# Retrieving Gap-Free Daily Soil Moisture Using Surface Flux Equilibrium Theory

**Pushpendra Raghav[1] and Mukesh Kumar[1*]**

[1] Department of Civil, Construction, and Environmental Engineering, University of Alabama, Tuscaloosa, AL, USA

[*] Corresponding Author, Email: mkumar4@eng.ua.edu, Phone: +1 205 348 0180

## Abstract

Soil moisture is a dominant control on crop productivity, land-atmosphere feedbacks, and the hydrologic response of watersheds. Despite its importance, obtaining gap-free daily moisture data remains challenging. For example, remote sensing-based soil moisture products often have gaps arising from limits posed by the presence of clouds and satellite revisit period. Here, we retrieve a proxy of daily root zone soil moisture (RZSM) using the Surface Flux Equilibrium theory. Our method is parsimonious, and only needs widely available meteorological data and standard land-surface parameters. Evaluation of the retrievals at Oklahoma Mesonet sites shows that our method, overall, matches or outperforms RZSM estimates from both the Variable Infiltration Capacity (VIC) model and the Atmosphere-Land EXchange Inversion (ALEXI) model. RZSM from our method could serve as a more accurate and temporally-complete alternative for a variety of applications including mapping of agricultural droughts, assimilation of RZSM for hydrometeorological forecasting, and design of optimal irrigation schedules.

Keywords: Soil moisture proxy, fraction of potential evapotranspiration, kernel regression, surface flux equilibrium theory, ALEXI

## 1. Introduction

Soil Moisture (SM) plays a critical role in the regional and global water cycle. The distribution of soil moisture influences the incidence and intensity of floods (Chen *et al* 2015, De Michele and Salvadori 2002, Norbiato *et al* 2008), and droughts (Samaniego *et al* 2018, Wang *et al* 2011), mediates water quality (Guo *et al* 2019, Zi *et al* 2016), and has a range of ecohydrological implications including on crop productivity (Bolten *et al* 2009, Chakrabarti *et al* 2014, Ines *et al* 2013) and the growth and sustainability of trees (Anderegg *et al* 2015, Liu *et al* 2017, Porporato *et al* 2002). SM also plays a vital role in the partitioning of water and energy fluxes between land and atmosphere (Lettenmaier and Famiglietti 2006, Liu *et al* 2020, Mintz and Serafini 1992, Trenberth *et al* 2007). Despite its influence on a range of ecohydrological and atmospheric processes, observed SM data at daily interval is not readily available over large domains. While remote sensing derived soil moisture products (Chauhan *et al* 2003, Entekhabi *et al* 2010, Kerr *et al* 2010, Njoku *et al* 2003, Torres *et al* 2012, Wagner *et al* 2013) do provide moisture information over large scales, they often suffer from temporal and spatial data gaps due to presence of cloud cover, narrow swath, and sparse revisit schedules (Anderson *et al* 2007a, Mao *et al* 2019, Sabaghy *et al* 2018, Walker and Houser 2004). For example, Anderson *et al* (2007a) reported that good quality thermal infrared (TIR) imagery, which is often used for moisture retrieval over the continental US, was only available around 30% of the time in their study area. Similarly, the temporal resolution of the latest 3-km SM product from the National Aeronautics and Space Administration Soil Moisture Active Passive (SMAP) mission, the SMAP/Sentinel-1 L2_SM_SP SM product (Das *et al* 2018) is

 

reported to vary between 3- to 12-days depending on the revisit schedules of backscatter measurements of Sentinel-1A and Sentinel-1B sensors (Das *et al* 2016, Entekhabi *et al* 2010, Mao *et al* 2019). The percentage of missing days for SMAP L3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture Version 6 (SPL3SMP) product and SMAP Enhanced L3 Radiometer Global Daily 9 km EASE-Grid Soil Moisture Version 3 (SPL3SMP_E) product at Oklahoma Mesonet sites, the study area under consideration, are over 50% for ascending and descending overpass and over 15% for composite data during 2015-2019 (see Figure S1 in Supporting Information).

Here, we propose a method based on the Surface Flux Equilibrium Theory (SFET) to retrieve Fraction of Available Water ($fAW$), a proxy for root-zone soil moisture. The proxy is then used to also obtain Volumetric Soil Moisture (VSM) in the root zone using data of soil properties. The remainder of this paper is organized as follows: Section 2 presents details of our methodology and a concise overview of the study area and datasets. Results are presented in section 3. Section 4 presents conclusions and related discussion.

## 2. Methods

### *2.1 Surface Flux Equilibrium Theory (SFET)*

McColl *et al* (2019) presented SFET where it was hypothesized that in the regions with no or minimal advective moisture convergence (e.g., inland continental regions), the near-surface atmosphere is in the state of "surface flux equilibrium", i.e., the surface heating and surface moistening terms in the near-surface relative humidity budget are in the state of equilibrium at daily to monthly time scale. McColl and Rigden (2020) provided physical explanations for why the hypothesis stands using a simple model of an idealized atmospheric boundary layer (ABL). The approach does not require an explicit parameterization of land surface conditions as it assumes that the turbulent fluxes at the land surface (latent and sensible heat fluxes) are encoded in the near-surface atmospheric states (near surface air temperature and specific humidity, respectively). Using this theory, the Bowen ratio ($B$) for a given location can be estimated as:

$$B \approx \frac{R_v C_p T_a^2}{\lambda^2 Q_a} \tag{1}$$

where $B$ is Bowen ratio ($= H/\lambda E$) [-], H is the sensible heat flux [W m$^{-2}$], $\lambda$ is the latent heat of vaporization [ J kg$^{-1}$], $\lambda E$ is the latent heat flux [W m$^{-2}$], $R_v = 461.5$ is the gas constant for water vapor [J kg$^{-1}$ K$^{-1}$], $C_p = 1005$ is the specific heat capacity of dry air at constant pressure [J kg$^{-1}$ K$^{-1}$], $T_a$ is the screen-level air temperature [K], and $Q_a$ is the screen-level specific humidity of the air [kg kg$^{-1}$]. The latent heat flux ($\lambda E$) can then be obtained using the following relation derived based on surface energy balance:

$$\lambda E = \frac{R_n - G}{1 + B} \tag{2}$$

where $R_n$ is the net solar radiation [W m$^{-2}$], and $G$ is the ground heat flux [W m$^{-2}$]. More details regarding the calculation of $R_n$ and $G$ are provided in Supporting Information Text S1. SFET based estimates of evapotranspiration have been shown to be remarkably accurate, with prediction errors comparable to errors in the eddy covariance measurements (McColl and Rigden 2020).

### *2.2 Soil moisture proxy retrieval*

Here we retrieve Fraction of Available Water ($fAW$), a commonly used proxy for soil moisture (Anderson *et al* 2007a, Hain *et al* 2011, Hain *et al* 2009). $fAW$ is defined as:

$$fAW = \frac{(\theta - \theta_{wp}) \times d}{(\theta_{fc} - \theta_{wp}) \times d} \tag{3}$$

where $\theta$ is the soil moisture content in the root zone soil layer [m$^3$ m$^{-3}$], $\theta_{wp}$ is the soil moisture content at wilting point [m$^3$ m$^{-3}$], $\theta_{fc}$ is the soil moisture content at field capacity [m$^3$ m$^{-3}$], and $d$ is the root-zone depth [m]. As latent heat flux is dominantly controlled by evapotranspiration of soil moisture, or its proxy, e.g., $fAW$, can be potentially retrieved based on evapotranspiration estimates. For example, land surface models (LSMs) (Camporese *et al* 2014, Ferguson *et al* 2016, Hanasaki *et al* 2013, Panday and Huyakorn 2004, Wang *et al* 2013, Wigmosta *et al* 1994, Wood and Lettenmaier 1996) often use a simple soil moisture stress function to relate the simulated available water fraction ($fAW$) to the ratio of actual evapotranspiration and potential evapotranspiration. This ratio, hereafter referred to as the fraction of potential evapotranspiration ($fPET$), is defined as:

$$fPET = \frac{ETa}{PET} \tag{4}$$

where $ETa$ is actual evapotranspiration [mm day$^{-1}$], and $PET$ is potential evapotranspiration [mm day$^{-1}$]. Our method obtains $ETa$ using SFET as described in the previous section. The $PET$ is estimated using the Penman-Monteith (PM) equation (Monteith 1965, Penman 1948) assuming the soil moisture conditions are at field capacity (see Equation 3 in Supporting Information Text S2). Once $fPET$ is evaluated,

the underlying relation between $fAW$ and $fPET$ is used to retrieve $fAW$.

The relation between $fAW$ and $fPET$ is usually derived using pre-defined functions (Anderson *et al* 2007a, Hain *et al* 2009, Mahrt and Pan 1984, Stewart and Verma 1992, Wetzel and Chang 1987). By intercomparing four different functions that relate $fAW$ and $fPET$, Hain *et al* (2009) reported better estimates of $fAW$ when it is derived as a nonlinear function of $fPET$ and $B_c^*$. $B_c^*$ is the plant factor that captures the effects of stomatal control on the plant transpiration under well-watered conditions (see Equation 7 in Supporting Information Text S2). Instead of using a pre-defined function structure (for which the results are shown in Figure S2 in Supporting Information), here we fit a single statistical relation between observed $fAW$ and $fPET$ over all sites using a nonparametric kernel-based regression method (Nadaraya 1965, Nadaraya 1964, Watson 1964). The data used for developing the regression relation is restricted to a training period, and excludes the model evaluation period (more details are in section 2.4). The kernel regression method (see Equation 9 in Supporting Information Text S2) has been demonstrated to successfully capture nonlinear relations effectively in numerous studies (Kannan and Ghosh 2013, Raghav *et al* 2020, Rubin *et al* 2010, Salvi and Ghosh 2013). Here, the predictors of the kernel regression are $fPET$ and $B_c^*$, and the predictand is $fAW$. The kernel regression algorithm described by Hayfield and Racine (2008) and implemented in the R-Package by (R Core Team 2013) is used. Notably, such a relation may be derived using any other observed soil moisture data sets existent within the region, such as SCAN and USCRN (Diamond *et al* 2013, Schaefer *et al* 2007). In regions where soil moisture data does not exist, as noted earlier, $fAW$ may be obtained using the pre-defined nonlinear functions.

### *2.3 Retrieving volumetric soil moisture ($\theta$)*

$fAW$ (derived in section 2.2) is converted to actual volumetric soil moisture ($\theta$, $m^3\ m^{-3}$) using the field capacity ($\theta_{fc}$) and wilting point ($\theta_{wp}$) data using:

$$\theta = (\theta_{fc} - \theta_{wp}) fAW + \theta_{wp} \quad (5)$$

Here we assume the field capacity to be the volumetric water content at -33 kPa and wilting point to be the water content at -1500 kPa. In absence of site-specific root zone depth data, $\theta_{fc}$ and $\theta_{wp}$ used here are the average values within 0-100 cm from the soil surface.

### *2.4 Data used for model implementation and validation*

The model is implemented in the state of Oklahoma and the results validated at the Oklahoma Mesonet sites (Brock *et al* 1995). Data from North American Land Data Assimilation System-Phase 2 (NLDAS-2) (Xia *et al* 2012a) such as air temperature, specific humidity, wind speed, shortwave downward radiation, longwave downward radiation, and near-surface atmospheric pressure are used to obtain estimates of potential evapotranspiration, actual evapotranspiration, $fPET$, and $fAW$. The data has a temporal resolution of 1 hour and spatial resolution of 1/8°, and so are the corresponding resolutions of our evapotranspiration estimates. The model also uses MODIS Global 500 m Collection 5 land cover (MCD12C1) (Friedl and Sulla-Menashe 2015) to define land surface parameters (e.g. surface albedo, $h_s$, and $R_{gl}$ used in Equation 6 in Supporting Information S2). LAI is set to a 5 based on Koren *et al* (2010) to ensure data parsimony. Soil properties used for retrieving VSM are obtained from the MesoSoil database (Scott *et al* 2013), which includes physical properties of 13 soil types for 545 individual soil layers across 117 Oklahoma Mesonet sites. MesoSoil provides the data (e.g., sand-silt-clay fraction, volumetric water content at -33kPa, -1500kPa, residual and saturation water content, saturated hydraulic conductivity etc.) at depths of 5cm, 10cm, 25cm, 45cm, 60cm, and 75cm.

We validate the estimated $fAW$ against the fractional water index (FWI) measurements at the Oklahoma Mesonet sites (Brock *et al* 1995). The observation data network is located in the south-central region of the United States and spans entire Oklahoma with an area of ~181,196 km$^2$. Dominant land cover types include grassland (~58%), croplands (~15%), and forests with Savannas (~15%), Woody Savannas (~5%), and Deciduous Broadleaf Forests (~3%) (see Figure 2a). The network has at least one gauging station in each of Oklahoma's 77 counties. The Mesonet network has been extensively used for validation of soil moisture or proxy products in previous studies (Drusch 2007, Fang *et al* 2013, Gu *et al* 2008, Hain *et al* 2009, Swenson *et al* 2008, Xia *et al* 2014). Notably, the Mesonet sites provide the opportunity for intercomparison not only with *in situ* data but also with retrievals from Atmosphere-Land EXchange Inversion (ALEXI) surface energy balance model (Anderson *et al* 1997, Anderson *et al* 2011, Mecikalski *et al* 1999). ALEXI is a state-of-the-science tool that has been frequently used to track soil moisture stress in crops and forests (Anderson *et al* 2007a, Anderson *et al* 2007b, Anderson *et al* 2016b, Knipper *et al* 2019, Mishra *et al* 2013) and forms the basis for next generation of moisture stress measurements (Anderson *et al* 2016a, Cawse-Nicholson *et al* 2020, Fisher *et al* 2020, Guan *et al* 2017).

The validation is performed for the warm period (April-September) of 2002-2004. The period allows model evaluation against both *in situ* data and ALEXI model estimates. Figure S3 in Supporting Information shows the locations of stations at which we validated our model results. The selected sites cover a variety of land covers across Oklahoma (see Figure 2a), and were also used in (Hain *et al* 2009) for evaluation of estimated $fAW$. For validation, estimated $fAW$ is compared against observed Fractional Water Index (FWI). FWI has been demonstrated to be equivalent to observed $fAW$ at Mesonet sites with a correlation coefficient of 0.97 between computed $fAW$ and FWI (Hain *et al* 2009). Although FWI observations are available at fine temporal resolution (~5 minutes) and at depths of 5 cm, 25 cm, 60 cm, and 75 cm (Illston *et al* 2008), due to the absence of site-specific root zone distribution data, following Hain *et al* (2009), the observations at all the sensor depths are averaged to obtain an average FWI. Notably, the nonparametric kernel regression that establishes a relation between $fAW$ and $fPET$ (as outlined in Section 2.2) is obtained using two years (2000-2001) of $fAW$ observation data and $fPET$ estimates at all the Mesonet sites used in this study. This training period for regression is mutually exclusive to the model validation period which ranges from 2002 to 2004. Derivation of a single regression relation that is then used for $fAW$ estimation at all the sites ensures the generality of the approach, as such a relation may be developed using any alternative soil moisture observation data as well.

Model results are also compared against other widely available temporally continuous root-zone soil moisture products. This includes Variable Infiltration Capacity (VIC) model-based soil moisture product, that was generated in the North American Land Data Assimilation System (NLDAS, (Xia *et al* 2012b)). The temporal resolution of the VIC simulated soil moisture within NLDAS is 1 hour and its spatial resolution is 1/8°. We use VIC-simulated 0-100 cm soil moisture for the evaluation of our results. Comparison is also performed against the moisture proxy retrieval from the ALEXI model. As data of ALEXI moisture estimates are not available, the comparison is directly performed against the ALEXI model performance statistics reported in (Hain *et al* 2009). In Hain *et al* (2009), the ALEXI model was executed at daily temporal resolution (but only on cloud-free days) and a spatial resolution of ~10 km. For evaluation, moisture estimates were averaged over composite periods spanning 2-5 days each due to data coverage gaps caused by clouds. The composite periods included 15-19 Jun 2002, 11-12 May 2003, 28-29 May 2003, 4-5 Jul 2003, 27-31 Jul 2003, 6-7 May 2004, 1-2 Jun 2004, and 1-3 Aug 2004. Although we simulate gap-free daily $fAW$ and VSM estimates and they are duly used for evaluation against *in situ* observations, composite estimates are obtained for intercomparison with ALEXI-derived $fAW$ estimates over the aforementioned composite periods.

Bias Error (BE), Mean Absolute Error (MAE), Root Mean Square Difference (RMSD), unbiased Root Mean Square Difference (ubRMSD), and Pearson correlation coefficient ($R$) are used as performance metrics to assess the accuracy of $fAW$ estimates against *in situ* data. As the standard $RMSD$ is sensitive to biases in the mean or high extreme values (outliers), here we also use the $ubRMSD$, a metric used by SMAP to quantify the product accuracy (Entekhabi *et al* 2010).

## 3. Results and Explanations

### *3.1 Evaluation of temporal variations in simulated $fAW$*

The retrieved $fAW$, using the method outlined in Section 2, captures more than 48% (R = 0.69) of the variations in the observed FWI (see Figure 1a). For the same eight composite periods as used in Hain *et al* (2009), the model's performance is better during the composite periods (Figure 1b) i.e., when clear sky conditions prevail, w.r.t. all periods (Figure 1a). The possible reason for this is that the estimates of net radiation (and therefore latent heat flux) are generally more accurate during the largely clear sky, non-raining days. This is demonstrated at selected FluxNet sites where observed net short wave radiation data is available for evaluation (see Figures S4 and S5 in Supporting Information). Overall, with respect to the observed, the presented method overestimates drier conditions and underestimates wetter conditions. This is in line with the conclusions in other studies (Akuraju *et al* 2017, Allen *et al* 1998, Scott *et al* 2003), where also it was observed that the relation between $fAW$ and $fPET$ is ineffective when soil moisture conditions are above (below) $\theta_{fc}$ ($\theta_{wp}$).

Comparison of our $fAW$ estimates against those obtained from ALEXI for the composite periods shows that our method, overall, matches or outperforms the ALEXI results (as reported in Table 2 of Hain *et al* (2009)) in the study area. ALEXI $fAW$ has a larger scatter with $R^2$ of 0.48 as compared to a $R^2$ of 0.60 or R of 0.77 (see Figure 1b and Table 1) from our method. Our method shows a slightly larger positive bias, with $BE$ of 8.7% as compared to BE of -4.3% for ALEXI model. The RMSE and MAE are found to be 18.86 % (21.3%)

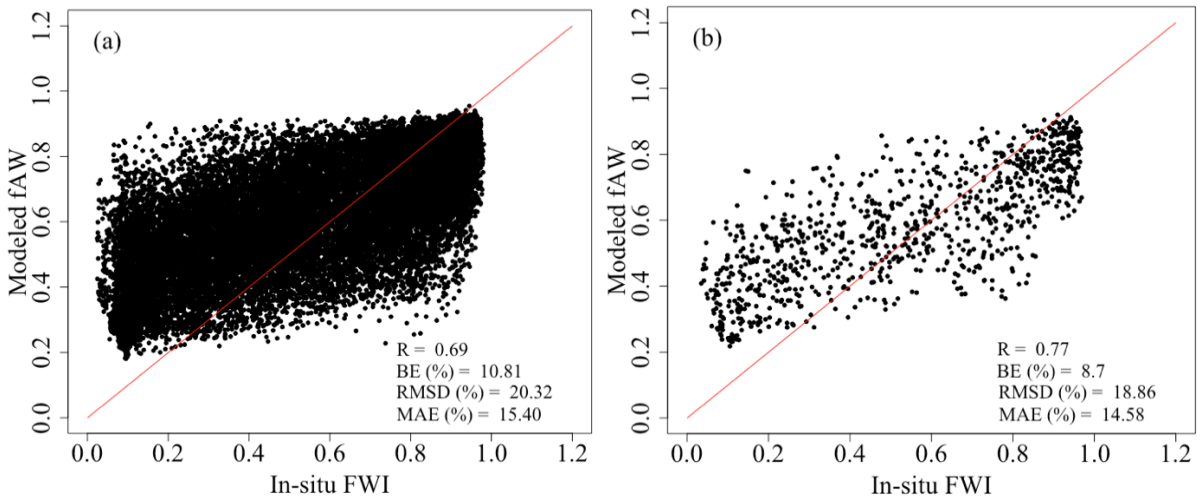


**Figure 1.** Comparison between daily observed and simulated fraction of available water ($fAW$) at selected Oklahoma Mesonet stations (identified in Figure S3 in Supplement Information) during April-September of 2002-2004. The red straight line is the 1:1 line. (a) shows the comparison for all days during April-September, 2002-2004, (b) shows comparisons for the eight composite periods (see section 2.4 for definition of composite periods).

**Table 1.** Error statistics for soil moisture proxy retrieved by our method and ALEXI during composite days at Mesonet stations. The error statistics for ALEXI shown here, have been obtained from Hain *et al* (2009).

| | **R** | **RMSD (% $fAW$)** | **BE (% $fAW$)** | **MAE (% $fAW$)** |
|---|---|---|---|---|
| **Our method** | 0.77 | 18.86 | 8.7 | 14.58 |
| **ALEXI** | 0.69 | 21.3 | -4.3 | 17.7 |

and 14.58% (17.7%) respectively for our method (ALEXI model). It is to be noted that Hain *et al* (2009) used a blended relation between $fAW$ and $fPET$ but we use a nonparametric kernel-based method for the same. To assess if better performance of $fAW$ estimates from our method w.r.t. the ALEXI estimates is due to the use of SFET or the kernel-based method, we regenerate the $fAW$ estimates using the blended relation used in Hain *et al* (2009). Results (Figure 1b and Figure S2b) show our model performance (R = 0.77) does not change for the composite periods depending on the use of the nonparametric kernel-based method and the blended relation, and the estimates from either are better than that from ALEXI (R = 0.69, see Table 1). Notably, the nonparametric kernel-based method yields better model performance when considering estimates from all the days (see Figure 1a vs. Figure S2a). These comparisons indicate better $fAW$ estimates from our method is a result of the use of both SFET to obtain evapotranspiration and the non-parametric kernel-based method used to obtain the relation between $fAW$ and $fPET$. Given that ALEXI derived $fAW$ estimates have been demonstrated to be more effective than Eta Data Assimilation System (EDAS) for accurate Numerical Weather Prediction (NWP) forecasts (Hain *et al* 2009), by extension, it may be claimed that estimates from our method will overperform the EDAS product as well. Notably, our method also provides temporally continuous daily estimates. In contrast, ALEXI estimates which uses TIR data suffer from large data gaps due to the presence of clouds and other satellite operational failures (Liou and Kar 2014).

Figure 2b (Figure 2c) shows the spatial variation of temporal correlations between daily $fAW$ estimates for April-September (composite periods) of 2002-2004 from our method and in-situ observations. The correlation is positive at all the sites with the highest correlation of 0.77 (0.99) and the lowest correlation of 0.18 (0.11) at 0.05 significant level (see Figures 2d and 2e). Between different land covers, the highest correlation is found in Woodland Savanna while the lowest correlation is observed in cropland areas (see Figure 2d). Relatively poor performance in croplands w.r.t. woodland is often attributed to the heterogeneity introduced by irrigation and is consistent with the conclusions of Naeimi *et al* (2009). Notably, the Bowen ratio obtained from SFET (see Equation 1) that is then used here for evaluation of $fAW$ (using

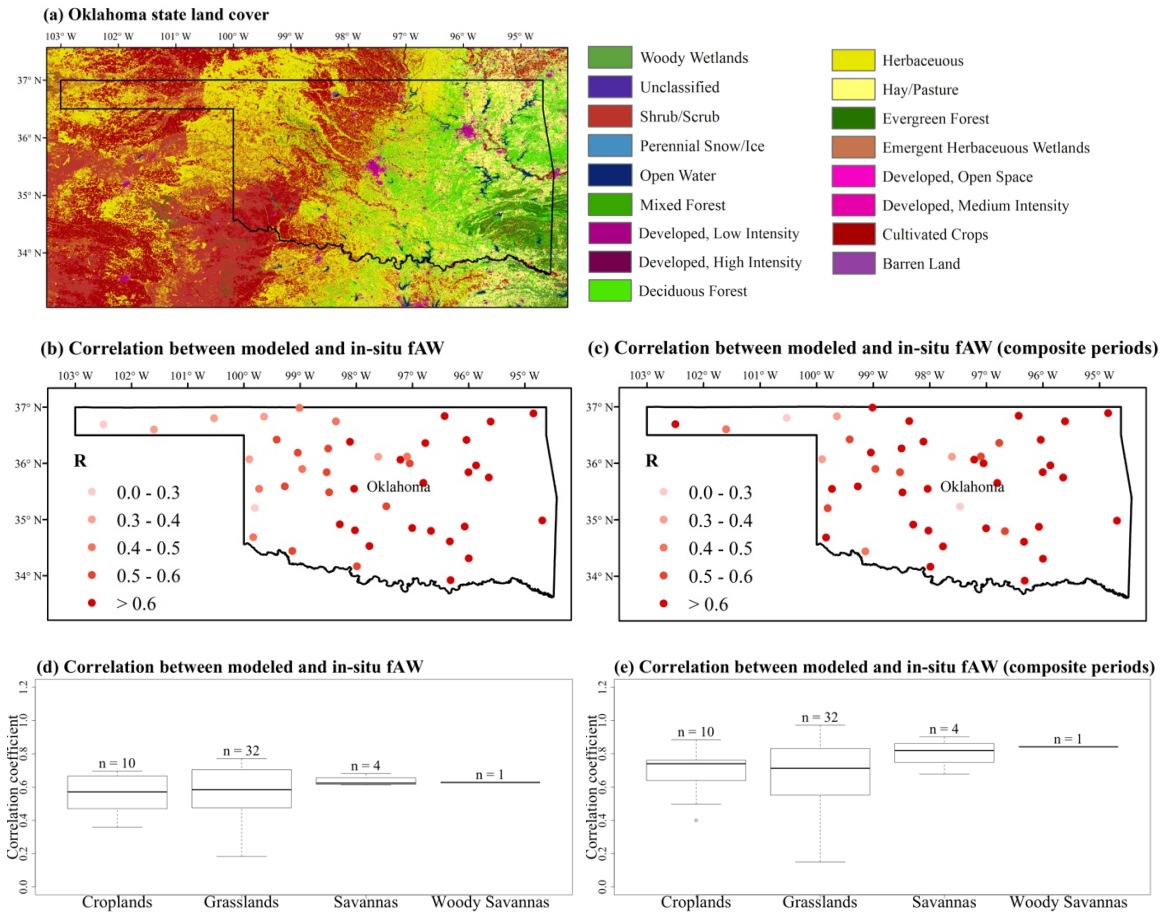


**Figure 2.** (a) Land cover map of Oklahoma, (b) correlation between daily modeled $fAW$ estimates and in-situ observations during warm periods of 2002-2004 at Oklahoma Mesonet sites, (c) correlation between modeled $fAW$ estimates and in-situ observations during the eight composite periods, (d) box plots of correlation coefficients shown in panel b for different land cover types, and (e) box plots of correlation coefficients shown in panel c for different land cover types. The dark black line in each boxplot shows the mean correlation for each land cover. The n-value at the top of each boxplot is the number of Mesonet sites within a particular land cover.

Equations 4 and S9), is an integrated land-atmosphere feedback response of areas surrounding the Mesonet station. Hence, $fAW$ derived from this method is likely to represent an effective soil moisture within the grid (of spatial resolution 1/8°), and may diverge from point estimates especially if the area experiences significant moisture heterogeneity.

### *3.2 Evaluation of spatial variations in simulated $fAW$*

The retrieved $fAW$, overall, captures the spatial gradient of root-zone soil moisture conditions within the study area (Figure 3). For example, for the composite period 15-19 June, 2002, observed dry soil moisture conditions in extreme western Oklahoma and wet soil moisture conditions in eastern Oklahoma are reflected in the model estimates as well (see Figure 3a). Similarly, our method also retrieves the dry soil moisture conditions all across Oklahoma during the composite period 27-31 July, 2003 (Figure 3e). Although, the overall spatial variability of soil moisture is captured by our model, there are disagreements. This could be due to a number of factors including scale mismatch between point measurements and our retrieval which is performed using meteorological data over a 0.125° x 0.125° grid. Additional sources of errors may be from varying $fAW$ vs. $fPET$ relations across different land covers, quality of input data, and errors in estimate of ET especially on cloudy days.

Next, the average of daily spatial correlations for all days during April-September of 2002-2004 is obtained. The average spatial correlation using data from all aforementioned days is equal to 0.62 (see Figure S6 in Supplement Information). The corresponding average correlation for the eight composite periods is 0.71. During the eight composite periods, the highest correlation is 0.84 for the composite period 15-19 June, 2002 when the soil moisture conditions are wettest (see Figure 3a) and the lowest correlation is 0.54

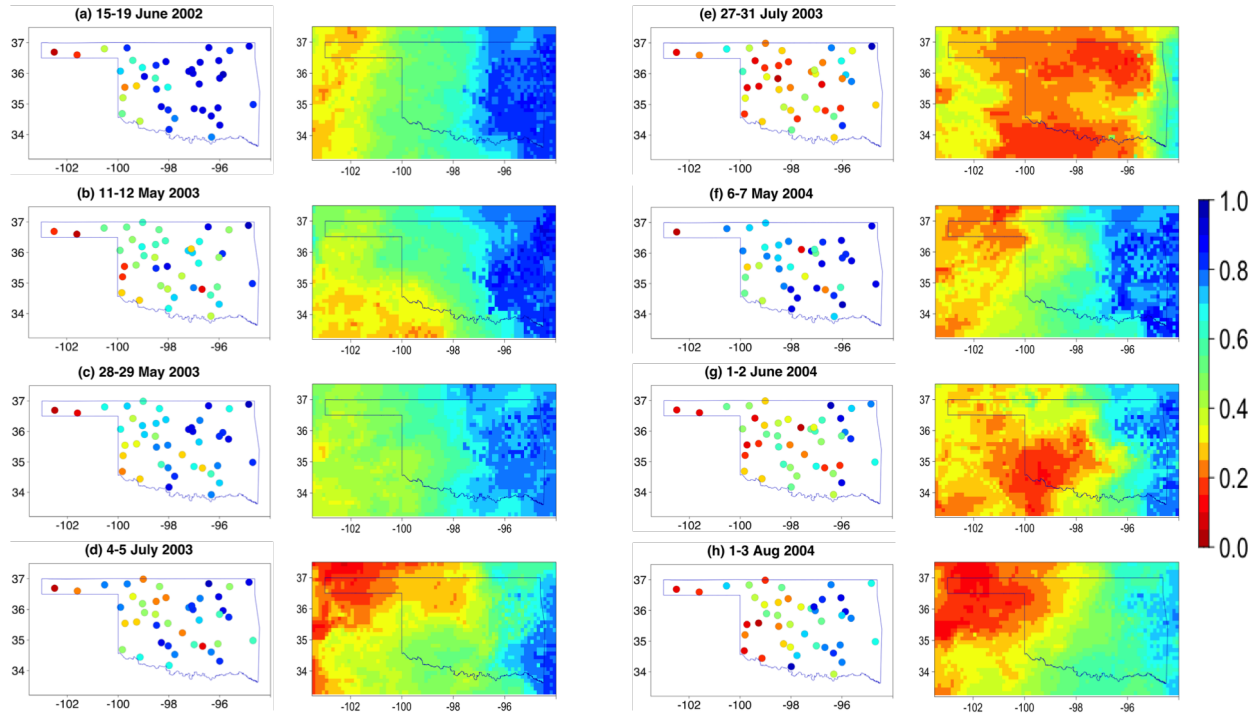


**Figure 3.** Spatial maps of $fAW$ over the state of Oklahoma for different composite periods. For each composite period, the left panel shows in-situ observations of $FWI$ at different mesonet sites and the right panel shows retrieved $fAW$ using our method.

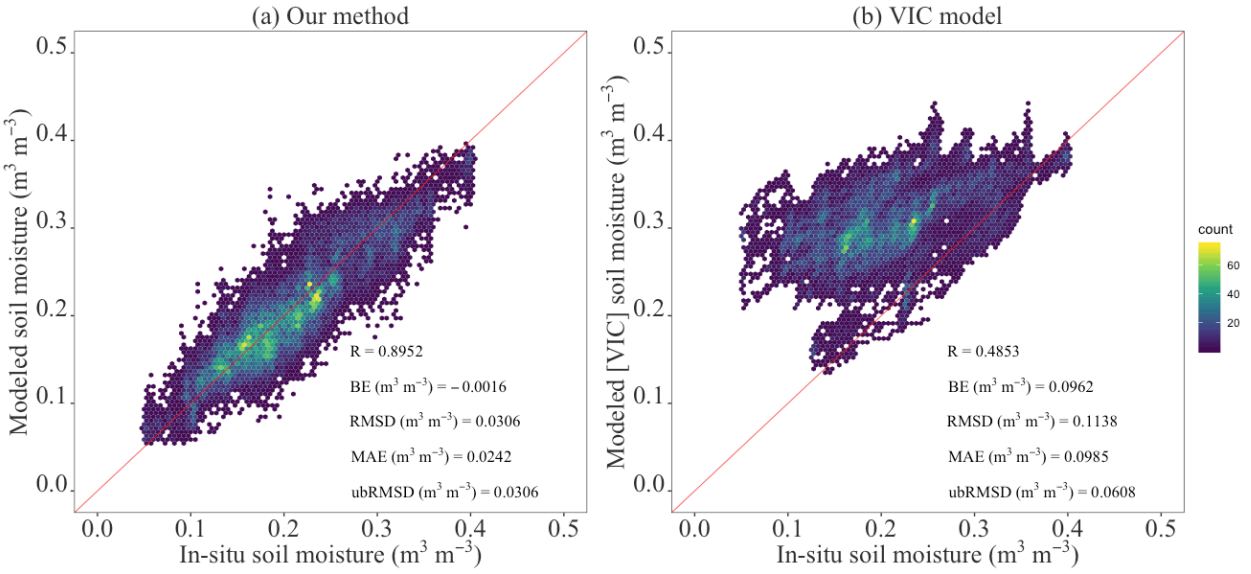


**Figure 4.** Comparison between daily observed and simulated volumetric soil moisture at the Oklahoma Mesonet sites (shown in Figure S3 of Supplementary Information). (a) SM estimates from our method, and (b) SM estimates from the VIC model, during April-September of 2002-2004. The red straight line is the 1:1 line.

during the composite period 27-31 July, 2003 when the moisture conditions are driest (see Figure 3e). The difference in model performance between wet and dry dates is likely due to multiple factors, including the existence of a larger spatial gradient in soil moisture during the wet composite period, and a smaller sensitivity of moisture dynamics on evapotranspiration when the ground is dry.

### *3.3 Retrieving actual soil moisture*

The retrieved $fAW$ estimates are converted to volumetric soil moisture (VSM) using the method outlined in Section 2.3.

**Table 2.** Error statistics for volumetric soil moisture ($m^3\ m^{-3}$) retrieved by our method, ALEXI, and VIC model during composite days at Mesonet stations. The error statistics for ALEXI shown here, have been taken from Hain *et al* (2009).

| | **RMSD** ($\boldsymbol{m^3\ m^{-3}}$) | **BE** ($\boldsymbol{m^3\ m^{-3}}$) | **MAE** ($\boldsymbol{m^3\ m^{-3}}$) | **ubRMSD** ($\boldsymbol{m^3\ m^{-3}}$) |
|---|---|---|---|---|
| **Our method** | 0.03 | -0.005 | 0.02 | 0.03 |
| **ALEXI** | 0.06 | -0.01 | 0.05 | 0.06 |
| **VIC** | 0.11 | 0.10 | 0.10 | 0.05 |

A comparison between VSM estimates from our method and observations is performed (Figure 4a). The correlation coefficient between the simulated daily soil moisture from our model and observed during the warm period of 2002-2004 is close to 0.90 when considering all the observation sites into the analysis. The significant increase in correlation for VSM w.r.t. that for $fAW$ (0.9 vs. 0.69 as seen in Figures 4a and 1a, respectively) is attributable to spatial heterogeneity in soil properties (see Supplement Figure S7). The model, overall, captures the temporal dynamics of VSM in diverse land cover types (Figure S8 in Supplementary information), although just like for $fAW$ (see Figure 1) lower (higher) moisture values are overpredicted (underpredicted). We also compare VSM estimates from Variable Infiltration Capacity model (VIC) (Figure 4b) that was run within the phase 2 of the North American Land Data Assimilation System (NLDAS-2) on a 1/8° grid over the continental U.S (Xia *et al* 2012a). It is to be noted that VIC derived soil moisture has been used in numerous studies for drought analysis (Mishra *et al* 2010, Sheffield *et al* 2004, Wang *et al* 2009), flood estimation (Lakshmi *et al* 2004), and climate change impact assessment (Mishra *et al* 2014). In contrast to VSM estimates from our model, the VIC model estimates show a larger scatter with a R of 0.49. Furthermore, the VIC model overestimates the VSM observations with a bias error of 46.46 % of the observations (see Figure 4b). These results are consistent with Xia *et al* (2015) where it was reported that VIC model overestimated soil moisture. Overall, VSM estimates from our model (VIC model) show correlation, bias, RMSD, MAE, and unbiased RMSD of 0.90 (0.49), -0.78% (46.46%), 14.79% (54.95%), 11.71% (47.59%), and 0.03 $m^3\ m^{-3}$ (0.06 $m^3\ m^{-3}$), respectively. We also evaluate the error statistics for VSM retrievals during the eight composite periods (Table 2). Results show our method outperformed both ALEXI and VIC model estimates. VIC model performance was worst among the three methods. Notably, our SM estimates adequately meet the SMAP mission requirement of $ubRMSD$ to be less than 0.04 $m^3 m^{-3}$ (Chan *et al* 2016).

## 4. Conclusions and synthesis

This study presented a new method to obtain daily estimates of root zone soil moisture proxy and volumetric root zone soil moisture. The method was based on surface flux equilibrium theory (SFET), and only needs readily available meteorological data and standard land surface parameterizations to obtain estimates of moisture proxy. Soil moisture proxy estimates from the method were in good agreement with the in-situ measurements at the Oklahoma Mesonet sites both temporally and spatially. The estimate of volumetric soil moisture in the root zone adequately met the SMAP soil moisture retrievals requirement of $ubRMSD < 0.04\ m^3\ m^{-3}$.

An intercomparison of our estimate with the ALEXI model, which forms the basis for the next generation of moisture stress measurements (Anderson *et al* 2016a, Cawse-Nicholson *et al* 2020, Fisher *et al* 2020, Guan *et al* 2017), showed our method matched or outperformed ALEXI derived estimates. Better performance was found to be due to two reasons, a better evapotranspiration estimate using SFET and use of the non-parametric kernel-based method to obtain the relation between $fAW$ and $fPET$. Another advantage of our method is its gap-free nature. In contrast, ALEXI or other thermal infrared (TIR) imagery-based retrievals of soil moisture provide estimates only during clear sky days as thermal imagery cannot be collected on cloudy days. Also, unlike our method, TIR-based methods for soil moisture proxy retrievals are dependent on a number of land surface parameters which are difficult to obtain in many cases. Given that ALEXI derived moisture proxy estimates have been demonstrated to be more effective than Eta Data Assimilation System (EDAS) for accurate Numerical Weather Prediction (NWP) forecasts (Hain *et al* 2009), by extension, it may be claimed that estimates from our method will overperform the EDAS product as well. Comparison against estimated moisture from VIC, a widely used land surface model, showed our results outperformed it as well. These results indicate the advantage of our method over several widely used land surface models.

As the presented method provides gap-free daily root-zone soil moisture estimates, it distinguishes itself from most remote-sensing based soil moisture retrievals methods that often suffer from data gaps due to being impacted by atmospheric conditions (such as the presence of clouds) and/or satellite revisit period. Furthermore, our method provides the root zone soil moisture estimate, while remote sensing-based soil moisture retrievals are often limited to moisture states in the surficial soil layer.

Although our validation results showed an overall satisfactory performance, it is to be noted that the performance is not competent across all soil moisture states and land covers. Our method especially fell short to capture extreme dry soil moisture conditions. Also, the performance was relatively poor in cropland settings. Subpar performance on occasions can be from multiple sources including (1) scale mismatch between the point measurements and model pixel, (2) quality and resolution of input meteorological data, (3) heterogeneity in soil properties, especially when converting moisture proxy to volumetric moisture, (4) absence of strong relation between fractional moisture content and the ratio of actual to potential evapotranspiration for extremely wet and dry moisture states, and (5) violation of assumptions that are inherent in SFET. Despite these limitations, this study highlights the advantages of our method over remote-sensing retrievals and land surface model predictions for root zone soil moisture retrievals. These advantages make the presented method apt for continuous assimilation of moisture in land surface and numerical weather prediction models. Gap-free moisture estimates from this method can be useful for many applications such as tracking crop stress, monitoring agriculture drought, irrigation management, estimation of groundwater recharge, etc. To further improve confidence in the applicability of the method for a wider range of settings, future work may focus on model evaluation in other settings. To assess the usefulness of the improved moisture estimates from this method, future studies may assimilate these estimates within the land surface and earth system models to evaluate the impacts on prediction accuracy of floods and droughts, and land-atmosphere feedbacks.

## Acknowledgments

This work is partially supported by NSF OIA-2019561, NSF EAR-1920425, and NSF EAR- 1856054. We thank Collin Mertz for providing the Oklahoma Mesonet soil moisture and soil properties data which was used for the validation of our results. The native NLDAS-2 and SMAP data used in this study are available from the respective data repositories at https://disc.gsfc.nasa.gov/ and https://www.wcc.nrcs.usda.gov/scan/. All the analysis has been carried out in R https://www.r-project.org/ and for this, we are very much thankful to R Core Team (2020).

## References

Akuraju, V. R., et al., 2017. Seasonal and inter-annual variability of soil moisture stress function in dryland wheat field, Australia. Agricultural and Forest Meteorology. 232**,** 489-499.

Allen, R. G., et al., 1998. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. Fao, Rome. 300**,** D05109.

Anderegg, W. R., et al., 2015. Tree mortality predicted from drought-induced vascular damage. Nature Geoscience. 8**,** 367-371.

Anderson, M., et al., Mapping evapotranspiration at multiple scales using multi-sensor data fusion. 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, 2016a, pp. 226-229.

Anderson, M., et al., 1997. A two-source time-integrated model for estimating surface fluxes using thermal infrared remote sensing.

Anderson, M. C., et al., 2011. Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery.

Anderson, M. C., et al., 2007a. A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 1. Model formulation. Journal of Geophysical Research: Atmospheres. 112.

Anderson, M. C., et al., 2007b. A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 2. Surface moisture climatology. Journal of Geophysical Research: Atmospheres. 112.

Anderson, M. C., et al., 2016b. The Evaporative Stress Index as an indicator of agricultural drought in Brazil: An assessment based on crop yield impacts. Remote Sensing of Environment. 174**,** 82-99.

Bolten, J. D., et al., 2009. Evaluating the utility of remotely sensed soil moisture retrievals for operational agricultural drought monitoring. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 3**,** 57-66.

Brock, F. V., et al., 1995. The Oklahoma Mesonet: a technical overview. Journal of Atmospheric and Oceanic Technology. 12**,** 5-19.

Camporese, M., et al., 2014. Simplified modeling of catchment-scale evapotranspiration via boundary condition switching. Advances in water resources. 69**,** 95-105.

Cawse-Nicholson, K., et al., 2020. Sensitivity and uncertainty quantification for the ECOSTRESS evapotranspiration algorithm–DisALEXI.

International Journal of Applied Earth Observation and Geoinformation. 89**,** 102088.

Chakrabarti, S., et al., 2014. Assimilation of SMOS soil moisture for quantifying drought impacts on crop yield in agricultural regions. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 7**,** 3867-3879.

Chan, S. K., et al., 2016. Assessment of the SMAP passive soil moisture product. IEEE Transactions on Geoscience and Remote Sensing. 54**,** 4994-5007.

Chauhan, N., et al., 2003. Spaceborne soil moisture estimation at high resolution: a microwave-optical/IR synergistic approach. International Journal of Remote Sensing. 24**,** 4599-4622.

Chen, X., et al., 2015. Variations in streamflow response to large hurricane-season storms in a southeastern US watershed. Journal of Hydrometeorology. 16**,** 55-69.

Das, N., et al., 2018. SMAP/sentinel-1 L2 radiometer/radar 30-second scene 3 km EASE-grid soil moisture version 2. NASA National Snow and Ice Data Center DAAC.

Das, N. N., et al., Combining SMAP and Sentinel data for high-resolution Soil Moisture product. 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, 2016, pp. 129-131.

De Michele, C., Salvadori, G., 2002. On the derived flood frequency distribution: analytical formulation and the influence of antecedent soil moisture condition. Journal of Hydrology. 262**,** 245-258.

Diamond, H. J., et al., 2013. US Climate Reference Network after one decade of operations: Status and assessment. Bulletin of the American Meteorological Society. 94**,** 485-498.

Drusch, M., 2007. Initializing numerical weather prediction models with satellite-derived surface soil moisture: Data assimilation experiments with ECMWF's Integrated Forecast System and the TMI soil moisture data set. Journal of Geophysical Research: Atmospheres. 112.

Entekhabi, D., et al., 2010. The soil moisture active passive (SMAP) mission. Proceedings of the IEEE. 98**,** 704-716.

Fang, B., et al., 2013. Passive microwave soil moisture downscaling using vegetation index and skin surface temperature. Vadose Zone Journal. 12**,** 1-19.

Ferguson, I. M., et al., 2016. Effects of root water uptake formulation on simulated water and energy budgets at local and basin scales. Environmental Earth Sciences. 75**,** 316.

Fisher, J. B., et al., 2020. ECOSTRESS: NASA's next generation mission to measure evapotranspiration from the International Space Station. Water Resources Research. 56**,** e2019WR026058.

Friedl, M., Sulla-Menashe, D., MCD12C1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 0.05Deg CMG V006 [Data set]. NASA EOSDIS Land Processes DAAC. Accessed 2020-04-04 from https://doi.org/10.5067/MODIS/MCD12C1.006. 2015.

Gu, Y., et al., 2008. Evaluation of MODIS NDVI and NDWI for vegetation drought monitoring using Oklahoma Mesonet soil moisture data. Geophysical Research Letters. 35.

Guan, K., et al., 2017. The shared and unique values of optical, fluorescence, thermal and microwave satellite data for estimating large-scale crop yields. Remote sensing of environment. 199**,** 333-349.

Guo, D., et al., 2019. Key factors affecting temporal variability in stream water quality. Water Resources Research. 55**,** 112-129.

Hain, C. R., et al., 2011. An intercomparison of available soil moisture estimates from thermal infrared and passive microwave remote sensing and land surface modeling. Journal of geophysical research: atmospheres. 116.

Hain, C. R., et al., 2009. Retrieval of an available water-based soil moisture proxy from thermal infrared remote sensing. Part I: Methodology and validation. Journal of Hydrometeorology. 10**,** 665-683.

Hanasaki, N., et al., 2013. A global water scarcity assessment under Shared Socio-economic Pathways–Part 1: Water use. Hydrology and Earth System Sciences. 17**,** 2375-2391.

Hayfield, T., Racine, J. S., 2008. Nonparametric econometrics: The np package. Journal of statistical software. 27**,** 1-32.

Illston, B. G., et al., 2008. Mesoscale monitoring of soil moisture across a statewide network. Journal of Atmospheric and Oceanic Technology. 25**,** 167-182.

Ines, A. V., et al., 2013. Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. Remote Sensing of Environment. 138**,** 149-164.

Kannan, S., Ghosh, S., 2013. A nonparametric kernel regression model for downscaling multisite daily precipitation in the Mahanadi basin. Water Resources Research. 49**,** 1360-1385.

Kerr, Y. H., et al., 2010. The SMOS mission: New tool for monitoring key elements ofthe global water cycle. Proceedings of the IEEE. 98**,** 666-687.

Knipper, K. R., et al., 2019. Evapotranspiration estimates derived using thermal-based satellite remote sensing and data fusion for irrigation management in California vineyards. Irrigation Science. 37**,** 431-449.

Koren, V., et al., 2010. Modification of Sacramento soil moisture accounting heat transfer component (SAC-HT) for enhanced evapotranspiration.

Lakshmi, V., et al., 2004. Soil moisture as an indicator of weather extremes. Geophysical research letters. 31.

Lettenmaier, D. P., Famiglietti, J. S., 2006. Water from on high. Nature. 444**,** 562-563.

Liou, Y.-A., Kar, S. K., 2014. Evapotranspiration estimation with remote sensing and various surface energy

balance algorithms—A review. Energies. 7**,** 2821-2849.

Liu, Y., et al., 2020. Plant hydraulics accentuates the effect of atmospheric moisture stress on transpiration. Nature Climate Change. 1-5.

Liu, Y., et al., 2017. Increasing atmospheric humidity and CO2 concentration alleviate forest mortality risk. Proceedings of the National Academy of Sciences. 114**,** 9918-9923.

Mahrt, L., Pan, H., 1984. A two-layer model of soil hydrology. Boundary-Layer Meteorology. 29**,** 1-20.

Mao, H., et al., 2019. Gap Filling of High-Resolution Soil Moisture for SMAP/Sentinel-1: A Two-Layer Machine Learning-Based Framework. Water Resources Research. 55**,** 6986-7009.

McColl, K. A., Rigden, A. J., 2020. Emergent simplicity of continental evapotranspiration. Geophysical Research Letters. 47**,** e2020GL087101.

McColl, K. A., et al., 2019. Surface flux equilibrium theory explains an empirical estimate of water-limited daily evapotranspiration. Journal of Advances in Modeling Earth Systems. 11**,** 2036-2049.

Mecikalski, J. R., et al., 1999. Estimating fluxes on continental scales using remotely sensed data in an atmospheric–land exchange model. Journal of Applied Meteorology. 38**,** 1352-1369.

Mintz, Y., Serafini, Y., 1992. A global monthly climatology of soil moisture and water balance. Climate Dynamics. 8**,** 13-27.

Mishra, V., et al., 2010. Assessment of drought due to historic climate variability and projected future climate change in the midwestern United States. Journal of Hydrometeorology. 11**,** 46-68.

Mishra, V., et al., 2013. A remote-sensing driven tool for estimating crop stress and yields. Remote Sensing. 5**,** 3331-3356.

Mishra, V., et al., 2014. Soil moisture droughts under the retrospective and projected climate in India. Journal of Hydrometeorology. 15**,** 2267-2292.

Monteith, J., The state and movement of water in living organisms. 19th Symposia of the Society for Experimental Biology. Cambridge University Press, London, 1965, 1965, pp. 205-234.

Nadaraya, E., 1965. On non-parametric estimates of density functions and regression curves. Theory of Probability & Its Applications. 10**,** 186-190.

Nadaraya, E. A., 1964. On estimating regression. Theory of Probability & Its Applications. 9**,** 141-142.

Naeimi, V., et al., 2009. An improved soil moisture retrieval algorithm for ERS and METOP scatterometer observations. IEEE Transactions on Geoscience and Remote Sensing. 47**,** 1999-2013.

Njoku, E. G., et al., 2003. Soil moisture retrieval from AMSR-E. IEEE transactions on Geoscience and remote sensing. 41**,** 215-229.

Norbiato, D., et al., 2008. Flash flood warning based on rainfall thresholds and soil moisture conditions: An assessment for gauged and ungauged basins. Journal of Hydrology. 362**,** 274-290.

Panday, S., Huyakorn, P. S., 2004. A fully coupled physically-based spatially-distributed model for evaluating surface/subsurface flow. Advances in water Resources. 27**,** 361-382.

Penman, H. L., 1948. Natural evaporation from open water, bare soil and grass. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences. 193**,** 120-145.

Porporato, A., et al., 2002. Ecohydrology of water-controlled ecosystems. Advances in Water Resources. 25**,** 1335-1348.

R Core Team, R., R: A language and environment for statistical computing. R foundation for statistical computing Vienna, Austria, 2013.

Raghav, P., et al., 2020. Revamping extended range forecast of Indian summer monsoon. Climate Dynamics. 1-15.

Rubin, Y., et al., 2010. A Bayesian approach for inverse modeling, data assimilation, and conditional simulation of spatial random fields. Water Resources Research. 46.

Sabaghy, S., et al., 2018. Spatially enhanced passive microwave derived soil moisture: Capabilities and opportunities. Remote sensing of environment. 209**,** 551-580.

Salvi, K., Ghosh, S., 2013. High-resolution multisite daily rainfall projections in India with statistical downscaling for climate change impacts assessment. Journal of Geophysical Research: Atmospheres. 118**,** 3557-3578.

Samaniego, L., et al., 2018. Anthropogenic warming exacerbates European soil moisture droughts. Nature Climate Change. 8**,** 421-426.

Schaefer, G. L., et al., 2007. The USDA natural resources conservation service soil climate analysis network (SCAN). Journal of Atmospheric and Oceanic Technology. 24**,** 2073-2077.

Scott, B. L., et al., 2013. New soil property database improves Oklahoma Mesonet soil moisture estimates. Journal of Atmospheric and Oceanic Technology. 30**,** 2585-2595.

Scott, C. A., et al., 2003. Mapping root zone soil moisture using remotely sensed optical imagery. Journal of Irrigation and Drainage Engineering. 129**,** 326-335.

Sheffield, J., et al., 2004. A simulated soil moisture based drought analysis for the United States. Journal of Geophysical Research: Atmospheres. 109.

Stewart, J., Verma, S., 1992. Comparison of surface fluxes and conductances at two contrasting sites within the FIFE area. Journal of Geophysical Research: Atmospheres. 97**,** 18623-18628.

Swenson, S., et al., 2008. Estimating profile soil moisture and groundwater variations using GRACE and Oklahoma Mesonet soil moisture data. Water Resources Research. 44.

Torres, R., et al., 2012. GMES Sentinel-1 mission. Remote Sensing of Environment. 120**,** 9-24.

Trenberth, K. E., et al., 2007. Estimates of the global water budget and its annual cycle using observational and model data. Journal of Hydrometeorology. 8**,** 758-769.

Wagner, W., et al., 2013. The ASCAT soil moisture product: A review of its specifications, validation results, and emerging applications. Meteorologische Zeitschrift. 22**,** 5-33.

Walker, J. P., Houser, P. R., 2004. Requirements of a global near-surface soil moisture satellite mission: accuracy, repeat time, and spatial resolution. Advances in water resources. 27**,** 785-801.

Wang, A., et al., 2009. Multimodel ensemble reconstruction of drought over the continental United States. Journal of Climate. 22**,** 2694-2712.

Wang, A., et al., 2011. Soil moisture drought in China, 1950–2006. Journal of Climate. 24**,** 3257-3271.

Wang, R., et al., 2013. Anomalous trend in soil evaporation in a semi-arid, snow-dominated watershed. Advances in water resources. 57**,** 32-40.

Watson, G. S., 1964. Smooth regression analysis. Sankhyā: The Indian Journal of Statistics, Series A. 359-372.

Wetzel, P. J., Chang, J.-T., 1987. Concerning the relationship between evapotranspiration and soil moisture. Journal of climate and applied meteorology. 26**,** 18-27.

Wigmosta, M. S., et al., 1994. A distributed hydrology-vegetation model for complex terrain. Water resources research. 30**,** 1665-1679.

Wood, E., Lettenmaier, D., 1996. Surface soil moisture parameterization of the VIC-2L model: Evaluation and modifications. Global Planet. Change. 13**,** 195-206.

Xia, Y., et al., 2015. Comparison of NLDAS-2 simulated and NASMD observed daily soil moisture. Part I: Comparison and analysis. Journal of Hydrometeorology. 16**,** 1962-1980.

Xia, Y., et al., 2012a. Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. Journal of Geophysical Research: Atmospheres. 117.

Xia, Y., et al., NCEP/EMC (2014), NLDAS VIC Land Surface Model L4 Hourly 0.125 x 0.125 degree V002, Edited by David Mocko, NASA/GSFC/HSL, Greenbelt, Maryland, USA. Goddard Earth Sciences Data and Information Services Center (GES DISC), 2012b.

Xia, Y., et al., 2014. Evaluation of multi-model simulated soil moisture in NLDAS-2. Journal of Hydrology. 512**,** 107-125.

Zi, T., et al., 2016. Simulating the spatio-temporal dynamics of soil erosion, deposition, and yield using a coupled sediment dynamics and 3D distributed hydrologic model. Environmental Modelling & Software. 83**,** 310-325.