2019_NLCD	--	0.149	0.195	0.184	0.213	0.297
	Silt	Silt	0-5	0.184	0.226	0.217	0.254	0.273
			0-10	0.186	0.230	0.218	0.255	0.272
			0-20	0.191	0.237	0.222	0.256	0.270
			0-50	0.175	0.221	0.197	0.239	0.260
			0-100	0.172	0.214	0.186	0.234	0.245
	Sand	Sand	0-5	-0.213	-0.267	-0.258	-0.262	-0.219
			0-10	-0.227	-0.285	-0.277	-0.278	-0.242
			0-20	-0.229	-0.292	-0.288	-0.291	-0.261
			0-50	-0.232	-0.297	-0.306	-0.300	-0.286
			0-100	-0.211	-0.270	-0.288	-0.275	-0.273
	Clay	Clay	0-5	0.187	0.229	0.258	0.222	0.224
			0-10	0.188	0.231	0.259	0.223	0.218
			0-20	0.190	0.236	0.266	0.229	0.207
			0-50	0.229	0.275	0.287	0.281	0.242
			0-100	0.246	0.290	0.292	0.288	0.245
	Bulk Density	BD	0-5	-0.019	-0.037	-0.054	-0.050	-0.116
			0-10	-0.013	-0.035	-0.050	-0.040	-0.106
			0-20	-0.011	-0.033	-0.050	-0.038	-0.100
			0-50	-0.017	-0.038	-0.059	-0.046	-0.112
			0-100	-0.022	-0.038	-0.060	-0.051	-0.110
	Organic Matter	OM	0-5	0.039	0.047	0.067	0.048	0.108
			0-10	0.043	0.049	0.065	0.040	0.091
			0-20	0.053	0.055	0.066	0.040	0.078
			0-50	0.048	0.046	0.026	0.042	0.068
			0-100	0.069	0.063	0.024	0.060	0.057
	Electrical Conductivity	EC	0-5	-0.052	-0.050	-0.048	-0.044	-0.038
			0-10	-0.053	-0.051	-0.050	-0.041	-0.031
			0-20	-0.073	-0.066	-0.064	-0.044	-0.019
			0-50	-0.080	-0.067	-0.063	-0.035	0.002
			0-100	-0.079	-0.064	-0.060	-0.029	0.012
	pH	pH	0-5	-0.136	-0.126	-0.099	-0.118	-0.100
			0-10	-0.142	-0.134	-0.102	-0.124	-0.114
			0-20	-0.144	-0.137	-0.106	-0.129	-0.127
			0-50	-0.137	-0.129	-0.098	-0.119	-0.119
			0-100	-0.123	-0.115	-0.086	-0.103	-0.109
	Depth to Restrictive Layer	DRL	--	0.022	0.021	0.010	0.017	0.015
Precipitation and Climate Demand	Antecedent 6-hour Precipitation	P6H	--	0.037	0.036	0.033	0.033	0.031
	Antecedent 1-Day Precipitation	P1DAY	--	0.104	0.111	0.104	0.096	0.088
	Antecedent 3-Day Precipitation	P3DAY	--	0.227	0.239	0.225	0.212	0.198
	Antecedent 7-Day Precipitation	P7DAY	--	0.311	0.331	0.315	0.297	0.278
	Antecedent 14-Day Precipitation	P14DAY	--	0.365	0.393	0.380	0.361	0.336
	Land Surface Temperature	LST	--	-0.109	-0.108	-0.083	-0.063	-0.041
	Evaporative Stress Index	ESI	--	0.248	0.267	0.286	0.232	0.180
	Potential Evapotranspiration	PET	--	-0.044	-0.048	-0.041	-0.029	-0.007
	Instantaneous Evapotranspiration	ETinst	--	0.148	0.160	0.177	0.148	0.132
	Aridity Index	AI	--	0.166	0.176	0.173	0.185	0.185
Topography	Elevation	elevation	--	-0.226	-0.259	-0.296	-0.278	-0.365
	Slope	slope	--	-0.169	-0.189	-0.207	-0.214	-0.181
	Aspect	aspect	--	0.025	0.023	0.031	0.046	-0.075
	Hillshade	hillshade	--	0.005	0.011	0.014	0.015	-0.001
	Multi-Scale Topographic Position Index	mTPI	--	0.039	0.042	0.055	0.055	0.055

Figure 18 shows the performance metrics for the machine learning methods' soil moisture predictions across all networks and depth ranges. This evaluation includes the entire dataset, including training and testing data, and the boxes and whiskers characterize the range in performance among different stations in the dataset. The results indicate that the XGBoost, CatBoost, and RF models exhibit better performance than the LightGBM and ANN models. In particular, the XGBoost, CatBoost, and RF models typically have lower RMSE and ubRMSE values and higher R, KGE, and NSE values. These same three models also tend to have smaller biases than the LightGBM and ANN models. Overall, the XGBoost method exhibits the best prediction accuracy among the methods. Previous studies have shown that XGBoost outperforms other machine learning algorithms for soil moisture prediction. For example, Jamei et al. (2022) found that XGBoost outperformed CatBoost for rootzone soil moisture prediction in a semiarid region of Iran. This was attributed to XGBoost's ability to learn complex relationships between the input and output variables, and to avoid overfitting. Karthikeyan and Mishra (2021) also found that XGBoost better captures the temporal dynamics of soil moisture and reduces bias and variance through gradient boosting. Similarly, Zhang et al. (2019) reported that the XGBoost model outperforms the ANN model due to its resistance to overfitting.

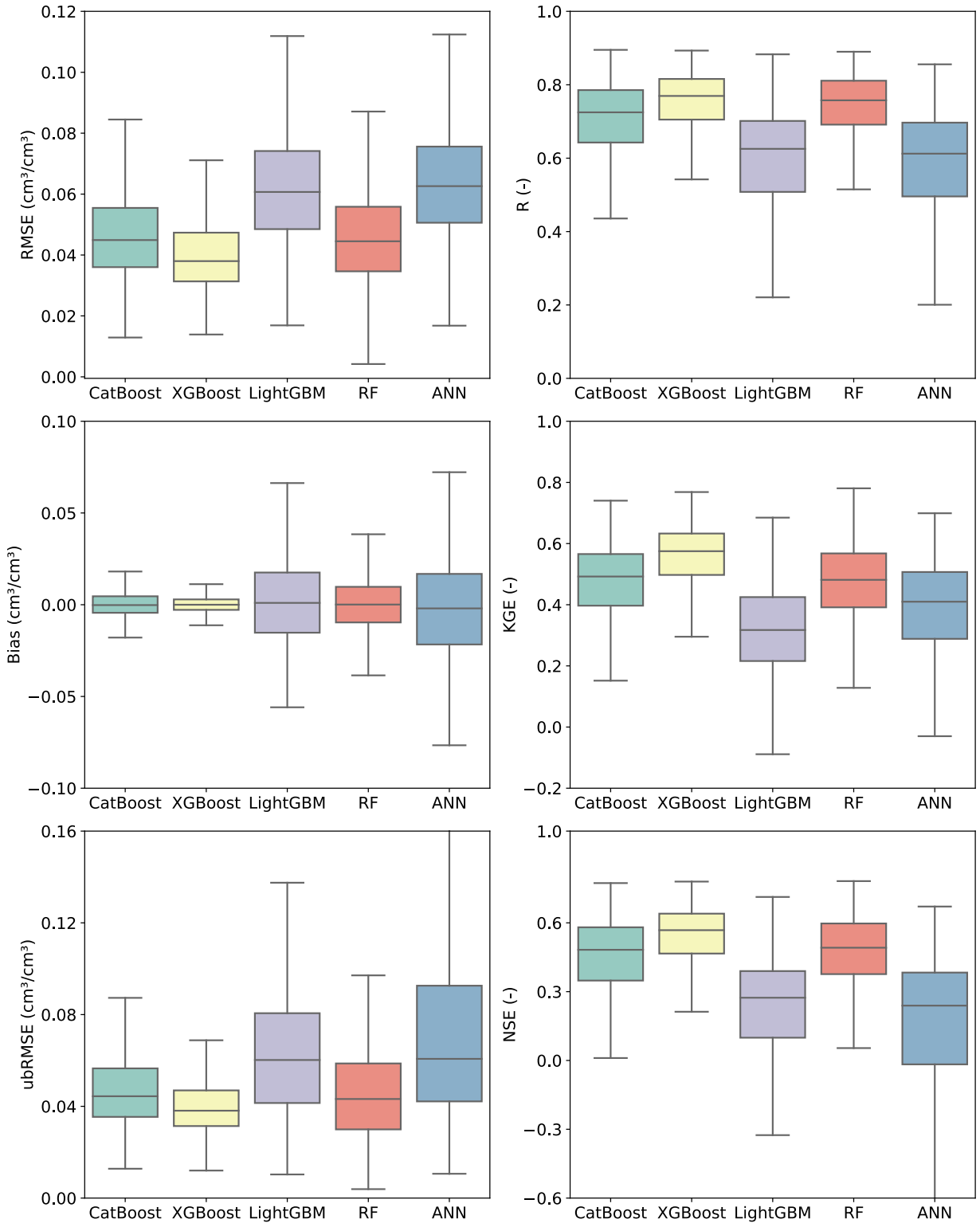

Figure 18. Boxplots of ubRMSE, R, Bias, and KGE for comparison of five machine learning algorithms using the modeled soil moisture predictions compared to observed data at all five depth layers across all stations.

Figure 19a examines the RMSE of the machine learning methods when the testing data are divided according to the in-situ soil moisture data collection networks (ARM, COSMOS, FLUXNET, SCAN, SNOTEL, TxSON, USCRN, and IRON). FLUXNET displays the lowest median RMSE values across all machine learning methods, but most data collection networks exhibit similar performance. The COSMOS network exhibits much higher RMSE values than the other networks. The poor performance at the COSMOS stations likely occurs because the spatial support for the cosmic ray neutron measurements (approx. 700 m diameter) is much larger (Zreda et al., 2008) than the support for the in-situ probes used in the other networks (centimeters at most). The machine learning methods are trained on all networks simultaneously, so the COSMOS data are inconsistent with the other datasets. Furthermore, no input variable allows the machine learning methods to identify whether an in-situ measurement is from the COSMOS dataset or the other networks. Finally, all input variables were represented at a 70 m resolution, so the machine learning methods lack information to characterize much of the spatial support for the COSMOS data.

Figure 19b compares the performance of the machine learning methods when the testing data are divided according to depth. As the depth increases, the median RMSE decreases for most methods. The largest improvement in performance occurs between 0-50 cm and 0-100 cm depth. The improved performance for deeper layers may occur due to the greater uniformity of soil moisture at greater depths, which facilitates model learning and prediction. XGBoost consistently has the lowest RMSE values across all depths, followed by RF, CatBoost, LightGBM, and ANN.

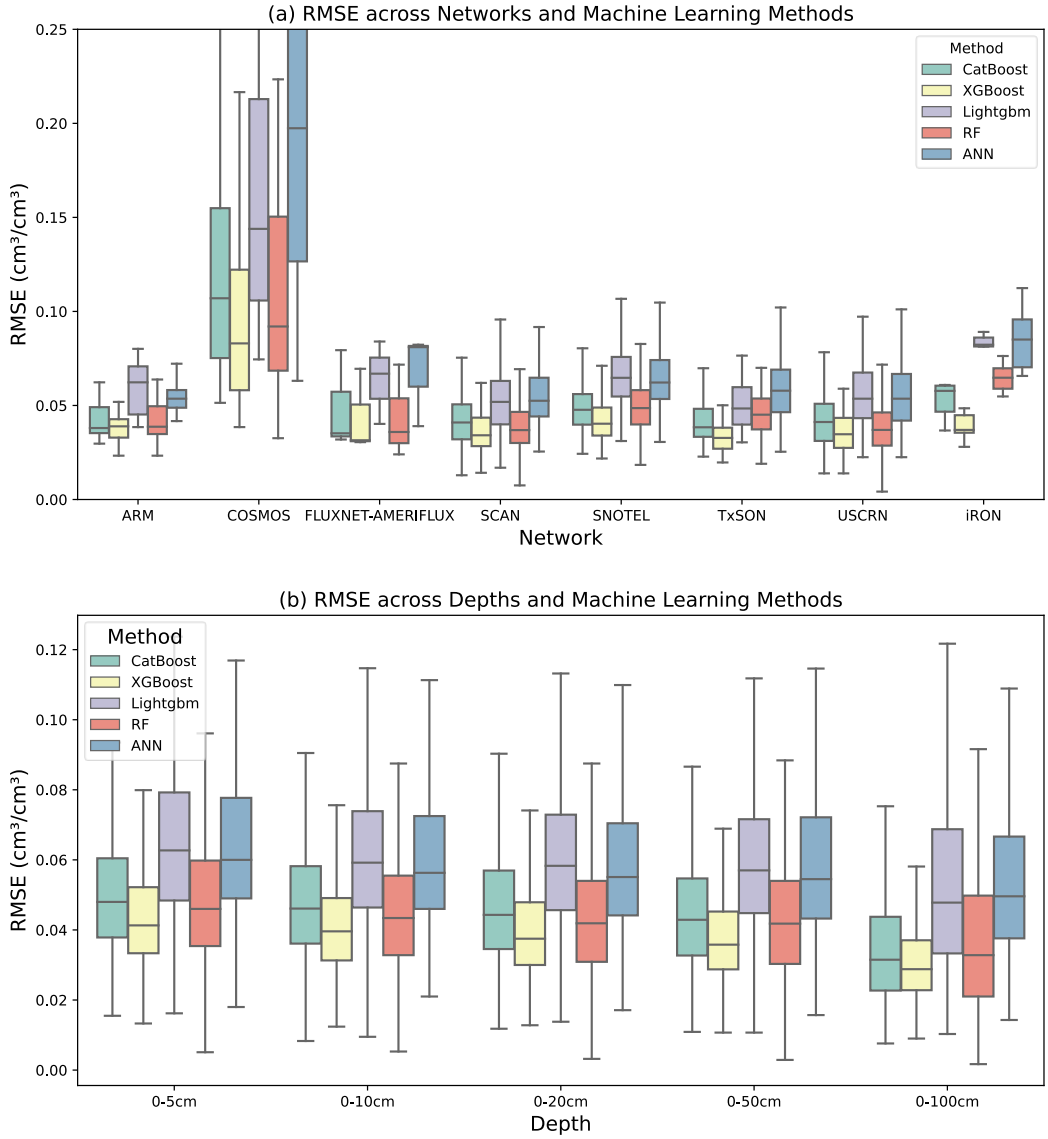

Figure 19. Boxplot of RMSE across networks and depths using in-situ vs. machine learning methods results.

Table 15 separately evaluates the performance of the ML methods for the training and testing datasets. The CatBoost, XGBoost, and RF methods again outperform the ANN and LightGBM methods, and the XGBoost method exhibits the best performance among the models. The ANN and LightGBM methods exhibit lower performance for the training datasets, which indicates

they have a weaker ability to identify and model the dependence of soil moisture on the input variables. However, the ANN and LightGBM exhibit the smallest changes in performance between the training and testing datasets, which suggests that their worse performance is not due to overfitting the data.

Table 15. Quantitative performance evaluation of machine learning models using deterministic metrics: R, ubRMSE, NSE, and KGE

Method	Label	R	RMSE	bias	ubRMSE	KGE	NSE
ANN	Training	0.733	0.065	0.002	0.064	0.623	0.528
	Testing	0.726	0.067	0.001	0.065	0.619	0.517
RandomForest	Training	0.824	0.047	0.000	0.047	0.630	0.660
	Testing	0.806	0.051	0.000	0.051	0.616	0.633
XGBoost	Training	0.841	0.040	0.000	0.040	0.692	0.696
	Testing	0.827	0.044	0.000	0.044	0.683	0.675
CatBoost	Training	0.816	0.048	0.000	0.048	0.647	0.653
	Testing	0.808	0.050	0.000	0.050	0.643	0.641
LightGBM	Training	0.747	0.064	0.000	0.064	0.530	0.539
	Testing	0.742	0.065	0.000	0.065	0.527	0.533

Figure 20 displays density plots comparing XGBoost soil moisture predictions to the in-situ soil moisture observations, utilizing the testing datasets across all depths within the four climatic regions. The results show that soil moisture predictions from the XGBoost model are relatively accurate across all climatic regions, with R exceeding 0.8 and RMSE below 0.05 for all regions. The weakest performance occurs in the semiarid region, which is the only region where the KGE and NSE values are below 0.70. This behavior contrasts with performance of optical/thermal soil moisture estimates, which performed most poorly for humid climates where soil moisture does not control the evapotranspiration rates (Sahaar et al., 2022; Sahaar and Niemann, 2020). To train and test the machine learning methods, 59 stations were used in the arid region, 401

stations in the semiarid region, 86 stations in the subhumid region, and 185 stations in the humid region. One reason for the lower RMSE in the semiarid region is that it may be more diverse than the other regions due to the large number of stations in that climate. This means that there may be a wider range of soil types, vegetation types, and other factors that affect soil moisture in the semiarid region, which can make it more difficult for the machine learning model to learn the patterns of soil moisture in the semiarid region.

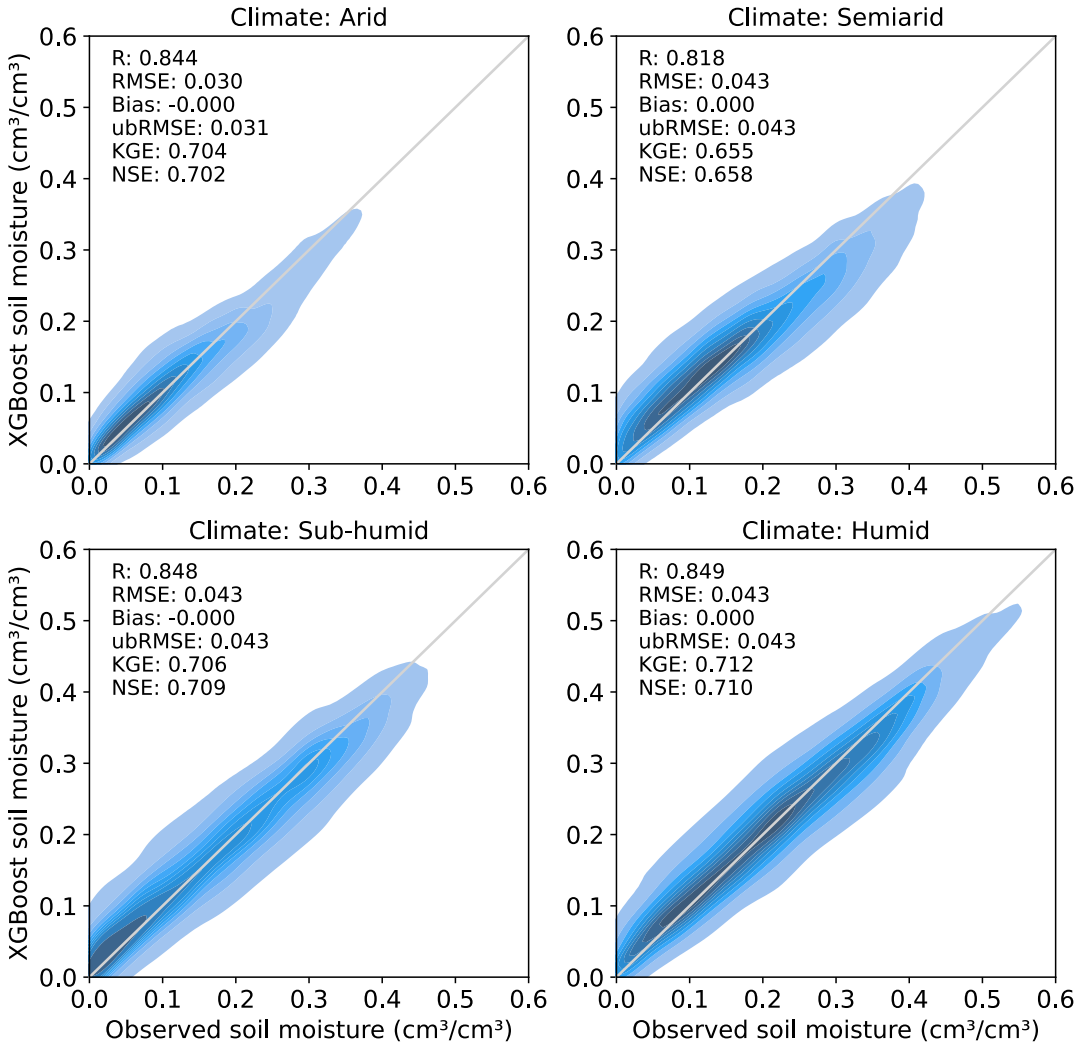

Figure 20. Density plot of observed vs. XGBoost soil moisture predictions for each depth across testing datasets for each climatic region.

Figures 21 and 22 provide example scatterplots illustrating the relationships between XGBoost and observed soil moisture for the surface (0-5 cm) and rootzone (0-100 cm), respectively. A single station from the testing dataset is shown for each climatic zone in each subplot. The figures also include the performance metrics at each station. The results show that the XGBoost model generally predicts surface (0-5 cm) soil moisture well for all four stations (Figure 21), with coefficients of correlation (R) ranging from 0.76 to 0.86. The RMSE values for the four stations range from 0.024 cm^3/cm^3 to 0.039 cm^3/cm^3, which indicates that the XGBoost model is able to predict soil moisture with a high degree of accuracy. The KGE and NSE values for the four stations range from 0.511 to 0.696 and from 0.551 to 0.694, respectively, which indicates that the XGBoost model is able to capture the overall variability of soil moisture well. The results show strong correlations between rootzone soil moisture and XGBoost soil moisture, with R values ranging from 0.76 to 0.86 for 0-100 cm soil moisture (Figure 22). The data also typically fall close to the 1:1 lines. At the arid and humid sites, XGBoost slightly underestimates soil moisture, with biases of -0.009 cm^3/cm^3 and -0.005 cm^3/cm^3, respectively. XGBoost slightly overestimates soil moisture at the sub-humid site (with a bias of 0.010 cm^3/cm^3).

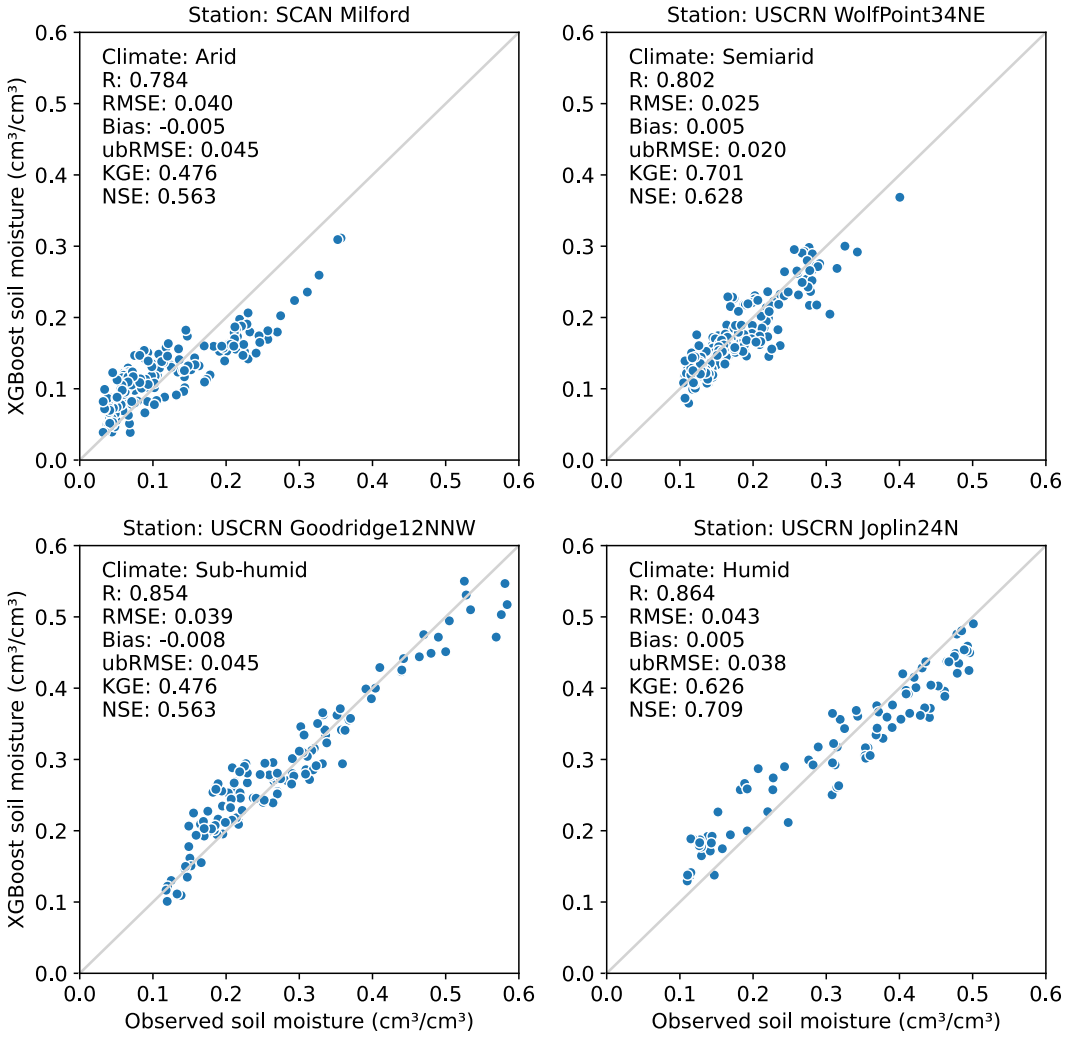

Figure 21. Comparison of XGBoost soil moisture predictions with observed soil moisture at depth of 0-5 cm (surface)

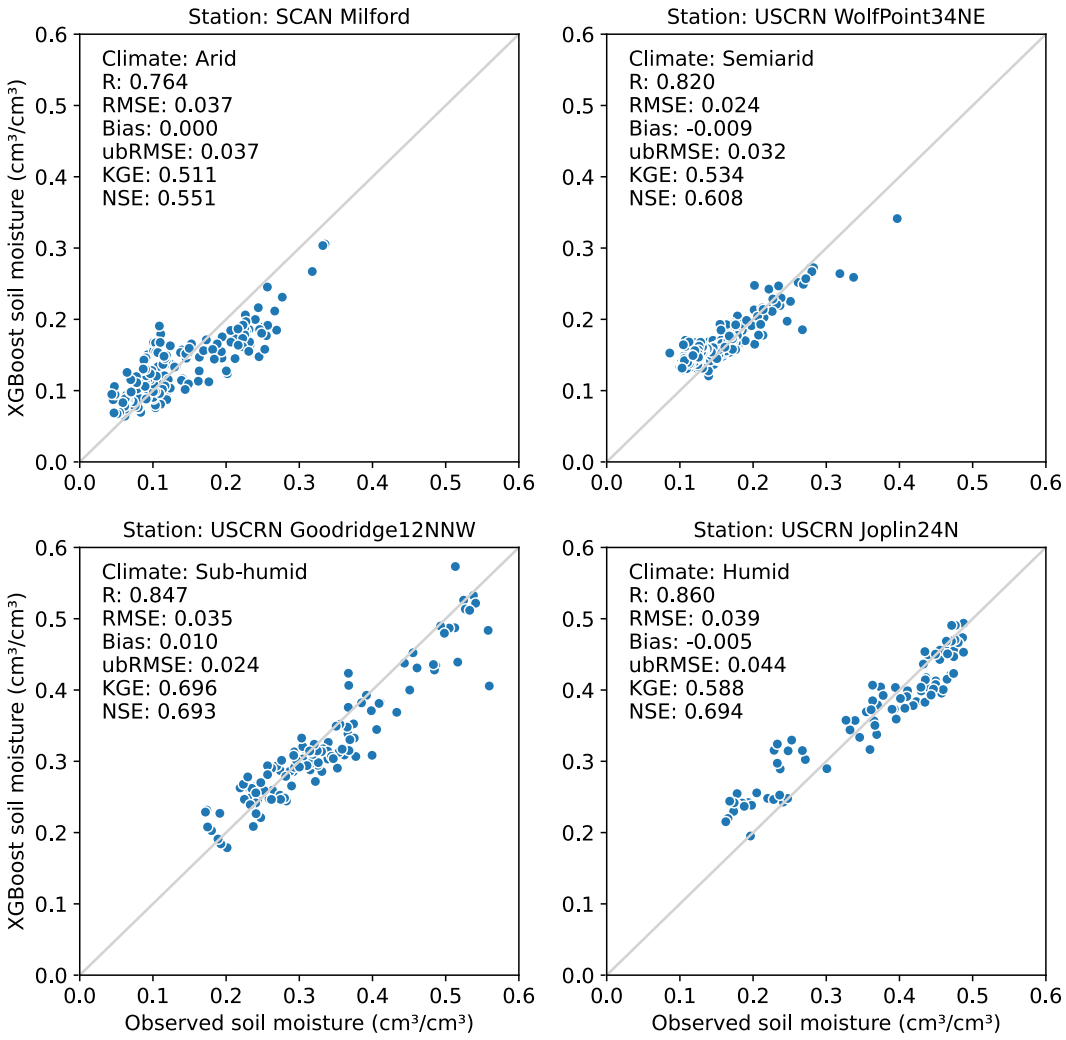

Figure 22. Comparison of XGBoost soil moisture predictions with observed soil moisture at depth of 0-100 cm (rootzone)

Figure 23 examines the temporal dynamics of soil moisture for the in-situ measurements, SMAP L4 product, and XGBoost model for the arid USCRN LasCruces20N site (a member of the testing dataset). The temporal variations in the XGBoost soil moisture closely follow the variations in the in-situ soil moisture at both depths. XGBoost typically overestimates the soil moisture at both depths for this site, but the magnitude of its wet bias is smaller than the

magnitude of dry bias if the SMAP L4 soil moisture product is used as an estimate of the in-situ soil moisture. The correlations between XGBoost and in-situ soil moisture are 0.79 and 0.75 for the surface and rootzone, respectively. The RMSE values are 0.016 cm^3/cm^3 for the surface and 0.017 cm^3/cm^3 for the rootzone. In contrast, SMAP has lower correlations (0.64 for surface and 0.70 for rootzone) and higher RMSE values (0.043 cm^3/cm^3 for surface and 0.054 cm^3/cm^3 for rootzone). Notably, small positive bias (0.012 cm^3/cm^3) is exhibited by XGBoost for both surface and rootzone soil moisture, whereas SMAP displays substantial negative bias (-0.035 cm^3/cm^3 for surface and -0.053 cm^3/cm^3 for rootzone).

Figure 24 examines the temporal variability of soil moisture at the humid USCRN Versailles3NNW site. The correlations between the XGBoost and in-situ soil moisture are 0.80 for the surface layer and 0.76 for the rootzone. Furthermore, the RMSE remains low at 0.043 cm^3/cm^3 for surface and 0.036 cm^3/cm^3 for rootzone predictions. Lower correlations (0.62 for the surface and 0.69 for the rootzone) and higher RMSE values (0.124 cm^3/cm^3 for both the surface and the rootzone) are observed for SMAP, indicating that SMAP provides less accurate estimates of the in-situ soil moisture than XGBoost. The XGBoost model exhibits small biases (0.007 cm^3/cm^3 for the surface and 0.013 cm^3/cm^3 for the rootzone), while substantial positive bias occur if SMAP is used as an estimate of the in-situ soil moisture (0.107 cm^3/cm^3 for the surface and 0.124 cm^3/cm^3 for the rootzone).

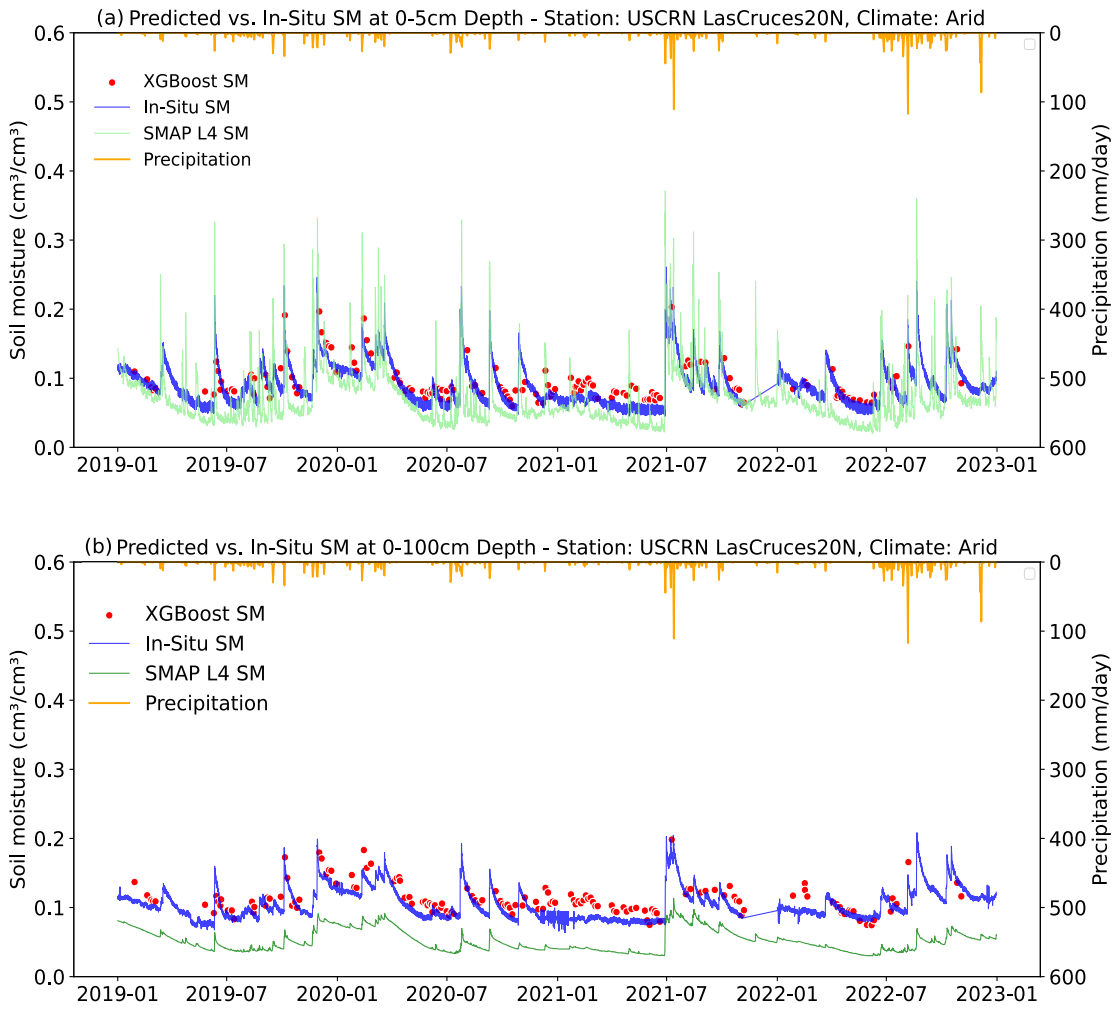

Figure 23.Time series of soil moisture at (a) 0-5 cm and (b) 0-100 cm depths at the arid USCRN Las Cruces 20N station. The plotted soil moisture data includes hourly in-situ measurements, predictions from the XGBoost model, and 3-hr SMAP L4 soil moisture data. Daily local Global Precipitation Measurement precipitation data are also shown.

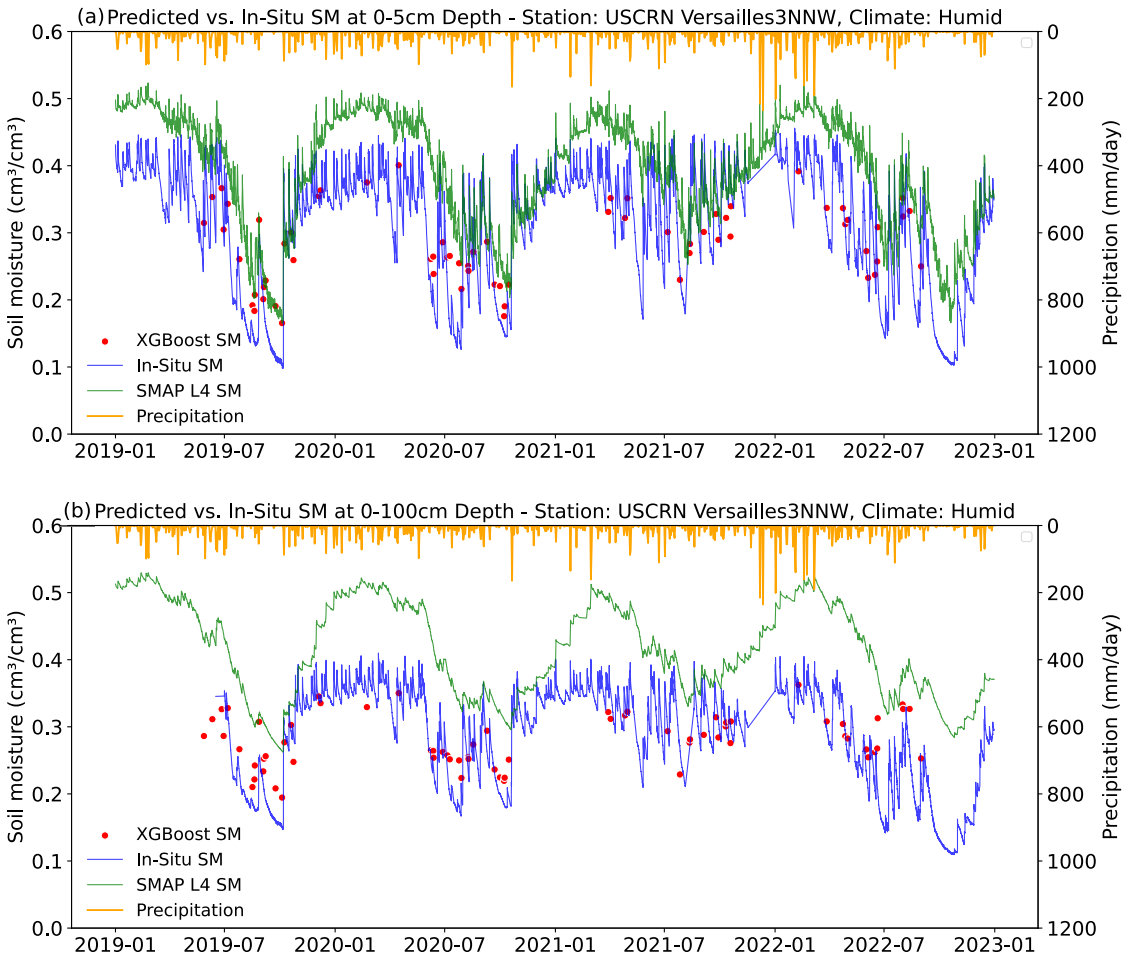

Figure 24. Time series of soil moisture at (a) 0-5 cm and (b) 0-100 cm depths at the humid USCRN Versailles 3NNW station. The plotted soil moisture data include hourly in-situ measurements, predictions from the XGBoost model, and 3-hr SMAP L4 soil moisture data. Daily local Global Precipitation Measurement precipitation data are also shown.

Table 16 presents the input variable importance scores for each machine learning method. These scores differ substantially between the methods. These differences likely suggest that different input variables contain analogous information, so different variables can be selected depending on the methods' algorithmic approaches and learning architectures (Sahin, 2020). Among topographic variables, elevation was most important to the model predictions. However, the

remaining topographic factors all typically had substantial importance (all were in the top half of the table in terms of average importance). Among soil texture variables, percent clay was most important for 3 of 4 models (LightGBM relied more heavily on silt and clay). Soil depth, organic matter, and bulk density were also somewhat important. Vegetation played a smaller role than topography and soil. Among vegetation variables, EVI was most important, followed by NDVI and LAI. All these were towards the middle in terms of importance. Land cover classification provided little value in predicting soil moisture. Among meteorological variables, the aridity index was by far the most important. Among temporally varying climate inputs, 14-day antecedent precipitation was most important. The other climate variables had only moderate importance (shorter antecedent precipitation as well as PET, ESI, and actual ET). Considering the SMAP soil moisture products, the most important products were all surface soil moisture. All RZSM and PSM products had low importance. Also, the most important SMAP products were 3-day and 14-day antecedent soil moisture products (except current SMAP surface soil moisture is used heavily by the RF model).

Table 16. Relative Importance of Predictors and Average Performance of Factors from Four Machine Learning Models

Feature	CatBoost Importance	XGBoost Importance	LightGBM Importance	Random Forest Importance	Average Importance
Elevation	1.00	0.58	0.79	0.94	0.83
Aridity Index	0.84	0.33	0.72	0.67	0.64
Clay	0.79	0.28	0.81	0.66	0.64
Depth	0.90	0.25	0.86	0.47	0.62
Antecedent 3-Day SMAP SSM	0.13	1.00	0.47	0.88	0.62
Antecedent 14-Day SMAP SSM	0.59	0.36	0.72	0.75	0.60
Sand	0.51	0.36	0.80	0.74	0.60
Silt	0.52	0.27	0.93	0.65	0.59
Multi-Scale Topographic Position Index	0.78	0.26	0.68	0.54	0.56
Organic Matter	0.61	0.18	0.95	0.51	0.56
Aspect	0.68	0.23	0.84	0.49	0.56
Antecedent 7-Day SMAP SSM	0.38	0.22	0.49	1.00	0.52
Hillshade	0.62	0.17	0.60	0.46	0.46
Bulk Density	0.41	0.18	0.76	0.50	0.46
Antecedent 14-Day Precipitation	0.29	0.10	1.00	0.42	0.45
Enhanced Vegetation Index	0.27	0.08	0.95	0.29	0.40
fraction of Photosynthetically Active Radiation	0.25	0.08	0.92	0.31	0.39
Antecedent 14-Day SMAP SSM	0.22	0.14	0.42	0.78	0.39
Normalized Difference Vegetation Index	0.18	0.11	0.93	0.31	0.38
pH	0.34	0.15	0.59	0.43	0.38
Slope	0.61	0.12	0.38	0.37	0.37
Evaporative Stress Index	0.23	0.07	0.81	0.26	0.34
Antecedent 7-Day Precipitation	0.27	0.06	0.76	0.26	0.34
Leaf Area Index	0.22	0.07	0.76	0.26	0.33
SMAP surface soil moisture (0-5 cm)	0.08	0.05	0.48	0.68	0.32
Land Surface Temperature	0.20	0.06	0.71	0.19	0.29
Depth to Restrictive Layer	0.42	0.08	0.29	0.24	0.26
Antecedent 6-hour Precipitation	0.03	0.01	0.85	0.03	0.23
Antecedent 14-Day SMAP RZSM	0.09	0.05	0.46	0.31	0.23
Antecedent 3-Day Precipitation	0.11	0.04	0.61	0.15	0.23
Antecedent 14-Day SMAP PSM	0.11	0.04	0.46	0.29	0.22
SMAP Root zone soil moisture (0-100 cm)	0.05	0.05	0.36	0.38	0.21
Instantaneous Evapotranspiration	0.06	0.03	0.55	0.16	0.20
Potential Evapotranspiration	0.09	0.02	0.52	0.13	0.19
Antecedent 1-Day Precipitation	0.04	0.01	0.62	0.07	0.19
SMAP profile soil moisture (0-bedrock depth)	0.05	0.03	0.32	0.34	0.19
Antecedent 7-Day SMAP RZSM	0.09	0.02	0.25	0.38	0.18
Antecedent 1-Day SMAP RZSM	0.06	0.01	0.15	0.49	0.18
Antecedent 3-Day SMAP RZSM	0.04	0.03	0.19	0.44	0.17
Antecedent 7-Day SMAP PSM	0.05	0.02	0.23	0.26	0.14
Antecedent 1-Day SMAP PSM	0.06	0.01	0.16	0.30	0.13
Antecedent 3-Day SMAP PSM	0.03	0.01	0.19	0.28	0.13
2019_NLCD_Shrub, scrub	0.32	0.02	0.02	0.11	0.12
2019_NLCD_Cultivated crops	0.16	0.07	0.03	0.09	0.09
2019_NLCD_Evergreen forest	0.14	0.02	0.02	0.06	0.06
2019_NLCD_Grassland, herbaceous	0.06	0.02	0.04	0.04	0.04
2019_NLCD_Deciduous forest	0.04	0.01	0.03	0.03	0.03
2019_NLCD_Developed, open space	0.02	0.01	0.02	0.03	0.02
2019_NLCD_Pasture, hay	0.01	0.00	0.02	0.02	0.01
2019_NLCD_Mixed forest	0.00	0.01	0.02	0.02	0.01
2019_NLCD_Emergent herbaceous wetlands	0.01	0.01	0.01	0.01	0.01
2019_NLCD_Woody wetlands	0.00	0.00	0.01	0.03	0.01
2019_NLCD_Developed, low intensity	0.00	0.00	0.00	0.01	0.00
2019_NLCD_Developed, medium intensity	0.00	0.00	0.00	0.00	0.00
2019_NLCD_Barren land (rock, sand, clay)	0.00	0.00	0.00	0.00	0.00

4.6 Conclusion

The primary purpose of this study was to evaluate and compare machine learning models for predicting soil moisture levels across different climates and depths. The study assesses the performance of several machine learning models, including XGBoost, CatBoost, RF, LightGBM, and ANN, in predicting soil moisture.

To summarize the main conclusions are as follows:

1. Antecedent soil moisture from SMAP is highly correlated with in-situ soil moisture, and vegetation indices such as EVI and NDVI exhibit moderate correlations to in-situ soil moisture values. Soil moisture also exhibits positive correlations with clay and silt content, and negative correlations with sand content. Topographic variables such as elevation and slope exhibited negative correlations with in-situ soil moisture.
2. XGBoost outperforms other machine learning algorithms for soil moisture prediction, including LightGBM, ANN, CatBoost, and RF. It has lower RMSE and ubRMSE values and higher R, KGE, and NSE values than the other models. XGBoost also has smaller biases than the other models.
3. The performance of machine learning methods for predicting soil moisture varies depending on the data collection network and the depth of the soil measurement. The COSMOS network exhibits the worst performance due to the large spatial support of its cosmic ray neutron measurements than the support for the in-situ probes used in the other networks. The performance of the machine learning methods improves as the depth of the soil measurement increases.
4. The XGBoost model exhibits high accuracy for predicting soil moisture, both at the surface and in the rootzone. The model performed well in all four climatic regions

considered, with correlation coefficient ranging from 0.76 to 0.86, RMSE values ranging from 0.024 cm^3/cm^3 to 0.039 cm^3/cm^3, and KGE and NSE values ranging from 0.511 to 0.696 and from 0.551 to 0.694, respectively. XGBoost model also outperforms the SMAP L4 product in predicting soil moisture timeseries patterns at both arid and humid sites.

5. Feature importance results show that elevation is the most important topographic variable for predicting soil moisture. Percent clay is the most important soil texture variable for predicting soil moisture. Vegetation plays a smaller role in predicting soil moisture than topography and soil, with EVI being the most important vegetation variable. Land cover classification provides little value in predicting soil moisture. The aridity index is by far the most important meteorological variable for predicting soil moisture. Among the SMAP soil moisture products, all surface soil moisture products are the most important, with RZSM and PSM products having low importance. The most important SMAP products are 3-day and 14-day antecedent soil moisture variables.

This study has some limitations that should be considered in future research. One limitation is the limited temporal data coverage of ECOSTRESS optical/thermal remote sensing data due to cloud presence. Another limitation is the need for feature selection and interpretability. This could be achieved by either removing land cover that showed limited impact on root zone soil moisture prediction, or by removing ECOSTRESS products that caused the removal of data from other input variables due to cloud presence. Finally, there is the potential for hyperparameter exploration, i.e., providing additional time for model hyperparameter tuning to increase model performance.

Future research directions could involve the integration of other remote sensing data, such as Sentinel-1, Sentinel-2, and Landsat data. Additionally, advanced machine learning techniques,

such as convolutional neural networks and recurrent neural networks, could be explored. Finally, ensemble models could be developed to combine predictions from multiple machine learning algorithms to enhance robustness and generalizability. These efforts will contribute to advancing our understanding of soil moisture dynamics and its applications in water resources management.

CHAPTER 5. CONCLUSIONS

5.1 Summary

This dissertation aimed to improve the estimation of soil moisture in the rootzone using remote sensing methods and machine learning techniques. It considered regional characteristics such as soil, topography, vegetation, and climate to enhance the accuracy of the estimates.

The main contributions of this dissertation include the following:

Chapter 2 evaluated the impact of regional soil, vegetation, and climatic conditions on the form and strength of the $\Lambda_{SEB} - \bar{\theta}$, $\Lambda_{PET} - \bar{\theta}$, $\Lambda_{SEB} - \bar{s}$, and $\Lambda_{PET} - \bar{s}$ relationships. The study used Extended Fourier Amplitude Sensitivity Test (eFAST) with a physically-based model (HYDRUS -1D) that simulates both evapotranspiration and soil moisture dynamics. Within a given climatic region, soil characteristics such as the percent clay and percent silt were most important in determining the shape of the relationships, while vegetation characteristics such as leaf area index and maximum rooting depth had the greatest effect on the strength of these relationships. The total annual precipitation, which helps determine the climatic region, also had a strong effect on both the form and strength of the relationships. The parameters that define the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships were also estimated

using the regional characteristics. Estimating the parameters in this way allows the methods to be adapted to local conditions and has the potential to improve the $\bar{\theta}$ and $\bar{s}$ estimates.

Chapter 3 evaluated the $\bar{\theta}$ estimates from remote sensing when the Λ_{SEB} vs. $\bar{\theta}$ and Λ_{PET} vs. $\bar{\theta}$ relationships were inferred from regional characteristics using the previously proposed relationships. Four study regions were considered including the Walnut Gulch Experimental Watershed in Arizona, the Piñon Canyon Maneuver Site and Lower Arkansas River Valley in Colorado, the Little Washita and Fort Cobb Experimental Watersheds in Oklahoma, and the Mississippi Delta region in Mississippi. The $\bar{\theta}$ estimates from the regionally adapted relationships were compared to $\bar{\theta}$ estimates from the single empirical relationship proposed by Bastiaanssen et al. (2000) and to in situ $\bar{\theta}$ measurements. The estimates from the regionally adapted relationships consistently outperformed the estimates from the single empirical relationship. The performance typically improved as more regional information were used in the relationships and reduced the root mean squared error of $\bar{\theta}$ by an average of 45% among the four regions. The regional method typically performed better for the arid and semiarid regions with root mean squared errors of 0.05 cm^3cm^{-3} and 0.04 cm^3cm^{-3}, respectively. The regionally adapted relationships better captured both spatial and temporal variations of soil moisture than the single empirical relationship.

Chapter 4 provided in-depth analysis of machine learning models for soil moisture prediction in different climate regions and soil depths. The study evaluated and compared the performance of XGBoost, CatBoost, RF, LightGBM, and ANN to predict soil moisture. These models, particularly XGBoost, provided accurate soil moisture predictions. For

example, XGBoost produced root means squared errors of 0.040 cm^3cm^{-3} (training) and 0.044 cm^3cm^{-3} (testing) in its soil moisture estimates. The study also examined the impact of climate regions, including arid and humid climates, on soil moisture predictions. The machine learning models can accurately capture soil moisture dynamics in different climates. Furthermore, the study explored the effect of the depth at which soil moisture is estimated on prediction accuracy and discovered that the models tend to perform better as the depth increased. Additionally, the study analyzed the relative importance of various predictors for rootzone soil moisture prediction. It identified key drivers of the models, such as elevation, the aridity index, and soil composition. The findings of this study could be used to improve soil moisture prediction, which could be useful for applications in water resource management, agriculture, hydrology, and environmental modeling.

5.2 Future Avenues of Research

Several avenues of research are available for future research:

1. In Chapter 2, it was assumed that the soil profile is always 100 cm deep, vertical, and homogeneous, and that the bottom boundary is freely draining. However, the depth of the soil profile can vary in nature and heterogeneity throughout the depth often occurs. Future models could consider using a 2D or 3D models to capture both spatial heterogeneity and simultaneously consider the effect of a shallow water table. The assumption of a freely draining bottom boundary could be one of the causes of the poor performance of the model in wetter regions in Chapter 3. Additionally, lateral flow was not taken into account, which

could lead to inaccurate soil moisture estimation where lateral flow is important. The impact of lateral flow on soil moisture can also be considered by utilizing 2D or 3D models.

2. Future studies could also incorporate the effects of particular crop types on the relationship between soil moisture and evapotranspiration. Variations in salinity, plant nutrition, and depth within the HYDRUS-1D model were also not considered.
3. In Chapter 2, the rectangular pulse model was used to simulate precipitation forcing for HYDRUS-1D. Future research could compare the results in this study with those obtained from other precipitation models, such as the exponential distribution model, gamma distribution model, and log-normal distribution model (Chapman, 1998). These models use different probability distributions, which could be selected to match the specific precipitation patterns of a given region. Also, the temperature forcing for HYDRUS-1D was assumed to be independent of the evapotranspiration. Future studies could explore this linkage.
4. In Chapter 3, extreme points of the thermal continuum need to be identified in the ReSET method, which are typically referred to as the hot and cold (or dry and wet) pixels. These points were chosen visually by the modeler and could introduce bias. Thus, it is worth considering an automatic way of selecting these points to reduce human bias. For instance, Molaei et al. (2022) conducted a study with the UAS-METRIC model that utilized such a method. Surface energy balance models that are more sophisticated than ReSET could also be considered.
5. In Chapter 3, the results from sub-humid and humid region showed that antecedent precipitation may impact the performance of the $\Lambda_{SEB} - \bar{\theta}$ and $\Lambda_{PET} - \bar{\theta}$ relationships by weakening the relationships. Future studies could add antecedent precipitation from remote sensing products to the proposed equations in Chapter 2.

6. Future research could also explore the effect of irrigated fields on the relationships and the accuracy of the soil moisture estimates. Irrigated fields often receive more water than natural grasslands. Applying the methods outlined in Chapter 3 to irrigated fields may yield additional insights into the strengths and weaknesses of the proposed models. One could also utilize MODIS or ECOSTRESS instead of Landsat to measure the components of the energy balance and assess the model performance using different optical and thermal products.
7. To simplify application of the machine learning methods, it might be beneficial to reduce the number of features considered. Feature selection in machine learning involves choosing a subset of the most significant features from the original feature set. This is done to enhance the machine learning model's performance and interpretability. In Chapter 4, no features were excluded from the input predictors. Future studies could explore which features could be discarded without affecting the model results. To do this, techniques such as neighborhood linear discriminant analysis (Zhu et al., 2022), locally linear embedding (Li et al., 2016), and t-distributed stochastic neighbor embedding (Cieslak et al., 2020) could be explored.
8. The in-situ soil moisture datasets used to evaluate the machine learning methods likely do not fully capture the spatial variations in soil moisture within their regions. Catchments and watersheds with dense in situ soil moisture measurements are may allow improved testing of the machine learning methods and an improved understanding of soil moisture prediction, as they typically observe more diverse combinations of topographic, vegetation, and soil conditions.

REFERENCES

Abbaszadeh, P., Moradkhani, H., Zhan, X., 2019. Downscaling SMAP Radiometer Soil Moisture Over the CONUS Using an Ensemble Learning Method. Water Resour. Res. 55, 324–344. https://doi.org/10.1029/2018WR023354

Abiy, A.Z., Wiederholt, R.P., Lagerwall, G.L., Melesse, A.M., Davis, S.E., 2022. Multilayer Feedforward Artificial Neural Network Model to Forecast Florida Bay Salinity with Climate Change. Water 14, 3495. https://doi.org/10.3390/w14213495

Abowarda, A.S., Bai, L., Zhang, C., Long, D., Li, X., Huang, Q., Sun, Z., 2021. Generating surface soil moisture at 30 m spatial resolution using both data fusion and machine learning toward better water resources management at the field scale. Remote Sens. Environ. 255, 112301. https://doi.org/10.1016/j.rse.2021.112301

Adab, H., Morbidelli, R., Saltalippi, C., Moradian, M., Ghalhari, G.A.F., 2020. Machine Learning to Estimate Surface Soil Moisture from Remote Sensing Data. Water 12, 3223. https://doi.org/10.3390/w12113223

Ahmad, M.U.D., Bastiaanssen, W.G.M., 2003. Retrieving Soil Moisture Storage in the Unsaturated Zone Using Satellite Imagery and Bi-Annual Phreatic Surface Fluctuations. Irrig. Drain. Syst. 17, 141–161. https://doi.org/10.1023/A:1025101217521

Ahmad, S., Kalra, A., Stephen, H., 2010. Estimating soil moisture using remote sensing data: A machine learning approach. Adv. Water Resour. 33, 69–80. https://doi.org/10.1016/j.advwatres.2009.10.008

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

Akuraju, V.R., Ryu, D., George, B., 2021. Estimation of root-zone soil moisture using crop water stress index (CWSI) in agricultural fields. GIScience Remote Sens. 58, 340–353. https://doi.org/10.1080/15481603.2021.1877009

Alburn, N.E., Niemann, J.D., Elhaddad, A., 2015. Evaluation of a surface energy balance method based on optical and thermal satellite imagery to estimate root-zone soil moisture. Hydrol. Process. 29, 5354–5368. https://doi.org/10.1002/hyp.10562

Ali, I., Greifeneder, F., Stamenkovic, J., Neumann, M., Notarnicola, C., 2015. Review of Machine Learning Approaches for Biomass and Soil Moisture Retrievals from Remote Sensing Data. Remote Sens. 7, 16398–16421. https://doi.org/10.3390/rs71215841

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evaporation–Guidelines for computing crop water requirements–FAO Irrigation and drainage paper 56. Food Agric. Organ. United Nations 300, D05109.

Allen, R.G., Tasumi, M., Morse, A., Trezza, R., Wright, J.L., Bastiaanssen, W., Kramber, W., Lorite, I., Robison, C.W., 2007a. Satellite-Based Energy Balance for Mapping

Evapotranspiration with Internalized Calibration (METRIC)—Applications. J. Irrig. Drain. Eng. 133, 395–406. https://doi.org/10.1061/(ASCE)0733-9437(2007)133:4(395)

Allen, R.G., Tasumi, M., Trezza, R., 2007b. Satellite-Based Energy Balance for Mapping Evapotranspiration with Internalized Calibration (METRIC)—Model. J. Irrig. Drain. Eng. 133, 380–394. https://doi.org/10.1061/(ASCE)0733-9437(2007)133:4(380)

Amani, M., Salehi, B., Mahdavi, S., Masjedi, A., Dehnavi, S., 2017. Temperature-Vegetation-soil Moisture Dryness Index (TVMDI). Remote Sens. Environ. 197, 1–14. https://doi.org/10.1016/j.rse.2017.05.026

Amazirh, A., Merlin, O., Er-Raki, S., Gao, Q., Rivalland, V., Malbeteau, Y., Khabba, S., Escorihuela, M.J., 2018. Retrieving surface soil moisture at high spatio-temporal resolution from a synergy between Sentinel-1 radar and Landsat thermal data: A study case over bare soil. Remote Sens. Environ. 211, 321–337. https://doi.org/10.1016/j.rse.2018.04.013

Andales, A.A., Bauder, T.A., Nolan, J.D., 2009. The Colorado Agricultural Meteorological Network (CoAgMet) and Crop ET Reports, Crop Series.

Anderson, W.B., Zaitchik, B.F., Hain, C.R., Anderson, M.C., Yilmaz, M.T., Mecikalski, J., Schultz, L., 2012. Towards an integrated soil moisture drought monitor for East Africa. Hydrol. Earth Syst. Sci. 16, 2893–2913. https://doi.org/10.5194/hess-16-2893-2012

Archer, C.L., Jacobson, M.Z., 2003. Spatial and temporal distributions of U.S. winds and wind power at 80 m derived from measurements. J. Geophys. Res. Atmos. 108. https://doi.org/10.1029/2002JD002076

Arguez, A., Durre, I., Applequist, S., Vose, R.S., Squires, M.F., Yin, X., Heim, R.R., Owen, T.W., 2012. NOAA's 1981–2010 U.S. Climate Normals: An Overview. Bull. Am. Meteorol. Soc. 93, 1687–1697. https://doi.org/10.1175/BAMS-D-11-00197.1

ASCE-EWRI, 2005. The ASCE Standardized Reference Evapotranspiration Equation, Technical Committee report to the Environmental and Water Resources Institute of the American Society of Civil Engineers from the Task Committee on Standardization of Reference Evapotranspiration. ASCE-EWRI, 1801 Alexander Bell Drive, Reston, VA 20191-440.

Babaeian, E., Sidike, P., Newcomb, M.S., Maimaitijiang, M., White, S.A., Demieville, J., Ward, R.W., Sadeghi, M., LeBauer, D.S., Jones, S.B., Sagan, V., Tuller, M., 2019. A New Optical Remote Sensing Technique for High-Resolution Mapping of Soil Moisture. Front. Big Data 2. https://doi.org/10.3389/fdata.2019.00037

Bai, L., Long, D., Yan, L., 2019. Estimation of Surface Soil Moisture With Downscaled Land Surface Temperatures Using a Data Fusion Approach for Heterogeneous Agricultural Land. Water Resour. Res. https://doi.org/10.1029/2018WR024162

Baier, W., 1969. Concepts of soil moisture availability and their effect on soil moisture estimates from a meteorological budget. Agric. Meteorol. 6, 165–178. https://doi.org/10.1016/0002-1571(69)90002-8

Barbeta, A., Mejía-Chang, M., Ogaya, R., Voltas, J., Dawson, T.E., Peñuelas, J., 2015. The combined effects of a long-term experimental drought and an extreme drought on the use of plant-water sources in a Mediterranean forest. Glob. Chang. Biol. 21, 1213–1225.

https://doi.org/10.1111/gcb.12785

Bartalis, Z., Wagner, W., Naeimi, V., Hasenauer, S., Scipal, K., Bonekamp, H., Figa, J., Anderson, C., 2007. Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT). Geophys. Res. Lett. 34, L20401. https://doi.org/10.1029/2007GL031088

Bartsch, A., Balzter, H., George, C., 2009. The influence of regional surface soil moisture anomalies on forest fires in Siberia observed from satellites. Environ. Res. Lett. 4, 045021. https://doi.org/10.1088/1748-9326/4/4/045021

Basara, J.B., Crawford, K.C., 2002. Linear relationships between root-zone soil moisture and atmospheric processes in the planetary boundary layer. J. Geophys. Res. Atmos. 107, 4274. https://doi.org/10.1029/2001JD000633

Bastiaanssen, W.G.M., 1998. Remote Sensing in Water Resources Management: the State of the Art., International Water Management Institute. International Water Management Institute, Colombo, Sri Lanka.

Bastiaanssen, W.G.M., Cheema, M.J.M., Immerzeel, W.W., Miltenburg, I.J., Pelgrum, H., 2012. Surface energy balance and actual evapotranspiration of the transboundary Indus Basin estimated from satellite measurements and the ETLook model. Water Resour. Res. 48, 1–16. https://doi.org/10.1029/2011WR010482

Bastiaanssen, W.G.M., Molden, D.J., Makin, I.W., 2000. Remote sensing for irrigated agriculture: examples from research and possible applications. Agric. Water Manag. 46, 137–155. https://doi.org/10.1016/S0378-3774(00)00080-9

Bastiaanssen, W.G.M., Pelgrum, H., Droogers, P., de Bruin, H.A.R., Menenti, M., 1997. Area-average estimates of evaporation, wetness indicators and top soil moisture during two golden days in EFEDA. Agric. For. Meteorol. 87, 119–137. https://doi.org/10.1016/S0168-1923(97)00020-8

Bastiaanssen, W.G.M., Pelgrum, H., Wang, J., Ma, Y., Moreno, J.F., Roerink, G.J., van der Wal, T., 1998. A remote sensing surface energy balance algorithm for land (SEBAL). J. Hydrol. 212–213, 213–229. https://doi.org/10.1016/S0022-1694(98)00254-6

Beck, H.E., Pan, M., Miralles, D.G., Reichle, R.H., Dorigo, W.A., Hahn, S., Sheffield, J., Karthikeyan, L., Balsamo, G., Parinussa, R.M., van Dijk, A.I.J.M., Du, J., Kimball, J.S., Vergopolan, N., Wood, E.F., 2021. Evaluation of 18 satellite- and model-based soil moisture products using in situ measurements from 826 sensors. Hydrol. Earth Syst. Sci. 25, 17–40. https://doi.org/10.5194/hess-25-17-2021

Bengtson, R.L., Carter, C.E., Fouss, J.L., Southwick, L.M., Willis, G.H., 1995. Agricultural Drainage and Water Quality in Mississippi Delta. J. Irrig. Drain. Eng. 121, 292–295. https://doi.org/10.1061/(ASCE)0733-9437(1995)121:4(292)

Bisong, E., 2019. Introduction to Scikit-learn, in: Building Machine Learning and Deep Learning Models on Google Cloud Platform. Apress, Berkeley, CA, pp. 215–229. https://doi.org/10.1007/978-1-4842-4470-8_18

Bot, A., Benites, J., 2015. The importance of soil organic matter. Key to drought-resistant soil

and sustained food and production, Fao Soils Bulletin 80.

Boudreau, J., Caron, J., Elrick, D., Fortin, J., Gallichand, J., 2009. Solute transport in sub-irrigated peat-based growing media. Can. J. Soil Sci. 89, 301–313. https://doi.org/10.4141/CJSS08023

Breiman, L., 2001. Random forests. Mach. Learn. 5–32. https://doi.org/10.1023/A:1010933404324

Brocca, L., Ciabatta, L., Massari, C., Camici, S., Tarpanelli, A., 2017. Soil Moisture for Hydrological Applications: Open Questions and New Opportunities. Water 9, 140. https://doi.org/10.3390/w9020140

Bronstert, A., Plate, E.J., 1997. Modelling of runoff generation and soil moisture dynamics for hillslopes and micro-catchments. J. Hydrol. 198, 177–195. https://doi.org/10.1016/S0022-1694(96)03306-9

Brunt, D., 1952. Physical and Dynamical Meteorology, Cambridge University Press.

Buchner, J.S., Šimůnek, J., Lee, J., Rolston, D.E., Hopmans, J.W., King, A.P., Six, J., 2008. Evaluation of CO2 fluxes from an agricultural field using a process-based numerical model. J. Hydrol. 361, 131–143. https://doi.org/10.1016/j.jhydrol.2008.07.035

Burkhalter, J.P., Gates, T.K., 2005. Agroecological Impacts from Salinization and Waterlogging in an Irrigated River Valley. J. Irrig. Drain. Eng. 131, 197–209. https://doi.org/10.1061/(ASCE)0733-9437(2005)131:2(197)

Burman, R.D., Jensen, M., Allen, R.G., 1987. Thermodynamic factors in evapotranspiration, in: In Steel Structures (Pp. 140-148). ASCE.

Çakir, R., 2004. Effect of water stress at different development stages on vegetative and reproductive growth of corn. F. Crop. Res. 89, 1–16. https://doi.org/10.1016/j.fcr.2004.01.005

Campbell, G.S., Campbell, M.D., 1982. Irrigation Scheduling Using Soil Moisture Measurements: Theory and Practice, in: Advances in Irrigation, 1. pp. 25–42. https://doi.org/10.1016/B978-0-12-024301-3.50008-3

Candido, C., Blanco, A.C., Medina, J., Gubatanga, E., Santos, A., Ana, R.S., Reyes, R.B., 2021. Improving the consistency of multi-temporal land cover mapping of Laguna lake watershed using light gradient boosting machine (LightGBM) approach, change detection analysis, and Markov chain. Remote Sens. Appl. Soc. Environ. 23, 100565. https://doi.org/10.1016/j.rsase.2021.100565

Carlson, T., 2007. An Overview of the “Triangle Method” for Estimating Surface Evapotranspiration and Soil Moisture from Satellite Imagery. Sensors 7, 1612–1629. https://doi.org/10.3390/s7081612

Carlson, T.N., Petropoulos, G.P., 2019. A new method for estimating of evapotranspiration and surface soil moisture from optical and thermal infrared measurements: the simplified triangle. Int. J. Remote Sens. 40, 7716–7729. https://doi.org/10.1080/01431161.2019.1601288

Celik, M.F., Isik, M.S., Yuzugullu, O., Fajraoui, N., Erten, E., 2022. Soil Moisture Prediction from Remote Sensing Images Coupled with Climate, Soil Texture and Topography via Deep Learning. Remote Sens. 14, 5584. https://doi.org/10.3390/rs14215584

Cerdà, A., 1999. Parent Material and Vegetation Affect Soil Erosion in Eastern Spain. Soil Sci. Soc. Am. J. 63, 362–368. https://doi.org/10.2136/sssaj1999.03615995006300020014x

Černý, R., 2009. Time-domain reflectometry method and its application for measuring moisture content in porous materials: A review. Measurement 42, 329–336. https://doi.org/10.1016/j.measurement.2008.08.011

Chan, S.K., Bindlish, R., O'Neill, P., Jackson, T., Njoku, E., Dunbar, S., Chaubell, J., Piepmeier, J., Yueh, S., Entekhabi, D., Colliander, A., Chen, F., Cosh, M.H., Caldwell, T., Walker, J., Berg, A., McNairn, H., Thibeault, M., Martínez-Fernández, J., Uldall, F., Seyfried, M., Bosch, D., Starks, P., Holifield Collins, C., Prueger, J., van der Velde, R., Asanuma, J., Palecki, M., Small, E.E., Zreda, M., Calvet, J., Crow, W.T., Kerr, Y., 2018. Development and assessment of the SMAP enhanced passive soil moisture product. Remote Sens. Environ. 204, 931–941. https://doi.org/10.1016/j.rse.2017.08.025

Chan, Y.K., Koo, V.C., 2008. AN INTRODUCTION TO SYNTHETIC APERTURE RADAR (SAR). Prog. Electromagn. Res. B 2, 27–60. https://doi.org/10.2528/PIERB07110101

Chandra Joshi, R., Ryu, D., Lane, P.N.J., Sheridan, G.J., 2023. Seasonal forecast of soil moisture over Mediterranean-climate forest catchments using a machine learning approach. J. Hydrol. 619, 129307. https://doi.org/10.1016/j.jhydrol.2023.129307

Chapman, T., 1998. Stochastic modelling of daily rainfall: the impact of adjoining wet days on the distribution of rainfall amounts. Environ. Model. Softw. 13, 317–324. https://doi.org/10.1016/S1364-8152(98)00036-X

Chauhan, N.S., Miller, S., Ardanuy, P., 2003. Spaceborne soil moisture estimation at high resolution: a microwave-optical/IR synergistic approach. Int. J. Remote Sens. 24, 4599–4622. https://doi.org/10.1080/0143116031000156837

Chávez, J.L., Neale, C.M.U., Prueger, J.H., Kustas, W.P., 2008. Daily evapotranspiration estimates from extrapolating instantaneous airborne remote sensing ET values. Irrig. Sci. 27, 67–81. https://doi.org/10.1007/s00271-008-0122-3

Chen, J.M., Liu, J., 2020. Evolution of evapotranspiration models using thermal and shortwave remote sensing data. Remote Sens. Environ. 237, 111594. https://doi.org/10.1016/j.rse.2019.111594

Chen, M., Willgoose, G.R., Saco, P.M., 2014. Spatial prediction of temporal soil moisture dynamics using HYDRUS-1D. Hydrol. Process. 28, 171–185. https://doi.org/10.1002/hyp.9518

Chen, T., Guestrin, C., 2016. XGBoost, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York, NY, USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785

Cho, E., Choi, M., Wagner, W., 2015. An assessment of remotely sensed surface and root zone soil moisture through active and passive sensors in northeast Asia. Remote Sens. Environ.

160, 166–179. https://doi.org/10.1016/j.rse.2015.01.013

Churkina, G., Running, S.W., 1998. Contrasting Climatic Controls on the Estimated Productivity of Global Terrestrial Biomes. Ecosystems 1, 206–215. https://doi.org/10.1007/s100219900016

Cieslak, M.C., Castelfranco, A.M., Roncalli, V., Lenz, P.H., Hartline, D.K., 2020. t-Distributed Stochastic Neighbor Embedding (t-SNE): A tool for eco-physiological transcriptomic analysis. Mar. Genomics 51, 100723. https://doi.org/10.1016/j.margen.2019.100723

Coleman, M.L., Niemann, J.D., 2013. Controls on topographic dependence and temporal instability in catchment-scale soil moisture patterns. Water Resour. Res. https://doi.org/10.1002/wrcr.20159

Colliander, A., Jackson, T.J., Bindlish, R., Chan, S., Das, N., Kim, S.B., Cosh, M.H., Dunbar, R.S., Dang, L., Pashaian, L., Asanuma, J., Aida, K., Berg, A., Rowlandson, T., Bosch, D., Caldwell, T., Caylor, K., Goodrich, D., al Jassar, H., Lopez-Baeza, E., Martínez-Fernández, J., González-Zamora, A., Livingston, S., McNairn, H., Pacheco, A., Moghaddam, M., Montzka, C., Notarnicola, C., Niedrist, G., Pellarin, T., Prueger, J., Pulliainen, J., Rautiainen, K., Ramos, J., Seyfried, M., Starks, P., Su, Z., Zeng, Y., van der Velde, R., Thibeault, M., Dorigo, W., Vreugdenhil, M., Walker, J.P., Wu, X., Monerris, A., O'Neill, P.E., Entekhabi, D., Njoku, E.G., Yueh, S., 2017. Validation of SMAP surface soil moisture products with core validation sites. Remote Sens. Environ. 191, 215–231. https://doi.org/10.1016/j.rse.2017.01.021

Corbari, C., Ravazzani, G., Galvagno, M., Cremonese, E., Mancini, M., 2017. Assessing Crop Coefficients for Natural Vegetated Areas Using Satellite Data and Eddy Covariance Stations. Sensors 17, 2664. https://doi.org/10.3390/s17112664

Cosh, M.H., Jackson, T.J., Starks, P., Heathman, G., 2006. Temporal stability of surface soil moisture in the Little Washita River watershed and its applications in satellite soil moisture product validation. J. Hydrol. 323, 168–177. https://doi.org/10.1016/j.jhydrol.2005.08.020

Coupland, R.T., Johnson, R.E., 1965. Rooting Characteristics of Native Grassland Species in Saskatchewan. J. Ecol. 53, 475. https://doi.org/10.2307/2257990

Cowley, G.S., Niemann, J.D., Green, T.R., Seyfried, M.S., Jones, A.S., Grazaitis, P.J., 2017. Impacts of precipitation and potential evapotranspiration patterns on downscaling soil moisture in regions with large topographic relief. Water Resour. Res. https://doi.org/10.1002/2016WR019907

Cox, J.R., Frasler, G.W., Renard, K.G., 1986. Biomass distribution at grassland and shrubland sites. Rangelands Arch. 8, 67–69.

Crago, R.D., 1996. Conservation and variability of the evaporative fraction during the daytime. J. Hydrol. 180, 173–194. https://doi.org/10.1016/0022-1694(95)02903-6

Crago, R.D., Brutsaert, W., 1994. The estimation of surface momentum flux under unstable conditions from the atmospheric pressure field. Water Resour. Res. 30, 617–623. https://doi.org/10.1029/93WR02921

Crow, W.T., Berg, A.A., Cosh, M.H., Loew, A., Mohanty, B.P., Panciera, R., de Rosnay, P.,

Ryu, D., Walker, J.P., 2012. Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. Rev. Geophys. 50. https://doi.org/10.1029/2011RG000372

D.K., T., B.G, P., Xiong, F., 2019. Auto-detection of epileptic seizure events using deep neural network with different feature scaling techniques. Pattern Recognit. Lett. 128, 544–550. https://doi.org/10.1016/j.patrec.2019.10.029

Das, N.N., Entekhabi, D., Dunbar, R.S., Chaubell, M.J., Colliander, A., Yueh, S., Jagdhuber, T., Chen, F., Crow, W., O'Neill, P.E., Walker, J.P., Berg, A., Bosch, D.D., Caldwell, T., Cosh, M.H., Collins, C.H., Lopez-Baeza, E., Thibeault, M., 2019. The SMAP and Copernicus Sentinel 1A/B microwave active-passive high resolution surface soil moisture product. Remote Sens. Environ. 233, 111380. https://doi.org/10.1016/j.rse.2019.111380

Davies, J.A., Allen, C.D., 1973. Equilibrium, Potential and Actual Evaporation from Cropped Surfaces in Southern Ontario. J. Appl. Meteorol. 12, 649–657. https://doi.org/10.1175/1520-0450(1973)012<0649:EPAAEF>2.0.CO;2

Davis, B.N., Werpy, J., Friesz, A., Impecoven, K., Quenzer, R.L., Maiersperger, T., Meyer, D.J., 2015. Interactive Access to LP DAAC Satellite Data Archives Through a Combination of Open-Source and Custom Middleware Web Services. IEEE Geosci. Remote Sens. Mag. 3, 8–20. https://doi.org/10.1109/MGRS.2015.2505999

Demir-Kavuk, O., Kamada, M., Akutsu, T., Knapp, E.-W., 2011. Prediction using step-wise L1, L2 regularization and feature selection for small data sets with large number of features. BMC Bioinformatics 12, 412. https://doi.org/10.1186/1471-2105-12-412

Denmead, O.T., Shaw, R.H., 1962. Availability of Soil Water to Plants as Affected by Soil Moisture Content and Meteorological Conditions 1. Agron. J. 54, 385–390. https://doi.org/10.2134/agronj1962.00021962005400050005x

Dewitz, J., 2021. NLCD 2019 Land Cover (CONUS) [WWW Document]. Multi-Resolution L. Charact. Consort.

Dewitz, J., USGS, 2021. National Land Cover Database (NLCD) 2019 Products (ver. 2.0, June 2021). U.S. Geol. Surv. data release.

Didan, K., 2021. MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V061 [Data set]. NASA EOSDIS L. Process. Distrib. Act. Arch. Center. Accessed 2023-09-15 from. https://doi.org/10.5067/MODIS/MOD13Q1.061

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

Dorigo, W., Himmelbauer, I., Aberer, D., Schremmer, L., Petrakovic, I., Zappa, L., Preimesberger, W., Xaver, A., Annor, F., Ardö, J., Baldocchi, D., Bitelli, M., Blöschl, G., Bogena, H., Brocca, L., Calvet, J.-C., Camarero, J.J., Capello, G., Choi, M., Cosh, M.C., van de Giesen, N., Hajdu, I., Ikonen, J., Jensen, K.H., Kanniah, K.D., de Kat, I., Kirchengast, G., Kumar Rai, P., Kyrouac, J., Larson, K., Liu, S., Loew, A., Moghaddam, M., Martínez Fernández, J., Mattar Bader, C., Morbidelli, R., Musial, J.P., Osenga, E.,

Palecki, M.A., Pellarin, T., Petropoulos, G.P., Pfeil, I., Powers, J., Robock, A., Rüdiger, C., Rummel, U., Strobel, M., Su, Z., Sullivan, R., Tagesson, T., Varlagin, A., Vreugdenhil, M., Walker, J., Wen, J., Wenger, F., Wigneron, J.P., Woods, M., Yang, K., Zeng, Y., Zhang, X., Zreda, M., Dietrich, S., Gruber, A., van Oevelen, P., Wagner, W., Scipal, K., Drusch, M., Sabia, R., 2021. The International Soil Moisture Network: serving Earth system science for over a decade. Hydrol. Earth Syst. Sci. 25, 5749–5804. https://doi.org/10.5194/hess-25-5749-2021

Dorigo, W.A., Xaver, A., Vreugdenhil, M., Gruber, A., Hegyiová, A., Sanchis-Dufau, A.D., Zamojski, D., Cordes, C., Wagner, W., Drusch, M., 2013. Global Automated Quality Control of In Situ Soil Moisture Data from the International Soil Moisture Network. Vadose Zo. J. 12, 1–21. https://doi.org/10.2136/vzj2012.0097

Douna, V., Barraza, V., Grings, F., Huete, A., Restrepo-Coupe, N., Beringer, J., 2021. Towards a remote sensing data based evapotranspiration estimation in Northern Australia using a simple random forest approach. J. Arid Environ. 191, 104513. https://doi.org/10.1016/j.jaridenv.2021.104513

Du, J., Kimball, J.S., Bindlish, R., Walker, J.P., Watts, J.D., 2022. Local Scale (3-m) Soil Moisture Mapping Using SMAP and Planet SuperDove. Remote Sens. 14, 3812. https://doi.org/10.3390/rs14153812

Duffie, J.A., Beckman, W.A., 2013. Solar Engineering of Thermal Processes, Solar Engineering of Thermal Processes: Fourth Edition. Wiley. https://doi.org/10.1002/9781118671603

Dumedah, G., P. Walker, J., Merlin, O., 2015. Root-zone soil moisture estimation from assimilation of downscaled Soil Moisture and Ocean Salinity data. Adv. Water Resour. 84, 14–22. https://doi.org/10.1016/j.advwatres.2015.07.021

Eldeiry, A.A., Waskom, R.M., Elhaddad, A., 2016. Using Remote Sensing to Estimate Evapotranspiration of Irrigated Crops Under Flood and Sprinkler Irrigation Systems. Irrig. Drain. 65, 85–97. https://doi.org/10.1002/ird.1945

Elhaddad, A., Garcia, L.A., 2011. ReSET-Raster: Surface Energy Balance Model for Calculating Evapotranspiration Using a Raster Approach. J. Irrig. Drain. Eng. 137, 203–210. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000282

Elhaddad, A., Garcia, L.A., 2008. Surface Energy Balance-Based Model for Estimating Evapotranspiration Taking into Account Spatial Variability in Weather. J. Irrig. Drain. Eng. 134, 681–689. https://doi.org/10.1061/(ASCE)0733-9437(2008)134:6(681)

Elhaddad, A., Garcia, L.A., Chávez, J.L., 2011. Using a Surface Energy Balance Model to Calculate Spatially Distributed Actual Evapotranspiration. J. Irrig. Drain. Eng. 137, 17–26. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000276

Eltahir, E.A.B., 1998. A Soil Moisture–Rainfall Feedback Mechanism: 1. Theory and observations. Water Resour. Res. 34, 765–776. https://doi.org/10.1029/97WR03499

Entekhabi, D., Njoku, E.G., Houser, P., Spencer, M., Doiron, T., Yunjin Kim, Smith, J., Girard, R., Belair, S., Crow, W., Jackson, T.J., Kerr, Y.H., Kimball, J.S., Koster, R., McDonald, K.C., O'Neill, P.E., Pultz, T., Running, S.W., Jiancheng Shi, Wood, E., van Zyl, J., 2004.

The hydrosphere State (hydros) Satellite mission: an Earth system pathfinder for global mapping of soil moisture and land freeze/thaw. IEEE Trans. Geosci. Remote Sens. 42, 2184–2195. https://doi.org/10.1109/TGRS.2004.834631

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

Entekhabi, D., Reichle, R.H., Koster, R.D., Crow, W.T., 2010b. Performance Metrics for Soil Moisture Retrievals and Application Requirements. J. Hydrometeorol. 11, 832–840. https://doi.org/10.1175/2010JHM1223.1

Entekhabi, D., Rodriguez-Iturbe, I., Castelli, F., 1996. Mutual interaction of soil moisture state and atmospheric processes. J. Hydrol. 184, 3–17. https://doi.org/10.1016/0022-1694(95)02965-6

Evans, J.G., Ward, H.C., Blake, J.R., Hewitt, E.J., Morrison, R., Fry, M., Ball, L.A., Doughty, L.C., Libre, J.W., Hitt, O.E., Rylett, D., Ellis, R.J., Warwick, A.C., Brooks, M., Parkes, M.A., Wright, G.M.H., Singer, A.C., Boorman, D.B., Jenkins, A., 2016. Soil water content in southern England derived from a cosmic-ray soil moisture observing system – COSMOS-UK. Hydrol. Process. 30, 4987–4999. https://doi.org/10.1002/hyp.10929

Evett, S.R., Tolk, J.A., Howell, T.A., 2003. A Depth Control Stand for Improved Accuracy with the Neutron Probe. Vadose Zo. J. 2, 642–649. https://doi.org/10.2136/vzj2003.6420

Falls, L.W., 1974. The beta distribution: A statistical model for world cloud cover. J. Geophys. Res. 79, 1261–1264. https://doi.org/10.1029/JC079i009p01261

Famiglietti, J.S., Rudnicki, J.W., Rodell, M., 1998. Variability in surface moisture content along a hillslope transect: Rattlesnake Hill, Texas. J. Hydrol. 210, 259–281. https://doi.org/10.1016/S0022-1694(98)00187-5

Famiglietti, J.S., Ryu, D., Berg, A.A., Rodell, M., Jackson, T.J., 2008. Field observations of soil moisture variability across scales. Water Resour. Res. 44. https://doi.org/10.1029/2006WR005804

Fang, K., Pan, M., Shen, C., 2019. The Value of SMAP for Long-Term Soil Moisture Estimation With the Help of Deep Learning. IEEE Trans. Geosci. Remote Sens. 57, 2221–2233. https://doi.org/10.1109/TGRS.2018.2872131

Farahani, A., Moradikhaneghahi, M., Ghayoomi, M., Jacobs, J.M., 2022. Application of Soil Moisture Active Passive (SMAP) Satellite Data in Seismic Response Assessment. Remote Sens. 14, 4375. https://doi.org/10.3390/rs14174375

Feng, H., Chen, C., Dong, H., Wang, J., Meng, Q., 2013. Modified Shortwave Infrared Perpendicular Water Stress Index: A Farmland Water Stress Monitoring Method. J. Appl. Meteorol. Climatol. 52, 2024–2032. https://doi.org/10.1175/JAMC-D-12-0164.1

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37, 4302–4315. https://doi.org/10.1002/joc.5086

Firat, M., Yurdusev, M.A., Turan, M.E., 2009. Evaluation of Artificial Neural Network Techniques for Municipal Water Consumption Modeling. Water Resour. Manag. 23, 617–632. https://doi.org/10.1007/s11269-008-9291-3

Fisher, J.B., Lee, B., Purdy, A.J., Halverson, G.H., Dohlen, M.B., Cawse-Nicholson, K., Wang, A., Anderson, R.G., Aragon, B., Arain, M.A., Baldocchi, D.D., Baker, J.M., Barral, H., Bernacchi, C.J., Bernhofer, C., Biraud, S.C., Bohrer, G., Brunsell, N., Cappelaere, B., Castro-Contreras, S., Chun, J., Conrad, B.J., Cremonese, E., Demarty, J., Desai, A.R., De Ligne, A., Foltýnová, L., Goulden, M.L., Griffis, T.J., Grünwald, T., Johnson, M.S., Kang, M., Kelbe, D., Kowalska, N., Lim, J., Maïnassara, I., McCabe, M.F., Missik, J.E.C., Mohanty, B.P., Moore, C.E., Morillas, L., Morrison, R., Munger, J.W., Posse, G., Richardson, A.D., Russell, E.S., Ryu, Y., Sanchez-Azofeifa, A., Schmidt, M., Schwartz, E., Sharp, I., Šigut, L., Tang, Y., Hulley, G., Anderson, M., Hain, C., French, A., Wood, E., Hook, S., 2020. ECOSTRESS: NASA's Next Generation Mission to Measure Evapotranspiration From the International Space Station. Water Resour. Res. 56. https://doi.org/10.1029/2019WR026058

Fleming, K., Hendrickx, J.M.H., Hong, S., 2005. Regional mapping of root zone soil moisture using optical satellite imagery, in: Watkins, W.R., Clement, D., Reynolds, W.R. (Eds.), Targets and Backgrounds XI: Characterization and Representation. p. 159. https://doi.org/10.1117/12.603397

Ford, T.W., Wulff, C.O., Quiring, S.M., 2014. Assessment of observed and model-derived soil moisture-evaporative fraction relationships over the United States Southern Great Plains. J. Geophys. Res. Atmos. 119, 6279–6291. https://doi.org/10.1002/2014JD021490

Foster, T., Mieno, T., Brozović, N., 2020. Satellite-Based Monitoring of Irrigation Water Use: Assessing Measurement Errors and Their Implications for Agricultural Water Management Policy. Water Resour. Res. 56. https://doi.org/10.1029/2020WR028378

French, A.N., Schmugge, T.J., Kustas, W.P., 2002. Estimating evapotranspiration over El Reno, Oklahoma with ASTER imagery. Agronomie 22, 105–106. https://doi.org/10.1051/agro:2001010

Fuentes, I., Padarian, J., Vervoort, R.W., 2022. Towards near real-time national-scale soil water content monitoring using data fusion as a downscaling alternative. J. Hydrol. 609, 127705. https://doi.org/10.1016/j.jhydrol.2022.127705

Fung, A.K., Li, Z., Chen, K.S., 1992. Backscattering from a randomly rough dielectric surface. IEEE Trans. Geosci. Remote Sens. 30, 356–369. https://doi.org/10.1109/36.134085

Gao, Q., Zribi, M., Escorihuela, M., Baghdadi, N., 2017. Synergetic Use of Sentinel-1 and Sentinel-2 Data for Soil Moisture Mapping at 100 m Resolution. Sensors 17, 1966. https://doi.org/10.3390/s17091966

Gao, X., Zhao, X., Brocca, L., Huo, G., Lv, T., Wu, P., 2017. Depth scaling of soil moisture content from surface to profile: multistation testing of observation operators. Hydrol. Earth Syst. Sci. Discuss. 1–25. https://doi.org/10.5194/hess-2017-292

Garcia, A., Torres, J.L., Prieto, E., de Francisco, A., 1998. Fitting wind speed distributions: a case study. Sol. Energy 62, 139–144. https://doi.org/10.1016/S0038-092X(97)00116-3

Gardner, B.R., Blad, B.L., Garrity, D.P., Watts, D.G., 1981. Relationships between crop temperature, grain yield, evapotranspiration and phenological development in two hybrids of moisture stressed sorghum. Irrig. Sci. 2, 213–224. https://doi.org/10.1007/BF00258375

Gentine, P., Entekhabi, D., Chehbouni, A., Boulet, G., Duchemin, B., 2007. Analysis of evaporative fraction diurnal behaviour. Agric. For. Meteorol. 143, 13–29. https://doi.org/10.1016/j.agrformet.2006.11.002

Geroy, I.J., Gribb, M.M., Marshall, H.P., Chandler, D.G., Benner, S.G., McNamara, J.P., 2011. Aspect influences on soil water retention and storage. Hydrol. Process. 25, 3836–3842. https://doi.org/10.1002/hyp.8281

Ghulam, A., Qin, Q., Zhan, Z., 2007. Designing of the perpendicular drought index. Environ. Geol. 52, 1045–1052. https://doi.org/10.1007/s00254-006-0544-2

Giglioli, N., Saltelli, A., 2003. Simlab 2.2, Software for sensitivity and uncertainty analysis, Simlab manual, Joint Research Centre European Commission.

Gillies, R.R., Kustas, W.P., Humes, K.S., 1997. A verification of the "triangle" method for obtaining surface soil water content and energy fluxes from remote measurements of the Normalized Difference Vegetation Index (NDVI) and surface e. Int. J. Remote Sens. 18, 3145–3166. https://doi.org/10.1080/014311697217026

Giteau, M., 1965. Une représentation de la Bhikṣâṭanamûrti de Çiva à Angkor Vat. Arts Asiat. 11, 131–137. https://doi.org/10.3406/arasi.1965.919

Gopalakrishnan, A., Biegler, L.T., 2013. Economic Nonlinear Model Predictive Control for periodic optimal operation of gas pipeline networks. Comput. Chem. Eng. 52, 90–99. https://doi.org/10.1016/j.compchemeng.2012.11.011

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Greifeneder, F., Notarnicola, C., Wagner, W., 2021. A Machine Learning-Based Approach for Surface Soil Moisture Estimations with Google Earth Engine. Remote Sens. 13, 2099. https://doi.org/10.3390/rs13112099

Grillakis, M.G., Koutroulis, A.G., Alexakis, D.D., Polykretis, C., Daliakopoulos, I.N., 2021. Regionalizing Root-Zone Soil Moisture Estimates From ESA CCI Soil Water Index Using Machine Learning and Information on Soil, Vegetation, and Climate. Water Resour. Res. 57. https://doi.org/10.1029/2020WR029249

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

Haghighi, E., Short Gianotti, D.J., Akbar, R., Salvucci, G.D., Entekhabi, D., 2018. Soil and Atmospheric Controls on the Land Surface Energy Balance: A Generalized Framework for Distinguishing Moisture-Limited and Energy-Limited Evaporation Regimes. Water Resour. Res. 54, 1831–1851. https://doi.org/10.1002/2017WR021729

Hain, C.R., Mecikalski, J.R., Anderson, M.C., 2009. Retrieval of an Available Water-Based Soil Moisture Proxy from Thermal Infrared Remote Sensing. Part I: Methodology and Validation. J. Hydrometeorol. 10, 665–683. https://doi.org/10.1175/2008JHM1024.1

Hancock, J.T., Khoshgoftaar, T.M., 2020. CatBoost for big data: an interdisciplinary review. J. Big Data 7, 94. https://doi.org/10.1186/s40537-020-00369-8

Hao, G., Su, H., Zhang, R., Tian, J., Chen, S., 2022. A Two-Source Normalized Soil Thermal Inertia Model for Estimating Field-Scale Soil Moisture from MODIS and ASTER Data. Remote Sens. 14, 1215. https://doi.org/10.3390/rs14051215

Hawk, K.L., Eagleson, P.S., 1992. Climatology of station storm rainfall in the continental United States: Parameters of the Bartlett-Lewis and Poisson rectangular pulses models., No. NASA-CR-192968.

He, X., Chaney, N.W., Schleiss, M., Sheffield, J., 2016. Spatial downscaling of precipitation using adaptable random forests. Water Resour. Res. 52, 8217–8237. https://doi.org/10.1002/2016WR019034

Hendrickx, J.M.H., Allen, R.G., Brower, A., Byrd, A.R., Hong, S., Ogden, F.L., Pradhan, N.R., Robison, C.W., Toll, D., Trezza, R., Umstot, T.G., Wilson, J.L., 2016. Benchmarking Optical/Thermal Satellite Imagery for Estimating Evapotranspiration and Soil Moisture in Decision Support Tools. JAWRA J. Am. Water Resour. Assoc. 52, 89–119. https://doi.org/10.1111/1752-1688.12371

Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J., Megown, K., 2015. Completion of the 2011 National Land Cover Database for the Conterminous United States – Representing A Decade Of Land Cover Change Information. Photogramm. Eng. Remote Sensing 81, 345–354. https://doi.org/10.1016/S0099-1112(15)30100-2

Hook, S., Fisher, J., 2019. ECOSTRESS Evapotranspiration PT-JPL Daily L3 Global 70 m V001 [Data set] [WWW Document]. NASA EOSDIS L. Process. Distrib. Act. Arch. Cent. https://doi.org/10.5067/ECOSTRESS/ECO3ETPTJPL.001

Hook, S., Hulley, G., 2022. ECOSTRESS Swath Land Surface Temperature and Emissivity Instantaneous L2 Global 70 m v002 [Data set] [WWW Document]. NASA EOSDIS L. Process. Distrib. Act. Arch. Cent. https://doi.org/10.5067/ECOSTRESS/ECO_L2_LSTE.002

Hou, X., Orth, R., 2020. Observational evidence of wildfire-promoting soil moisture anomalies. Sci. Rep. 10, 11008. https://doi.org/10.1038/s41598-020-67530-4

Hu, T., Renzullo, L.J., van Dijk, A.I.J.M., He, J., Tian, S., Xu, Z., Zhou, J., Liu, T., Liu, Q., 2020. Monitoring agricultural drought in Australia using MTSAT-2 land surface temperature retrievals. Remote Sens. Environ. 236, 111419. https://doi.org/10.1016/j.rse.2019.111419

Huang, S., Zhang, X., Chen, N., Ma, H., Zeng, J., Fu, P., Nam, W.H., Niyogi, D., 2022. Generating high-accuracy and cloud-free surface soil moisture at 1 km resolution by point-surface data fusion over the Southwestern U.S. Agric. For. Meteorol.

https://doi.org/10.1016/j.agrformet.2022.108985

Huang, S., Zhang, X., Wang, C., Chen, N., 2023. Two-step fusion method for generating 1 km seamless multi-layer soil moisture with high accuracy in the Qinghai-Tibet plateau. ISPRS J. Photogramm. Remote Sens. 197, 346–363. https://doi.org/10.1016/j.isprsjprs.2023.02.009

Huffman, G.J., Stocker, E.F., Bolvin, D.T., Nelkin, E.J., Jackson, T., 2019. GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V06, Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC). Goddard Earth Sci. Data Inf. Serv. Cent. (GES DISC).

Itenfisu, D., Elliott, R.L., Allen, R.G., Walter, I.A., 2003. Comparison of Reference Evapotranspiration Calculations as Part of the ASCE Standardization Effort. J. Irrig. Drain. Eng. 129, 440–448. https://doi.org/10.1061/(ASCE)0733-9437(2003)129:6(440)

Iverson, L.R., Prasad, A.M., Rebbeck, J., 2002. A comparison of the integrated moisture index and the topographic wetness index as related to two years of soil moisture monitoring in Zaleski State Forest, Ohio. Proc. 14th Cent. Hardwoods For. Conf.

Jackson, R.B., Canadell, J., Ehleringer, J.R., Mooney, H.A., Sala, O.E., Schulze, E.D., 1996a. A global analysis of root distributions for terrestrial biomes. Oecologia 108, 389–411. https://doi.org/10.1007/BF00333714

Jackson, R.B., Canadell, J., Ehleringer, J.R., Mooney, H.A., Sala, O.E., Schulze, E.D., 1996b. A global analysis of root distributions for terrestrial biomes. Oecologia 108, 389–411. https://doi.org/10.1007/BF00333714

Jamali, S., Seaquist, J.W., Ardo, J., Eklundh, L., 2011. Investigating temporal relationships between rainfall, soil moisture and MODIS-derived NDVI and EVI for six sites in Africa, in: 34th International Symposium on Remote Sensing of Environment - The GEOSS Era: Towards Operational Environmental Monitoring.

Jamei, Mehdi, Karbasi, M., Malik, A., Jamei, Mozhdeh, Kisi, O., Yaseen, Z.M., 2022. Long-term multi-step ahead forecasting of root zone soil moisture in different climates: Novel ensemble-based complementary data-intelligent paradigms. Agric. Water Manag. 269, 107679. https://doi.org/10.1016/j.agwat.2022.107679

James, S.E., Pärtel, M., Wilson, S.D., Peltzer, D.A., 2003. Temporal heterogeneity of soil moisture in grassland and forest. J. Ecol. 91, 234–239. https://doi.org/10.1046/j.1365-2745.2003.00758.x

Jawson, S.D., Niemann, J.D., 2007. Spatial patterns from EOF analysis of soil moisture at a large scale and their dependence on soil, land-use, and topographic properties. Adv. Water Resour. https://doi.org/10.1016/j.advwatres.2006.05.006

Jefferson, J.L., Gilbert, J.M., Constantine, P.G., Maxwell, R.M., 2015. Active subspaces for sensitivity analysis and dimension reduction of an integrated hydrologic model. Comput. Geosci. 83, 127–138. https://doi.org/10.1016/j.cageo.2015.07.001

Jensen, M.E., Allen, R.G., R.D., B., 1990. Evapotranspiration and Irrigation Water Requirements, Evapotranspiration and Irrigation Water Requirements. ASCE Manuals and Reports on Engineering Practice. No.70, New York.

Jiang, D., Lin, W., Raghavan, N., 2020. A Novel Framework for Semiconductor Manufacturing Final Test Yield Classification Using Machine Learning Techniques. IEEE Access 8, 197885–197895. https://doi.org/10.1109/ACCESS.2020.3034680

Jiménez-Martínez, J., Skaggs, T.H., van Genuchten, M.T., Candela, L., 2009. A root zone modelling approach to estimating groundwater recharge from irrigated areas. J. Hydrol. 367, 138–149. https://doi.org/10.1016/j.jhydrol.2009.01.002

Jobson, H.E., 1982. Evaporation Into the Atmosphere: Theory, History, and Applications. Eos, Trans. Am. Geophys. Union 63, 1223–1224. https://doi.org/10.1029/EO063i051p01223-04

Karthikeyan, L., Kumar, D.N., 2016. A novel approach to validate satellite soil moisture retrievals using precipitation data. J. Geophys. Res. Atmos. 121. https://doi.org/10.1002/2016JD024829

Karthikeyan, L., Mishra, A.K., 2021. Multi-layer high-resolution soil moisture estimation using machine learning over the United States. Remote Sens. Environ. 266, 112706. https://doi.org/10.1016/j.rse.2021.112706

Karthikeyan, L., Pan, M., Wanders, N., Kumar, D.N., Wood, E.F., 2017. Four decades of microwave satellite soil moisture observations: Part 1. A review of retrieval algorithms. Adv. Water Resour. 109, 106–120. https://doi.org/10.1016/j.advwatres.2017.09.006

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y., 2017. LightGBM: A highly efficient gradient boosting decision tree, in: Advances in Neural Information Processing Systems. p. 30.

Kebede, H., Fisher, D.K., Sui, R., Reddy, K.N., 2014. Irrigation Methods and Scheduling in the Delta Region of Mississippi: Current Status and Strategies to Improve Irrigation Efficiency. Am. J. Plant Sci. 05, 2917–2928. https://doi.org/10.4236/ajps.2014.520307

Keefer, T.O., Moran, M.S., Paige, G.B., 2008. Long-term meteorological and soil hydrology database, Walnut Gulch Experimental Watershed, Arizona, United States. Water Resour. Res. 44, 1–8. https://doi.org/10.1029/2006WR005702

Kerr, Y.H., Waldteufel, P., Richaume, P., Wigneron, J.P., Ferrazzoli, P., Mahmoodi, A., Al Bitar, A., Cabot, F., Gruhier, C., Juglea, S.E., Leroux, D., Mialon, A., Delwart, S., 2012. The SMOS Soil Moisture Retrieval Algorithm. IEEE Trans. Geosci. Remote Sens. 50, 1384–1403. https://doi.org/10.1109/TGRS.2012.2184548

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

Kim, S., Zhang, R., Pham, H., Sharma, A., 2019. A Review of Satellite-Derived Soil Moisture and Its Usage for Flood Estimation. Remote Sens. Earth Syst. Sci. 2, 225–246. https://doi.org/10.1007/s41976-019-00025-7

King, D.M., Skirvin, S.M., Holifield Collins, C.D., Moran, M.S., Biedenbender, S.H., Kidwell, M.R., Weltz, M.A., Diaz-Gutierrez, A., 2008. Assessing vegetation change temporally and spatially in southeastern Arizona. Water Resour. Res. 44, 1–17. https://doi.org/10.1029/2006WR005850

Kisekka, I., Peddinti, S.R., Kustas, W.P., McElrone, A.J., Bambach-Ortiz, N., McKee, L., Bastiaanssen, W., 2022. Spatial–temporal modeling of root zone soil moisture dynamics in a vineyard using machine learning and remote sensing. Irrig. Sci. 40, 761–777. https://doi.org/10.1007/s00271-022-00775-1

Kornelsen, K.C., Coulibaly, P., 2014. Root-zone soil moisture estimation using data-driven methods. Water Resour. Res. 50, 2946–2962. https://doi.org/10.1002/2013WR014127

Koster, R.D., Suarez, M.J., 2003. Impact of Land Surface Initialization on Seasonal Precipitation and Temperature Prediction. J. Hydrometeorol. 4, 408–423. https://doi.org/10.1175/1525-7541(2003)4<408:IOLSIO>2.0.CO;2

Koster, R.D., Suarez, M.J., 2001. Soil Moisture Memory in Climate Models. J. Hydrometeorol. 2, 558–570. https://doi.org/10.1175/1525-7541(2001)002<0558:SMMICM>2.0.CO;2

Krishnan, S., Indu, J., 2023. Assessing the potential of temperature/vegetation index space to infer soil moisture over Ganga Basin. J. Hydrol. 621, 129611. https://doi.org/10.1016/j.jhydrol.2023.129611

Krishnan, S., Pradhan, A., Indu, J., 2022. Estimation of high-resolution precipitation using downscaled satellite soil moisture and SM2RAIN approach. J. Hydrol. 610, 127926. https://doi.org/10.1016/j.jhydrol.2022.127926

Kühnlein, M., Appelhans, T., Thies, B., Nauss, T., 2014. Improving the accuracy of rainfall rates from optical satellite sensors with machine learning — A random forests-based approach applied to MSG SEVIRI. Remote Sens. Environ. 141, 129–143. https://doi.org/10.1016/j.rse.2013.10.026

Kurc, S.A., Small, E.E., 2004. Dynamics of evapotranspiration in semiarid grassland and shrubland ecosystems during the summer monsoon season, central New Mexico. Water Resour. Res. 40, 1–15. https://doi.org/10.1029/2004WR003068

Kustas, W.P., Schmugge, T.J., Humes, K.S., Jackson, T.J., Parry, R., Weltz, M.A., Moran, M.S., 1993. Relationships between Evaporative Fraction and Remotely Sensed Vegetation Index and Microwave Brightness Temperature for Semiarid Rangelands. J. Appl. Meteorol. 32, 1781–1790. https://doi.org/10.1175/1520-0450(1993)032<1781:RBEFAR>2.0.CO;2

Laipelt, L., Henrique Bloedow Kayser, R., Santos Fleischmann, A., Ruhoff, A., Bastiaanssen, W., Erickson, T.A., Melton, F., 2021. Long-term monitoring of evapotranspiration using the SEBAL algorithm and Google Earth Engine cloud computing. ISPRS J. Photogramm. Remote Sens. 178, 81–96. https://doi.org/10.1016/j.isprsjprs.2021.05.018

Leng, P., Li, Z.L., Liao, Q.Y., Gao, M.F., Duan, S.B., Zhang, X., Shang, G.F., 2019. Determination of all-sky surface soil moisture at fine spatial resolution synergistically using optical/thermal infrared and microwave measurements. J. Hydrol. https://doi.org/10.1016/j.jhydrol.2019.124167

Li, B., Liu, K., Wang, M., Wang, Y., He, Q., Zhuang, L., Zhu, W., 2023. High-spatiotemporal-resolution dynamic water monitoring using LightGBM model and Sentinel-2 MSI data. Int. J. Appl. Earth Obs. Geoinf. 118, 103278. https://doi.org/10.1016/j.jag.2023.103278

Li, L., Dai, Y., Shangguan, W., Wei, Z., Wei, N., Li, Q., 2022. Causality-Structured Deep

Learning for Soil Moisture Predictions. J. Hydrometeorol. 23, 1315–1331. https://doi.org/10.1175/JHM-D-21-0206.1

Li, M., Luo, X., Yang, J., Sun, Y., 2016. Applying a Locally Linear Embedding Algorithm for Feature Extraction and Visualization of MI-EEG. J. Sensors 2016, 1–9. https://doi.org/10.1155/2016/7481946

Liu, C., Sun, G., McNulty, S.G., Noormets, A., Fang, Y., 2017. Environmental controls on seasonal ecosystem evapotranspiration/potential evapotranspiration ratio as determined by the global eddy flux measurements. Hydrol. Earth Syst. Sci. 21, 311–322. https://doi.org/10.5194/hess-21-311-2017

Liu, E., Zhu, Y., Lü, H., Horton, R., Gou, Q., Wang, X., Ding, Z., Xu, H., Pan, Y., 2023. Estimation and Assessment of the Root Zone Soil Moisture from Near-Surface Measurements over Huai River Basin. Atmosphere (Basel). 14, 124. https://doi.org/10.3390/atmos14010124

Liu, W., Baret, F., Gu, X., Zhang, B., Tong, Q., Zheng, L., 2003. Evaluation of methods for soil surface moisture estimation from reflectance data. Int. J. Remote Sens. 24, 2069–2083. https://doi.org/10.1080/01431160210163155

Liu, Y., Jing, W., Wang, Q., Xia, X., 2020. Generating high-resolution daily soil moisture by using spatial downscaling techniques: a comparison of six machine learning algorithms. Adv. Water Resour. 141, 103601. https://doi.org/10.1016/j.advwatres.2020.103601

Lobell, D.B., Asner, G.P., 2002. Moisture Effects on Soil Reflectance. Soil Sci. Soc. Am. J. 66, 722–727. https://doi.org/10.2136/sssaj2002.7220

Long, D., Bai, L., Yan, L., Zhang, C., Yang, W., Lei, H., Quan, J., Meng, X., Shi, C., 2019. Generation of spatially complete and daily continuous surface soil moisture of high spatial resolution. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2019.111364

Lu, S., Ju, Z., Ren, T., Horton, R., 2009. A general approach to estimate soil water content from thermal inertia. Agric. For. Meteorol. 149, 1693–1698. https://doi.org/10.1016/j.agrformet.2009.05.011

Luo, P., Song, Y., Huang, X., Ma, H., Liu, J., Yao, Y., Meng, L., 2022. Identifying determinants of spatio-temporal disparities in soil moisture of the Northern Hemisphere using a geographically optimal zones-based heterogeneity model. ISPRS J. Photogramm. Remote Sens. 185, 111–128. https://doi.org/10.1016/j.isprsjprs.2022.01.009

Lyons, D.S., Dobrowski, S.Z., Holden, Z.A., Maneta, M.P., Sala, A., 2021. Soil moisture variation drives canopy water content dynamics across the western U.S. Remote Sens. Environ. 253, 112233. https://doi.org/10.1016/j.rse.2020.112233

Ma, H., Zeng, J., Chen, N., Zhang, X., Cosh, M.H., Wang, W., 2019. Satellite surface soil moisture from SMAP, SMOS, AMSR2 and ESA CCI: A comprehensive assessment using global ground-based observations. Remote Sens. Environ. 231, 111215. https://doi.org/10.1016/j.rse.2019.111215

Machado, R., Serralheiro, R., 2017. Soil Salinity: Effect on Vegetable Crop Growth. Management Practices to Prevent and Mitigate Soil Salinization. Horticulturae 3, 30.

https://doi.org/10.3390/horticulturae3020030

Manabe, S., Wetherald, R.T., 1986. Reduction in Summer Soil Wetness Induced by an Increase in Atmospheric Carbon Dioxide. Science (80-.). 232, 626–628. https://doi.org/10.1126/science.232.4750.626

Martínez-Fernández, J., González-Zamora, A., Almendra-Martín, L., 2021. Soil moisture memory and soil properties: An analysis with the stored precipitation fraction. J. Hydrol. https://doi.org/10.1016/j.jhydrol.2020.125622

Martínez-Fernández, J., González-Zamora, A., Sánchez, N., Gumuzzio, A., Herrero-Jiménez, C.M., 2016. Satellite soil moisture for agricultural drought monitoring: Assessment of the SMOS derived Soil Water Deficit Index. Remote Sens. Environ. 177, 277–286. https://doi.org/10.1016/j.rse.2016.02.064

Matsushima, D., Kimura, R., Shinoda, M., 2012. Soil Moisture Estimation Using Thermal Inertia: Potential and Sensitivity to Data Conditions. J. Hydrometeorol. 13, 638–648. https://doi.org/10.1175/JHM-D-10-05024.1

Matsushita, B., Yang, W., Chen, J., Onda, Y., Qiu, G., 2007. Sensitivity of the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) to Topographic Effects: A Case Study in High-density Cypress Forest. Sensors 7, 2636–2651. https://doi.org/10.3390/s7112636

McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. https://doi.org/10.1007/BF02478259

McPherson, R.A., Fiebrich, C.A., Crawford, K.C., Kilby, J.R., Grimsley, D.L., Martinez, J.E., Basara, J.B., Illston, B.G., Morris, D.A., Kloesel, K.A., Melvin, A.D., Shrivastava, H., Wolfinbarger, J.M., Bostic, J.P., Demko, D.B., Elliott, R.L., Stadler, S.J., Carlson, J.D., Sutherland, A.J., 2007. Statewide Monitoring of the Mesoscale Environment: A Technical Update on the Oklahoma Mesonet. J. Atmos. Ocean. Technol. 24, 301–321. https://doi.org/10.1175/JTECH1976.1

Melillo, J.M., Steudler, P.A., Aber, J.D., Newkirk, K., Lux, H., Bowles, F.P., Catricala, C., Magill, A., Ahrens, T., Morrisseau, S., 2002. Soil Warming and Carbon-Cycle Feedbacks to the Climate System. Science (80-.). 298, 2173–2176. https://doi.org/10.1126/science.1074153

Melliger, J.J., Niemann, J.D., 2010. Effects of gullies on space–time patterns of soil moisture in a semiarid grassland. J. Hydrol. 389, 289–300. https://doi.org/10.1016/j.jhydrol.2010.06.006

Méndez-Barroso, L.A., Vivoni, E.R., Watts, C.J., Rodríguez, J.C., 2009. Seasonal and interannual relations between precipitation, surface soil moisture and vegetation dynamics in the North American monsoon region. J. Hydrol. 377, 59–70. https://doi.org/10.1016/j.jhydrol.2009.08.009

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., Nauss, T., 2018. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. Environ. Model. Softw. 101, 1–9. https://doi.org/10.1016/j.envsoft.2017.12.001

Mhawej, M., Faour, G., 2020. Open-source Google Earth Engine 30-m evapotranspiration rates retrieval: The SEBALIGEE system. Environ. Model. Softw. 133, 104845. https://doi.org/10.1016/j.envsoft.2020.104845

Mishra, V., Cruise, J.F., Hain, C.R., Mecikalski, J.R., Anderson, M.C., 2018. Development of soil moisture profiles through coupled microwave-thermal infrared observations in the southeastern United States. Hydrol. Earth Syst. Sci. https://doi.org/10.5194/hess-22-4935-2018

Mo, T., Schmugge, T.J., 1987. A Parameterization of the Effect of Surface Roughness on Microwave Emission. IEEE Trans. Geosci. Remote Sens. GE-25, 481–486. https://doi.org/10.1109/TGRS.1987.289860

Molaei, B., Peters, R.T., Khot, L.R., Stöckle, C.O., 2022. Assessing Suitability of Auto-Selection of Hot and Cold Anchor Pixels of the UAS-METRIC Model for Developing Crop Water Use Maps. Remote Sens. 14, 4454. https://doi.org/10.3390/rs14184454

Monin, A.S., Obukhov, A.M., 1954. Osnovnye zakonomernosti turbulentnogo peremeshivanija v prizemnom sloe atmosfery (Basic Laws of Turbulent Mixing in the Atmosphere Near the Ground). Tr. Geofiz. inst. AN SSSR 24, 163–187.

Monteith, J.L., 1965. Evaporation and environment. Symp. Soc. Exp. Biol. 19, 205–34.

Morales, P., Sykes, M.T., Prentice, I.C., Smith, P., Smith, B., Bugmann, H., Zierl, B., Friedlingstein, P., Viovy, N., Sabaté, S., Sánchez, A., Pla, E., Gracia, C.A., Sitch, S., Arneth, A., Ogee, J., 2005. Comparing and evaluating process-based ecosystem model predictions of carbon and water fluxes in major European forest biomes. Glob. Chang. Biol. 11, 2211–2233. https://doi.org/10.1111/j.1365-2486.2005.01036.x

Moriasi, D.N., Starks, P.J., Steiner, J.L., Guzman, J.A., Allen, P.B., Naney, J.W., 2014. Upper Washita River Experimental Watersheds: Physiography Data. J. Environ. Qual. 43, 1298–1309. https://doi.org/10.2134/jeq2013.08.0337

Mostovoy, G. V., Anantharaj, V.G., 2008. Observed and Simulated Soil Moisture Variability over the Lower Mississippi Delta Region. J. Hydrometeorol. 9, 1125–1150. https://doi.org/10.1175/2008JHM999.1

Muller, E., Décamps, H., 2001. Modeling soil moisture–reflectance. Remote Sens. Environ. 76, 173–180. https://doi.org/10.1016/S0034-4257(00)00198-X

Murray, F.W., 1967. On the Computation of Saturation Vapor Pressure. J. Appl. Meteorol. 6, 203–204. https://doi.org/10.1175/1520-0450(1967)006<0203:OTCOSV>2.0.CO;2

Myneni, R., Knyazikhin, Y., Park, T., 2021. MODIS/Terra+Aqua Leaf Area Index/FPAR 4-Day L4 Global 500m SIN Grid V061 [Data set]., in: NASA EOSDIS Land Processes Distributed Active Archive Center. https://doi.org/10.5067/MODIS/MCD15A3H.061

Nagler, P.L., Cleverly, J., Glenn, E., Lampkin, D., Huete, A., Wan, Z., 2005. Predicting riparian evapotranspiration from MODIS vegetation indices and meteorological data. Remote Sens. Environ. 94, 17–30. https://doi.org/10.1016/j.rse.2004.08.009

NASA-JPL, 2013. NASA Shuttle Radar Topography Mission Global 1 arc second number

[WWW Document]. Nasa Lp Daac. https://doi.org/10.5067/MEaSUREs/SRTM/SRTMGL1N.003

Nash, J.E., Sutcliffe, J. V., 1970. River flow forecasting through conceptual models part I - A discussion of principles. J. Hydrol. https://doi.org/10.1016/0022-1694(70)90255-6

Nath, T.N., 2014. Soil texture and total organic matter content and its influences on soil water holding capacity of some selected tea growing soils in Sivasagar district of Assam, India. Int. J. Chem. Sci. 12, 1419–1429.

Ngo vThi, D., Ha, N.T.T., Tran Dang, Q., Koike, K., Trong, N.M., 2019. Effective Band Ratio of Landsat 8 Images Based on VNIR-SWIR Reflectance Spectra of Topsoils for Soil Moisture Mapping in a Tropical Region. Remote Sens. 11, 716. https://doi.org/10.3390/rs11060716

Nguyen, T.T., Ngo, H.H., Guo, W., Chang, S.W., Nguyen, D.D., Nguyen, C.T., Zhang, J., Liang, S., Bui, X.T., Hoang, N.B., 2022. A low-cost approach for soil moisture prediction using multi-sensor data and machine learning algorithm. Sci. Total Environ. 833, 155066. https://doi.org/10.1016/j.scitotenv.2022.155066

Ni, L., Wang, D., Wu, J., Wang, Y., Tao, Y., Zhang, J., Liu, J., 2020. Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. J. Hydrol. 586, 124901. https://doi.org/10.1016/j.jhydrol.2020.124901

Nie, Y., Tan, Y., Deng, Y., Yu, J., 2020. Suitability Evaluation of Typical Drought Index in Soil Moisture Retrieval and Monitoring Based on Optical Images. Remote Sens. 12, 2587. https://doi.org/10.3390/rs12162587

Niemann, J.D., Eltahir, E.A.B., 2004. Prediction of regional water balance components based on climate, soil, and vegetation parameters, with application to the Illinois River Basin. Water Resour. Res. 40, 1–14. https://doi.org/10.1029/2003WR002806

Njoku, E.G., Jackson, T.J., Lakshmi, V., Chan, T.K., Nghiem, S.V., 2003. Soil moisture retrieval from AMSR-E. IEEE Trans. Geosci. Remote Sens. 41, 215–229. https://doi.org/10.1109/TGRS.2002.808243

Norman, J.M., Kustas, W.P., Humes, K.S., 1995. Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature. Agric. For. Meteorol. 77, 263–293. https://doi.org/10.1016/0168-1923(95)02265-Y

NRCS, 2018. Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture [WWW Document]. Web Soil Surv. Available online https//websoilsurvey.nrcs.usda.gov/. Accessed [2018-10-12].

Ntukamazina, N., Onwonga, R.N., Sommer, R., Mukankusi, C.M., Mburu, J., Rubyogo, J.C., 2017. Effect of excessive and minimal soil moisture stress on agronomic performance of bush and climbing bean (Phaseolus vulgaris L.). Cogent Food Agric. 3, 1373414. https://doi.org/10.1080/23311932.2017.1373414

Nuñez-Olivieri, J., Muñoz-Barreto, J., Tirado-Corbalá, R., Lakhankar, T., Fisher, A., 2017. Comparison and Downscale of AMSR2 Soil Moisture Products with In Situ Measurements from the SCAN–NRCS Network over Puerto Rico. Hydrology 4, 46. https://doi.org/10.3390/hydrology4040046

Ochsner, T.E., Cosh, M.H., Cuenca, R.H., Dorigo, W.A., Draper, C.S., Hagimoto, Y., Kerr, Y.H., Larson, K.M., Njoku, E.G., Small, E.E., Zreda, M., 2013. State of the Art in Large-Scale Soil Moisture Monitoring. Soil Sci. Soc. Am. J. 77, 1888–1919. https://doi.org/10.2136/sssaj2013.03.0093

Orth, R., Seneviratne, S.I., 2012. Analysis of soil moisture memory from observations in Europe. J. Geophys. Res. Atmos. https://doi.org/10.1029/2011JD017366

Osakabe, Y., Osakabe, K., Shinozaki, K., Tran, L.-S.P., 2014. Response of plants to water stress. Front. Plant Sci. 5, 1–8. https://doi.org/10.3389/fpls.2014.00086

Ould Mohamed, S., Bruand, A., Raison, L., Bertuzzi, P., Bruckler, L., 1997. Field Evaluation and Error Analysis of Soil Water Content Measurement using the Capacitance Probe Method. Soil Sci. Soc. Am. J. 61, 399–408. https://doi.org/10.2136/sssaj1997.03615995006100020006x

Owe, M., De Jeu, R., Walker, J., 2001. A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index. IEEE Trans. Geosci. Remote Sens. https://doi.org/10.1109/36.942542

Park, Seonyoung, Im, J., Park, Sumin, Rhee, J., 2015. AMSR2 soil moisture downscaling using multisensor products through machine learning approach, in: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, pp. 1984–1987. https://doi.org/10.1109/IGARSS.2015.7326186

Pastor, J., Post, W.M., 1986. Influence of climate, soil moisture, and succession on forest carbon and nitrogen cycles. Biogeochemistry 2, 3–27. https://doi.org/10.1007/BF02186962

Patel, N.R., Anapashsha, R., Kumar, S., Saha, S.K., Dadhwal, V.K., 2009. Assessing potential of MODIS derived temperature/vegetation condition index (TVDI) to infer soil moisture status. Int. J. Remote Sens. 30, 23–39. https://doi.org/10.1080/01431160802108497

Paulik, C., Dorigo, W., Wagner, W., Kidd, R., 2014. Validation of the ASCAT Soil Water Index using in situ data from the International Soil Moisture Network. Int. J. Appl. Earth Obs. Geoinf. 30, 1–8. https://doi.org/10.1016/j.jag.2014.01.007

Peng, J., Albergel, C., Balenzano, A., Brocca, L., Cartus, O., Cosh, M.H., Crow, W.T., Dabrowska-Zielinska, K., Dadson, S., Davidson, M.W.J., de Rosnay, P., Dorigo, W., Gruber, A., Hagemann, S., Hirschi, M., Kerr, Y.H., Lovergine, F., Mahecha, M.D., Marzahn, P., Mattia, F., Musial, J.P., Preuschmann, S., Reichle, R.H., Satalino, G., Silgram, M., van Bodegom, P.M., Verhoest, N.E.C., Wagner, W., Walker, J.P., Wegmüller, U., Loew, A., 2021. A roadmap for high-resolution satellite soil moisture applications – confronting product characteristics with user requirements. Remote Sens. Environ. 252, 112162. https://doi.org/10.1016/j.rse.2020.112162

Peng, J., Borsche, M., Liu, Y., Loew, A., 2013. How representative are instantaneous evaporative fraction measurements of daytime fluxes? Hydrol. Earth Syst. Sci. 17, 3913–3919. https://doi.org/10.5194/hess-17-3913-2013

Peng, J., Loew, A., Merlin, O., Verhoest, N.E.C., 2017. A review of spatial downscaling of satellite remotely sensed soil moisture. Rev. Geophys. 55, 341–366.

https://doi.org/10.1002/2016RG000543

Peng, T., Wang, S., 2012. Effects of land use, land cover and rainfall regimes on the surface runoff and soil loss on karst slopes in southwest China. CATENA 90, 53–62. https://doi.org/10.1016/j.catena.2011.11.001

Penman, H.L., 1948. Natural evaporation from open water, bare and grass. Proc. R. Soc. Lond. Ser. A 193, 120–145.

Penna, D., Tromp-van Meerveld, H.J., Gobbi, A., Borga, M., Dalla Fontana, G., 2011. The influence of soil moisture on threshold runoff generation processes in an alpine headwater catchment. Hydrol. Earth Syst. Sci. 15, 689–702. https://doi.org/10.5194/hess-15-689-2011

Perez, F.L., 1987. Soil Moisture and the Upper Altitudinal Limit of Giant Paramo Rosettes. J. Biogeogr. 14, 173. https://doi.org/10.2307/2845071

Petropoulos, G., Carlson, T.N., Wooster, M.J., Islam, S., 2009. A review of Ts/VI remote sensing based methods for the retrieval of land surface energy fluxes and soil surface moisture. Prog. Phys. Geogr. Earth Environ. 33, 224–250. https://doi.org/10.1177/0309133309338997

Petropoulos, G., Srivastava, P., Piles, M., Pearson, S., 2018. Earth Observation-Based Operational Estimation of Soil Moisture and Evapotranspiration for Agricultural Crops in Support of Sustainable Water Management. Sustainability 10, 181. https://doi.org/10.3390/su10010181

Piles, M., Entekhabi, D., Camps, A., 2009. A Change Detection Algorithm for Retrieving High-Resolution Soil Moisture From SMAP Radar and Radiometer Observations. IEEE Trans. Geosci. Remote Sens. 47, 4125–4131. https://doi.org/10.1109/TGRS.2009.2022088

Piles, M., Petropoulos, G.P., Sánchez, N., González-Zamora, Á., Ireland, G., 2016. Towards improved spatio-temporal resolution soil moisture retrievals from the synergy of SMOS and MSG SEVIRI spaceborne observations. Remote Sens. Environ. 180, 403–417. https://doi.org/10.1016/j.rse.2016.02.048

Portal, G., Vall-Llossera, M., Piles, M., Camps, A., Chaparro, D., Pablos, M., Rossato, L., Aabouch, K., 2018. Microwave and optical data fusion for global mapping of soil moisture at high resolution, in: International Geoscience and Remote Sensing Symposium (IGARSS). https://doi.org/10.1109/IGARSS.2018.8518129

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2017. CatBoost: unbiased boosting with categorical features. Proc. 32nd Int. Conf. Neural Inf. Process. Syst. Montréal, QC, Canada 31, 6638–6648.

Qiu, J., Crow, W.T., Wagner, W., Zhao, T., 2019. Effect of vegetation index choice on soil moisture retrievals via the synergistic use of synthetic aperture radar and optical remote sensing. Int. J. Appl. Earth Obs. Geoinf. 80, 47–57. https://doi.org/10.1016/j.jag.2019.03.015

Rahimzadeh-Bajgiran, P., Berg, A.A., Champagne, C., Omasa, K., 2013. Estimation of soil moisture using optical/thermal infrared remote sensing in the Canadian Prairies. ISPRS J. Photogramm. Remote Sens. 83, 94–103. https://doi.org/10.1016/j.isprsjprs.2013.06.004

Rani, A., Kumar, N., Kumar, J., Kumar, J., Sinha, N.K., 2022. Machine learning for soil moisture assessment, in: Deep Learning for Sustainable Agriculture. Elsevier, pp. 143–168. https://doi.org/10.1016/B978-0-323-85214-2.00001-X

Ranney, K.J., Niemann, J.D., Lehman, B.M., Green, T.R., Jones, A.S., 2015. A method to downscale soil moisture to fine resolutions using topographic, vegetation, and soil data. Adv. Water Resour. 76, 81–96. https://doi.org/10.1016/j.advwatres.2014.12.003

Rawat, K.S., Sehgal, V.K., Singh, S.K., Ray, S.S., 2022. Soil moisture estimation using triangular method at higher resolution from MODIS products. Phys. Chem. Earth, Parts A/B/C 126, 103051. https://doi.org/10.1016/j.pce.2021.103051

Reichle, R., Lannoy, G. De, Koster, R.D., Crow, W.T., Kimball, J.S., Liu, Q., Bechtold, M., 2022. SMAP L4 Global 3-hourly 9 km EASE-Grid Surface and Root Zone Soil Moisture Analysis Update, Version 7. NASA Natl. Snow Ice Data Cent. Distrib. Act. Arch. Cent. https://doi.org/10.5067/LWJ6TF5SZRG3

Reichle, R.H., De Lannoy, G.J.M., Liu, Q., Ardizzone, J. V., Colliander, A., Conaty, A., Crow, W., Jackson, T.J., Jones, L.A., Kimball, J.S., Koster, R.D., Mahanama, S.P., Smith, E.B., Berg, A., Bircher, S., Bosch, D., Caldwell, T.G., Cosh, M., González-Zamora, Á., Holifield Collins, C.D., Jensen, K.H., Livingston, S., Lopez-Baeza, E., Martínez-Fernández, J., McNairn, H., Moghaddam, M., Pacheco, A., Pellarin, T., Prueger, J., Rowlandson, T., Seyfried, M., Starks, P., Su, Z., Thibeault, M., van der Velde, R., Walker, J., Wu, X., Zeng, Y., 2017. Assessment of the SMAP Level-4 Surface and Root-Zone Soil Moisture Product Using In Situ Measurements. J. Hydrometeorol. 18, 2621–2645. https://doi.org/10.1175/JHM-D-17-0063.1

Reusser, D.E., Buytaert, W., Zehe, E., 2011. Temporal dynamics of model parameter sensitivity for computationally expensive models with the Fourier amplitude sensitivity test. Water Resour. Res. 47. https://doi.org/10.1029/2010WR009947

Ritchie, J.T., 1972. Model for predicting evaporation from a row crop with incomplete cover. Water Resour. Res. 8, 1204–1213. https://doi.org/10.1029/WR008i005p01204

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS J. Photogramm. Remote Sens. 67, 93–104. https://doi.org/10.1016/j.isprsjprs.2011.11.002

Rodriguez-Iturbe, I., D'Odorico, P., Porporato, A., Ridolfi, L., 1999. On the spatial and temporal links between vegetation, climate, and soil moisture. Water Resour. Res. 35, 3709–3722. https://doi.org/10.1029/1999WR900255

Roerink, G.., Su, Z., Menenti, M., 2000. S-SEBI: A simple remote sensing algorithm to estimate the surface energy balance. Phys. Chem. Earth, Part B Hydrol. Ocean. Atmos. 25, 147–157. https://doi.org/10.1016/S1464-1909(99)00128-8

Rong, L., Duan, X., Feng, D., Zhang, G., 2017. Soil moisture variation in a farmed dry-hot valley catchment evaluated by a redundancy analysis approach. Water (Switzerland). https://doi.org/10.3390/w9020092

Rosenbaum, U., Bogena, H.R., Herbst, M., Huisman, J.A., Peterson, T.J., Weuthen, A., Western, A.W., Vereecken, H., 2012. Seasonal and event dynamics of spatial soil moisture patterns at the small catchment scale. Water Resour. Res. 48. https://doi.org/10.1029/2011WR011518

Rossiter, D.G., Poggio, L., Beaudette, D., Libohova, Z., 2022. How well does digital soil mapping represent soil geography? An investigation from the USA. SOIL 8, 559–586. https://doi.org/10.5194/soil-8-559-2022

Rouse, W.R., Mills, P.F., Stewart, R.B., 1977. Evaporation in high latitudes. Water Resour. Res. 13, 909–914. https://doi.org/10.1029/WR013i006p00909

Ruark, M.D., Niemann, J.D., Greimann, B.P., Arabi, M., 2011. Method for Assessing Impacts of Parameter Uncertainty in Sediment Transport Modeling Applications. J. Hydraul. Eng. 137, 623–636. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000343

S. G. Wangemann, R. A. Kohl, P. A. Molumeli, 2000. INFILTRATION AND PERCOLATION INFLUENCED BY ANTECEDENT SOIL WATER CONTENT AND AIR ENTRAPMENT. Trans. ASAE 43, 1517–1523. https://doi.org/10.13031/2013.3051

Sabaghy, S., Walker, J.P., Renzullo, L.J., Jackson, T.J., 2018. Spatially enhanced passive microwave derived soil moisture: Capabilities and opportunities. Remote Sens. Environ. 209, 551–580. https://doi.org/10.1016/j.rse.2018.02.065

Saber, M., Boulmaiz, T., Guermoui, M., Abdrabo, K.I., Kantoush, S.A., Sumi, T., Boutaghane, H., Nohara, D., Mabrouk, E., 2022. Examining LightGBM and CatBoost models for wadi flash flood susceptibility prediction. Geocarto Int. 37, 7462–7487. https://doi.org/10.1080/10106049.2021.1974959

Sadeghi, M., Babaeian, E., Tuller, M., Jones, S.B., 2017. The optical trapezoid model: A novel approach to remote sensing of soil moisture applied to Sentinel-2 and Landsat-8 observations. Remote Sens. Environ. 198, 52–68. https://doi.org/10.1016/j.rse.2017.05.041

Sahaar, S.A., Niemann, J.D., 2020. Impact of regional characteristics on the estimation of root-zone soil moisture from the evaporative index or evaporative fraction. Agric. Water Manag. 238, 106225. https://doi.org/10.1016/j.agwat.2020.106225

Sahaar, S.A., Niemann, J.D., Elhaddad, A., 2022. Using regional characteristics to improve uncalibrated estimation of rootzone soil moisture from optical/thermal remote-sensing. Remote Sens. Environ. 273, 112982. https://doi.org/10.1016/j.rse.2022.112982

Sahin, E.K., 2020. Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. SN Appl. Sci. 2, 1308. https://doi.org/10.1007/s42452-020-3060-1

Saito, H., Šimůnek, J., 2009. Effects of meteorological models on the solution of the surface energy balance and soil temperature variations in bare soils. J. Hydrol. 373, 545–561. https://doi.org/10.1016/j.jhydrol.2009.05.019

Saltelli, A., Chan, K.., Scott, E.M., 2000. Sensitivity Analysis. Probability and statistics series, in: John Wiley and Sons Ltd: West Sussex. pp. 1–475.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M.,

Tarantola, S., 2007. Global Sensitivity Analysis. The Primer, Global Sensitivity Analysis: The Primer. Wiley. https://doi.org/10.1002/9780470725184

Saltelli, A., Tarantola, S., Chan, K.P.S., 1999. A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output. Technometrics 41, 39–56. https://doi.org/10.1080/00401706.1999.10485594

Sanadhya, P., Gironás, J., Arabi, M., 2014. Global sensitivity analysis of hydrologic processes in major snow-dominated mountainous river basins in Colorado. Hydrol. Process. 28, 3404–3418. https://doi.org/10.1002/hyp.9896

Santos, W.J.R., Silva, B.M., Oliveira, G.C., Volpato, M.M.L., Lima, J.M., Curi, N., Marques, J.J., 2014. Soil moisture in the root zone and its relation to plant vigor assessed by remote sensing at management scale. Geoderma 221–222, 91–95. https://doi.org/10.1016/j.geoderma.2014.01.006

Saxton, K.E., Rawls, W.J., 2006. Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions. Soil Sci. Soc. Am. J. 70, 1569–1578. https://doi.org/10.2136/sssaj2005.0117

Schaap, M.G., Leij, F.J., van Genuchten, M.T., 2001. rosetta : a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. J. Hydrol. 251, 163–176. https://doi.org/10.1016/S0022-1694(01)00466-8

Schaefer, G.L., Cosh, M.H., Jackson, T.J., 2007. The USDA Natural Resources Conservation Service Soil Climate Analysis Network (SCAN). J. Atmos. Ocean. Technol. 24, 2073–2077. https://doi.org/10.1175/2007JTECHA930.1

Schenk, H.J., Jackson, R.B., 2002. Rooting depths, lateral root spreads and below-ground/above-ground allometries of plants in water-limited ecosystems. J. Ecol. 90, 480–494. https://doi.org/10.1046/j.1365-2745.2002.00682.x

Schmugge, T.J., Jackson, T.J., McKim, H.L., 1980. Survey of methods for soil moisture determination. Water Resour. Res. 16, 961–979. https://doi.org/10.1029/WR016i006p00961

Scholze, M., Kaminski, T., Knorr, W., Blessing, S., Vossbeck, M., Grant, J.P., Scipal, K., 2016. Simultaneous assimilation of SMOS soil moisture and atmospheric CO2 in-situ observations to constrain the global terrestrial carbon cycle. Remote Sens. Environ. 180, 334–345. https://doi.org/10.1016/j.rse.2016.02.058

Scott, C.A., Bastiaanssen, W.G.M., Ahmad, M.-D., 2003. Mapping Root Zone Soil Moisture Using Remotely Sensed Optical Imagery. J. Irrig. Drain. Eng. 129, 326–335. https://doi.org/10.1061/(ASCE)0733-9437(2003)129:5(326)

Sechidis, K., Tsoumakas, G., Vlahavas, I., 2011. On the Stratification of Multi-label Data, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp. 145–158. https://doi.org/10.1007/978-3-642-23808-6_10

Seleiman, M.F., Al-Suhaibani, N., Ali, N., Akmal, M., Alotaibi, M., Refay, Y., Dindaroglu, T., Abdul-Wajid, H.H., Battaglia, M.L., 2021. Drought stress impacts on plants and different approaches to alleviate its adverse effects. Plants. https://doi.org/10.3390/plants10020259

Senay, G.B., Bohms, S., Singh, R.K., Gowda, P.H., Velpuri, N.M., Alemu, H., Verdin, J.P., 2013. Operational Evapotranspiration Mapping Using Remote Sensing and Weather Datasets: A New Parameterization for the SSEB Approach. JAWRA J. Am. Water Resour. Assoc. 49, 577–591. https://doi.org/10.1111/jawr.12057

Seneviratne, S.I., Corti, T., Davin, E.L., Hirschi, M., Jaeger, E.B., Lehner, I., Orlowsky, B., Teuling, A.J., 2010. Investigating soil moisture–climate interactions in a changing climate: A review. Earth-Science Rev. 99, 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004

Seneviratne, S.I., Lüthi, D., Litschi, M., Schär, C., 2006. Land–atmosphere coupling and climate change in Europe. Nature 443, 205–209. https://doi.org/10.1038/nature05095

Senyurek, V., Lei, F., Boyd, D., Kurum, M., Gurbuz, A.C., Moorhead, R., 2020. Machine Learning-Based CYGNSS Soil Moisture Estimates over ISMN sites in CONUS. Remote Sens. 12, 1168.

Shuttleworth, W., Gurney, R., Hsu, A., Ormsby, J., 1989. FIFE: the variation in energy partition at surface flux sites. Remote Sens. Large-Scale Glob. Porc. (IAHS Publ.) 186, 67–74.

Sims, A.P., Niyogi, D.D.S., Raman, S., 2002. Adopting drought indices for estimating soil moisture: A North Carolina case study. Geophys. Res. Lett. 29. https://doi.org/10.1029/2001GL013343

Šimůnek, J., M. Šejna, A., Saito, H., Sakai, M., Genuchten, M.T. Van, 2013. The HYDRUS-1D software package for simulating the movement of water, heat, and multiple solutes in variably saturated media, version 4.17. HYDRUS Softw. Ser. 3, Dep. Environ. Sci. Univ. Calif. Riverside, Riverside, California, USA pp.343.

Šimůnek, J., van Genuchten, M.T., Šejna, M., 2008. Development and Applications of the HYDRUS and STANMOD Software Packages and Related Codes. Vadose Zo. J. 7, 587–600. https://doi.org/10.2136/vzj2007.0077

Singh, A., Gaurav, K., 2023. Deep learning and data fusion to estimate surface soil moisture from multi-sensor satellite images. Sci. Rep. 13, 2251. https://doi.org/10.1038/s41598-023-28939-9

Smith, E.A., Hsu, A.Y., Crosson, W.L., Field, R.T., Fritschen, L.J., Gurney, R.J., Kanemasu, E.T., Kustas, W.P., Nie, D., Shuttleworth, W.J., Stewart, J.B., Verma, S.B., Weaver, H.L., Wesely, M.L., 1992. Area-averaged surface fluxes and their time-space variability over the FIFE experimental domain. J. Geophys. Res. Atmos. 97, 18599–18622. https://doi.org/10.1029/91JD03060

Smith, J.L., Doran, J.W., 2015. Measurement and Use of pH and Electrical Conductivity for Soil Quality Analysis, in: Methods for Assessing Soil Quality. pp. 169–185. https://doi.org/10.2136/sssaspecpub49.c10

Soil Survey Staff, 2020. Gridded National Soil Survey Geographic (gNATSGO) Database for the Conterminous United States [WWW Document]. United States Dep. Agric. Nat. Resour. Conserv. Serv. URL https://nrcs.app.box.com/v/soils (accessed 1.5.23).

Soliman, A., Heck, R., Brenning, A., Brown, R., Miller, S., 2013. Remote Sensing of Soil Moisture in Vineyards Using Airborne and Ground-Based Thermal Inertia Data. Remote

Sens. 5, 3729–3748. https://doi.org/10.3390/rs5083729

Song, Y.M., Wang, Z.F., Qi, L.L., Huang, A.N., 2019. Soil Moisture Memory and Its Effect on the Surface Water and Heat Fluxes on Seasonal and Interannual Time Scales. J. Geophys. Res. Atmos. https://doi.org/10.1029/2019JD030893

Starks, P. J., Fiebrich, C.A., Grimsley, D.L., Garbrecht, J.D., Steiner, J.L., Guzman, J.A., Moriasi, D.N., 2014. Upper Washita River Experimental Watersheds: Meteorologic and Soil Climate Measurement Networks. J. Environ. Qual. 43, 1239–1249. https://doi.org/10.2134/jeq2013.08.0312

Starks, Patrick J., Steiner, J.L., Stern, A.J., 2014. Upper Washita River Experimental Watersheds: Land Cover Data Sets (1974-2007) for Two Southwestern Oklahoma Agricultural Watersheds. J. Environ. Qual. 43, 1310–1318. https://doi.org/10.2134/jeq2013.07.0292

Steiner, J.L., Starks, P.J., Daniel, J.A., Garbrecht, J.D., Moriasi, D., McIntyre, S., Chen, J.-S., 2008. Environmental effects of agricultural conservation: A framework for research in two watersheds in Oklahoma's Upper Washita River Basin. J. Soil Water Conserv. 63, 443–452. https://doi.org/10.2489/jswc.63.6.443

Stevens, M.R., Dupree, J.A., Kuzmiak, J.M., 2008. Temporal and Spatial Variations in Precipitation, Streamflow, Suspended-Sediment Loads and Yields, and Land-Condition Trend Analysis at the U.S. Army Pinon Canyon Maneuver Site, Las Animas County, Colorado, 1983 through 2007, USGS Numbered Series. https://doi.org/10.3133/sir20085111

Stisen, S., Sandholt, I., Nørgaard, A., Fensholt, R., Jensen, K.H., 2008. Combining the triangle method with thermal inertia to estimate regional evapotranspiration — Applied to MSG-SEVIRI data in the Senegal River basin. Remote Sens. Environ. 112, 1242–1255. https://doi.org/10.1016/j.rse.2007.08.013

Su, Z., 2002. The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. Hydrol. Earth Syst. Sci. 6, 85–100. https://doi.org/10.5194/hess-6-85-2002

Sumargo, E., McMillan, H., Weihs, R., Ellis, C.J., Wilson, A.M., Ralph, F.M., 2021. A soil moisture monitoring network to assess controls on runoff generation during atmospheric river events. Hydrol. Process. 35. https://doi.org/10.1002/hyp.13998

Sun, H., Zhang, X., Zhao, X., 2022. Series or Parallel? An Exploration in Coupling Physical Model and Machine Learning Method for Disaggregating Satellite Microwave Soil Moisture. IEEE Trans. Geosci. Remote Sens. 60, 1–15. https://doi.org/10.1109/TGRS.2022.3216343

Sun, Z., Wei, B., Su, W., Shen, W., Wang, C., You, D., Liu, Z., 2011. Evapotranspiration estimation based on the SEBAL model in the Nansi Lake Wetland of China. Math. Comput. Model. 54, 1086–1092. https://doi.org/10.1016/j.mcm.2010.11.039

Sutanto, S.J., Wenninger, J., Coenders-Gerrits, A.M.J., Uhlenbrook, S., 2012. Partitioning of evaporation into transpiration, soil evaporation and interception: a comparison between isotope measurements and a HYDRUS-1D model. Hydrol. Earth Syst. Sci. 16, 2605–2616. https://doi.org/10.5194/hess-16-2605-2012

Tagesson, T., Horion, S., Nieto, H., Zaldo Fornies, V., Mendiguren González, G., Bulgin, C.E., Ghent, D., Fensholt, R., 2018. Disaggregation of SMOS soil moisture over West Africa using the Temperature and Vegetation Dryness Index based on SEVIRI land surface parameters. Remote Sens. Environ. 206, 424–441. https://doi.org/10.1016/j.rse.2017.12.036

Takagi, K., Lin, H.S., 2011. Temporal Dynamics of Soil Moisture Spatial Variability in the Shale Hills Critical Zone Observatory. Vadose Zo. J. 10, 832–842. https://doi.org/10.2136/vzj2010.0134

Tao, S., Zhang, X., Feng, R., Qi, W., Wang, Y., Shrestha, B., 2023. Retrieving soil moisture from grape growing areas using multi-feature and stacking-based ensemble learning modeling. Comput. Electron. Agric. 204, 107537. https://doi.org/10.1016/j.compag.2022.107537

Tasumi, M., Trezza, R., Allen, R.G., Wright, J.L., 2005. Operational aspects of satellite-based energy balance models for irrigated crops in the semi-arid U.S. Irrig. Drain. Syst. 19, 355–376. https://doi.org/10.1007/s10795-005-8138-9

Theobald, D.M., Harrison-Atlas, D., Monahan, W.B., Albano, C.M., 2015. Ecologically-Relevant Maps of Landforms and Physiographic Diversity for Climate Adaptation Planning. PLoS One 10, e0143619. https://doi.org/10.1371/journal.pone.0143619

Thirumalaiah, K., Deo, M.C., 1998. River Stage Forecasting Using Artificial Neural Networks. J. Hydrol. Eng. 3, 26–32. https://doi.org/10.1061/(ASCE)1084-0699(1998)3:1(26)

Thornthwaite, C., Mather, J.R., 1955. The water balance. Publ. Climatol. Inst. Tech., Lab. Clim 8, 104.

Thornthwaite, C.W., 1948. An Approach Toward a Rational Classification of Climate. Soil Sci. 66, 77. https://doi.org/10.1097/00010694-194807000-00007

Tong, C., Wang, H., Magagi, R., Goïta, K., Zhu, L., Yang, M., Deng, J., 2020. Soil Moisture Retrievals by Combining Passive Microwave and Optical Data. Remote Sens. 12, 3173. https://doi.org/10.3390/rs12193173

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., Traver, I.N., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'Abbate, M., Croci, R., Pietropaolo, A., Huchler, M., Rostan, F., 2012. GMES Sentinel-1 mission. Remote Sens. Environ. 120, 9–24. https://doi.org/10.1016/j.rse.2011.05.028

Trabucco, a., Zomer, R.J., 2009. Global aridity index (global-aridity) and global potential evapo-transpiration (global-PET) geospatial database. CGIAR Consort. Spat. Inf.

Trabucco, A., Zomer, R.J., 2019. Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v2. CGIAR Consort. Spat. Inf. https://doi.org/https://doi.org/10.6084/m9.figshare.7504448.v3

UNEP, 1997. World atlas of desertification 2ED. United Nations Environ. Program.

Uribeetxebarria, A., Castellón, A., Aizpurua, A., 2023. Optimizing Wheat Yield Prediction Integrating Data from Sentinel-1 and Sentinel-2 with CatBoost Algorithm. Remote Sens.

15, 1640. https://doi.org/10.3390/rs15061640

Van doninck, J., Peters, J., De Baets, B., De Clercq, E.M., Ducheyne, E., Verhoest, N.E.C., 2011. The potential of multitemporal Aqua and Terra MODIS apparent thermal inertia as a soil moisture indicator. Int. J. Appl. Earth Obs. Geoinf. 13, 934–941. https://doi.org/10.1016/j.jag.2011.07.003

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

Van Genuchten, M.T., 1987. A numerical model for water and solute movement in and below the root zone. United States Dep. Agric. Agric. Res. Serv. US Salin. Lab.

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

Vergopolan, N., Sheffield, J., Chaney, N.W., Pan, M., Beck, H.E., Ferguson, C.R., Torres-Rojas, L., Eigenbrod, F., Crow, W., Wood, E.F., 2022. High-Resolution Soil Moisture Data Reveal Complex Multi-Scale Spatial Variability Across the United States. Geophys. Res. Lett. 49. https://doi.org/10.1029/2022GL098586

Verstraeten, W.W., Veroustraete, F., van der Sande, C.J., Grootaers, I., Feyen, J., 2006. Soil moisture retrieval using thermal inertia, determined with visible and thermal spaceborne data, validated for European forests. Remote Sens. Environ. 101, 299–314. https://doi.org/10.1016/j.rse.2005.12.016

Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., Figa-Saldaña, J., de Rosnay, P., Jann, A., Schneider, S., Komma, J., Kubu, G., Brugger, K., Aubrecht, C., Züger, J., Gangkofner, U., Kienberger, S., Brocca, L., Wang, Y., Blöschl, G., Eitzinger, J., Steinnocher, K., 2013. The ASCAT Soil Moisture Product: A Review of its Specifications, Validation Results, and Emerging Applications. Meteorol. Zeitschrift 22, 5–33. https://doi.org/10.1127/0941-2948/2013/0399

Walter, I.A., Allen, R.G., Elliott, R., Jensen, M.E., Itenfisu, D., Mecham, B., Howell, T.A., Snyder, R., Brown, P., Echings, S., Spofford, T., Hattendorf, M., Cuenca, R.H., Wright, J.L., Martin, D., 2001. ASCE's Standardized Reference Evapotranspiration Equation, in: Watershed Management and Operations Management 2000. American Society of Civil Engineers, Reston, VA, pp. 1–11. https://doi.org/10.1061/40499(2000)126

Wang, H., Li, Z., Zhang, W., Ye, X., Liu, X., 2022. A Modified Temperature-Vegetation Dryness Index (MTVDI) for Assessment of Surface Soil Moisture Based on MODIS Data. Chinese Geogr. Sci. 32, 592–605. https://doi.org/10.1007/s11769-022-1288-y

Wang, L., Qu, J.J., 2009. Satellite remote sensing applications for surface soil moisture monitoring: A review. Front. Earth Sci. China 3, 237–247. https://doi.org/10.1007/s11707-009-0023-7

Wang, S., Fu, B.J., Gao, G.Y., Yao, X.L., Zhou, J., 2012. Soil moisture and evapotranspiration of different land cover types in the Loess Plateau, China. Hydrol. Earth Syst. Sci. 16, 2883–

2892. https://doi.org/10.5194/hess-16-2883-2012

Wang, T., Zlotnik, V.A., 2012. A complementary relationship between actual and potential evapotranspiration and soil effects. J. Hydrol. 456–457, 146–150. https://doi.org/10.1016/j.jhydrol.2012.03.034

Wang, W., Gong, Y., Xing, X., 2020. Groundwater Evaporation for Salt-Affected Soil Under Plastic Film-Covered Cultivation Condition: a Review. J. Soil Sci. Plant Nutr. 20, 1229–1237. https://doi.org/10.1007/s42729-020-00207-z

Wang, W., Huang, D., Wang, X.-G., Liu, Y.-R., Zhou, F., 2011. Estimation of soil moisture using trapezoidal relationship between remotely sensed land surface temperature and vegetation index. Hydrol. Earth Syst. Sci. 15, 1699–1712. https://doi.org/10.5194/hess-15-1699-2011

Wang, Y., Leng, P., Peng, J., Marzahn, P., Ludwig, R., 2021. Global assessments of two blended microwave soil moisture products CCI and SMOPS with in-situ measurements and reanalysis data. Int. J. Appl. Earth Obs. Geoinf. 94, 102234. https://doi.org/10.1016/j.jag.2020.102234

Wang, Y., Yang, J., Chen, Y., Fang, G., Duan, W., Li, Y., De Maeyer, P., 2019. Quantifying the Effects of Climate and Vegetation on Soil Moisture in an Arid Area, China. Water 11, 767. https://doi.org/10.3390/w11040767

Wei, L., Yang, M., Li, Z., Shao, J., Li, L., Chen, P., Li, S., Zhao, R., 2022. Experimental Investigation of Relationship between Infiltration Rate and Soil Moisture under Rainfall Conditions. Water 14, 1347. https://doi.org/10.3390/w14091347

Wei, Z., Meng, Y., Zhang, W., Peng, J., Meng, L., 2019. Downscaling SMAP soil moisture estimation with gradient boosting decision tree regression over the Tibetan Plateau. Remote Sens. Environ. 225, 30–44. https://doi.org/10.1016/j.rse.2019.02.022

Weidong, L., Baret, F., Xingfa, G., Qingxi, T., Lanfen, Z., Bing, Z., 2002. Relating soil surface moisture to reflectance. Remote Sens. Environ. 81, 238–246. https://doi.org/10.1016/S0034-4257(01)00347-9

West, N.E., 1983. Great Basin- Colorado Plateau Sagebrush Semi-Desert. Temp. deserts semi-deserts 5, 331–369.

Western, A.W., Grayson, R.B., Blöschl, G., Willgoose, G.R., McMahon, T.A., 1999. Observed spatial organization of soil moisture and its relation to terrain indices. Water Resour. Res. 35, 797–810. https://doi.org/10.1029/1998WR900065

Wight, J.R., Hanson, C.L., 1990. Crop Coefficients for Rangeland. J. Range Manag. 43, 482. https://doi.org/10.2307/4002349

Wilson, G.W., Fredlund, D.G., Barbour, S.L., 1997. The effect of soil suction on evaporative fluxes from soil surfaces. Can. Geotech. J. 34, 145–155. https://doi.org/10.1139/t96-078

Wu, M., Scholze, M., Kaminski, T., Voßbeck, M., Tagesson, T., 2020. Using SMOS soil moisture data combining CO2 flask samples to constrain carbon fluxes during 2010–2015 within a Carbon Cycle Data Assimilation System (CCDAS). Remote Sens. Environ. 240,

111719. https://doi.org/10.1016/j.rse.2020.111719

Wu, X., Walker, J.P., Rudiger, C., Panciera, R., Gao, Y., 2017. Intercomparison of Alternate Soil Moisture Downscaling Algorithms Using Active–Passive Microwave Observations. IEEE Geosci. Remote Sens. Lett. 14, 179–183. https://doi.org/10.1109/LGRS.2016.2633521

Xiang, K., Li, Y., Horton, R., Feng, H., 2020. Similarity and difference of potential evapotranspiration and reference crop evapotranspiration – a review. Agric. Water Manag. 232, 106043. https://doi.org/10.1016/j.agwat.2020.106043

Xu, B., Park, T., Yan, K., Chen, C., Zeng, Y., Song, W., Yin, G., Li, J., Liu, Q., Knyazikhin, Y., Myneni, R., 2018. Analysis of Global LAI/FPAR Products from VIIRS and MODIS Sensors for Spatio-Temporal Consistency and Uncertainty from 2012–2016. Forests 9, 73. https://doi.org/10.3390/f9020073

Xu, J., Ma, X., Logsdon, S.D., Horton, R., 2012. Short, Multineedle Frequency Domain Reflectometry Sensor Suitable for Measuring Soil Water Content. Soil Sci. Soc. Am. J. 76, 1929–1937. https://doi.org/10.2136/sssaj2011.0361

Xu, L., Abbaszadeh, P., Moradkhani, H., Chen, N., Zhang, X., 2020. Continental drought monitoring using satellite soil moisture, data assimilation and an integrated drought index. Remote Sens. Environ. 250, 112028. https://doi.org/10.1016/j.rse.2020.112028

Yadav, D.S., Jaiswal, B., Gautam, M., Agrawal, M., 2020. Soil Acidification and its Impact on Plants, in: Plant Responses to Soil Pollution. Springer Singapore, Singapore, pp. 1–26. https://doi.org/10.1007/978-981-15-4964-9_1

Yang, Yuting, Guan, H., Long, D., Liu, B., Qin, G., Qin, J., Batelaan, O., 2015. Estimation of Surface Soil Moisture from Thermal Infrared Remote Sensing Using an Improved Trapezoid Method. Remote Sens. 7, 8250–8270. https://doi.org/10.3390/rs70708250

Yang, Yongmin, Su, H., Zhang, R., Tian, J., Li, L., 2015. An enhanced two-source evapotranspiration model for land (ETEML): Algorithm and evaluation. Remote Sens. Environ. 168, 54–65. https://doi.org/10.1016/j.rse.2015.06.020

Yang, Z., He, Q., Miao, S., Wei, F., Yu, M., 2023. Surface Soil Moisture Retrieval of China Using Multi-Source Data and Ensemble Learning. Remote Sens. 15, 2786. https://doi.org/10.3390/rs15112786

Yao, A.Y.M., 1974. Agricultural potential estimated from the ratio of actual to potential evapotranspiration. Agric. Meteorol. 13, 405–417. https://doi.org/10.1016/0002-1571(74)90081-8

Zachos, L.G., Swann, C.T., Altinakar, M.S., McGrath, M.Z., Thomas, D., 2016. Flood vulnerability indices and emergency management planning in the Yazoo Basin, Mississippi. Int. J. Disaster Risk Reduct. 18, 89–99. https://doi.org/10.1016/j.ijdrr.2016.03.012

Zakeri, F., Mariethoz, G., 2021. A review of geostatistical simulation models applied to satellite remote sensing: Methods and applications. Remote Sens. Environ. 259, 112381. https://doi.org/10.1016/j.rse.2021.112381

Zeng, J., Li, Z., Chen, Q., Bi, H., 2015. Method for Soil Moisture and Surface Temperature

Estimation in the Tibetan Plateau Using Spaceborne Radiometer Observations. IEEE Geosci. Remote Sens. Lett. 12, 97–101. https://doi.org/10.1109/LGRS.2014.2326890

Zeng, N., Neelin, J.D., Lau, K.M., Tucker, C.J., 1999. Enhancement of interdecadal climate variability in the Sahel by vegetation interaction. Science (80-.). 286, 1537–1540. https://doi.org/10.1126/science.286.5444.1537

Zhan, X., Zheng, W., Fang, L., Liu, J., Hain, C., Yin, J., Ek, M., 2016. A preliminary assessment of the impact of SMAP Soil Moisture on numerical weather Forecasts from GFS and NUWRF models, in: 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, pp. 5229–5232. https://doi.org/10.1109/IGARSS.2016.7730362

Zhang, D., Zhou, G., 2016. Estimation of Soil Moisture from Optical and Thermal Remote Sensing: A Review. Sensors 16, 1308. https://doi.org/10.3390/s16081308

Zhang, J., Liu, K., Wang, M., 2021. Downscaling Groundwater Storage Data in China to a 1-km Resolution Using Machine Learning Methods. Remote Sens. 13, 523. https://doi.org/10.3390/rs13030523

Zhang, R., Chen, Z.-Y., Xu, L.-J., Ou, C.-Q., 2019. Meteorological drought forecasting based on a statistical model with machine learning techniques in Shaanxi province, China. Sci. Total Environ. 665, 338–346. https://doi.org/10.1016/j.scitotenv.2019.01.431

Zhang, Y.-K., Schilling, K.E., 2006. Effects of land cover on water table, soil moisture, evapotranspiration, and groundwater recharge: A Field observation and analysis. J. Hydrol. 319, 328–338. https://doi.org/10.1016/j.jhydrol.2005.06.044

Zhang, Y.-Y., Wu, W., Liu, H., 2019. Factors affecting variations of soil pH in different horizons in hilly regions. PLoS One 14, e0218563. https://doi.org/10.1371/journal.pone.0218563

Zhang, Y., Zhao, Z., Zheng, J., 2020. CatBoost: A new approach for estimating daily reference crop evapotranspiration in arid and semi-arid regions of Northern China. J. Hydrol. 588, 125087. https://doi.org/10.1016/j.jhydrol.2020.125087

Zhao, W., Li, Z.-L., 2013. Sensitivity study of soil moisture on the temporal evolution of surface temperature over bare surfaces. Int. J. Remote Sens. 34, 3314–3331. https://doi.org/10.1080/01431161.2012.716532

Zhao, Y., Tang, J., Graham, C., Zhu, Q., Takagi, K., Lin, H., 2012. Hydropedology in the ridge and valley: Soil moisture patterns and preferential flow dynamics in two contrasting landscapes, in: Hydropedology: Synergistic Integration of Soil Science and Hydrology. https://doi.org/10.1016/B978-0-12-386941-8.00012-5

Zhu, F., Gao, J., Yang, J., Ye, N., 2022. Neighborhood linear discriminant analysis. Pattern Recognit. 123, 108422. https://doi.org/10.1016/j.patcog.2021.108422

Zhu, W., Wei, J., Xiu, H., Jia, S., Lv, A., 2021. Coupled and continuous estimation of soil moisture and evaporative fraction within the remotely sensed surface temperature-vegetation index framework. J. Hydrol. 592, 125827. https://doi.org/10.1016/j.jhydrol.2020.125827

Zohaib, M., Kim, H., Choi, M., 2017. Evaluating the patterns of spatiotemporal trends of root

zone soil moisture in major climate regions in East Asia. J. Geophys. Res. https://doi.org/10.1002/2016JD026379

Zomer, R.J., Xu, J., Trabucco, A., 2022. Version 3 of the Global Aridity Index and Potential Evapotranspiration Database. Sci. Data 9, 409. https://doi.org/10.1038/s41597-022-01493-1

Zreda, M., Desilets, D., Ferré, T.P.A., Scott, R.L., 2008. Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons. Geophys. Res. Lett. 35, L21402. https://doi.org/10.1029/2008GL035655

Zribi, M., 2003. Surface soil moisture estimation from the synergistic use of the (multi-incidence and multi-resolution) active microwave ERS Wind Scatterometer and SAR data. Remote Sens. Environ. 86, 30–41. https://doi.org/10.1016/S0034-4257(03)00065-8