

Downscaling satellite soil moisture for landscape applications: A case study in Delaware, USA

Daniel L. Warner^{a,*}, Mario Guevara^{b,2,3}, John Callahan^{a,4}, Rodrigo Vargas^{b,5}

^a Delaware Geological Survey, University of Delaware, Newark, DE, USA

^b Department of Plant and Soil Sciences, University of Delaware, Newark, DE, USA

ARTICLE INFO

Keywords:

Soil moisture
Remote sensing
Machine learning
Spatiotemporal
Soil moisture networks

ABSTRACT

Study region: Delaware, USA and its surrounding watersheds.

Study focus: An ensemble using multiple Kernel K-nearest neighbors (KNN) models was trained to predict daily grids of SSM at 100-meter resolution based on SSM estimates from the European Space Agency's Climate Change Initiative Soil Moisture Product, terrain data, soil maps, and local meteorological network data. Estimated SSM was evaluated against independent in situ SSM observations and were investigated for relationships with land cover class and vegetation phenology (i.e., NDVI).

New hydrological insights for the region: Downscaled daily mean SSM estimates had lower error in space (27%) and greater predictive performance over time compared to the raw, coarse resolution remotely sensed SSM dataset when calibrated to field observed values. Downscaled SSM identified stronger and more widespread temporal relationships with NDVI than other estimation methods. However, both coarse and fine resolution datasets greatly underestimated SSM in wetland areas. The findings highlight the need for enhanced in situ SSM monitoring across diverse settings to improve landscape-level downscaled SSM. The downscaling methodology developed in this study was able to produce daily SSM estimates, providing a framework that can support future SSM modeling efforts, hydroecological investigations, and agricultural studies in this and other regions around the world when used in conjunction with ground-based monitoring networks.

1. Introduction

Soil moisture content is a critical driver of many ecosystems due to its influence on climate, plant productivity, soil biogeochemistry, and surface water hydrology (Seneviratne et al. 2010, McColl et al. 2017). Climate induced changes in the water cycle are expected to alter soil moisture regimes and resulting feedback mechanisms across local to global scales (Bates et al. 2008, Green et al. 2019, Huntington, 2006). Soil moisture variability may limit carbon uptake by terrestrial ecosystems (Green et al. 2019), and

* Correspondence to: University of Delaware, Delaware Geological Survey, Rm 225, Newark, DE 19716, USA.

E-mail address: warnerdl@udel.edu (D.L. Warner).

¹ 0000-0002-6602-1416

² 0000-0002-9788-9947

³ Present address: Centro de Geociencias, Universidad Nacional Autónoma de México, Campus Juriquilla, Qro. MX

⁴ 0000-0002-6299-9388

⁵ 0000-0001-6829-5333

<https://doi.org/10.1016/j.ejrh.2021.100946>

Received 8 December 2020; Received in revised form 4 October 2021; Accepted 10 October 2021

Available online 15 October 2021

2214-5818/© 2021 The Authors.

Published by Elsevier B.V. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

reductions in crop productivity due to deficient or excessive soil moisture may have major societal and economic consequences (Bates et al. 2008). Thus, understanding and mapping the spatial and temporal variability of soil moisture in our environment is of great interest to a broad audience in academic, agricultural, and economic communities.

Soil moisture in the top few centimeters of the soil profile, known as surface soil moisture (SSM), is of particular interest to researchers as it is relatively easier to measure than, and often well correlated to, moisture conditions deeper in the soil profile (McColl et al. 2017). Measuring in situ SSM at single points is made possible by using time-domain reflectometry probes, which respond to changes in the dielectric constant of soil (Topp et al., 1980). Though very useful, these probes may require calibration to soil type and present additional challenges when used to assess SSM dynamics with multiple probes across large spatial scales in heterogeneous environments. Consequently, assessing the distribution of saturated areas across a landscape or agricultural fields is financially and logistically challenging with soil moisture probes alone.

Advances in airborne and satellite remote sensing have provided an alternative approach for assessing spatial patterns of SSM. Active and passive microwave imaging is also affected by the dielectric constant of the soil surface and can be used to estimate SSM from a remote sensing perspective (Dubois et al. 1995, Wagner et al. 1999). Microwave imagers operating in the decimeter wavelength range are powerful tools for remote sensing of SSM measurement as they are relatively unaffected by atmospheric conditions, moderately resistant to vegetation interference and day and night cycles, and can determine SSM in the top 0–5 cm of the soil surface (Schmugge et al. 2002). Remote sensing is advantageous for assessing large-scale spatial and temporal SSM patterns as measurements are an average of a target land surface (which is dependent on the specific instrument used (Mohanty et al. 2017)), while in situ measurements can provide more accurate local SSM values (with a sensor footprint of a few cm³) but lack broad spatial coverage (Ford and Quiring, 2019). Global SSM products generated from satellite microwave information are now widely available and have proved valuable for assessing large regional-to-global scale soil moisture trends with high temporal frequency (Dong et al. 2020). These global products have been used to examine the impacts of large-scale weather and climate patterns, biogeochemical cycles, and terrestrial hydrology in new and insightful ways (Dorigo and de Jeu, 2016) but have limited application for addressing the spatiotemporal variability of soil moisture at more localized landscape scales (e.g., ≤ 100 -m grids) due to their coarse spatial resolution. Higher spatial resolution SSM estimates may be achieved using remote sensors such as infrared imagers and synthetic aperture radars, but the resulting SSM products from these remote sensors still have reduced spatial coverage and limited temporal frequency (Dorigo et al. 2017). Thus, there is a need for SSM products with both high temporal and spatial resolution, but current satellite remote sensing technology cannot meet both needs alone.

Recent research has aimed to address these spatial resolution limitations by attempting to “downscale” coarse resolution global soil moisture datasets using a wide variety of algorithms and statistical approaches. Downscaling approaches (e.g., methods and algorithms) are diverse, and developing and testing various downscaling approaches is a pressing need for improving the quality and spatial detail of current and future satellite soil moisture missions (Liu et al. 2017, Liu et al. 2020, Peng et al. 2017). Some process-based approaches are based on the relationship between soil moisture and other environmental data such as vegetation indexes and land surface temperature (Fang et al. 2018, Gillies et al. 1997). These relationships can be represented by fitting nonlinear models using a triangle or trapezoid feature space to derive soil moisture estimates with increased granularity with the spatial support of land surface or vegetation index datasets (Carlson, 2007). This is because localized land surface characteristics and vegetation indexes can be more easily represented with higher spatial resolution (fine grained) compared to global satellite soil moisture datasets. Beyond process-based models, empirical downscaling approaches treat the data mechanism regulating the SSM variability as unknown (Liu et al. 2017). These empirical approaches can be based in probability theory or may employ machine learning approaches including regression trees, deep learning methods, or combined approaches such as ensemble or multiple regressors or classifiers. Across all downscaling approaches, many studies leverage linear and nonlinear models for predicting or downscaling satellite soil moisture estimates using a variety of ancillary information available with increased granularity (Guevara and Vargas, 2019, Merlin et al. 2013, Montzka et al. 2018, Piles et al. 2011, Ranney et al. 2015, Sharma et al. 2021, Yan and Bai, 2020, Zappa et al. 2019). Arguably, most common predictors for soil moisture are vegetation indexes and climate datasets, although there are a large number of environmental layers potentially useful for predicting satellite soil moisture values with higher granularity and accuracy than satellite estimates. While using ancillary vegetation data may improve downscaling performance, it also increases the risk of confounding problems (e.g., spurious correlations) in downstream investigations into vegetation-SSM relationships. Spatiotemporal relationships between SSM and vegetation phenology are of great importance to ecological and agricultural sciences, as SSM is an important environmental variable for developing agricultural drought forecasting systems (CEC, 2021, Ford and Quiring, 2019, Han et al. 2014). As an example, a researcher investigating landscape scale drought patterns may want to examine correlations between a spectral vegetation index and a downscaled SSM grid, but if the downscaled SSM grid used this vegetation index as an ancillary predictor dataset, the resulting investigation would be inherently biased towards the downscaling model relationship and results could be potentially influenced by spurious correlations. Recent research has demonstrated that a purely geomorphometric (i.e., terrain is the only ancillary dataset employed) downscaling approach is viable for continental scale application at a moderate resolution of one kilometer, though the usefulness of such an approach at landscape scale resolutions (≤ 100 m) has not been explored (Guevara and Vargas, 2019). While a few downscaling approaches have targeted resolutions in the 30–100 m range capable of assessing landscape scale SSM variations (Abowarda et al. 2021, Merlin et al. 2013, Ranney et al. 2015), all have employed vegetation datasets as ancillary predictors.

This study presents a landscape-scale SSM downscaling methodology employing a kernel-based approach that relies on spatial information (e.g., spatial structure of satellite soil moisture values), topographic land surface characteristics, and meteorological data rather than vegetation or land cover data. The overarching goal of this study was to develop a methodology for downscaling a coarse spatial resolution global daily SSM dataset to a spatial resolution suitable for evaluating landscape scale SSM variability (i.e., 100 m

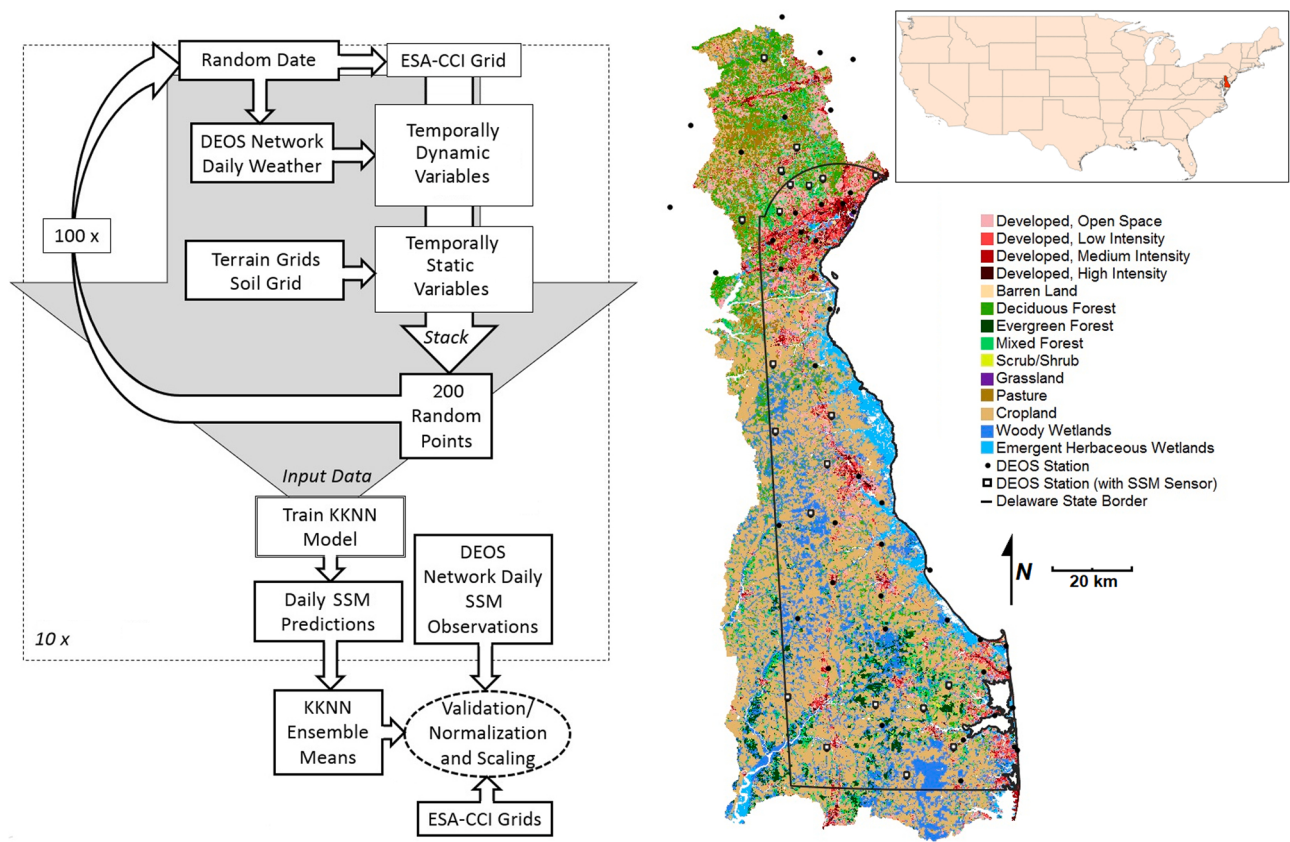


Fig. 1. General workflow of the downscaling method (left) and overview of study area with land cover classifications and locations of DEOS network meteorological stations noted (right). The location of the study area within the United States is noted in the inset.

resolution) that minimizes potential confounding bias in subsequent investigations of downscaled SSM with relation to local vegetation indices (e.g., NDVI) and land cover classes. The specific goals of this study were to evaluate the performance of this downscaling approach compared to a coarse resolution remotely sensed product against a local SSM monitoring network, explore potential relationships between SSM estimates and vegetation and land cover data, and identify limitations and areas of poor downscaling performance to target in future downscaling efforts. This study focuses on the state of Delaware, USA and its surrounding watersheds. This study area is anticipated to experience a complex shift of hydrologic regimes in the coming decades, with potentially higher mean annual precipitation but also higher occurrence of droughts and intense storm events that may impact agricultural activities (Najjar et al. 2000, Huntington et al., 2009). The study area contains a relatively dense meteorological monitoring network that includes 21 SSM monitoring stations, allowing for testing and evaluation of downscaling methodologies.

2. Methods

2.1. Study area

The area considered in this study included the state of Delaware, USA and portions of major watersheds (USGS HUC12 delineations) flowing into or out of the state, constituting a total land area of roughly 8130 km². The study area was elected because it has a dense meteorological network with in situ soil moisture information (see Section 2.2) with about 1 station per 390 km². Contained within this study area are distinct subregions of the rugged Piedmont, the heavily agricultural mid-Atlantic coastal plain, and large networks of coastal bays and wetlands (Fig. 1). The study area lies in a transition zone between continental and subtropical climate zones, experiencing a four-season climate mostly based on temperature with warm, humid summers and cold, relatively dry winters and nearly constant precipitation throughout the year. The 40-year mean average annual temperature is roughly 13 °C, reaching a minimum monthly mean of 1.8 °C in January and maximum monthly mean of 25 °C in July. Average annual precipitation is roughly 1143 mm (Office of the Delaware State Climatologist, <http://www.climate.udel.edu/>).

2.2. Data sources: ESA CCI and DEOS network

The satellite-based estimates of daily soil moisture were taken from the soil moisture program of the European Space Agency's Climate Change Initiative (ESA-CCI SM version 4.5; Dorigo et al. 2017, Gruber et al. (2017,2019)). This dataset consists of daily global soil moisture estimates at 0.25 degree spatial resolution (approximately 27 km vertical and 21 km horizontal in our study area), which are generated by combining multiple satellite soil moisture records available since the late 1970 s (Dorigo et al. 2015, Dorigo et al. 2017). This dataset is constantly being updated and improved (Chung et al. 2018). Soil moisture records contained in the ESA-CCI are estimated using passive and active microwave remote sensors (e.g., spaceborne radiometers and radar sensors) able to measure the dielectric constant in the top layer of soil, from 0 to roughly 5 cm depth (Llamas et al. 2020). Specifically, ESA-CCI SM was developed by combining data from C-band scatterometers from European remote sensing satellite missions (e.g., ERS-1/2, METOP) and data from multi-frequency radiometers such as the Scanning Multichannel Microwave Radiometer (SMMR), Special Sensor Microwave Imager (SSM/I), Microwave Imager (TMI), Advanced Microwave Scanning Radiometer (AMSR-E), and Windsat (Llamas et al. 2020).

Ground-based in situ soil moisture measurements were taken from the Delaware Environmental Observing System (DEOS) network. The DEOS maintains a dense network of meteorological monitoring stations throughout the State of Delaware, extending northward into southeastern Pennsylvania to cover the upper basins of the Brandywine River, White Clay Creek, and Red Clay Creek. The network has been operating for over 15 years and now consists of over 60 monitoring stations. Recently, DEOS has begun to add soil moisture and soil temperature monitoring capabilities across its many stations using probes capturing the upper 0–5 cm of the soil. In total, 21 of the DEOS stations had consistent SSM data for 2018, the study year of interest, and these station observations were used as an independent benchmark for evaluating and calibrating both the ESA datasets and downscaled SSM estimates (Fig. 1). Weather

Table 1

Descriptions of terrain attributes considered as potential predictors in the downscaling model.

Terrain Attribute	Description
Aspect	The direction that a slope faces relative to due south
Catchment Slope	The mean slope of all grid cells within a given grid cell's upslope accumulation area
Channel Network Base Level	The interpolated elevation of a channel network with a sufficiently large upslope accumulation area (Conrad et al. 2015)
Closed Depressions	The depth of a grid cell identified as a topographic depression relative to its lowest surrounding grid cell (Conrad et al. 2015)
Cross-Sectional Curvature	The surface curvature perpendicular to aspect and slope normal line
Flow Accumulation	The number of grid cells that drain through a given grid cell
Flow Line Curvature	The curvature of the nearest grid cells draining through a given grid cell (Conrad et al. 2015)
Slope	The maximum slope angle at a given grid cell
Topographic Position Index	The difference of grid cell elevation and local mean elevation, this index is indicative of the local high and low points (Gallant and Wilson, 2000)
Topographic Wetness Index	The upslope accumulation area of a grid cell divided by its slope, this index is indicative of where water tends to accumulate on a landscape (Beven and Kirkby, 1979)
Valley Depth	The difference of grid cell elevation beneath an interpolated local ridge line (Conrad et al. 2015)

data obtained from DEOS included ambient air temperature, soil temperature, and precipitation. For all variables, 5-min observations were averaged to produce daily mean values, except precipitation, which was considered as the sum of the 5-min intervals. Because of gaps in the data, at least 12 h of soil moisture data, or 18 h for weather data, must have been present for daily values to be valid. One of the 21 stations, located near Dover, Delaware towards the center of the study area, had a large gap in its SSM record during spring and early summer 2018 (200 days), but was still used for evaluation purposes in other seasons.

2.3. Ancillary data processing

Terrain attributes were extracted from a 100-m resolution DEM, which was generated by resampling a LIDAR-derived 3-m resolution DEM of the study area. A resolution of 100-m was selected as it captured general landscape features (i.e., hills and valleys) while still allowing for relatively rapid computation. LIDAR for the State of Delaware was collected during the leaf off period of winter 2014 with a minimum return density of 2 points m^{-2} and hydroflattened water bodies, while small portions of the northern section of the study area were taken from 1-m resolution county level elevation datasets in the states of Pennsylvania and Maryland. A set of terrain attributes were derived from the DEM using several SAGA GIS terrain analysis modules (SAGA GIS 6.4) and are described in Table 1. Surface layer saturated conductivity data was sourced from the National Resources Conservation Service's Soil Survey Geographic Database (Soil Survey Staff, National Resources Conservation Service, US Department of Agriculture, 2015), which was converted to a gridded format at 100-m resolution. An additional set of six oblique, or rotated, geographic coordinate axes were generated and converted to gridded values to be considered as potential predictors. Oblique coordinates are used as spatially explicit predictors in the downscaling model as one may use latitude and longitude, but unlike 2-dimensional latitude and longitude, oblique coordinates are less likely to introduce orthogonal artifacts in resulting prediction grids made by nonlinear models (Møller et al. 2020). The elevation attributes, oblique coordinates, and soil data were referred to as temporally static predictors.

Because SSM is a temporally dynamic variable, our downscaling methodology also incorporated daily grids of antecedent mean temperature, antecedent cumulative precipitation, and day length grids as predictors. Daily temperature and precipitation at the DEOS network stations were interpolated to a 100 m grid spatially aligned with the elevation grid using a regularized spline with an n value (i.e., the number of points considered for local approximation) of 10 and smoothing weight of 0.2. The resulting daily grids were stacked and processed to generate precipitation and temperature variables to establish a meteorological context of current and antecedent conditions across the study area. Antecedent precipitation and temperature conditions were considered as potential predictors as they have shown to be related to SSM dynamics in previous research (Buczko et al., 2007, Han et al. 2014). Day length grids were based on a single value estimated from the day of year and the latitude of the centroid of our study area. This set of predictors were referred to as temporally dynamic predictors.

2.4. Model training and variable selection

The downscaling method used in this study employed a Kernel K-Nearest Neighbors (KNN) regression algorithm for fitting ESA-CCI SSM values to terrain attributes and temporal predictors. This algorithm was chosen based on its relative speed, simplicity, and demonstrated performance in previous SSM downscaling efforts at larger scales (Guevara and Vargas, 2019; Guevara et al. 2021). The methodology of training KNN models in the variable selection process and in the final model training followed the same general approach shown in Fig. 1. A set of 10 KNN models were trained based on a set of 100 randomly selected dates in 2018. On each date, a stack of predictor grids was built consisting of ESA-CCI SSM estimates and grids of predictor variables. Temporally static predictor variables were held constant, while temporally dynamic variables were specific for each date. Spatial points ($n = 200$) were then randomly sampled from each date's predictor stack, yielding a dataset of 20,000 input points. A KNN model was then fit to these points using 10 repetitions of 5-fold cross validation to select the optimal kernel function and k (number of surrounding points to consider) that minimized cross validated mean absolute error (MAE; Fig. 1). We selected MAE as the reference criteria over root mean square error as it provides a general measure of model error but is less sensitive to infrequent, large magnitude errors.

To select which of the predictor variables will be included in the model, each predictor variable was sequentially excluded from repetitions of the described KNN model training process and evaluated for its importance based on the change in model MAE before and after its exclusion. Variables that had consistently low-importance across this process were then eliminated from the predictor set, yielding a set of influential predictor variables that also reduced the dimensionality of the predictor variable space.

A final ensemble of 10 KNN models were then trained using the final set of predictor variables using the workflow illustrated in Fig. 1. Soil moisture predictions were then extrapolated from each model across the study area on each date using gridded data of each predictor variable. Final downscaled estimated daily SSM values represent the mean prediction of all ten model outputs at each 100 m grid cell.

2.5. Evaluation against ground network observations

We evaluated SSM predictions using methods similar to those used in global scale SSM studies (Gruber et al. 2020). Daily estimated SSM was compared to daily DEOS network SSM observations in grid cells corresponding to station locations. Given the narrow and conservative range of values modeled from the ESA-CCI soil moisture product, daily predicted grids (both coarse resolution and downscaled) were then normalized from zero to one over the entire prediction domain. These grids were then linearly rescaled based on the highest and lowest SSM values observed by the DEOS network on each day (i.e., values taken from two of the 21 stations). The purpose of this rescaling, or calibration, was to better account for spatiotemporal variability in SSM values that is lost due to the

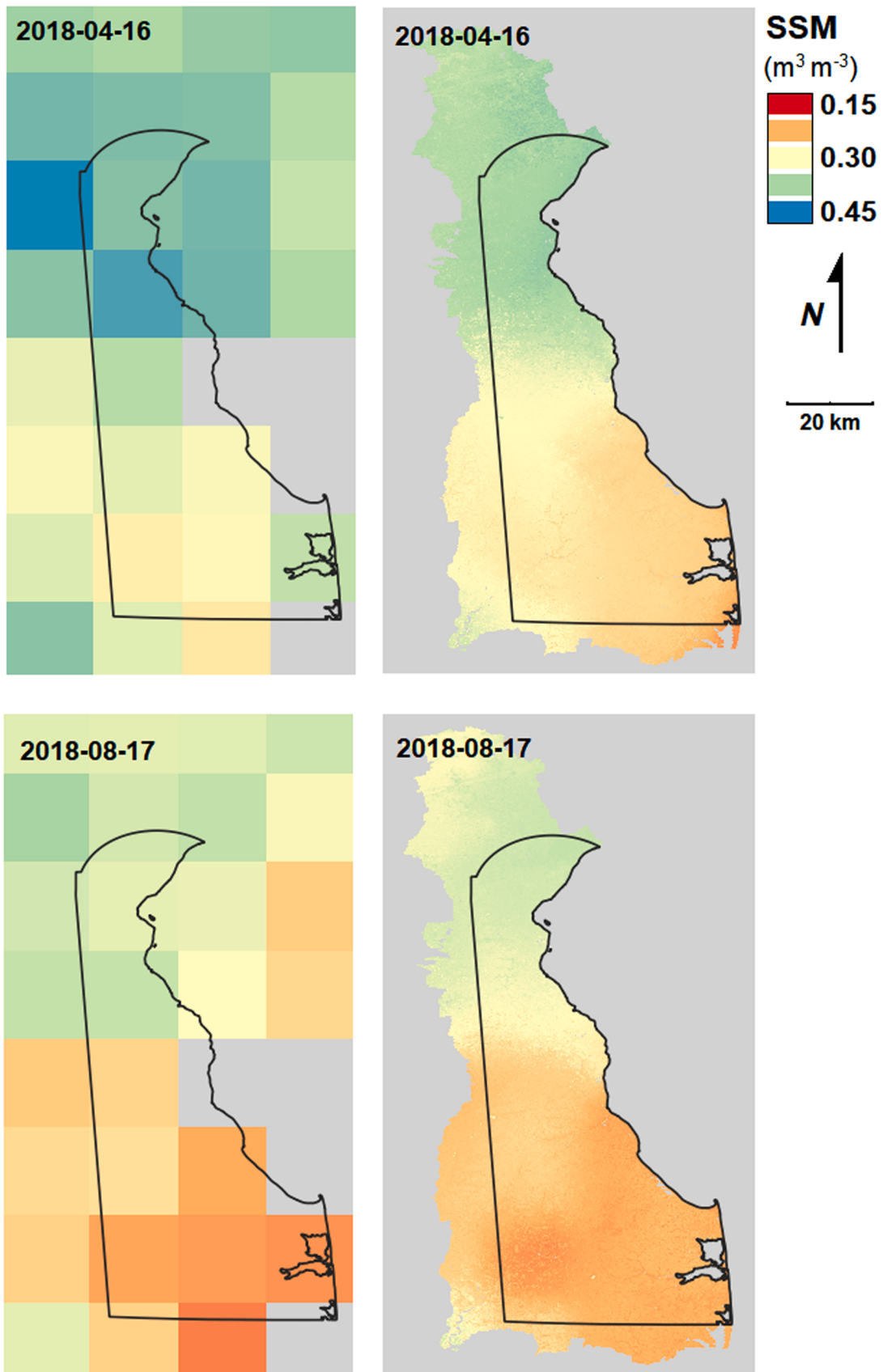


Fig. 2. Illustrated example of two raw ESA-CCI SSM datasets at 0.25 degree resolution (left) and datasets produced by our KKNn downscaling method at 100-m resolution (right), which reflects the large scale variations of the coarse resolution data as well as small scale variations in underlying ancillary datasets. The dates pictured are 2018-04-16 and 2018-08-17.

averaging effects of the large footprint ESA-CCI product grid cells. While it is possible that this calibration may introduce a small degree of confounding bias into later evaluations of downscaling model performance, beyond the two stations used for daily calibration, the remaining DEOS network stations were not used to calibrate SSM values in any way. For the remainder of this manuscript, the following terms will be used:

- $DEOS_{OBS}$: SSM observed by the DEOS network
- ESA_{RAW} : SSM values from original ESA-CCI soil moisture grid cells
- ESA_{SC} : ESA_{RAW} values that are calibrated to daily minimum and maximum SSM observed by the DEOS network
- $KKNn_{RAW}$: modeled SSM from our downscaling method
- $KKNn_{SC}$: $KKNn_{RAW}$ values that are calibrated to daily minimum and maximum SSM observed by the DEOS network.

Raw and calibrated values of both the ESA and KKNn products were evaluated spatially based on the mean absolute error (MAE), and mean bias error (MBE) between predicted values and $DEOS_{OBS}$. These performance metrics were evaluated across all network stations on a daily basis between predicted SSM at grid cells corresponding to DEOS station observations. An overall coefficient of determination (R^2) was calculated for all predictions and corresponding observations over the whole study period. To assess the predictive performance of SSM estimates temporally (i.e., at individual network stations over time) Nash-Sutcliffe Efficiency (NSE) was calculated for each station on a seasonal and annual basis (Nash and Sutcliffe, 1970). Values of NSE theoretically range from negative infinity to one, with values greater than zero indicating that the predictive performance of a given model for a given site is greater than simply using the mean of observed data. These evaluations allowed us to assess the general performance of our downscaling method relative to the coarser ESA product, as well as the performance of both products after being calibrated to observed daily maximum and minimum values in the ground-based network.

2.6. Comparison across land cover classes and vegetation indices

We investigated how different SSM estimates varied across land cover classes and in relation to the Normalized Difference Vegetation Index (NDVI) with the intent of identifying strengths and weaknesses of the estimates in the context of observations from previous research.

Distributions of predicted soil moisture across different land cover classes were assessed using the 2016 National Land Cover Dataset (NLCD; Yang et al. 2018). Zonal means, standard deviations (S.D.), and coefficients of variation (C.V.; standard deviation expressed as a percentage of the mean) of SSM were extracted for each date and each land cover class in the published NLCD 2016 grid after resampling from 30 m to 100 m resolution based on the majority class of each cell.

Spatiotemporal relationships between SSM estimates from each of the four methods and NDVI were investigated for each NDVI retrieval in 2018. The NDVI product used was a MODIS-Terra 16-day composite NDVI, generated at 250-m resolution and processed to minimize influences from clouds, aerosols, and view angles (Didan, 2015). Pixel-wise linear regressions were used to investigate temporal relationships at each individual point across the 22 NDVI retrieval dates, which were evaluated based on the regression slope, p-values, and R^2 values for all individual grid cells and within different land cover classes. Correlations between NDVI and SSM across space were also examined for each NDVI retrieval date and were evaluated based on the same metrics for each retrieval date.

2.7. Software and packages

All terrain attributes and daily climate surfaces were produced in SAGA GIS (version 6.4; Conrad et al. 2015). All statistical analyses and modeling were performed in R statistical software (version 3.5; R Core Team, 2019), using packages “geosphere” (Hijmans 2017a), “raster” (Hijmans 2017b), “knn” (Schliep and Hechenbichler, 2016), “RSAGA” (Brenning et al. 2018), and “caret” (Kuhn et al. 2018). An example code and small set of test data is available on GitHub (Warner, 2020), and the downscaled SSM grids produced in this study are available on the Hydroshare repository (Warner et al. 2020).

3. Results

3.1. Variable selection and model ensemble performance

The final set of high-importance variables selected for training the ensemble of KKNn models consisted of slope, cross-sectional curvature, channel network base level, catchment slope, topographic wetness index, valley depth, topographic position index, the set of six oblique geographic coordinate grids, three-day cumulative antecedent precipitation, and two-day antecedent mean temperature. A visualization of a daily ESA_{RAW} dataset and resulting downscaled $KKNn_{RAW}$ dataset is shown in Fig. 2.

After training, model testing set MAE values ranged 0.019–0.020 $m^3 m^{-3}$. In this instance, values of MAE reflect how well testing set predictions matched the values of the ESA grid cells used as inputs rather than in situ SSM observations. Model optimal k ranged from

13 to 16, and the optimal kernel across all models was triangular.

3.2. Evaluation of SSM estimation methods against ground network observations

Over the course of 2018, DEOS_{OBS} had an annual mean (± 1 S.D.) of $0.26 \pm 0.11 \text{ m}^3 \text{ m}^{-3}$. Across individual dates, mean DEOS_{OBS} ranged $0.09\text{--}0.38 \text{ m}^3 \text{ m}^{-3}$ and standard deviations ranged $0.02\text{--}0.13 \text{ m}^3 \text{ m}^{-3}$, with generally higher values in spring, fall, and winter, and several network-wide decreases in SSM during July, August, and September (Fig S1). In early January 2018, several dates with very low observed SSM led to high errors to corresponding modeled grid cell values. These erroneous dates correspond to a series of cold snaps with daily ambient temperatures well below 0°C across the DEOS network, potentially causing frozen soil conditions that decrease accuracy of measurements by soil moisture probes.

Fig. 3 illustrates MAE and MBE from spatial evaluations (i.e., comparisons of SSM estimates in grid cells corresponding to DEOS SSM monitoring station locations) of the four SSM estimates relative to DEOS_{OBS} calculated of each date in 2018, and a summary of these values is provided in Table 2. MAE for KKNNSC was relatively low throughout the year, while both ESA_{RAW} and KKNNSC had large positive bias and absolute error in the summer and early fall (Fig. 3) corresponding to periods of low SSM based on DEOS_{OBS} (Fig S1). Although ESA_{SC} and KKNNSC had relatively lower MAE and MBE than ESA_{RAW} and KKNNSC in the summer months, both estimates had a “wet bias” and consistently overestimated SSM in the spring and winter months (Fig. 3b). As a result, the annual means of MBE values for ESA_{RAW}, KKNNSC, and ESA_{SC} were roughly double that of KKNNSC, at 0.031, 0.030, and 0.032 compared to $0.015 \text{ m}^3 \text{ m}^{-3}$, respectively. Evaluating across all daily predictions corresponding to DEOS_{OBS} station locations, SSM estimates from ESA_{RAW}, KKNNSC, ESA_{SC}, and KKNNSC had R^2 values of 0.48, 0.52, 0.66, and 0.73, respectively (Fig S2).

Temporal evaluations of SSM estimates against DEOS_{OBS} were calculated on an annual and seasonal basis for each SSM monitoring station in the DEOS network. When evaluated on an annual basis, NSE from KKNNSC was greater than zero for 62% of DEOS network stations, compared to 43%, 29%, and 19% for ESA_{SC}, KKNNSC, and ESA_{RAW}, respectively, with noticeably improved performance in the southern portion of the study area (Fig. 4). KKNNSC SSM estimates generally had higher NSE across network stations when evaluated on a seasonal basis, though ESA_{RAW} and KKNNSC estimate NSE exceeded the calibrated estimates only in winter when performance of all estimation methods was poor (Fig S3).

3.3. Spatial and temporal variability of SSM with land cover class and NDVI

Estimates of daily zonal mean SSM varied significantly across major land cover classes although the relative differences among land cover classes were similar for each estimation method (Fig. 5). Variability in daily zonal mean SSM was much higher in ESA_{SC} and KKNNSC, and lowest in KKNNSC. Cropland, barren, evergreen forest, and woody wetland land cover classes had consistently low daily mean SSM, while pasture land, developed, and deciduous forest land cover classes were consistently high. Although emergent wetlands had generally higher SSM than woody wetlands, neither wetland class was predicted to have high mean daily SSM relative to the other classes in any of the estimation methods. Developed land cover classes had very little separation among daily mean values. Daily zonal C.V. varied within each estimation method as well, with particularly high C.V. values in barren land cover and particularly low values in pasture lands.

The four methods (i.e., ESA_{RAW}, ESA_{SC}, KKNNSC, and KKNNSC) yielded significant differences (pairwise Wilcoxon test, $p < 0.05$) in daily mean SSM over the entire year for several land cover classes. KKNNSC estimated lower SSM than all other methods in barren lands, evergreen and mixed forests, croplands, and both wetland land cover classes. In emergent wetlands, both KKNNSC datasets had significantly lower estimated SSM than both ESA products. In pasture lands, both ESA_{SC} and KKNNSC estimated higher SSM than the other products. Daily zonal C.V. values also varied across the different estimation methods and were generally higher for ESA_{SC} and KKNNSC.

Over the course of 2018, significant ($p < 0.05$) pixel-wise correlations between NDVI and estimated SSM were identified at 2% and 7% of grid cells using the ESA_{RAW} and KKNNSC methods, and 77% and 91% of grid cells using the ESA_{SC} and KKNNSC methods, respectively (Fig. 6, top row). The occurrence of these significant relationships varied across land cover class. Notably, significant NDVI and SSM relationships were less common in barren, cropland, and evergreen forest grid cells (63%, 71%, and 79% of grid cells for ESA_{SC}, and 66%, 85%, and 85% of grid cells for KKNNSC, respectively) than for pastureland, deciduous forest, and mixed forest grid cells (83%, 96%, and 96% for ESA_{SC}, and 98%, 99%, and 99% for KKNNSC, respectively). Slopes and R^2 values of pixel-wise NDVI-SSM relationships showed similar variability among land cover classes for both ESA_{SC} and KKNNSC, with both datasets identifying the highest mean pixel-wise R^2 values and steeply negative slopes in deciduous forests, mixed forests, emergent wetlands, and pasture lands. Barren lands, croplands, and evergreen forest classes generally showed weaker relationships less steeply negative slopes between NDVI and SSM from ESA_{SC} and KKNNSC. Coarse resolution grid cell-specific artifacts were evident in the relationships between NDVI and SSM derived from ESA_{SC} data (Fig. 6), with rectangular zones exhibiting particularly strong or weak relationships relative to other coarse cell footprints. We did not identify any clear spatial correlations between NDVI and SSM on each retrieval date.

4. Discussion

4.1. Strengths and limitations of the KKNNSC-based downscaling method

This study demonstrates a geomorphometry based downscaling method that provides estimates of daily SSM at much higher spatial resolution than is available in daily remotely sensed products using ancillary terrain, spatial, and meteorological data from publicly

available data sources. The performance indicators (i.e., MAE and MBE) calculated between our best-performing high resolution downscaled SSM estimates ($KKNN_{SC}$) and in situ observations were comparable to those identified in continental scale comparisons of observations and coarse remotely sensed (Draper et al. 2012, Albergel et al., 2012, Gruber et al. 2020) and downscaled (Guevara and Vargas, 2019) SSM products. A novel benefit of this approach is that by not using ancillary vegetation or land cover data in the downscaling process, we are able to examine spatiotemporal relationships between downscaled SSM estimates, vegetation distributions, and land cover maps while minimizing the risk of confounding factors and misleading interpretations (e.g., spurious correlations). It should be noted, however, that land cover and NDVI are not wholly independent of terrain features. In the northern portion of our study region, for example, developed areas and pasturelands primarily occupy the relatively flat hilltops in the landscape while steep hillslopes and valley bottoms contain dense forested and scrub areas that are unsuitable for development or agriculture. Thus, some confounding effects are inevitable in any geospatial research (Kedron et al., 2021).

Our results highlight the importance of comparing and calibrating remotely sensed environmental products with field observations when assessing SSM at specific regions or localities. Although $KKNN_{RAW}$ showed slightly higher temporal predictive performance for DEOS network stations than ESA_{RAW} (Fig. 4), daily calculated MAE and MBE was similar for both estimates, reflecting the limited ability of the raw values to capture the pronounced seasonal decline of SSM during the warm summer months (Fig. 3). This lack of seasonality resulted in a narrow range of SSM values across all land cover classes (Fig. 5) and a lack of significant temporal correlations between NDVI and SSM estimates from ESA_{RAW} and $KKNN_{RAW}$ (Fig. 6). Calibrating daily SSM datasets to the daily minima and maxima of $DOES_{OBS}$ helped capture these seasonal patterns and revealed widespread significant relationships between NDVI and SSM in both the coarse and downscaled datasets (Fig. 6). Both ESA_{SC} and $KKNN_{SC}$ outperformed their non-calibrated counterparts in both the daily calculated performance metrics (Table 2, Fig. 3) and in their predictive performance (i.e., NSE) for DEOS network stations over the study year (Fig. 4). The higher NSE values observed from $KKNN_{SC}$ in the southern part of the study area may indicate a benefit of the increased resolution of the downscaled estimates, as this area has a large degree of spatial heterogeneity in network station observations that may be challenging to resolve using low resolution satellite products. The need for calibration is not unexpected, as previous research has highlighted the need for calibration as a way to overcome the hurdle of systematic differences between large scale remotely sensed SSM products and field observations (Reichle et al. 2004, Draper et al. 2012). Thus, the accuracy of SSM estimates in our study region appears to be highly dependent on the availability of in situ data, and our calibrated SSM estimates are influenced by the environmental monitoring network design. The DEOS network is an excellent data source, but it is fundamentally designed as a meteorological network with stations located in open fields and roadsides to provide exposure to wind and solar

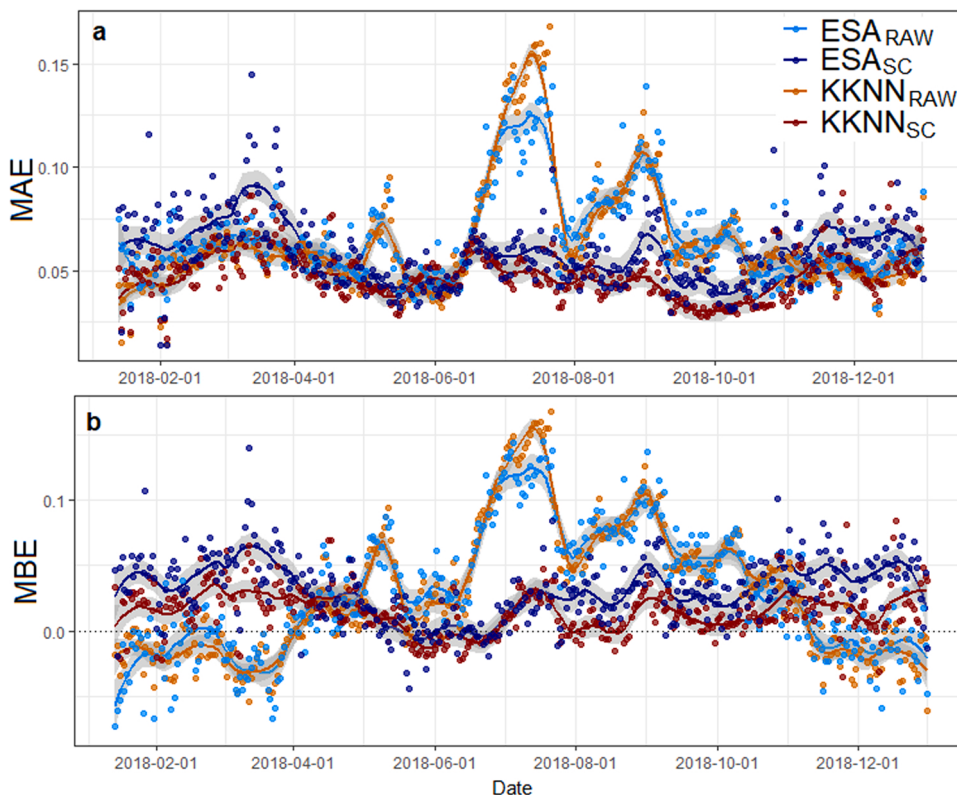


Fig. 3. Spatial evaluations of mean absolute error (MAE) and mean bias error (MBE) between $DOES_{OBS}$ and the four SSM estimates. MAE and MBE were calculated on each date during 2018 at grid cells corresponding to DEOS network site locations and plotted over time in units of $m^3 m^{-3}$. Points represent daily values while lines represent the generalized trend using a LOESS smoother with span = 0.1.

Table 2

Summary of mean absolute error (MAE) and mean bias error (MBE) values from SSM estimation methods. Values are reported as means and standard deviations for MAE and MBE calculated for each date ($n = 365$) between DEOS_{Obs} and SSM estimates at corresponding grid cells. All units are in m^3 .

Source	Season	MBE.mean	MBE.sd	MAE.mean	MAE.sd
ESA _{RAW}	Annual	0.031	0.077	0.066	0.050
	Spring	0.012	0.072	0.059	0.043
	Summer	0.080	0.067	0.087	0.058
	Fall	0.036	0.067	0.062	0.044
	Winter	-0.010	0.073	0.058	0.046
ESA _{SC}	Annual	0.032	0.067	0.059	0.045
	Spring	0.029	0.072	0.060	0.048
	Summer	0.019	0.064	0.054	0.040
	Fall	0.036	0.057	0.054	0.040
	Winter	0.044	0.073	0.070	0.049
KKN _{RAW}	Annual	0.030	0.076	0.065	0.051
	Spring	0.009	0.069	0.056	0.041
	Summer	0.084	0.070	0.091	0.061
	Fall	0.036	0.065	0.060	0.044
	Winter	-0.012	0.066	0.051	0.043
KKN _{SC}	Annual	0.016	0.060	0.048	0.039
	Spring	0.018	0.064	0.051	0.042
	Fall	0.017	0.051	0.042	0.033
	Summer	0.007	0.059	0.048	0.036
	Winter	0.021	0.065	0.053	0.044

radiation. It is unlikely that this distribution captures the full spectrum of spatial heterogeneity of SSM, terrain, and land cover across the study region. Consequently, this network design limits our downscaling and calibration efforts. Improved SSM monitoring capabilities in this region and others will facilitate the development of high-resolution data products that better capture the spatio-temporal variability of SSM.

While the downscaling and calibration method used in this study showed promise when evaluated against in situ SSM observations, there are major limitations of this method in some areas. One limitation is the challenge of model interpretability that results from an approach using a randomly generated ensemble of empirical machine learning models, which makes assessing variable dependence challenging. We clarify that this is not necessarily a weakness of our approach, but we recognize that this is a challenge for interpretation of machine learning approaches (Gilpin et al. 2018). A more practical limitation was identified because downscaled SSM estimates in wetland land cover classes were similar to (or lower than) many other land cover classes across all seasons (Fig. 5). This is a clear underestimation, as wetland soils are, by definition, saturated or near saturation for most of the year. However, wetland SSM was also similarly low in both coarse resolution ESA datasets (Fig. 5). Wetlands have long been known to present challenges for coarse scale remotely sensed SSM products due their small spatial extent but high SSM within a given grid cell (Wagner et al. 1999). As a result, the coarse resolution SSM products used as inputs to our downscaling model cannot reflect the very high SSM of wetland features within the landscape due to averaging effects within each grid cell. Within the multivariate space of our ancillary data alone, the extensive coastal wetland networks within our study area likely appear similar to nearby coastal plain areas (i.e., flat to mild slopes, low elevation, low relief), which generally sit atop well-drained sandy soils containing agricultural fields. As a result, the model was unable to distinguish the coastal wetlands from the neighboring dry soils. Thus, the inputs to our geomorphological downscaling methodology (i.e., ESA_{RAW} values and ancillary terrain and spatial datasets) inherently limit the ability of its outputs (i.e., KKN_{SC} values) to represent these high SSM features. Accounting for this remains a major challenge to prioritize in future downscaling efforts across landscapes containing wetland areas.

Our results demonstrated a consistent wet bias towards higher values of both ESA_{SC} and KKN_{SC} in the spring, fall, and winter (Fig. 3), which may explain why calibrated estimates had lower NSE at many sites in winter and fall (Fig S3). This bias may be explained by the relatively small scale of the study area relative to ESA_{RAW} grid resolution. Our downscaling methodology is based on daily information derived from ESA_{RAW} values, where each grid cell provides a single SSM value representing all the heterogeneity within its 27-km footprint. The spatial extent of our study area is defined by the spatial extent of the DEOS network, which limits the number of ESA_{RAW} grid cells from which our randomized sampling could draw on any given day and in turn limits the number of possible values that could be modeled in the KKN_{RAW} output on any given day. This resulted in limited variance among higher range SSM estimates, which, when calibrated to DEOS network minima and maxima, yielded wide swaths of the study area with values skewed towards the daily maximum SSM observed by the DEOS network. This highlights the challenges of accounting for small scale spatial SSM heterogeneity, particularly in separating localized, high SSM features from the greater landscape.

4.2. Spatial and temporal relationships of SSM with land cover and NDVI

We were able to examine how different estimation methods lead to differences in SSM among land cover classes as well as the variability of SSM values within each class by comparing zonal means and C.V. values of daily SSM estimates. Mean daily estimated SSM showed similar variations among land cover classes for all four estimation methods, suggesting that both coarse and downscaled

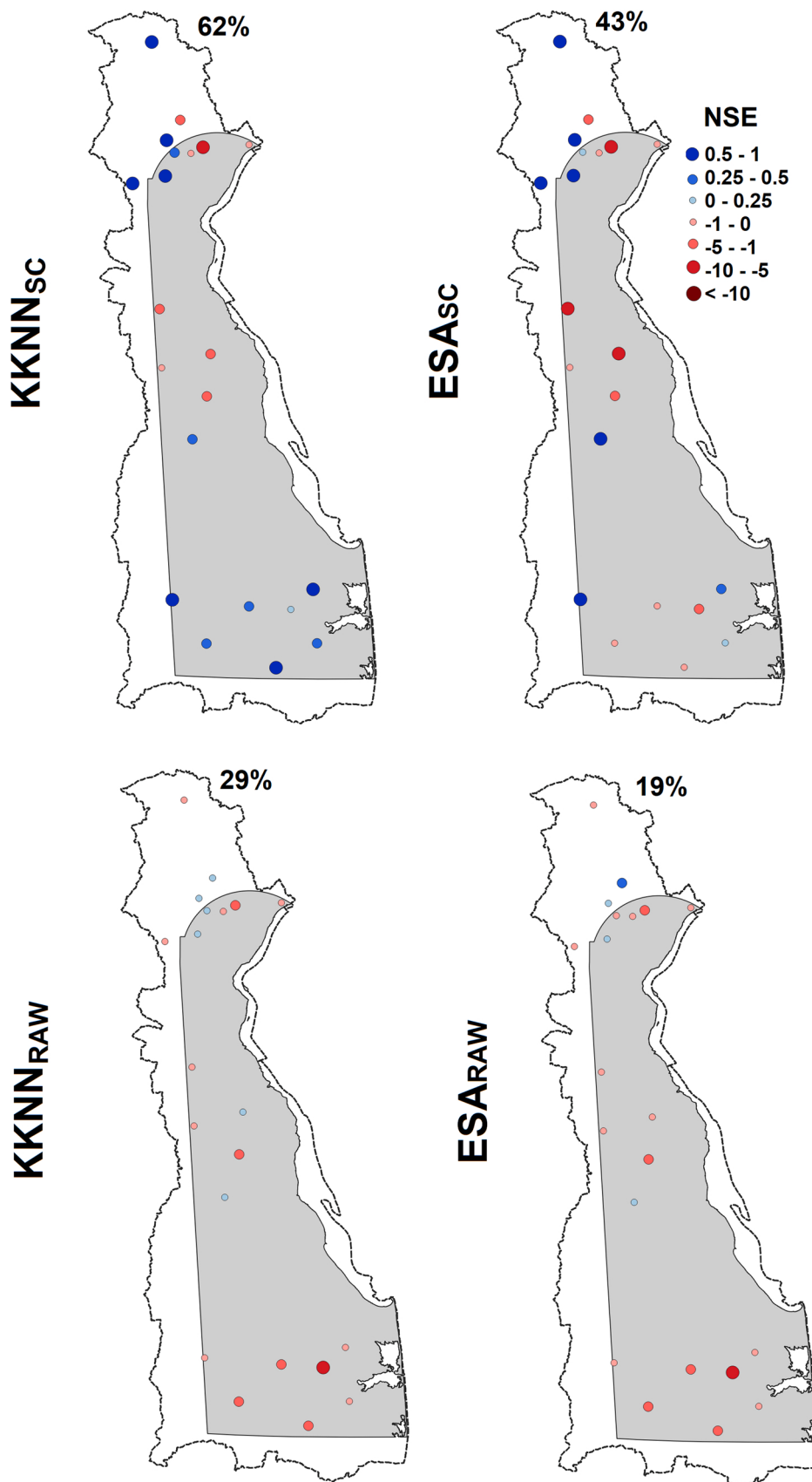


Fig. 4. Annual map of Nash-Sutcliffe Efficiency (NSE) across all DEOS monitoring stations for each of the four estimation techniques. Increasingly dark and large circles indicate increasingly negative (red) and positive (blue) NSE. Percentages indicate the percent of stations above the threshold of 0, indicating that the modeled values outperformed the observational mean.

resolution methods were able to distinguish general differences in SSM between land cover over the course of the year across our study region (Fig. 5). However, the downscaled grids have the advantage of the ability to resolve sub-grid cell variability that is lost in the coarser ESA datasets. In fact, the combination of downscaling and calibration (i.e., KNN_{SC}) led to the greatest degree of SSM variability throughout the study region compared to other estimation methods, as SSM C.V. (i.e., the zonal standard deviation expressed as a percentage of the zonal mean) was consistently higher from KNN_{SC} than other estimation methods for most land cover classes (Fig. 5).

Except for wetland land cover classes, all SSM estimation methods provided values that generally agreed with observed SSM in major land cover classes in this region. Unfortunately, the DEOS network SSM monitoring locations were only located in cropland, open spaces, and reported SSM observations from many other land cover classes are scarce. SSM values in the cropland class fell within ranges of surface water content reported in previous work in various types of croplands in the mid-Atlantic coastal plain (McCann et al. 2007, Roygard et al. 2002) which contains most croplands in the study area. Of the forest classes, mixed and deciduous forests are concentrated in the northern Piedmont region and scattered throughout the rest of the study area, while evergreen forests are almost exclusively located on sandy coastal plain soils in the southern portion of the study area (Fig. 1). Estimated ranges of SSM for deciduous and mixed forests fell within ranges of in situ seasonal observations from a recent study in a typical Piedmont forest in this region (Warner et al. 2018). Barren and grassland land cover classes occupy a very small fraction of the study area and, though they generally correspond to beaches and dune areas, we found no in situ SSM data for comparison in this region. Reported in situ SSM observations for the study region were also lacking for developed land cover classes. For all methods, estimated SSM was similar across all developed land cover intensities, which may reflect a limitation of our methodology to capture the well-established heterogeneity of urban SSM driven by human influenced factors like soil disturbances and the extent of impermeable surfaces (Wiesner et al. 2016, Zhang et al. 2019). The ancillary datasets used in our downscaling methodology could not account for such factors, limiting its ability to resolve fine scale variation of soil moisture in developed areas. We highlight the importance of collecting and cataloging in situ SSM data in all heterogeneous features and varieties of land cover to support future SSM downscaling and modeling efforts, echoing the sentiments of other recent research on soil moisture downscaling (Zappa et al. 2019).

Estimates from ESA_{RAW} appeared unsuitable for evaluating temporal relationships between NDVI and SSM across the study region based on the pixel-wise linear regressions, of which only 2% suggested significant correlations ($p < 0.05$). Downscaling alone (i.e., KNN_{RAW}) only slightly increased the number of significant temporal relationships (7%), most of which featured very steep slopes resulting from the narrow range of KNN_{RAW} SSM values throughout the year. The additional calibration to daily DEOS network minima and maxima (i.e., ESA_{SC} and KNN_{SC}) greatly improved our ability to identify temporal relationships between SSM and NDVI (77% and 91%). Spatial evaluations of relationships between NDVI and KNN_{SC} had the additional benefit over ESA_{SC} of being free from artifacts introduced by the coarse resolution ESA grids (Fig. 6), which arise from the averaging of estimated SSM across all land cover and vegetation heterogeneity within each ESA grid cell footprint.

The pixel-wise temporal correlations between NDVI and KNN_{SC} indicated that SSM and NDVI are generally negatively correlated through time across the study area (Fig. 6). This reflects seasonal transitions of vegetation greening and corresponding increases in root zone water use resulting in decreased SSM. The R-squared values suggested that these relationships were particularly strong in deciduous forests, mixed forests, and pasture lands. Conversely, these relationships were relatively weak in evergreen forests, barren lands, and croplands. These weaker correlations may be a result of limited seasonal variations in evergreen forest canopy foliage, a lack of foliage in barren lands, and irregular seasonality of NDVI in croplands due to agricultural practices (i.e., irrigation regimes, harvest, land tillage) for which our downscaling approach cannot account (Hmimina et al. 2013, Moulin et al. 1997). In fact, 71% of grid cells lacking significant temporal correlations between NDVI and KNN_{SC} corresponded to croplands, though croplands accounted for only 41% of the study area. Agricultural land management may also explain the lack of clear spatial correlations between NDVI and SSM estimates across the study area for each date of NDVI retrieval, as the modeled SSM cannot account for irrigation practices that may keep cropland NDVI values high throughout the growing season. The spatial distribution of agricultural land management regimes for individual crop fields is unknown in this region, and coordinated studies with agricultural stakeholders may help improve future SSM downscaling efforts for the benefit of the agricultural community. Without including such land management strategies, high spatial resolution SSM downscaling methods employing primarily terrain data may have limited application for targeted field or farm-specific studies, though they may still be useful for assessing mesoscale SSM dynamics in agricultural areas.

5. Conclusions

This study demonstrates the strengths and limitations of a geomorphometric kernel K-nearest neighbors downscaling method for estimating high resolution distributions of SSM based on remote sensing products from the ESA-CCI soil moisture program using meteorological data from a ground network, spatial coordinates, and terrain attributes as predictors. Our downscaled estimates were evaluated against ground network data and showed improvement over the coarse resolution product, although calibration of both coarse and downscaled SSM was necessary for capturing seasonal SSM variations and identifying temporal relationships to NDVI. All SSM estimates indicated a consistent wet bias, and subsequent analysis of SSM estimates in relation to land cover classes found clear limitations of all estimation methods in wetland areas. Our methodology provided high spatiotemporal resolution SSM maps with

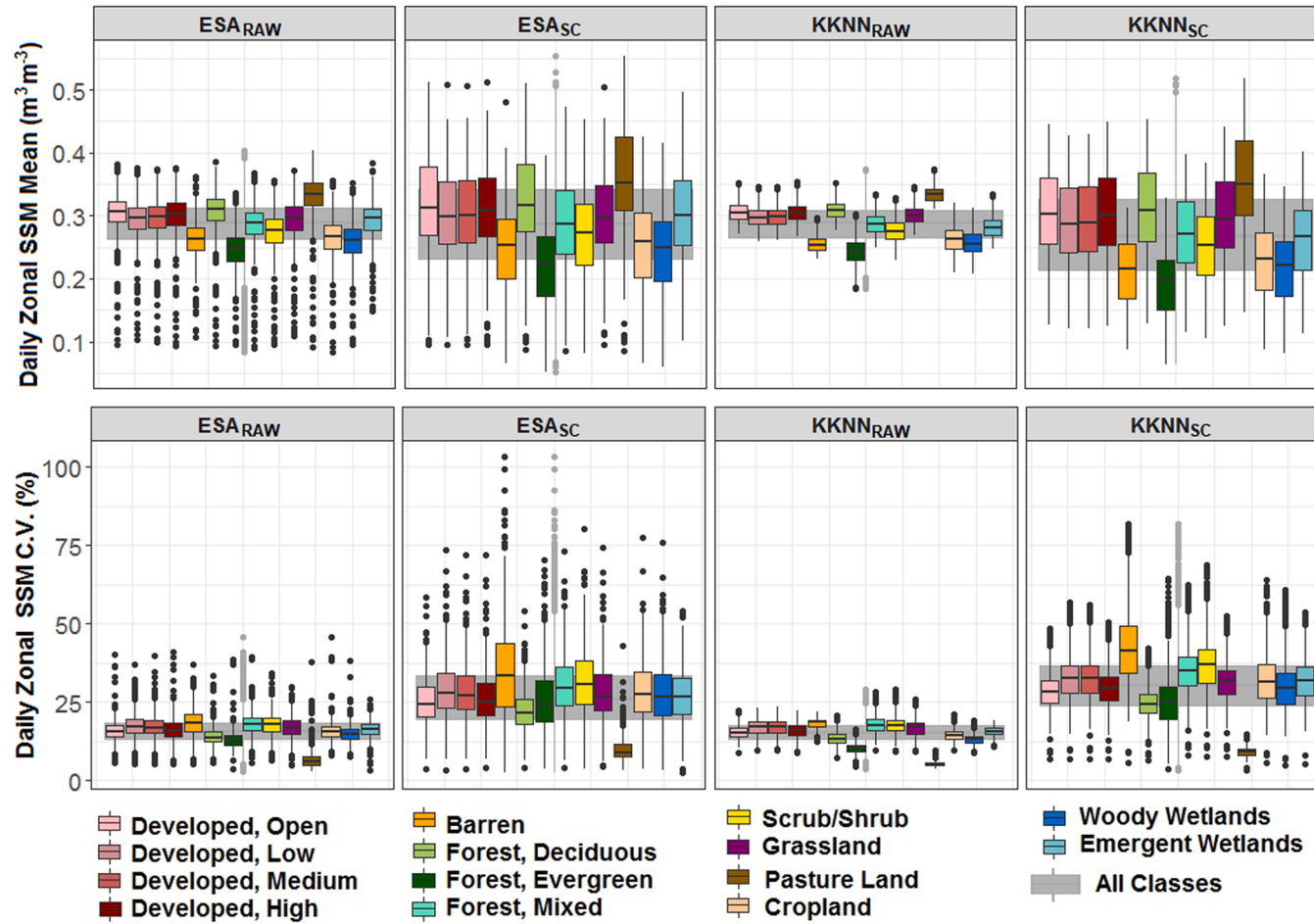


Fig. 5. Boxplots of daily zonal means (top) and daily zonal coefficients of variation (C.V., bottom, expressed as the standard deviation as a percentage of the mean) for daily SSM estimates and land cover classes over the year 2018. Colored bars represent different land cover classes, while the wide grey bar in the background represents all classes combined.

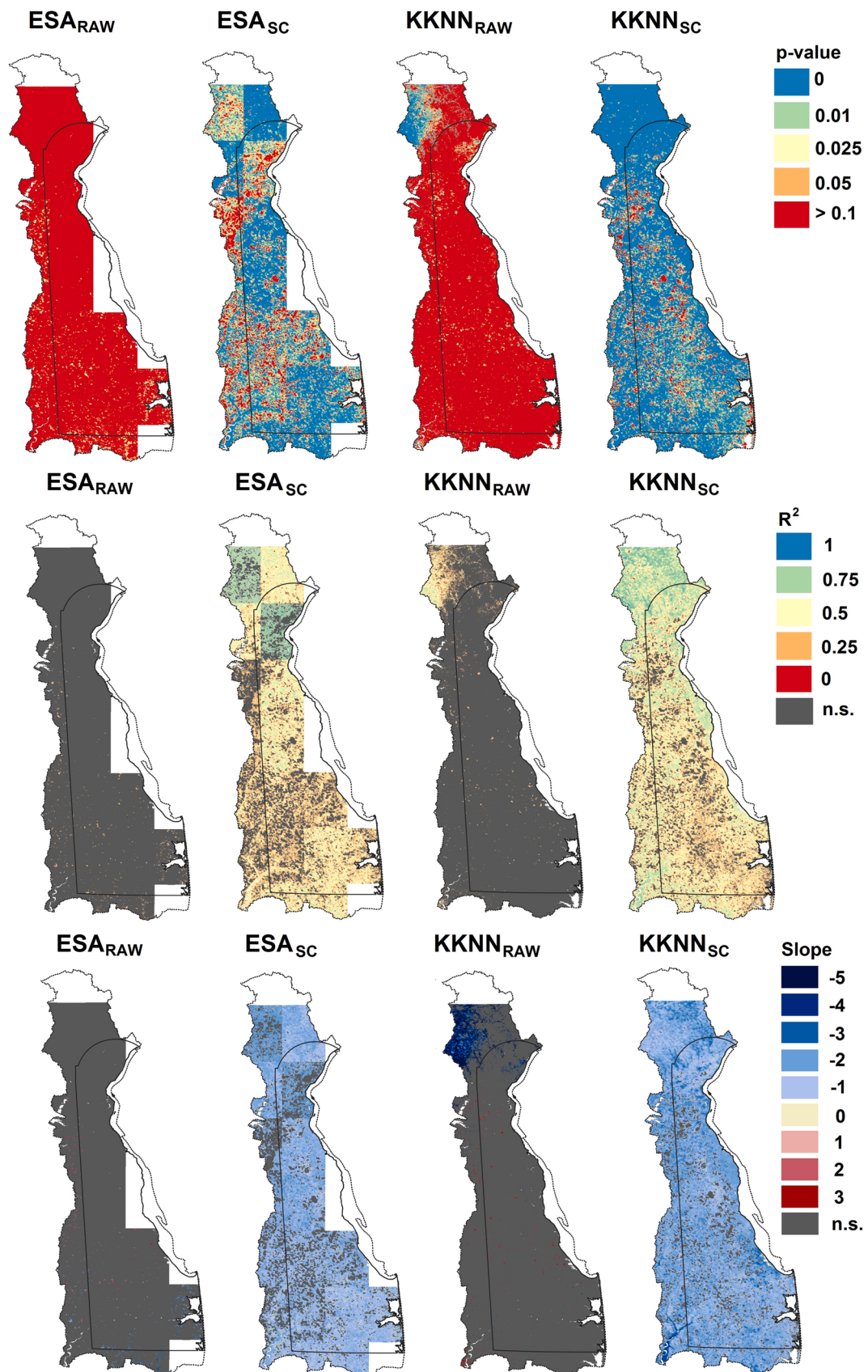


Fig. 6. P-values (top row), R-squared values (middle row), and slopes (bottom row) of pixel-wise linear regressions of SSM estimates and NDVI over time within the study area. A small sliver of the northern portion of the study area was removed due to missing data. Grey areas indicate grid cells where correlations were not significant at an alpha level of 0.05.

lower error than coarse resolution datasets and showed higher predictive performance through time when evaluated against monitoring network observations. This data driven downscaling methodology may be applied to other (coarse spatial scale/high temporal scale) remotely sensed SSM datasets (e.g., SMAP; Entekhabi et al., 2010a) and in other regions provided the necessary spatial and monitoring network data. We highlight the need for more spatially representative ground based SSM monitoring networks that will enhance our ability to evaluate and refine downscaled SSM estimates. Future work should address the challenges posed by downscaling remotely sensed SSM in wetland areas and attempt to account for causes of wet bias, perhaps with greater interaction between ground networks and satellite data. Finally, we emphasize the importance of developing downscaling techniques that allow for investigations into spatiotemporal relationships between SSM and remotely sensed vegetation data, which will support investigations into ecological processes, drought forecasting, and agricultural management strategies at localized landscape scales.

CRedit authorship contribution statement

Daniel L. Warner: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – review & editing. **Mario Guevara:** Conceptualization, Data curation, Methodology, Writing – review & editing. **John Callahan:** Data curation, Supervision, Writing – review & editing. **Rodrigo Vargas:** Conceptualization, Funding acquisition, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We acknowledge Kevin Brinson and Chris Hughes from the University of Delaware Center for Environmental Monitoring and Analysis for their help in acquiring DEOS network soil moisture and meteorological data. RV acknowledges support from NSF (#1724843 and #2103836).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.ejrh.2021.100946](https://doi.org/10.1016/j.ejrh.2021.100946).

References

- 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 (January), 112301 <https://doi.org/10.1016/j.rse.2021.112301>.
- Albergel, C., de Rosnay, P., Gruhier, C., Muñoz-Sabater, J., Hasenauer, S., Isaksen, L., Kerr, Y., Wagner, W., 2012. Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations. *Remote Sens. Environ.* 118, 215–226. <https://doi.org/10.1016/j.rse.2011.11.017>.
- Bates, B., Kundzewicz, Z., Wu, S., Palutikof, J., 2008. Climate change and water. IPCC Technical Paper VI.
- Brenning, A., Bangs, D., Becker, M., 2018. RSAGA: SAGA geoprocessing and terrain analysis. R package version 1.3.0. <https://CRAN.R-project.org/package=RSAGA>.
- Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.* 24, 43–69. <https://doi.org/10.1080/02626667909491834>.
- Buczko, U., Bens, O., Hüttl, R.F., 2007. Changes in soil water repellency in a pine-beech forest transformation chronosequence: influence of antecedent rainfall and air temperatures. *Ecol. Eng.* 31, 154–164. <https://doi.org/10.1016/j.ecoleng.2007.03.006>.
- Carlson, T., 2007. An overview of the “Triangle Method” for estimating surface evapotranspiration and soil moisture from satellite imagery. *Sensors* 7 (8), 1612–1629. <https://doi.org/10.3390/s7081612>.
- CEC, 2021. Guide to drought indices and indicators used in North America. Montreal, Canada: Commission for Environmental Cooperation, p. 62.
- Chung, D., Dorigo, W., De Jeu, R., Kidd, R., Wagner, W., 2018. Alzheimer’s disease and the autophagic-lysosomal system, ESA Climate Change Initiative Phase II – Soil moisture, product specification document (PSD); D.1.2.1 Version 4.4; Earth Observation Data Centre for Water Resources Monitoring (EODC) GmbH: Vienna, Austria, 697; p. 49–58.
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4, Geoscientific Model Development, 8, 1991–2007, doi:10.5194/gmd-8-1991-2015.
- Didan, K., 2015. MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006. NASA EOSDIS Land Processes DAAC, <https://doi.org/10.5067/MODIS/MOD13Q1.006>.
- Dong, J., Crow, W.T., Tobin, K.J., Cosh, M.H., Bosch, D.D., Starks, P.J., Seyfried, M., Collins, C.H., 2020. Comparison of microwave remote sensing and land surface modeling for surface soil moisture climatology estimation. *Remote Sens. Environ.* 242. <https://doi.org/10.1016/j.rse.2020.111756>.
- Dorigo, W.A., de Jeu, R., 2016. Satellite soil moisture for advancing our understanding of earth system processes and climate change. *Int. J. Appl. Earth Obs. Geoinf.* 48. <https://doi.org/10.1016/j.jag.2016.02.007>.

- Dorigo, W.A., Gruber, A., De Jeu, R.A.M., Wagner, W., Stacke, T., Loew, A., Albergel, C., Brocca, L., Chung, D., Parinussa, R.M., Kidd, R., 2015. Evaluation of the ESA CCI soil moisture product using ground-based observations. *Remote Sens. Environ.* 162, 380–395.
- Dorigo, W.A., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, D.P., Hirschi, M., Ikonen, J., De Jeu, R., Kidd, R., Lahoz, W., Liu, Y.Y., Miralles, D., Lecomte, P., 2017. ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sens. Environ.* 2017. <https://doi.org/10.1016/j.rse.2017.07.001>.
- Draper, C.S., Reichle, R.H., De Lannoy, G.J.M., Liu, Q., 2012. Assimilation of passive and active microwave soil moisture retrievals. *Geophys. Res. Lett.* 39 (4), 1–5. <https://doi.org/10.1029/2011GL050655>.
- Dubois, P.C., van Zyl, J., Engman, T., 1995. Measuring soil moisture with imaging radars. *IEEE Trans. Geosci. Remote Sens.* 33, 915–926. <https://doi.org/10.1109/TGRS.1995.477194>.
- 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 (5), 704–716. <https://doi.org/10.1109/JPROC.2010.2043918>.
- Fang, B., Lakshmi, V., Bindlish, R., Jackson, T.J., 2018. AMSR2 soil moisture downscaling using temperature and vegetation data. *Remote Sens.* 10 (10), 1575. <https://doi.org/10.3390/rs10101575>.
- Ford, T.W., Quiring, S.M., 2019. Comparison of contemporary in situ, model, and satellite remote sensing soil moisture with a focus on drought monitoring. *Water Resour. Res.* 55, 1565–1582. <https://doi.org/10.1029/2018WR024039>.
- Gallant, J.C., Wilson, J.P., 2000. Primary topographic attributes. In: Wilson, J.P., Gallant, J.C. (Eds.), *Terrain Analysis: Principles and Applications*. Wiley, New York, pp. 51–85.
- 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. *Int. J. Remote Sens.* 18 (15), 3145–3166. <https://doi.org/10.1080/014311697217026>.
- Gilpin, L.H., Bau, D., Yuan, B.Z., Bajwa, A., Specter, M., Kagal, L., 2018. Explaining Explanations: An Overview of Interpretability of Machine Learning. Pages 80–89 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA).
- Green, J.K., Seneviratne, S.I., Berg, A.M., Findell, K.L., Hagemann, S., Lawrence, D.M., Gentile, P., 2019. Large influence of soil moisture on long-term terrestrial carbon uptake. *Nature* 565 (7740), 476–479. <https://doi.org/10.1038/s41586-018-0848-x>.
- Gruber, A., De Lannoy, G., Albergel, C., Al-Yaari, A., Brocca, L., Calvet, J.C., Colliander, A., Cosh, M., Crow, W., Draper, C., Hirschi, M., Kerr, Y., Konings, A., Lahoz, W., McColl, K., Montzka, C., Muñoz-Sabater, J., Peng, J., Reichle, R., Richaume, P., Rüdiger, C., Scanlon, T., van der Schalie, R., Wigneron, J.P., Wagner, W., 2020. Validation practices for satellite soil moisture retrievals: What are (the) errors? *Remote Sens. Environ.* 244. <https://doi.org/10.1016/j.rse.2020.111806>.
- Gruber, A., Dorigo, W.A., Crow, W., Wagner, W., 2017. Triple collocation-based merging of satellite soil moisture retrievals. *IEEE Trans. Geosci. Remote Sens.* 1–13. <https://doi.org/10.1109/TGRS.2017.2734070>.
- Gruber, A., Scanlon, T., van der Schalie, R., Wagner, W., Dorigo, W., 2019. Evolution of the ESA CCI Soil Moisture Climate Data Records and their underlying merging methodology. *Earth System Sci. Data* 11, 717–739. <https://doi.org/10.5194/essd-11-717-2019>.
- Guevara, M., Tauffer, M., Vargas, R., 2021. Gap-free global annual soil moisture: 15 km grids for 1991–2018. *Earth System Sci. Data* 13, 1711–2021. <https://doi.org/10.5194/essd-13-1711-2021>.
- Guevara, M., Vargas, R., 2019. Downscaling satellite soil moisture using geomorphometry and machine learning. *PLoS One* 1–41. <https://doi.org/10.1371/journal.pone.0219639>.
- Han, E., Crow, W.T., Holmes, T., Bolten, J., 2014. Benchmarking a soil moisture data assimilation system for agricultural drought monitoring. *J. Hydrometeorol.* 15 (3), 1117–1134. <https://doi.org/10.1175/JHM-D-13-0125.1>.
- Hijmans, R.J., 2017a. geosphere: Spherical Trigonometry. R package version 1.5–7. <https://CRAN.R-project.org/package=geosphere>.
- Hijmans, R.J., 2017b. raster: Geographic Data Analysis and Modeling. R package version 2.6–7. <https://CRAN.R-project.org/package=raster>.
- Hmimina, G., Dufrène, O., Pontauzier, J.Y., Delpierre, N., Aubinet, N., Caquet, B., de Grandcourt, A., Burban, B., Flechard, C., Granier, A., Gross, P., Heinesch, B., Longdoz, B., Moureaux, C., Ourcival, J.M., Rambal, S., Saint André, L., Soudani, K., 2013. Evaluation of the potential of MODIS satellite data to predict vegetation phenology in different biomes: an investigation using ground-based NDVI measurements. *Remote Sens. Environ.* 132, 145–158. <https://doi.org/10.1016/j.rse.2013.01.010>.
- Huntington, T.G., 2006. Evidence for intensification of the global water cycle: review and synthesis. *J. Hydrol.* 319 (1–4), 83–95. <https://doi.org/10.1016/j.jhydrol.2005.07.003>.
- Huntington, T.G., Richardson, A.D., McGuire, K.J., Hayhoe, K., 2009. Climate and hydrological changes in the northeastern United States: recent trends and implications for forested and aquatic ecosystems. *Can. J. For. Res.* 39, 199–212. <https://doi.org/10.1139/X08-116>.
- Kedron, P., Li, W., Fotheringham, S., Goodchild, M., 2021. Reproducibility and replicability: opportunities and challenges for geospatial research. *Int. J. Geogr. Inf. Sci.* 35, 427–445.
- Kuhn, M. Contributions from Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team, Michael Benesty, Reynald Lescarbeau, Andrew Ziem, Luca Scrucca, Yuan Tang, Can Candan and Tyler Hunt. 2018. caret: Classification and Regression Training. R package version 6.0–81. <https://CRAN.R-project.org/package=caret>.
- Liu, Y., Yang, Y., Jing, W., Yue, X., 2017. Comparison of different machine learning approaches for monthly satellite-based soil moisture downscaling over northeast china. *Remote Sens.* 10 (1), 31. <https://doi.org/10.3390/rs10010031>.
- 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>.
- Llomas, R.M., Guevara, M., Rorabaugh, D., Tauffer, M., Vargas, R., 2020. Spatial gap-filling of ESA CCI satellite-derived soil moisture based on geostatistical techniques and multiple regression. *Remote Sens.* 12 (4), 665. <https://doi.org/10.3390/rs12040665>.
- McCann, I., Kee, E., Adkins, J., Ernest, E., Ernest, J., 2007. Effect of irrigation rate on yield of drip-irrigated seedless watermelon in a humid region. *Sci. Hortic.* 113 (2), 155–161. <https://doi.org/10.1016/j.scienta.2007.03.008>.
- McColl, K.A., Alemohammad, S.H., Akbar, R., Konings, A.G., Yueh, S., Entekhabi, D., 2017. The global distribution and dynamics of surface soil moisture. *Nat. Geosci.* 10 (2), 100–104. <https://doi.org/10.1038/ngeo2868>.
- Merlin, O., Escorihuela, M.J., Mayoral, M.A., Hagolle, O., Al Bitar, A., Kerr, Y., 2013. Self-calibrated evaporation-based disaggregation of SMOS soil moisture: an evaluation study at 3km and 100m resolution in Catalunya, Spain. *Remote Sens. Environ.* 130 (2013), 25–38. <https://doi.org/10.1016/j.rse.2012.11.008>.
- Mohanty, B.P., Cosh, M.H., Lakshmi, V., Montzka, C., 2017. Soil moisture remote sensing: state-of-the-science. *Vadose Zone J.* 16 (1) <https://doi.org/10.2136/vzj2016.10.0105>.
- Montzka, C., Rötzer, K., Bogen, H.R., Sanchez, N., Vereecken, H., 2018. A new soil moisture downscaling approach for SMAP, SMOS, and ASCAT by predicting sub-grid variability. *Remote Sens.* 10. <https://doi.org/10.3390/rs10030427>.
- Møller, A.B., Beucher, A.M., Pouliad, N., Greve, M.H., 2020. Oblique geographic coordinates as covariates for digital soil mapping. *Soil* 6, 269–289. <https://doi.org/10.5194/soil-6-269-2020>.
- Moulin, S., Kergoat, L., Viovy, N., Dedieu, G., 1997. Global-scale assessment of vegetation phenology using NOAA/AVHRR satellite measurements. *J. Clim.* 10 (6), 1154–1170. [https://doi.org/10.1175/1520-0442\(1997\)010<1154:GSAOVP>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<1154:GSAOVP>2.0.CO;2).
- Najjar, R.G., Walker, H.A., Anderson, P.J., Barron, E.J., Bord, R.J., Gibson, J.R., Kennedy, V.S., Knight, C.G., Megonigal, J.P., O'Connor, R.E., Polsky, C.D., Psuty, N.P., Richards, B.A., Sorenson, L.G., Steele, E.M., Swanson, R.S., 2000. The potential impacts of climate change on the mid-Atlantic coastal region. *Clim. Res.* 7, 219–233. <https://doi.org/10.3354/cr014219>.
- Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual model. Part 1—a discussion of principles. *J. Hydrol.* 10, 282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6).

- Peng, J., Loew, A., Merlin, O., Verhoest, N.E., 2017. A review of spatial downscaling of satellite remotely sensed soil moisture. *Rev. Geophys.* 55 (2), 341–366. <https://doi.org/10.1002/2016RG000543>.
- Piles, M., Camps, A., Vall-Llossera, M., Corbella, I., Panciera, R., Rudiger, C., Kerr, Y.H., Walker, J., 2011. Downscaling SMOS-derived soil moisture using MODIS visible/infrared data. *IEEE Trans. Geosci. Remote Sens.* 49 (9), 3156–3166. <https://doi.org/10.1109/TGRS.2011.2120615>.
- R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>.
- 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>.
- Reichle, R.H., Koster, R.D., Dong, J., Berg, A.A., 2004. Global soil moisture from satellite observations, land surface models, and ground data: implications for data assimilation. *J. Hydrometeorol.* 5 (3), 430–442. [https://doi.org/10.1175/1525-7541\(2004\)005<0430:GSMFSO>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0430:GSMFSO>2.0.CO;2).
- Roygard, J.K.F., Alley, M.M., Khosla, R., 2002. No-till corn yields and water balance in the mid-Atlantic coastal plain. *Agron. J.* 94 (3), 612–623. <https://doi.org/10.2134/agronj2002.0612>.
- Schliep, K., Hechenbichler, K., 2016. kkn: Weighted k-Nearest Neighbors. R package version 1.3.1. <https://CRAN.R-project.org/package=kkn>.
- Schmugge, T.J., Kustas, W.P., Ritchie, J.C., Jackson, T.J., Rango, A., 2002. Remote sensing in hydrology. *Adv. Water Resour.* 25, 1367–1385. [https://doi.org/10.1016/S0309-1708\(02\)00065-9](https://doi.org/10.1016/S0309-1708(02)00065-9).
- 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 Sci. Rev.* 99 (3–4), 125–161. <https://doi.org/10.1016/j.earscirev.2010.02.004>.
- Sharma, J., Prasad, R., Srivastava, P.K., Singh, S.K., Yadav, S.A., Yadav, V.P., 2021. Roughness characterization and disaggregation of coarse resolution SMAP soil moisture using single-channel algorithm. *J. Appl. Remote Sens.* 15 (01), 1–14. <https://doi.org/10.1117/1.jrs.15.014514>.
- Soil Survey Staff, National Resources Conservation Service, US Department of Agriculture. 2015. Soil Survey Geographic Database for DE, MD, NJ, and PA. Accessed: 05–30–2019.
- Topp, G.C., Davis, J.L., Annan, A.P., 1980. Electromagnetic determination of soil water content: measurements in coaxial transmission lines. *Water Resour. Res.* 16, 574–582.
- Wagner, W., Lemoine, G., Rott, H., 1999. A method for estimating soil moisture from ERS Scatterometer and soil data. *Remote Sens. Environ.* 70, 191–207. [https://doi.org/10.1016/S0034-4257\(99\)00036-X](https://doi.org/10.1016/S0034-4257(99)00036-X).
- Warner, D.L. 2020. warnerdl/DownscalingDelawareSSM: Downscaling Soil Moisture: Code and Data example (Version v1.0). Zenodo. <http://doi.org/10.5281/zenodo.4304678>.
- Warner, D.L., Guevara, M., Callahan, J., Vargas, R. 2020. 2018 Daily Downscaled ESA-CCI Soil Moisture Grids for Delaware, USA, HydroShare, <http://www.hydroshare.org/resource/81867bb55f134d8d89ec2250ae6c1014>.
- Warner, D.L., Vargas, R., Seyfferth, A., Inamdar, S., 2018. Transitional slopes act as hotspots of both soil CO₂ emission and CH₄ uptake in a temperate forest landscape. *Biogeochemistry* 138 (2), 121–135. <https://doi.org/10.1007/s10533-018-0435-0>.
- Wiesner, S., Gröngroft, A., Ament, F., Eschenbach, A., 2016. Spatial and temporal variability of urban soil water dynamics observed by a soil monitoring network. *J. Soils Sedim.* 16 (11), 2523–2537. <https://doi.org/10.1007/s11368-016-1385-6>.
- Yan, R., Bai, J., 2020. A new approach for soil moisture downscaling in the presence of seasonal difference. *Remote Sens.* 12 (17), 1–19. <https://doi.org/10.3390/rs12172818>.
- Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Grannemann, B., Rigge, M., Xian, G., 2018. A new generation of the United States National Land Cover Database: requirements, research priorities, design, and implementation strategies. *ISPRS J. Photogramm. Remote Sens.* 146, 108–123.
- Zappa, L., Forkel, M., Xaver, A., Dorigo, W., 2019. Deriving field scale soil moisture from satellite observations and ground measurements in a Hilly Agricultural Region. *Remote Sens.* 11 (22), 2596. <https://doi.org/10.3390/rs11222596>.
- Zhang, J., Li, S., Sun, X., Tong, J., Fu, Z., Li, J., 2019. Sustainability of urban soil management: analysis of soil physicochemical properties and bacterial community structure under different green space types. *Sustainability* 11 (5), 1–17. <https://doi.org/10.3390/su11051395>.