6-6-2019

Mapping Soil Moisture from Remotely Sensed and In-situ Data with Statistical Methods

Yaping Xu
Louisiana State University and Agricultural and Mechanical College, yxu53@lsu.edu

MAPPING SOIL MOISTURE FROM REMOTELY SENSED AND IN-SITU DATA WITH STATISTICAL METHODS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Geography and Anthropology

by
Yaping Xu
B.E., Guizhou University, 2009
M.S., Chinese Academy of Sciences, 2013
August 2019

ACKNOWLEDGEMENTS

The successful completion of my dissertation is a significant moment to me. I would like to acknowledge the individuals and organizations who have provided major supports toward this process.

First, I would like to give many thanks to my advisor, Dr. Lei Wang, for his devotion and support. Dr. Wang is such a knowledgeable and kind advisor. He always encouraged me to communicate with him and other fellow graduates. Several of my research ideas were from our discussion. Many thanks to my committee members, Dr. Robert Rohli, Dr. Bin Li, Dr. Barry Keim, and Dr. Edward Bush. I appreciate their invaluable services as my committee members. Their constructive advices in my general exam have helped me a great deal in improving my dissertation. This research is funded and supported by Louisiana State University Dissertation Fellowship, South Central Climate Science Center, and Louisiana Sea Grant. Thank Wesley Burgett from Texas Tech University, the West Texas Mesonet operations manager and Todd Caldwell from University of Texas at Austin, the Principal Investigator for Texas Soil Observation Network (TxSON) for the help with the data cleaning and processing. Thank the NASA DEVELOP and USDA Southeast Regional Climate Hub (SERCH) for their time and support to make this project possible. Finally, I would like to thank my dear parents and my wife, Dr. Cuiling Liu for their continued support and selfless devotion to my career and life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Soil moisture is an important factor for accurate prediction of agricultural productivity and rainfall runoff with hydrological models. Remote sensing satellites such as Soil Moisture Active Passive (SMAP) offer synoptic views of soil moisture distribution at a regional-to-global scale. To use the soil moisture product from these satellites, however, requires a downscaling of the data from an usually large instantaneous field of view (i.e. 36 km) to the watershed analysis scales ranging from 30 m to 1 km. In addition, validation of the soil moisture products using the ground station observations without an upscaling treatment would lead to cross-level fallacy. In the literature of geographical analysis, scale is one of the top research concens because of the needs for multi-source geospatial data fusion. This dissertation research introduced a multi-level soil moisture data assimilation and processing methodology framework based on spatial information theories. The research contains three sections: downscaling using machine learning and geographically weighted regression, upscaling ground network observation to calibrate satellite data, and spatial and temporal multi-scale data assimilation using spatio-temporal interpolation.

(1) Soil moisture downscaling

In the first section, a downscaling method is designed using 1-km geospatial data to obtain subpixel soil moisture from the 9-km soil moisture product of the SMAP satellite. The geospatial data includes normalized difference vegetation index (NDVI), land surface temperature (LST), gross primary productivity (GPP), topographical moisture index (TMI), with all resampled to 1-km resolution. The machine learning algorithm – random forest was used to create a prediction model of the soil moisture at a 1-km resolution. The 1-km soil moisture product was compared with the ground samples from the West Texas Mesonet (WTM) station data. The residual was then interpolated to compensate the unpredicted variability of the model. The entire process was based

on the concept of regression kriging- where the regression was done by the random forest model. Results show that the downscaling approach was able to achieve better accuracy than the current statistical downscaling methods.

(2) Station network data upscaling

The Texas Soil Observation Network (TxSON) network was designed to test the feasibility of upscaling the in-situ data to match the scale of the SMAP data. I advanced the upscaling method by using the Voronoi polygons and block kriging with a Gaussian kernel aggregation. The upscaling algorithm was calibrated using different spatial aggregation parameters, such as the fishnet cell size and Gaussian kernel standard deviation. The use of the kriging can significantly reduce the spatial autocorrelation among the TxSON stations because of its declustering ability. The result proved the new upscaling method was better than the traditional ones.

(3) Multi-scale data fusion in a spatio-temporal framework

None of the current works for soil moisture statistical downscaling honors time and space equally. It is important, however, that the soil moisture products are consistent in both domains. In this section, the space-time kriging model for soil moisture downscaling and upscaling computation framework designed in the last two sections is implemented to create a spatio-temporal integrated solution to soil moisture multi-scale mapping.

The present work has its novelty in using spatial statistics to reconcile the scale difference from satellite data and ground observations, and therefore proposes new theories and solutions for dealing with the modifiable areal unit problem (MAUP) incurred in soil moisture mapping from satellite and ground stations.

CHAPTER 1. INTRODUCTION

1.1. Background

Soil moisture is important for both atmospheric and water cycles. Along with temperature, and organic matter concentrations, it is a major player for soil respiration and carbon transportation in the environment. Knowledge of soil is important to hydrological and agricultural studies. An oversaturated soil at its maximum water-holding capacity will be more likely to cause a flood than its unsaturated status because of the temporal retaining of rainfall by the soil can drastically flatten the peak flow in the channels. Soil Moisture is an important indicator of near-surface hydrologic conditions and is commonly used for drought monitoring (Oglesby & Erickson III, 1989; Quiring & Papakryiakou, 2003) and climate forecasting (Koster & Suarez, 2001). Soil moisture also influences evapotranspiration and water content available to plants and thus affects the productivity of crops. Therefore it is considered to be an important parameter in climate models (Sheikh *et al.*, 2009).

Currently, drought and flood prediction rely heavily on the humidity in the atmosphere, instead of observing the moisture in soils; yet this is mostly due to the lack of soil moisture data (Entekhabi *et al.*, 2010). Generally, in-situ soil moisture measurement (e.g. gravimetric, time-domain reflectometry, etc.) is at a scale of square meters (point scale), while satellite measurement is at a scale of several square kilometers (Sheikh *et al.*, 2009). However, in-situ measurement is too sparse and irregular in space, and satellite observations are at low resolutions and prone to large measurement errors (Meng & Quiring, 2008; Sheikh *et al.*, 2009; Yee *et al.*, 2016).

The Soil Moisture Active Passive (SMAP) satellite was launched on January 31, 2015. The Level 3 (36-km resolution) soil moisture data from the SMAP satellite are used here. The accuracy of using SMAP to monitor soil moisture content generally displayed a good statistical correlation

between soil climate analysis network (SCAN) data. However, an exceptionally low value during the previous accuracy assessment emphasized the importance of performing further validation (Xu *et al.*, 2018). It may reveal new points about SCAN, SMAP, and the methods utilized, help raise new questions, and provide an opportunity to learn more about the SMAP satellite and its data.

In this research, a soil moisture modeling methodology based on satellite data and ground truth data was proposed. The accuracy of the soil moisture modeling was assessed and compared to the satellite measurement. The drought model was applied to climatology through its correlation analysis with the extreme weather, with the objective to determine a spatial pattern and temporal pattern of drought that might be caused by extreme weather, such as El Nino, or La Nina.

1.2. Objectives

This dissertation aims to reconcile the discrepancy between the scales of satellite soil moisture observation and station soil moisture observation using geostatistical models and machine learning approaches, correct the spatial bias and therefore increase the availability of satellite soil moisture observations through soil moisture modeling. A further objective based on the model is to update the current conventional drought monitoring mechanism by incorporating SMAP Level 3 data and in-situ data throughout the southeastern states.

To achieve this main objective, the following objectives have been set:

(1) Obtain satellite soil moisture products from passive microwave sources such as SMAP, and evaluate its accuracy and uncertainty with ground truth data.

(2) Develop a soil moisture model that assimilates satellite data, probe data, and GIS spatial statistics to make appropriate predictions of soil moisture.

(3) Develop a soil moisture model that upscale station observation data based on GIS spatial statistics methods to make reliable validation of the satellite observations.

(4) Apply the soil moisture observations to an agricultural simulation for the application into agriculture droughts.

1.3. Research Questions

Upon finishing this research, the following questions should be answered:

(1) Is the soil moisture product from SMAP-radiometer accurate everywhere in the study area?

(2) Can in-situ network data match the satellite pixels? If not, can I apply the spatial statistical method to rectify them?

(3) Can I use multi-sensor data fusion/assimilation to improve the SMAP Level 3 product, which have better spatial resolution?

(4) Can I upscale station observation data based on GIS spatial statistics methods to make reliable validation of the satellite observations?

(5) How can I use the soil moisture prediction model into agricultural droughts?

1.4. Contributions and Significance

Soil moisture modeling based on satellite data combines the best feature of in-situ measurement and satellite measurement. It provides a more accurate measurement than satellite-based measurements, with also a much larger coverage than in-situ point measurements. Soil moisture based on statistics can avoid those redundant and ambiguous parameters generally required by physical models, which simplifies the simulation and thus provide a more reliable model.

The dissertation has its novelty in using the spatial statistical method to reconcile the scale difference from satellite data and ground observations, and therefore propose new theories and solutions for dealing with MAUP incurred in soil moisture mapping from satellite and ground stations.

As a result, the integration of soil moisture based on satellite monitoring and enhanced ground stations will increase the end user's water management capabilities in response to drought conditions, and help discover the correlations between extreme climatic events, such as flooding, and soil moisture change.

1.5. Methodology Overview

1.5.1 Data Collection

The primary data component was acquired from the NASA satellite Soil Moisture Active Passive (SMAP). The data products were Level 3 Soil Moisture (SMAP L3-SM-P, and SMAP L3-SM-P-E) data from L-Band Radiometer (NASA, 2014). The Level 3 data were a daily global composite of the Level 2, a soil moisture product based on brightness temperature measurements that were sensitive to soil moisture. The Level 2 accuracy was equal to or better than 0.04 cm^3 (NASA, 2014). SMAP Level 3 products were downloaded from the NASA's Earth Observing System Data and Information System (EOSDIS) Reverb Echo portal on EARTHDATA as GeoTIFFs in WGS 1984 Geographic Coordinate System.

The in-situ data used were the Soil Climate Analysis Network (SCAN), Texas Soil Moisture Observation Network (TxSON), and West Texas Mesonet (WTM). SCAN data retrieved from testing stations in the study area were downloaded from the USDA National Resources Conservation Service (NRCS) as csv files. The sensor was a dielectric constant measuring device, at a depth of 5.08 cm (USDA, 2016). TxSON is an intensively-monitored 36 km (1300 km^2) grid cell soil moisture network that consists of 40 monitoring stations that measure in situ soil moisture and precipitation. TxSON uses nested design to replicate soil moisture at 3, 9 and 36 km satellite pixels in support of NASA's Soil Moisture Active/Passive (SMAP) mission and its Calibration

and Validation Program. WTM is a sparse network that measures soil moisture and other meteorological and agricultural information in West Texas.

Table 1.1. Data Description for Soil Moisture Datasets

Platform & Sensor	Parameter	Use
Soil Moisture Active Passive (SMAP), Passive Radiometer	Soil Moisture, Level-3, 9km/36 km Resolution	Daily measurement of soil moisture Provides direct sensing of soil moisture at 6 a.m. local solar time in the top 5 cm of the soil column in units of cm3/cm3 .
NASA North American Land Data Assimilation System (NLDAS)	Soil Moisture	Historical mean and standard deviation of soil moisture
USDA Soil Climate Analysis Network (SCAN)	Soil Moisture	Used for validation. The soil moisture sensor was a dielectric constant measuring device, at a depth of 5.08cm.
West Texas Mesonet	Soil Moisture	**Soil Water Content (SWC) collected using Campbell Scientific CS615/616 Water Content Reflectometer at 40 stations.**

Specifically, soil moisture data from NASA's North American Land Data Assimilation System (NLDAS) were used as historical reference in models. The NLDAS Noah Land Surface Model L4 Hourly 0.125 x 0.125 degree V002 data were downloaded from Goddard Earth Sciences (GES) Data and Information Center data portal, Mirador(USDA, 2016). NLDAS 36 years of the hourly data files were downloaded in NetCDF format and in the WGS 1984 Geographic Coordinate System. The files were processed to extract the 1200 UTC data for each day over the 36 years (Xia, 2016).

Several types of ancillary dataset were used for the research. DEM was available from USGS Earth Explorer. Topography is an important factor of soil moisture modeling (Meng & Quiring, 2008; Sheikh *et al.*, 2009; Westenbroek *et al.*, 2010). The SSURGO database contains information about soil as collected by the National Cooperative Soil Survey over the course of a century. Examples of information include available water capacity, soil reaction, electrical

conductivity, and frequency of flooding; yields for cropland, woodland, rangeland, and pastureland; and limitations affecting recreational development, building site development, and other engineering uses. Land use and vegetation cover data were collected as ancillary data, for the purpose of an accurate soil moisture upscaling methodology.

Table 1.2. Data Description for Ancillary Datasets

Data/Database	Parameter	Use
DEM (Digital Elevation Model)	Topography	soil moisture modeling
Gross Primary Productivity (GPP)	MODIS MOD17A2H	soil moisture modeling
Land Surface Temperature (LST)	MODIS MOD11A1	soil moisture modeling
NDVI (normalized difference vegetation index)	Vegetation cover	soil moisture downscaling.
Topographical Water Index (TWI)	Calculated from SRTM DEM	soil moisture downscaling.
NLCD (National Land Cover Database)	Land use	soil moisture upscaling.
SSURGO (Soil Survey Geographic Database)	Soil information	water capacity, electrical conductivity and wilting point are important factors for soil moisture modeling

1.5.2 Soil Moisture Downscaling

Soil moisture downscaling aims to improve satellite soil moisture accuracy and extend ground truth soil moisture spatial availability through modeling. SMAP level 3 data as a support layer, ground truth soil moisture data from WTM were collected for this project. All the above datasets were provided in the raster format, therefore rigorous data preprocessing were required before modeling in R.

The modeling methods used in this dissertation include random forest and statistical models. Two methods were used, and hence compared.

1.5.3 Upscaling Sparse In-situ Soil Moisture Data

SCAN in-situ data can provide sufficient observations for upscaling on a scale of 11 states, while on a regional scale of East Baton Rouge Parish, SCAN data are too sparse (Crow *et al.*, 2012), because SCAN data as a point measurement might not be representative for the footprint of SMAP (L3-SM-P 36 *36 km, or L3-SM-P-E 9 * 9 km). To calibrate SMAP data on a regional scale, 78 days of soil moisture observations were collected from TxSON soil moisture network that consists of 40 monitoring stations that measure in-situ soil moisture and precipitation, and upscaled the 15 stations within the 9-km grid to compare with the 9-km SMAP data.

The algorithm used was block kriging - a kriging method in which the average expected value in an area around an unsampled point is generated rather than the estimated exact value of an unsampled point (Burgess & Webster, 1980a). Block kriging is commonly used to provide better variance estimates and smooth interpolated results. In block kriging, the semi-variances between the data points and the interpolated point are replaced by the average semi-variances between the data points and all points in the region, resulting in smaller estimation variances and smoother maps (Burgess & Webster, 1980a).

1.5.4 Analysis with Soil Moisture Index

A Standardized Soil Moisture Index (SSI) was used to simplify the national analysis. Using a relative measurement by calculating the normal conditions would compare current soil conditions to historical averages efficiently. The SSI was calculated with Equation 1-1 for the first two years that SMAP data were available.

$$\text{SSI} = \frac{x_{SMAP} - \mu_{NLDAS}}{\sigma_{NLDAS}} \quad (1\text{-}1)$$

Where x_{SMAP} is the soil moisture content from SMAP Level 3 data for a single day, μ_{NLDAS} is the mean value of soil moisture content for corresponding day, and σ_{NLDAS} the standard deviation of soil moisture content for corresponding day.

Because SMAP data were only available for April 1, 2015 to current, and data were acquired from April 1, 2015 to March 31, 2016, SCAN ground station data (USDA, 2016) were downloaded for this date range and yearly average were calculated for each station (point data). The study area contains 11 states of southeastern United States: Alabama, Arkansas, Georgia, Florida, Kentucky, Texas, Mississippi, North Carolina, South Carolina, Tennessee, and Virginia. Each state has various stations, and 7 observations (stations) from the above states were finally processed for modeling.

The raster calculation tool from ArcGIS was used to calculate the SSI for each day. The SSI indicates how many standard deviations a daily input soil moisture is from the mean, and therefore is an ideal relative measure of the soil moisture content.

Other drought indices to be considered are listed in Table 1.3.

Table 1.3. A Comparison between Drought Indices

Drought Index	Advantage	Disadvantage
PDSI (Palmer Drought Severity index)	Basic data can be used for calculation: precipitation and air temperature for which extensive historical records exist; most effective where impacts sensitive to soil moisture; evapotranspiration and soil moisture are also considered	Limited applicability in locations with climatic extremes, mountainous terrain, or snow-pack unless calibrated; Less transparency because of more sophisticated computation; uses the Thornthwaite method to estimate potential evapotranspiration. Although this index has had wide acceptance, it is still considered an approximation
Keetch-Byram Drought Index (KBDI)	Analyzes precipitation and soil moisture in the water budget model	Use a simple water balance model that evaporation removes water from soil. Used to determine forest fire potential.
Standardized Precipitation Evapotranspiration Index (SPEI)	SPEI was required to overcome the shortcomings of SPI in addressing the consequences of climate change on drought behavior; incorporates temperature data; considers water balance and evapotranspiration	When there are no apparent temporal trends in temperature, SPEI is nearly equivalent to SPI

1.5.5 Time Series Analysis

Time series analysis is a useful technique to character changes in soil moisture over time and to relate these changes to other observations (Albertson & Kiely, 2001; Nash *et al.*, 1991). Autocorrelations and cross-correlations were used to examine the relationships between the soil moisture content and precipitation. Each ground station was checked and nearly three months data were collected to see a daily change.

CHAPTER 2. SPATIALLY EXPLICIT MODEL FOR STATISTICAL DOWNSCALING OF SATELLITE PASSIVE MICROWAVE SOIL MOISTURE

2.1. Introduction

Soil Moisture Active Passive (SMAP) is a polar-orbiting satellite that offers a synoptic view of the earth surface with a combination of radar and radiometer measurements that are associated with the soil moisture content. Microwave satellite observation can provide soil moisture of a surface soil layer. For example, SMAP can sense 5 cm soil layers at the L-band. It is because L-band microwave signals can only penetrate a thin layer of soil (Hong *et al.*, 2016), whereas the penetration depth is of the order of one-tenth of the wavelength (Ulaby *et al.*, 1986). Nonetheless, surface soil moisture is an important link between the land surface and the atmosphere. It is a key component linking energy and water exchanges at the land surface and the atmosphere interface, and therefore an important element in the global circulation process (Hassan-Esfahani *et al.*, 2015; Owe *et al.*, 2001). Due to the lower soil moisture sensitivity of the radar observations and coarse spatial resolution of the radiometer observations (Entekhabi *et al.*, 2010), the soil moisture products from the satellites must have large footprints (i.e. radiometer at 36 km or radar/ radiometer at 10 km). A downscaling is required before the soil moisture products can be applied to hydrological or agricultural studies.

Literature show that soil moisture maps at 1 km are required for agriculture management purposes, and especially important under conditions of extreme drought (Colliander *et al.*, 2017a; Fang *et al.*, 2018; Montzka *et al.*, 2018; Peng *et al.*, 2017). There are four types of soil moisture downscaling methods: physical model-based methods, multi-source data merge methods, Kalman-filter data assimilation, and statistical model-based methods. Merlin *et al.* (Merlin *et al.*, 2008) used a physical-based soil evaporative efficiency model to downscale the 40-km resolution SMOS

data by using the 1-km resolution Moderate Resolution Infrared Spectroradiometer (MODIS) brightness temperature (Merlin *et al.*, 2008). Merlin *et al.* (Merlin *et al.*, 2012) further included soil temperature data to develop the Disaggregation based on Physical And Theoretical scale Change (DisPATCh) algorithm. However, the DisPATCh algorithm was defined as "semi-physical" due to the limitation of remote sensing in direct estimation of the physical properties of soils (Merlin *et al.*, 2012). Remote sensing data fusion methods are typically those using active (radar) to improve the passive (radiometer) microwave images. Several algorithms such as the change detection method (Njoku *et al.*, 2002) and the Bayesian merging method (Zhan *et al.*, 2006) have been proposed to merge radiometer and radar data to provide high-resolution soil moisture data. Das *et al.* (2011) improved the change detection algorithm and developed a method that was adopted as the baseline algorithm for the SMAP combined active/passive soil moisture product. Piles *et al.* (2009), and Akbar and Moghaddam (2015) found that the active-passive disaggregation algorithm presented much more spatial details, and the average soil moisture retrieval error outperformed active or passive only methods. Das *et al.* (2011) showed another approach of data fusion-based soil moisture downscaling by merging the brightness temperature data with the radar-backscatter data to retrieve high-resolution soil moisture from the downscaled brightness temperature. However, it is difficult to obtain the passive and active microwave observations at the same time, and the time lapse affected the data fusion results (Peng *et al.*, 2017). The SMAP satellite is designed to obtain soil moisture data from an L-band radar and an L-band radiometer. However, due to an unrepairable hardware failure six months after the launch in January 2015, the paired image product discontinued. Data assimilation downscaling approach aims at optimizing soil moisture prediction based on both in situ and remotely sensed soil moisture observations. For example, Reichle *et al.* (2001) integrated the uncertainties of model prediction and observation in

a data assimilation model. The model was able to make a reasonable prediction based on the calibrated physical model of soil moisture and energy flux. However, not only was the model difficult to implement, but also the computation cost was high. To this end, many researchers attempted to improve the efficiency of data assimilation systems. For example, Crow and Yilmaz (2014) proposed the Auto-Tuned Land Assimilation System (ATLAS) as an efficient solution for land data assimilation systems. Whereas data assimilation could reduce the observation errors of soil moisture, there has always been a compromise between realistic physical representations and computational feasibility, and yet this type of methodology requires the ground measurements as well as parameters of the physical processes (Reichle *et al.*, 2008; Reichle *et al.*, 2001). Statistical downscaling has been widely adopted in soil moisture mapping in recent years. Proxy environment variables that are related to soil moisture are acquired at a relatively finer scale. A model using the environmental variables can estimate the spatial distribution of soil moisture observation from the satellites at higher resolutions (Park, 2013). Land surface temperature and vegetation data are proven as effective proxy environment variables (Colliander *et al.*, 2017a; Fang *et al.*, 2018). Land surface models were used in the literature to explain the environment variables at the finer scales and soil moisture observed from the satellites at coarse scales (Atkinson, 2013; Kerry *et al.*, 2012; Kyriakidis, 2004; Wang *et al.*, 2015b). For statistical downscaling, the change of support in the statistical models leads to the downscaling or upscaling of the data. However, such change should be justifiable, or otherwise, it might subject to the modifiable areal unit problem (MAUP, (Openshow, 1979)), which could fundamentally invalidate the statistical methods. The other problem with the current statistical models is spatial autocorrelation in the observations. Regression models calibrated from correlated observations could be biased if the autocorrelation was not considered as part of the regression model. Local regression models such as geographically

weighted regression (GWR) were used to construct the statistical downscaling model (Jin *et al.*, 2018). However, Wheeler and Tiefelsdorf (Wheeler & Tiefelsdorf, 2005) pointed out high-level local multicollinearity could be observed in GWR. Griffith (Griffith, 2008) also concluded that whereas GWR accounts for some of the spatial autocorrelation by transferring it to the spatially varying coefficients, the presence of positive spatial autocorrelation could not be neglected. Therefore, although GWR and kriging were used in existing research to downscale the soil moisture, the results still tends to have a large bias. Furthermore, the GWR model does not find the trend component because of the use of localized samples. A regression kriging based on GWR (Jin *et al.*, 2018) was not appropriate because the second-order stationarity assumption by kriging requires there is no trend in the data.

I propose a spatially explicit statistical downscaling model (SESD) by introducing spatial computation components to the regression model based on the random forest machine learning method. Although linear regression models were quite common in the existing research (Loew & Mauser, 2008; Park, 2013; Piles *et al.*, 2014), the multicollinearity problem will decrease the model performance when there exists correlation among the geospatial variables, which is not unusual (Fotheringham & Oshan, 2016). Machine learning algorithms have been developed to solve the non-linearity and multicollinearity problems (Im *et al.*, 2016; Kwon *et al.*, 2018; Liu *et al.*, 2018; Srivastava *et al.*, 2013; Xing *et al.*, 2017) since a random subset of features is chosen for each split of a random tree in a forest (Breiman, 2001; Segal, 2004; Svetnik *et al.*, 2003). The random forest predicts the large-scale variation in the soil moisture. The un-modeled small-scale variation is then mapped using a regression kriging so that the data comparison across multiple scales becomes reasonable.

2.2. Study Area and Datasets

The study area is located in north Texas, which has a relatively dense soil moisture sampling network. There are 40 stations in the study area published as the West Texas Mesonet (Figure 2.1).

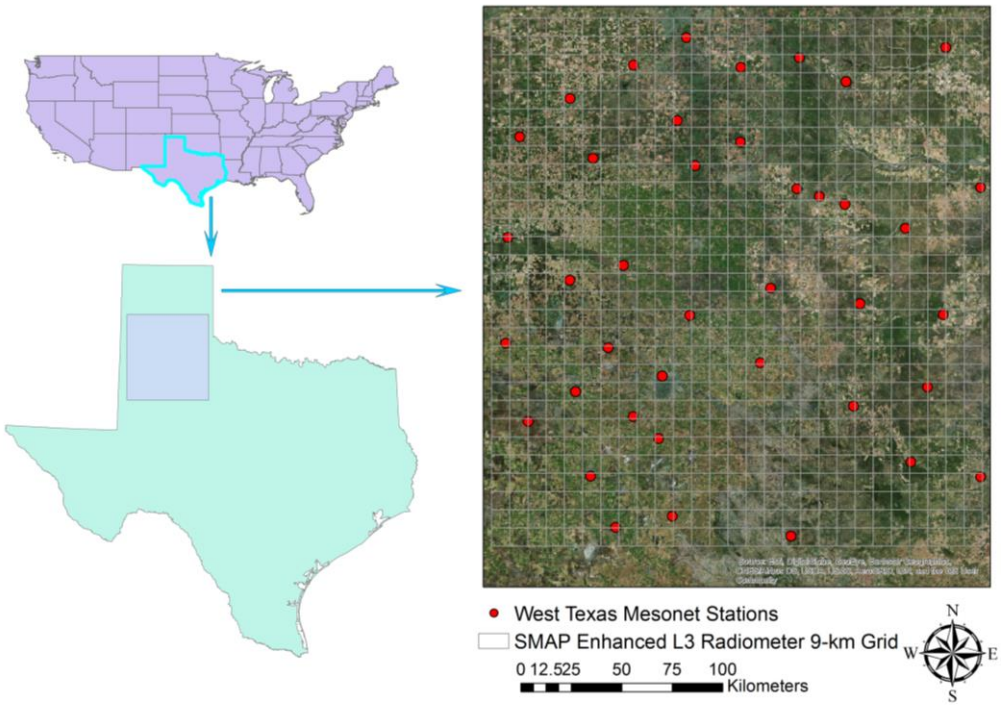


Figure 2.1. Study Area with West Texas Mesonet Stations

Ground observation data collected from West Texas Mesonet were measured at a depth of 5 cm from 40 stations. Twenty-four stations were randomly assigned as training stations for model calibration and 16 as testing stations for validation. Satellite observation soil moisture data products were collected from SMAP Enhanced L3 Radiometer Global Daily 9 km EASE-Grid Soil Moisture V001 (SMAP L3-SM-P-E), downloaded as the Level 3 products from the NASA EARTHDATA. The SMAP satellite was launched on January 31, 2015 (Entekhabi *et al.*, 2010). SMAP provides direct sensing of soil moisture in the unit of cm^3/cm^3 in the top 5 cm soil layers at 6 a.m./p.m. local solar time. The SMAP satellite has a payload comprised of an L-band radar (VV,

HH, and HV polarizations) and an L-band radiometer for soil moisture retrieval at areas with moderate vegetation cover. However, due to an unrepairable hardware failure on July 7, 2015, the radar/radiometer fusion product at the 10-km resolution and the radar-only product at the 3-km resolution were officially discontinued (Chan *et al.*, 2016). The accuracy of the radiometer-only (36-km resolution) soil moisture data generally displayed a good statistical correlation with Soil Climate Analysis Network (SCAN) data, according to past research (Xu *et al.*, 2018). The SMAP Enhanced Level-3 Radiometer Global Daily 9 km EASE-Grid Soil Moisture product was based on the Backus-Gilbert (BG) optimal interpolation technique (Poe, 1990) applied to the original standard brightness temperature (TB) data. The BG optimal interpolation takes advantage of the overlapped radiometer footprints on its orbits. Standard correction/calibration procedures were done on the L1B_TB product to produce the SMAP Level 1C Enhanced Brightness Temperature (L1C_TB_E) Product, which was then fed into the current operational SMAP baseline soil moisture retrieval algorithm to produce L2_SM_P_E as the final output (Colliander *et al.*, 2018b). It was reported that the enhanced SMAP soil moisture products at 9-km spatial resolutions displayed an overall 0.039 m^3/m^3 unbiased root mean square deviation (Colliander *et al.*, 2018b). In this study, the a.m. data were used because the thermal gradient along the vertical soil profile is often more uniform in the morning (Peng *et al.*, 2017).

The following geospatial datasets were selected for downscaling the SMAP data: digital elevation model (DEM), land surface temperature (LST), normalized difference vegetation index (NDVI), and gross primary productivity (GPP). These are important explanatory variables for soil moisture modeling (Fang & Lakshmi, 2014; Meng & Quiring, 2008; Mohanty *et al.*, 2017; Peng *et al.*, 2017). The topographical wetness index (TWI) was also incorporated into the statistical model (Ågren *et al.*, 2014; Beven & Kirkby, 1979b). TWI is essentially landform-based calculator

for soil water content concentration (Beven & Kirkby, 1979a; Wilson & Gallant, 2000). Because the slope is a local calculator, I created the TWI from the high resolution (30 m) DEM and aggregated it to the target resolution (1 km) to match other datasets (Table 2.1). (Beven & Kirkby, 1979a; Wilson & Gallant, 2000). Because the slope is a local calculator, TWI was computed from the high resolution (30 m) DEM and aggregated it to the target resolution (1 km) to match other datasets (Table 2.1).

Table 2.1. Datasets Description for Soil Moisture Statistical Downscaling

Data Type	Data	Sensor/Product	Temporal Coverage	Sample Size/Pixel Size
Ground Observation Data	Soil Water Content (SWC) from West Texas Mesonet	Campbell Scientific CS615/616 Water Content Reflectometer	15-minutes 20 October 2017	40 stations
Satellite Observation Data	SMAP Enhanced L3 Radiometer Global Daily 9 km EASE-Grid Soil Moisture V001	SMAP L3-SM-P-E	Daily 20 October 2017	9 km
	Land Surface Temperature (LST)	MODIS MOD11A1	Daily 20 October 2017	1 km
	Normalized Difference Vegetation Index (NDVI)	Advanced Very High Resolution Radiometer (AVHRR)	Weekly 17 to 23 October 2017	1 km
	Gross Primary Productivity (GPP)	MODIS MOD17A2H	8-Day 16 to 25 October 2017	500 m
	Digital Elevation Model (DEM)	USGS GMTED2010		1 km
	Topographical Water Index (TWI)	Calculated from SRTM DEM		30

2.3. Methods and Data Processing

2.3.1 Design of the Multiscale Soil Moisture Model

It is assumed that the spatial structure of soil moisture distribution follows the general definition

in geostatistics (Krivoruchko, 2011) (page 290):

$$SM = trend + small\text{-}scale + micro\text{-}scale + error \quad (2\text{-}1)$$

where the true soil moisture (*SM*) is the sum of the trend component that can be predicted from a regression model, the *small-scale* variation and *micro-scale* variation that were not captured by the regression model, and the random *error*. Therefore, the model is comprised of two segments: the regression model supported by a random forest algorithm and a regression kriging model, which is based on the residual of the regression model to reconstruct the small-scale and micro-scale components. The first segment is similar to other statistical downscaling works except for the use of the random forest. The second segment is the innovative part of our approach, which will be compared to the result of the first segment only model to demonstrate the superiority of our model.

2.3.2 Overall Data Flow and Algorithms

Statistical downscaling incorporates fine-resolution geospatial data to downscale the coarse resolution SMAP data. As shown in Figure 2.2, the coarse resolution SMAP soil moisture product was resampled to the 1-km resolution using a Gaussian spatial filter as spatial weights to determine the pixel values at the edge between adjacent 9-km pixels. Then, the machine-learning algorithm - RF - was used to derive statistical relationships between the explanatory geospatial data and the SMAP soil moisture at the 1-km scale. The calibrated RF model is able to make a prediction (downscaled) of soil moisture using the geospatial datasets at the 1-km resolution.

After the statistical downscaling, the unexplained data variability is treated as the small-scale and micro-scale component of soil moisture in the geostatistical model. This is equivalent to the regression kriging, whereas the residuals from the regression of the last step compared with the ground observations from West Texas Mesonet stations are modeled by simple kriging.

According to equation (2-1), the true soil moisture signal was reconstructed by adding the small-scale and micro-scale components from the simple kriging to the trend component modeled by the regression.

The validation used the unbiased ubRMSD, root mean square deviation (RMSD) and bias of the test set separated from the calibration set from the West Texas Mesonet (WTM) station data. To overcome the problem of the low number of ground stations for validation, Monte-Carlo cross-validation was used to randomly split the 40 stations to 24 training sites and 16 validation sites. The simulation was repeated 100 times to make a reliable assessment of the model accuracy.

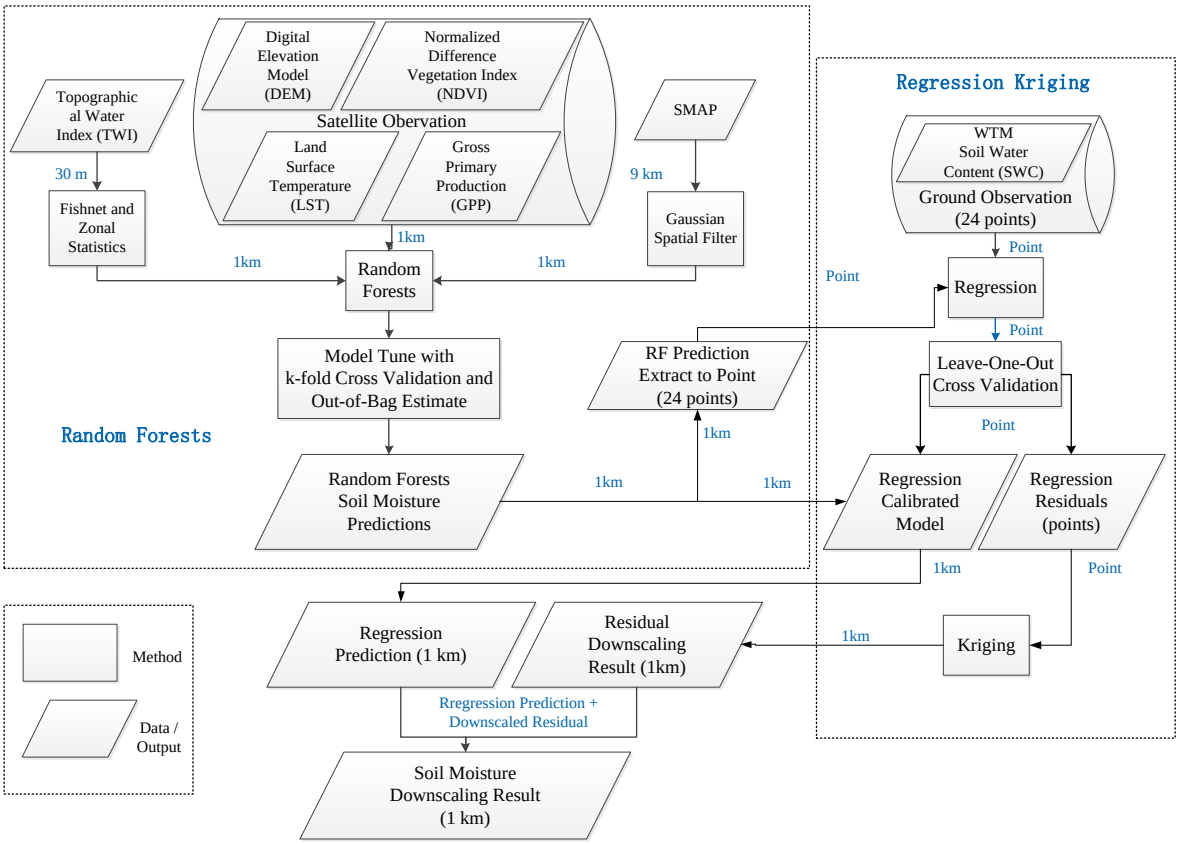


Figure 2.2. Research Scheme for the Spatially Explicit Model

2.3.3 1-km Fishnet as the Support of the Statistical Analysis

The support of the downscaling is selected as the 1-km fishnet. The reason is 1 km is an acceptable unit size to most regional and continent studies. Many global satellite products, such as MODIS land surface characteristics and Advanced Very High Resolution Radiometer (AVHRR) normalized difference vegetation index (NDVI) are provided at the 1-km grid spacing. All the geospatial datasets were aggregated to the 1-km fishnet grid. The statistics were calculated, and models were constructed on the fishnet structure. This explicit spatial structure enables the use of many spatial statistical methods. The fishnet grids were aligned with the 9-km pixel boundaries to ensure the enclosure of 9 by 9 fishnet grids within each 9-km SMAP soil moisture cell.

2.3.4 Geospatial Data Processing

The aforementioned geospatial data products, including digital elevation model (DEM), gross primary productivity (GPP), land surface temperature (LST), normalized difference vegetation index (NDVI), and topographical wetness index (TWI) at a fine scale are used as predictors for modeling soil moisture data. Data processing includes Gaussian kernel smoothing filter applied to SMAP, TWI calculation, satellite products reprojection and resampling. Random forest is then trained to derive statistical relationships between Gaussian kernel filtered SMAP and auxiliary geospatial data at the 1-km fishnet structure.

Decomposing the 9 km soil moisture to 1 km

The SMAP soil moisture product at the 9 km resolution was still inconsistent with the 1-km support of the statistical model. The first step is to decompose the 9 km product to 1 km units. The decomposition assumes the 9 km product is a weighted average of the soil moisture at 1 km, which is unknown and to be predicted by the model. In the first step of our model, the goal is to predict the trend component of the true soil moisture. The trend component is defined as the large-

scale variation that can be represented as a certain polynomial equation (first order, second order, etc.). The trend component must be continuously differentiable (smooth). The raw 9-km product, however, does not meet the requirement. Therefore, to approximate the soil moisture trend, the 9-km soil moisture was decomposed to the 1 km fishnets by spatial convolution. The spatial convolution weights were modeled as the 2D Gaussian kernel function, which is higher if the location is closer to the center of the 9 km square unit, and vice versa. Gaussian kernel smoothing is an image filtering technique widely used in signal detection, noise removal, and feature detection (Chung, 2012; Lopez-Molina *et al.*, 2013; Wink & Roerdink, 2004):

$$Y(t) = \int K(t,s)X(s)\,ds \tag{2-2}$$

where K is the kernel of the integral. Given the input signal X, Y represents the output signal. The smoothness of the output depends on the smoothness of the kernel (Chung, 2012). The n-dimensional isotropic Gaussian kernel is defined as the product of n 1D kernels. Let t = (t1, · · · , tn)' ∈ Rn. Then the n-dimensional kernel is given by:

$$K\sigma(t) = K_\sigma(t_1)K_\sigma(t_2)\cdots K_\sigma(t_n) = \frac{1}{(2\pi)^{n/2}\sigma^n}\exp\left(\frac{1}{2\sigma^2}\sum_{i=1}^{n} t_i^2\right) \tag{2-3}$$

The Gaussian kernel was adapted for decomposing the 9-km soil moisture product from SMAP to the 1-km fishnet (Figure 2.3). Without Gaussian smoothing, the 1-km soil moisture values form steps at the boundary of each 9 km SMAP pixel. The Gaussian filter replaced the step-like edges with gradual values, which can better present the trend component of the true soil moisture. (Figure 2.3b).

(a) (b)

Figure 2.3. Gaussian Kernel Applied to SMAP: (a) 9-km Original SMAP data, (b) 1-km data with Gaussian Kernel Applied to SMAP, the step-like edges in the original data were replaced with gradual values.

TWI Calculation

The TWI is an important predictor for soil moisture modeling. TWI is a local index because it is very sensitive to the micro-scale landforms (Beven & Kirkby, 1979a). Therefore, TWI must be computed at the landform scales (i.e. ~30 m). In the model, the TWI was derived from the 30-m DEM using the equation:

$$TWI = \ln(\frac{a}{\tan(\beta)}) \quad (2\text{-}4)$$

where a is the specific upslope catchment area, which is computed as A/L ($m^2\ m^{-1}$, and A is the upslope area and L is the contour length); β is the slope gradient (in degrees). This equation assumes uniform soil properties (i.e., the soil transmissivity is constant throughout the landscape). Then, the TWI was averaged within the 1-km fishnet unit range. The GPP variable was originally at the 500 m resolution. It was also aggregated to 1 km using the bilinear resampling method.

2.3.5 Random Forest Machine Learning with Geospatial Data

Random forest improves the prediction of tree-based models using randomization and ensemble voting (Breiman, 2001). In standard regression tree models, each node uses the best split variable

among all variables. In a random forest, each node was split by the best variable among a subset of variables randomly chosen at that node, rather than all variables (Liaw & Wiener, 2002). With such a strategy, overfitting was not a problem anymore (Breiman, 2001). Actually, each tree in a random forest was intendedly over fit to achieve the best prediction ability. These design features make random forest one of the most effective machine learning methods (Svetnik *et al.*, 2003). Random forest algorithm can be described as below (Liaw & Wiener, 2002; Segal, 2004):

1) Draw ntree bootstrap samples from the original data.

2) For each of the bootstrap samples, grow an unpruned regression tree, with the following modification: Specify m < p (the number of covariates). At each node of every tree, random forest algorithm randomly selects m variables out of all M possible variables (independently for each node) and find the best split from the randomly selected m variables.

3) Predict new data by aggregating the predictions of the ntree trees (average for regression).

4) In addition to excellent prediction performance, random forest also measure variable importance.

The parameter *m* of the random forest models was tuned using package 'caret' in R for the best model performance. Random forest algorithm grows many trees. For a complex model, large *m* usually means high correlation and small bias. However, when $m = p$, the random forest algorithm becomes bagging. Therefore, the model is tuned to find the best *m* that has the higher correlation and lowest bias. Cross-Validation (CV) is a common approach for evaluating the robustness and accuracy of the models when choosing the final tuning parameters in machine learning (Colliander *et al.*, 2017c; Segal, 2004). A random forest possesses a special cross-validation. Cross-Validation (CV) is a common approach for evaluating the robustness and

accuracy of the models when choosing the final tuning parameters in machine learning (Colliander *et al.*, 2017c; Segal, 2004). A random forest possesses a special cross-validation. When the training set for the current tree is drawn by sampling with replacement, about one-third of the cases are left out of the sample (Breiman, 2001). These data left out are “out-of-bag" (“OOB") for those trees. In a random forest, the predictions of OOB samples from these trees provide an estimate of generalization error on new data, which act as a counterpart of n-fold CV. This study used both n-fold CV as well as OOB for the model performance assessment. The model with the highest R-squared and lowest RMSD was considered as the most optimal. The 10-fold CV showed that when m = 3, the model yielded the best R-squared (0.7255) and RMSD (0.0163), and the OOB result also showed that when m = 3, model yielded the best R-squared (0.7257) and RMSD (0.0163). The results do not indicate a significant difference between the two models (Table 2.2).

Table 2.2. Best M for Cross Validation and Oob Estimate

	10-fold CV		**OOB**	
m	RMSD	R-squared	RMSD	R-squared
2	0.0163	0.7250	0.0163	0.7249
3	0.0163	0.7255	0.0163	0.7257
5	0.0164	0.7221	0.0163	0.7229

2.3.6 Regression Kriging with Ground Data

Based on the regression prediction of the trend component in Equation 2-1, the regression kriging approach was then used to model the small-scale and micro-scale variations and remove the noise component. The residuals of between the random forest prediction and the West Texas Mesonet ground station data were interpolated using a kriging model to create a 1-km resolution residual raster.

A total of 24 stations were selected randomly to compare with the model prediction. The residual was modeled by an exponential semi-variogram equation:

0.0015014 * Nugget + 0.0024994 * Exponential (204000) (2-5)

The semi-variogram is shown in Figure 2.5.

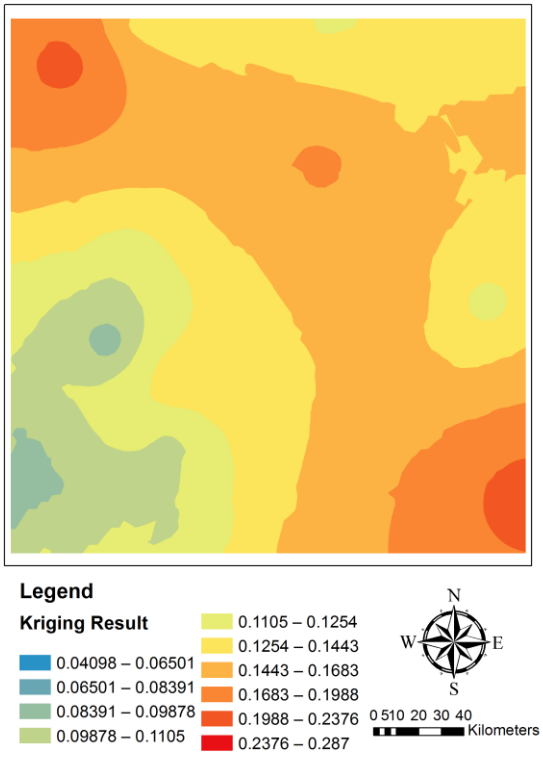


Figure 2.4. Map of Residuals from Regression Kriging

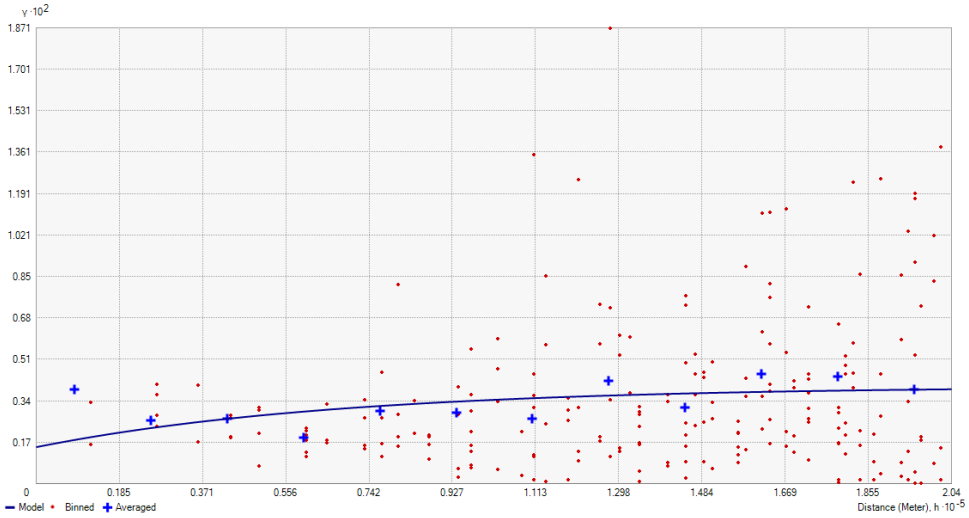


Figure 2.5. Semivariogram for the Kriging Model

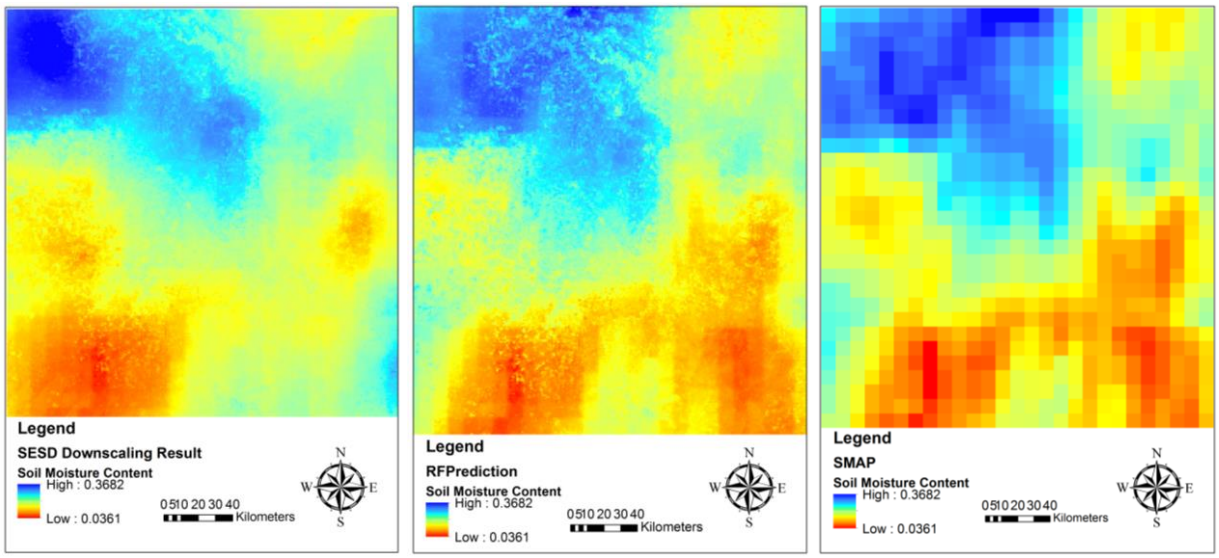
2.3.7 Downscaled Soil Moisture Data Validation

As only 40 West Texas Mesonet stations are included, one critical mission of the model validation is to utilize this limited number of soil moisture stations to obtain an unbiased correlation validation metric for satellite surface soil moisture retrieval products.

To overcome the low sample problem, a Monte-Carlo CV method is adopted. For each of the 100 iterations, 24 stations were randomly chosen for the kriging model construction and the other 16 stations were used for validation. In each iteration, three metrics were computed each time: root mean square deviation (RMSD), unbiased RMSD (ubRMSD), and bias. RMSD is widely used in accuracy assessment, yet it might not be accurate if biases exist in either the mean or the amplitude of fluctuations in the retrieval. In this situation, the ubRMSD can be used to reflect the RMSD of soil moisture anomalies that are computed by removing the mean of soil moisture (Piles *et al.*, 2014).

2.4. Results

Figure 2.6 shows the map of the downscaled soil moisture using the SESD method compared with the random forest-only method, as well as the SMAP data at 9-km resolution. The comparison shows that the SESD method significantly improves the spatial details of SMAP soil moisture. A further validation is to check the accuracy of the method.



(a) (b) (c)

Figure 2.6. Downscaled Soil Moisture Map and Original Map: (a) downscaled result with SESD, (b) downscaled result with random forest, c. SMAP L3 at 9 km

The 100-iteration average for the downscaled soil moisture with SESD method, downscaled soil moisture with random forest, as well as the original SMAP data were compared with the station data (Table 2.3). Table 2.3 lists the RMSD, ubRMSD, and bias for the validation results.

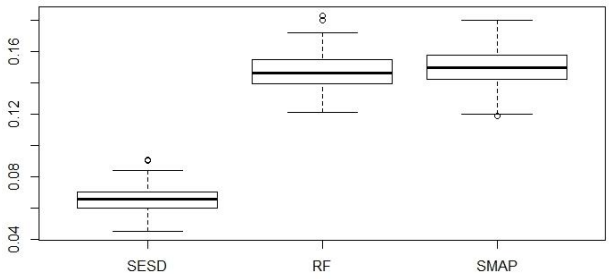
Table 2.3 shows that the RMSD for the 9-km SMAP data was 0.1500 cm^3/cm^3, which is much higher than our downscaled soil moisture RMSD (0.0656 cm^3/cm^3). This implies that the downscaling improved the comparability of the soil moisture product from the satellite and the

ground stations. Nevertheless, it is important to acknowledge that the accuracy here was still higher than the average accuracy of the SMAP enhanced soil moisture product (0.039 cm^3/cm^3, ubRMSD, reported by (Colliander *et al.*, 2018a)). The accuracy could be higher in some certain conditions, such as low vegetation coverage, urban fraction ≤ 0.25, water fraction ≤ 0.1, and DEM slope standard deviation ≤ 3 degrees, etc. (Colliander *et al.*, 2017c). None of these were included as prerequisites in the model because it would limit its generalization ability. Therefore, downscaled soil moisture data here are reasonably accurate given the complex environment of the study area.

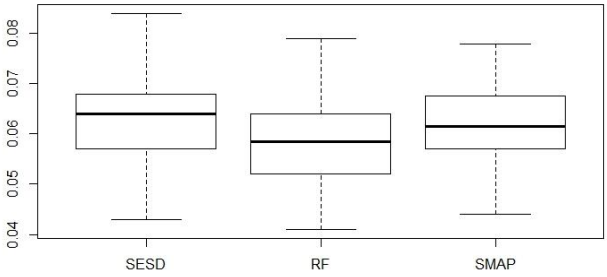
The box plot based on the 100 iterations of CV is shown in Figure 2.7. Overall, the SESD method proposed in this study performed better, based on its low RMSD (a) and bias (c) on average. The ubRMSD (b) plot was comparable to the SMAP data, indicating the bias component has been removed by the downscaling process.

Table 2.3. Average Accuracy Matrices for the Downscaled Results in Comparison with Random Forest-Only Method and Original SMAP Data

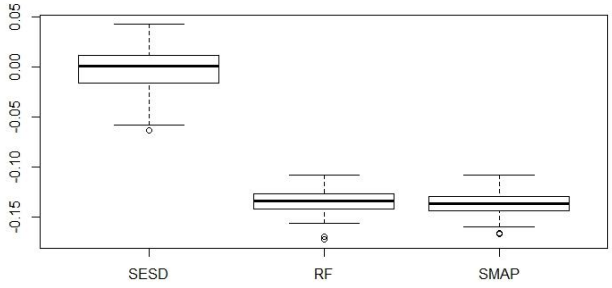
	Downscaled Result with SESD	**Random Forest-only Method**	**SMAP L3 At 9 Km**
RMSD	0.0656	0.1467	0.1500
ubRMSD	0.0627	0.0586	0.0617
bias	-0.0011	-0.1342	-0.1366



(a)



(b)



(c)

Figure 2.7. Box Plots for the SESD Downscaled Soil Moisture Map, Competed with Random Forest and Original Map: (a) RMSD, (b) ubRMSD, (c) bias

2.5. Discussion

2.5.1 Variable Importance

Random forest measures relative variable importance by using a permutation approach. According to literature (Breiman, 2001), the squared relative importance of variable X is the sum of the

estimated squared improvements (weighted by node size) over all internal nodes for which it was chosen as the splitting variable. Other than relative variable importance, OOB samples are also used by random forest to estimate relative variable importance. The principles are that when the *b*th tree is grown, the OOB samples are passed down the tree, and the prediction accuracy is recorded. The values for variable *X* in the OOB samples are randomly permuted (values randomly shuffled) to record the prediction accuracy for a second time. The decrease in accuracy is averaged over all trees and is then used as a measure of the importance of variable *X*. The advantage of this permutation approach is that the variable importance is often more uniform over the variables.

According to the OOB variable importance analysis, the LST is the most important predictor for the soil moisture model, which agrees with Fang et al. (2018), and Colliander et al. (2017a), whereas NDVI is the least important (Table 2.4). However, the random forest relative variable importance gives a different scenario, with DEM as the most important predictor. When the two standards do not agree with each other, OOB estimate will be the most robust and informative measure, because it is the increase in MSE of predictions (estimated with OOB-CV) as the result of variable *X* being permuted.

As both methods suggested that NDVI was the least important predictor, the effect of removal of NDVI from the prediction model was tested. The model comparison suggested that although NDVI was the least important predictor, the model performed better when NDVI is included (R-squared = 0.7257 versus 0.7058, and RMSD = 0.0162 cm^3/cm^3 versus 0.0168 cm^3/cm^3 when m =3, Table 2.5). Therefore, NDVI remained in the model.

Table 2.4 also shows that TWI played an important role in the machine learning algorithm. TWI is sensitive to micro-scale landform variation and therefore effective in the prediction for soil water content concentration.

Table 2.4. Variable Importance by Oob Estimate and Relative Variable Importance

Importance	IncMSE (OOB Estimate)		IncNodePurity (relative variable importance)	
1	LST	622.5	DEM	25.90
2	DEM	547.4	LST	19.29
3	TWI	467.3	TWI	7.55
4	GPP	236.5	GPP	4.30
5	NDVI	204.4	NDVI	3.06

Table 2.5. Model Comparison With/Without NDVI

	With NDVI		Without NDVI	
m	RMSD	R-squared	RMSD	R-squared
2	0.0162	0.7249	0.0168	0.7076
3	0.0162	0.7257	0.0168	0.7058
4	0.0163	0.7229	0.0169	0.7033

2.5.2 RMSD or ubRMSD

The difference between RMSD and ubRMSD is the bias component. The RMSE is comprised of bias, variance, and irreducible errors, while ubRMSD only includes variance and irreducible errors. If the bias were not treated, there would be an inconsistency between RMSD and ubRMSD. Previous works, such as (Colliander *et al.*, 2017a; Colliander *et al.*, 2017c) used ubRMSD as the error assessment index because the existing models were not able to remove the bias between the satellite measurement and the ground truth data. Table 2.3 showed the RMSD of the SMAP data and the random forest model were almost double of their ubRMSD. Therefore, ubRMSD cannot explain the uncertainty distributed in the spatial data. Colliander et al. (Colliander *et al.*, 2017c) showed that in the time series, the ubRMSD and RMSD of the testing sites using the weighted average were almost identical. This suggests that the soil moisture product of SMAP has little bias in the time domain but contains a large bias in the spatial domain. If the bias in the spatial domain was not removed, the spatially explicit models, such as watershed hydrological simulate models, will be biased by the soil moisture data input.

The un-modeled bias component in the SMAP soil moisture product and the random forest

product were due to the across-scale inconsistency with the ground station data. The bias, which includes both small-scale and local-scale variation, was modeled in the kriging semi-variogram and interpolation. Therefore, the bias component of soil moisture in the spatial domain was removed by the SESD model. The proof is the consistency of ubRMSD and RMSD of the SESD model.

2.5.3 Factors That Impact the Model

The in-situ soil moisture data may be affected by outliers. For example, the Hart 3N station had a problem when irrigation water flooded into the station and had not been working correctly afterward; Memphis station had new water content reflectometers, and the readings needed to be recalibrated for proper comparison with other stations. Because these outliers could affect model calibration and prediction accuracy, and because these problems were known a priori, the data with known problems were removed. In addition, Monte-Carlo CV was employed to reduce the impact of those outliers that were not handled in the data preprocessing.

High-density ground soil moisture network would help improve the downscaling accuracy. Thus, ground truth data were obtained by the state-of-the-art sensor network, even though there are only 40 stations covering the study area. The purpose of the downscaling research is to make use of the satellite observations to obtain soil moisture measurement comparable to ground station data. Results show that the regression kriging method is a viable solution to the fusion of multi-scale soil moisture data. The SESD model was able to achieve desirable accuracy with no bias.

2.5.4 Time Series versus Spatial Distribution

The SMAP satellite covers the study area on a descending (a.m.) or ascending (p.m.) node every three days (8-day exact repeat). That enables near daily (including both am and pm) update of the soil moisture information with strong coherent in both space and time. However, between time

and space, soil moisture is highly sampled in time but scarcely sampled in space. Most of the previous works focused on how to remove the bias in the time series but neglected the bias in the spatial domain (such as the use of ubRMSD). Previous works have demonstrated that the SMAP data had very little bias when validated in the time series (Colliander *et al.*, 2017c). It can be assumed that the bias in the time domain should not be a concern of the research. Therefore, this research focused more on the computation in the spatial domain and introduced a spatial explicit downscaling model. In addition, this research only used the validation data sampled over the spatial domain to study the variable importance and accuracy, while most previous works relied on time series data for accuracy assessment. It is argued that since there is strong autocorrelation in time series data and the accuracy assessment requires the samples to be acquired independently, the accuracy assessment that included consecutive days may have inflated the accuracy and confidence.

On a side note, this study did not consider rainfall as one of the factors that impact the soil moisture distribution, as the study area is within an arid climate zone. However, for the moist climate zones, under the circumstances of heavy rainfall, precipitation might be a dominant variable that can control the soil moisture distribution. I will further explore this pattern in future research.

2.6. Conclusions

A spatially explicit statistical downscaling (SESD) method was proposed using random forest and regression kriging for soil moisture downscaling based on satellite passive microwave observations, public domain geospatial data, and limited ground observations. The validation through West Texas Mesonet showed that our downscaling method significantly improved the spatial details of SMAP soil moisture and achieved higher accuracy than the original SMAP Level

3 data. Compared with the reference data and random forest-only method, the SESD method-based downscaled soil moisture product was accurate (RMSD = 0.0656 cm^3/cm^3, ubRMSD = 0.0627 cm^3/cm^3, and bias = -0.0011), given the study area include both developed and vegetated area.

The machine learning result showed that LST, elevation, and topographic moisture conditions are the most effective predictors that impact soil moisture content. This study confirmed that through using geospatial data, in-situ data, and spatial models, soil moisture product resolution and accuracy covering large domains can be improved. I hope the work helps improve the usability of the radiometric-only SMAP data.

CHAPTER 3. SOIL MOISTURE UPSCALING USING THIESSEN POLYGON-BASED BLOCK KRIGING

3.1. Introduction

Soil moisture plays an important role in both atmospheric circulation and water cycle. Soil moisture is the key component of understanding global and regional climate change impact on the agriculture and hydrological applications. Currently, droughts prediction relies more heavily on observing the moisture levels in the atmosphere than observing the moisture levels of soil; yet this is mostly due to the lack of soil moisture data available. Having continuous soil moisture measurements will improve crop yield forecasting, and irrigation planning. The status of soil is vital to environmental studies. Not only that the capacity of surface soils to store water (surface moisture) can be a good indicator of the surface stability for embankment slopes with respect to the sliding failures, but also oversaturated soil in the levees is prone to collapse when there is a large amount of water pushing on the levee. From the environmentalist's perspective, when the soil is saturated, seepage and infiltration is barely possible, triggering the surface runoff, which will increase the potential of flooding to occur. What makes it worse is that in the fertilized farmland, fertilizer running off to the water bodies will increase the nitrogen and other nutrients, and therefore provide an ideal environment for algae blooms. When this fertilizer runoff was cut off later, the decomposition of algae blooms will absorb the oxygen in that water body, and therefore cause a series of environmental problems. Therefore, knowing the soil status, especially the soil moisture status, is critical to reducing such environmental issues.

Typically, two methods can be used to measure soil moisture: station and satellite. Generally, in situ soil moisture measurement methods (e.g. gravimetric, TDR, etc.) are observed at a scale of meters (point scale), while satellite measurements are observed at a scale of several square kilometers (pixel size) (Sheikh *et al.*, 2009). In other words, soil moisture observations

from the satellite and ground are inconsistent in their scales. Study showed that soil moisture varies significantly from small scales (< 10 m) to field scale and larger (> 1 km) scale (Famiglietti *et al.*, 1999). Current statistical models that tried to link these two types of observations directly in model calibration usually generate calibrated models under the condition of modifiable area unit problem (MAUP): the statistical models calibrated at one scale do not usually work at other scales when individual samples (ground) are aggregated (satellite).

There are two approaches toward matching the scales between satellite and station soil moisture datasets: downscaling (Colliander *et al.*, 2017a; Knipper *et al.*, 2017; Piles *et al.*, 2014) and upscaling (Crow *et al.*, 2012; Crow *et al.*, 2005; Wang *et al.*, 2015a). Soil moisture downscaling aims to downscale the soil moisture products from satellites. This operation can bring down the soil moisture from a coarse spatial resolution (i.e., ~10 km) to a medium spatial resolution (~1 km, typically), which significantly decrease the bias caused by the scale difference between the two types of observations. However, as a station can only measure the soil moisture at a local scale of square meters, it does not represent the large footprint measured from the satellite, even if it is downscaled to 1-km scale. Direct comparison between these two types of measurement will oftentimes cause large bias.

Upscaling involves the operation that upscale the points level to match the satellite grid. The challenge is in finding a soil moisture network that meet the measurement requirement. Soil moisture can be categorized into dense network and sparse network. Sparse network means typically 1 to 2 ground observations per satellite footprint, whereas dense or "core" networks consist of about 5 to 10 ground observations per satellite footprint.

Commonly used upscaling strategies in the literature include four types. The first strategy is using a time-stability concept, which means locating the soil moisture stations at 'representative'

landscape locations (Crow *et al.*, 2012), and predict large-scale moisture averages from these few sensors located at particular sites. It was originally developed to sample a 2000 m^2 sparsely instrumented grass field based on stable measurement sites that predict the large-scale average over long time scales (Vachaud *et al.*, 1985), and later extended to a larger network (Cosh *et al.*, 2004). This requires static vegetation type, soil type, and topography (Brocca *et al.*, 2007; Cosh *et al.*, 2004; Grayson *et al.*, 1997; Mohanty & Skaggs, 2001). It is reported that only 6 points are required within a 75 km^2 footprint of a satellite (Cosh *et al.*, 2006). However, a great challenge is that direct identification of time-stable sites typically requires very dense spatial sampling of a coarse-scale area over an extended period. The second strategy is spatial statistical models that include arithmetic average (Colliander *et al.*, 2017b); Thiessen polygon (Bhuiyan *et al.*, 2018; Colliander *et al.*, 2017b; Rowlandson *et al.*, 2015); IDW (Perry & Niemann, 2008; Yuan & Quiring, 2017); kriging (Kang *et al.*, 2015; Qin *et al.*, 2015). The fundamental difference among these methods lie in the approach of how to weight the stations to obtain an accurate estimate of the average soil moisture. With regard to spatial up-scaling, unless the in-situ measurements are dense and evenly spread across the site, using the arithmetic average of the measured values does not guarantee an accurate estimate of the grid average soil moisture (Colliander *et al.*, 2017b). The third strategy is using land surface models (Crow *et al.*, 2005; De Rosnay *et al.*, 2009a) and geostatistical models (Clewley *et al.*, 2017; Wang *et al.*, 2015a). Land surface models and geostatistical models are generally required for sparse network. However, in the case of dense networks, for example, TxSON, autocorrelation exists in the dataset, and therefore special technics might be necessary to upscale the dataset. The fourth strategy is based on field campaign. Intensive soil moisture measurements are collected at different times and stations and therefore create the

dense soil moisture network. However, this strategy is very expensive, and are also open to combination with any of the above-mentioned strategies for a better upscaling result.

This study proposes to use Thiessen polygon-based block kriging for the upscaling. A kriging method in which the average expected value in an area around an unsampled point is generated rather than the estimated exact value of an unsampled point. In block kriging, the semi-variances between the data points and the interpolated point, which was used in kriging, are replaced by the average semi-variances between the data points and all points in the region, and that is the main reason that block kriging is commonly used to provide better variance estimates and smooth interpolated results (Burgess & Webster, 1980b; Crow *et al.*, 2012). Compared with point based kriging, block kriging can yield smaller estimation variances and smoother maps (Burgess & Webster, 1980b). Block-kriging methods provide a more sophisticated approach to aggregating these observations based on observed auto-correlation structure (Crow *et al.*, 2005).

It was generally assumed in the literature that the soil moisture data measured in the morning are more accurate than in the afternoon from the radar or radiometer because the land surface temperature and soil dielectric properties are likely to be more uniform in the vertical profiles of soil (Basharinov & Shutko, 1975; Entekhabi *et al.*, 2014). However, to my knowledge, other than an analysis conducted by Chan et al. (Chan *et al.*, 2018), no further studies have confirmed this through experiments. In this study, the a.m./p.m. data obtained from SMAP will be compared for better performance.

Specific research objectives include:

1. validate the statistical models by comparing the experimental data with the predicted values.

2. analyze the soil moisture change pattern using am./p.m. soil moisture data from satellite and station observations.

3.2. Material and Study Area

The datasets include satellite data from SMAP, and ground station data collected from the Texas Soil Observation Network (TxSON).

The primary data component was acquired from the NASA satellite Soil Moisture Active Passive (SMAP). The data products were Level 3 Soil Moisture data from L-Band Radiometer (SMAP L3-SM-P-E, NASA, 2017). The Level 3 data were a daily global composite of the Level 2, a soil moisture product based on brightness temperature measurements that were sensitive to soil moisture. The Level 2 accuracy was equal to or better than 0.04 cm^3 (Entekhabi *et al.*, 2014). SMAP Level 3 products were downloaded from the NASA's Earth Observing System Data and Information System (EOSDIS) Reverb Echo portal on EARTHDATA as GeoTIFFs in WGS 1984 Geographic Coordinate System.

Texas Soil Moisture Observation Network (TxSON) is an intensively-monitored 36-km (1300 km^2) grid cell soil moisture network that consists of 40 monitoring stations that measure in situ soil moisture and precipitation. The 40 Station on real-time data collection includes 27 Micro-stations, 6 Weather Stations, and 7 Partner Stations with Lower Colorado River Authority (LCRA). The network is located near Fredericksburg, Texas, along the Pedernales River and within the middle reaches of the lower Colorado River. TxSON uses nested design to replicate soil moisture at 3, 9 and 36 km satellite pixels in support of NASA's Soil Moisture Active/Passive (SMAP) mission and its Calibration and Validation Program. The measured depth includes 5 cm, 10 cm, 20 cm, and 50 cm. The network provides 5-minute sampling

intervals, which are then averaged hourly and updated hourly. The soil moisture sensors are from Campbell CS-655, which is the recent model and provides the best accuracy.

For this dissertation research, I used the TxSON data at the 9 km pixel for the upscaling. To match the time interval of SMAP, I collected the hourly dataset from July 1 to September 16, and then filtered the temporal resolution of the data to daily 6 a.m. and 6 p.m..

Table 3.1. TxSON Stations within the Study Area

Site Name	Soil Unit	F_loggerID	Latitude	Longitude	Elevation	Vegetation Type
TEAG_1	Bastrop loamy fine sand, 1 to 5 percent slopes	'CR200_1'	30.4376	-98.8059	520	Pasture/Hay
TEAG_2	Heaton loamy fine sand	'CR200_3'	30.4283	-98.8065	529	Grassland/Herbaceous
RABK_1	Luckenbach clay loam	'CR200_4'	30.4298	-98.7792	512	Pasture/Hay
TEAG_3	Loneoak fine sand	'CR200_9'	30.4319	-98.8133	535	Shrub/Scrub
OWEN_1	Luckenbach clay loam	'CR200_13'	30.4327	-98.8583	556	Evergreen Forest
ECKE_2	Heaton loamy fine sand	'CR200_14'	30.4151	-98.8025	520	Grassland/Herbaceous
WALT	Purves soils	'CR200_19'	30.4175	-98.8542	603	Shrub/Scrub
RABK_2	Brackett soils	'CR200_21'	30.4218	-98.7839	542	Evergreen Forest
OWEN_2	Heaton loamy fine sand	'CR200_22'	30.4315	-98.8604	553	Shrub/Scrub
ECKE_3	Brackett soils	'CR200_26'	30.4193	-98.8046	557	Grassland/Herbaceous
TOMF_1	Tarrant soils	'CR200_28'	30.4613	-98.8451	605	Shrub/Scrub
TOMF_3	Purves soils	'CR200_29'	30.4487	-98.8480	597	Shrub/Scrub
ECKE_1	Krum silty clay	'CR1000_1'	30.4205	-98.8033	543	Shrub/Scrub
TOMF_2	Oakalla silty clay loam	'CR1000_6'	30.4421	-98.8427	534	Shrub/Scrub
2847	Purves	'LCRA_2'	30.4206	-98.8519	613	Evergreen Forest

The vegetations types are obtained from the 2011 version of National Land Cover Database (NLCD), elevations are from DEM data, and soil units are obtained from Soil Survey Geographic database (SSURGO).

Table 3.1 show that the soil type and vegetation cover vary in the study area, which violates the basic requirement of the time-stability concept as mentioned in the introduction. The main vegetation type is shrub/scrub, the elevation ranges between 512 and 613 meters.

3.3. Methods

3.3.1 Data Pre-processing

The SMAP L3-SM-P-E 9-km soil moisture datasets were downloaded as .hdf file, and the Extract Subdataset tool were used to get the soil moisture layers. They are layer 12 (a.m. Soil Moisture) and layer 41 (p.m. Soil Moisture). The TxSON data were originally provided in Matlab file. I used csv to store the soil moisture information for 78 days and extracted the 6 a.m./6 p.m. data.

3.3.2 Arithmetic Average and Thiessen Polygon

Arithmetic average, or arithmetic mean, is simply the mean or average of the stations being measured. The arithmetic mean of a set of station observation is defined as being equal to the sum of the numerical values of each station observation divided by the total number of station observations. As shown in the Equation 3.1,

$$\theta = \frac{\sum_1^{15} \theta i}{15} \quad (3\text{-}1)$$

where θi is the soil moisture of each station observation.

The Thiessen polygon method is selected as the default approach by SMAP science team, which used a Voronoi diagram to find the weighting of the stations (Colliander *et al.*, 2017b). This Voronoi diagram is a partitioning of a plane into regions that used to divide the area covered by the input point features into polygons. The polygons are named Thiessen polygons. Each Thiessen polygon contains only a single point input feature, i.e., a soil moisture station, as shown in Figure 3.2. Any location within a Thiessen polygon is closer to its associated station than to any other stations.

The procedure to create Thiessen polygons are as follows (Brassel & Reif, 1979):

1) All points are triangulated into a triangulated irregular network (TIN) that are performed on Delaunay conforming triangulations.

2) The perpendicular bisectors for each triangle edge are generated, forming the edges of the Thiessen polygons. The location at which the bisectors intersect determine the locations of the Thiessen polygon vertices.

The Thiessen polygon method is a weighted average method, where the area of each polygon is considered as the weight for each soil moisture station. The Thiessen polygons were created using ArcGIS.

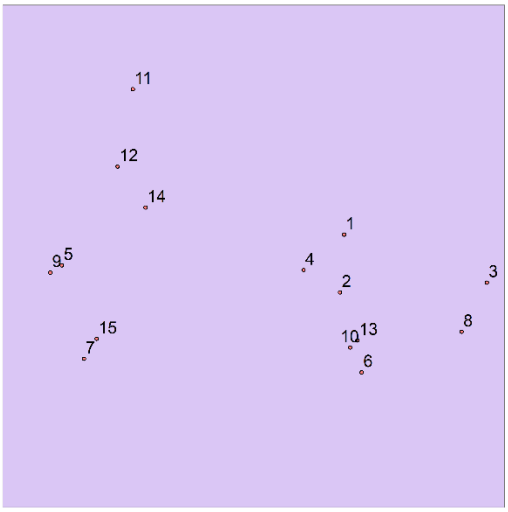


Figure 3.1. TxSON Stations

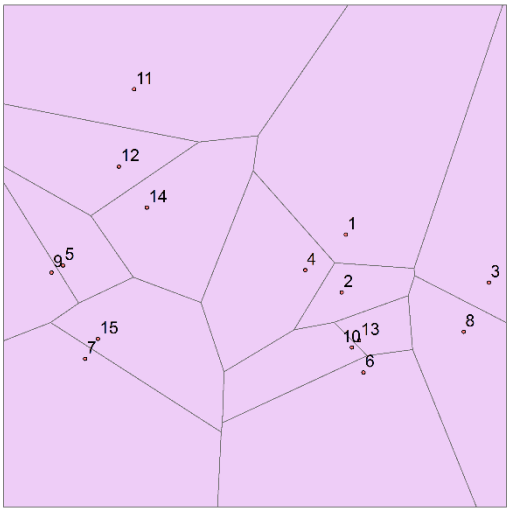


Figure 3.2. Thiessen Polygon Generated from Stations

3.3.3 Gaussian Weighted Average

The SMAP soil moisture produce, L3_SM_P_E, was retrieved from the brightness temperature. Thermally generated radiometric sources have an amplitude probability distribution function that is Gaussian in nature (Entekhabi *et al.*, 2014; Piepmeier, 2014).

The spatial convolution weights were modeled as the 2D Gaussian kernel function, which is higher if the location of the station is closer to the center of the 9 km square unit, and vice versa. Gaussian kernel The 1D Gaussian Kernel is defined as (Babaud *et al.*, 1986):

$$Y(t) = \int K(t,s)X(s)\,ds \qquad (3\text{-}2)$$

where K is the kernel of the integral. Given the input signal X, Y represents the output signal.

To test the performance of Gaussian weighted average method, Gaussian kernel was applied to the TxSON soil moisture product as the weight of average, and compared that with the SMAP 9-km grid.

3.3.4 Thiessen Polygon Based Block Kriging

Block interpolation is an interpolation method that predicts the average value of a phenomenon within a specified area. Block kriging provides an estimate for a discrete area around an interpolation point. The block is defined as the rectangular area around a point that is not included in an adjacent block. Block kriging performs an estimate not for an unknown point, but for a block or area. A major disadvantage of ordinary kriging is that when sample values change very fast within short distances, ordinary point kriging may results in surfaces that have many sharp spikes or pits at the data points. Under such circumstances, block kriging acts as a method for smoothing such structures by dividing the whole area into several blocks and calculate a simple local average for each of them.

In this study, I applied the block kriging method using the Thiessen polygon as the input blocks and the 9-km fishnet polygon as the output blocks. The results were then averaged to compare with the satellite measurement. Block kriging works by calculating predictions for a number of specified locations within an area; the values are averaged, and the average is assigned as the prediction for the entire area (Burgess & Webster, 1980b).

3.3.5 Time Series Analysis

Time series analysis is useful technique to characterize changes in soil moisture over time and to relate these changes to other observations (Albertson & Kiely, 2001). Autocorrelations and cross-correlations were used to examine the relationships between the soil moisture content and precipitation. Daily data for each ground station will be examined to see a daily change.

3.3.6 Accuracy Assessment

The accuracy was assessed based on comparison between the SMAP 9-km grids and the upscaled results from the four algorithms. Two matrices were used for the evaluation: RMSD and ubRMSD (Entekhabi *et al.*, 2010).

$$\text{RMSD} = \sqrt{E[(\theta est - \theta true)^2]} \quad (3\text{-}3)$$

where $\theta\, est$ is the estimated soil moisture from satellites and $\theta\, true$ is the soil moisture ground truth data.

$$\text{ubRMSD} = \sqrt{\frac{\sum_{i=1}^{N}((x_i-\bar{x})-(y_i-\bar{y}))^2}{N-1}} \quad (3\text{-}4)$$

where *x* and *y* are two sources of soil moisture observations.

The relationship between RMSD and ubRMSD is:

$$\text{RMSD}^2 = \text{ubRMSD}^2 + \text{bias}^2 \quad (3\text{-}5)$$

3.4. Results

The results include the daily trend analysis for both SMAP and different upscaling methods, as well as their correlation plot and matrix table.

3.4.1 Daily Trend Analysis

Figure 3.3 to 3.6 show the daily trend analysis in the morning based on both SMAP and the results of arithmetic average, Thiessen polygon, Gaussian weighted average, and Thiessen polygon-based block kriging.

Figure 3.7 to 3.10 show the daily trend analysis in the afternoon based on both SMAP and the results of arithmetic average, Thiessen polygon, Gaussian weighted average, and Thiessen polygon-based block kriging.

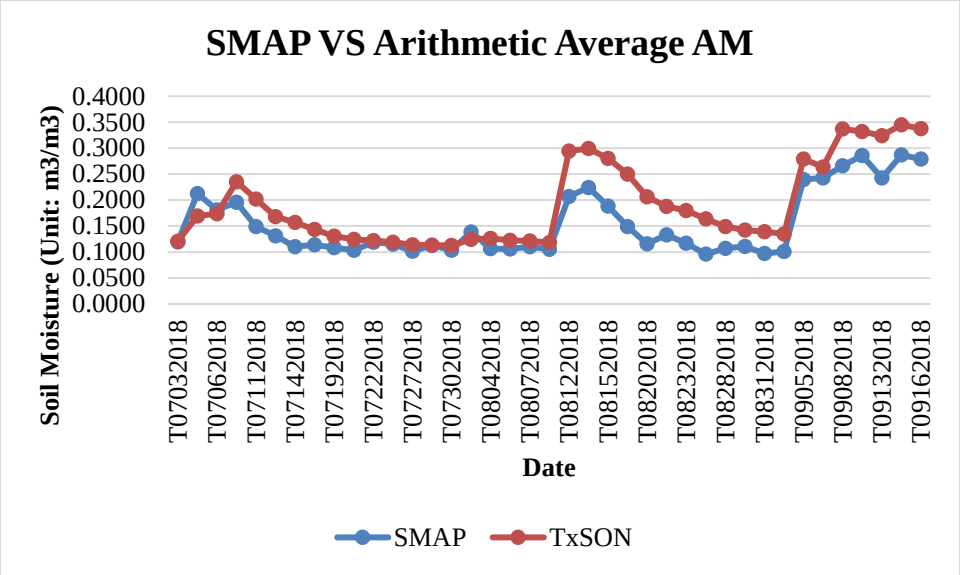


Figure 3.3. Time Plot of SMAP and Arithmetic Average for the AM Data for the TxSON Stations from June to September 2018

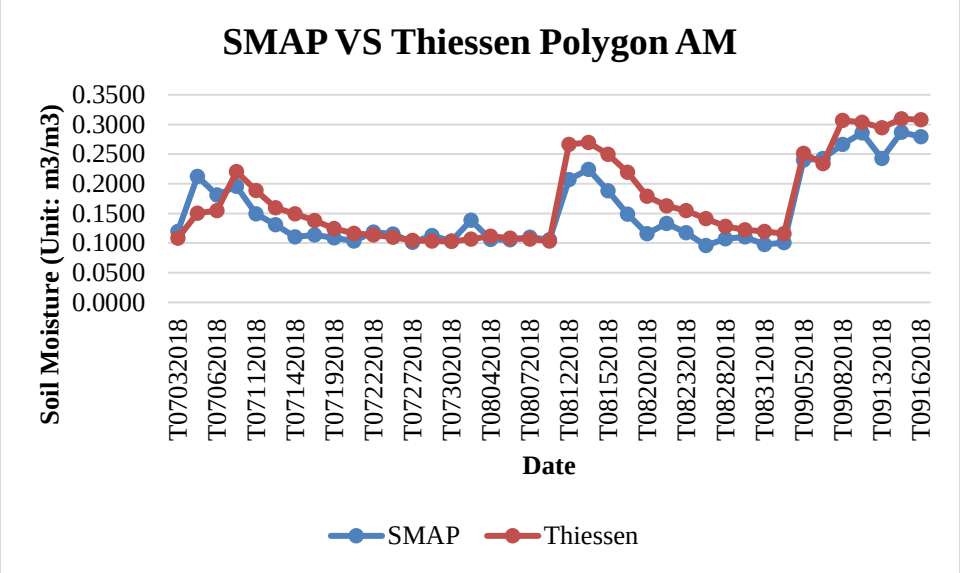


Figure 3.4. Time Plot of SMAP and Thiessen Polygon Method for the AM Data for the TxSON Stations from June to September 2018

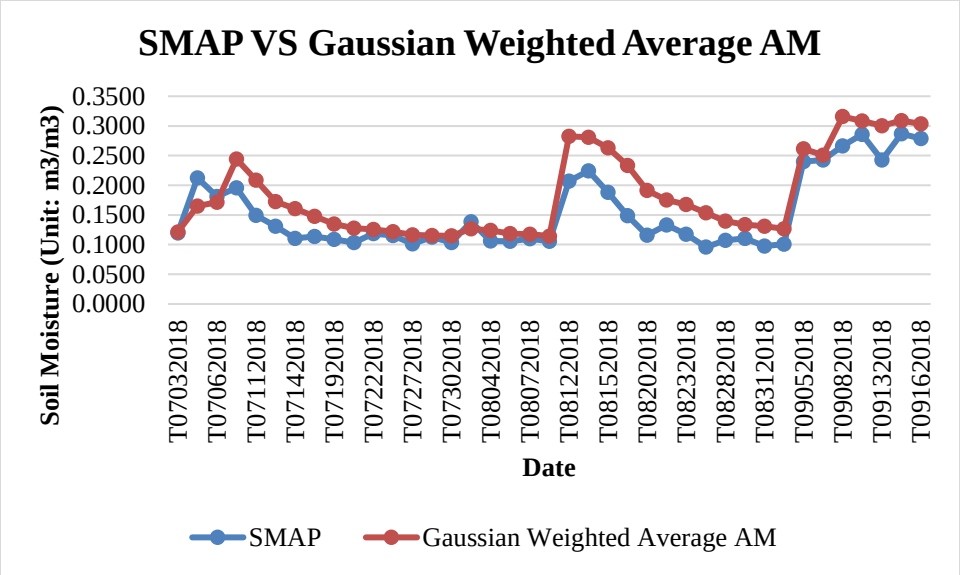


Figure 3.5. Time Plot of SMAP and Gaussian Weighted Average for the AM Data for the TxSON Stations from June to September 2018

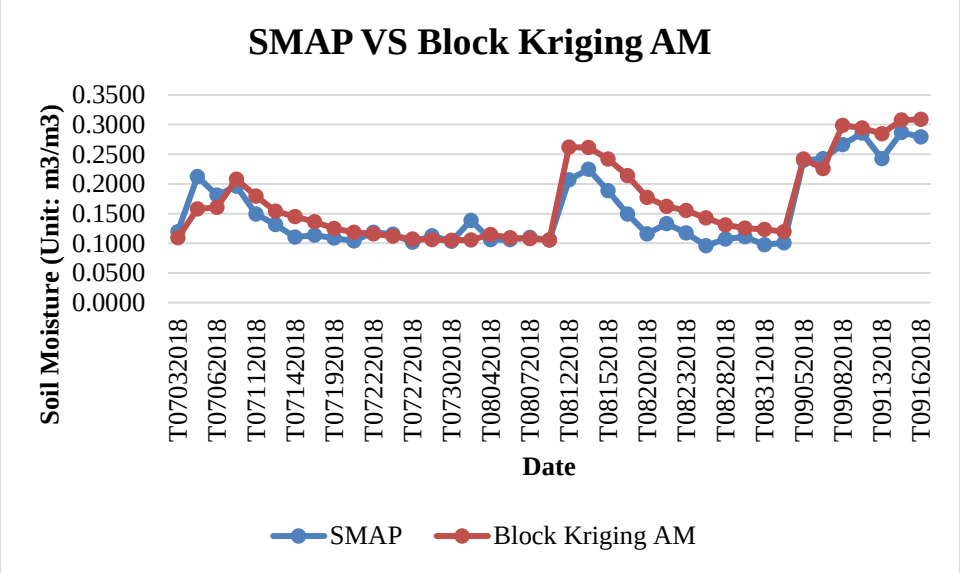


Figure 3.6. Time Plot of SMAP and Block Kriging for the AM Data for the TxSON Stations from June to September 2018

The daily soil moisture trend in the morning shows that the soil moisture from July 1 to July 5 (SAMP) or July 8 (TxSON) has an increasing trend, followed by a decreasing trend until August 08. Another increase was observed from August 08 to August 12 (TxSON) or 14 (SMAP), and

generally decrease until September 1. The trend thereafter had several ups and downs. From July 1 to September 16, three soil moisture peaks occurred: July 5 or 8; August 12 or 14, and September 15.

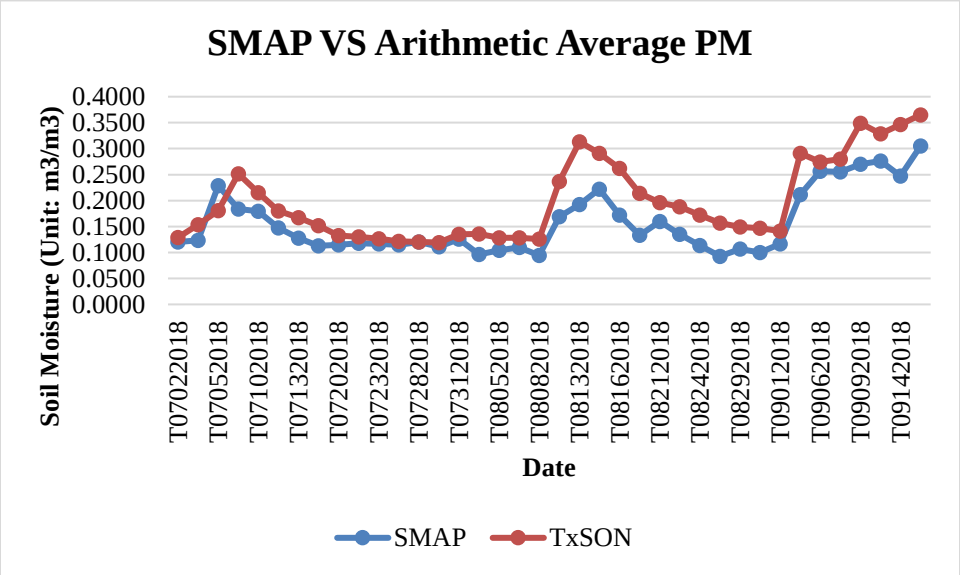


Figure 3.7. Time Plot of SMAP and Arithmetic Average for the PM Data for the TxSON Stations from June to September 2018

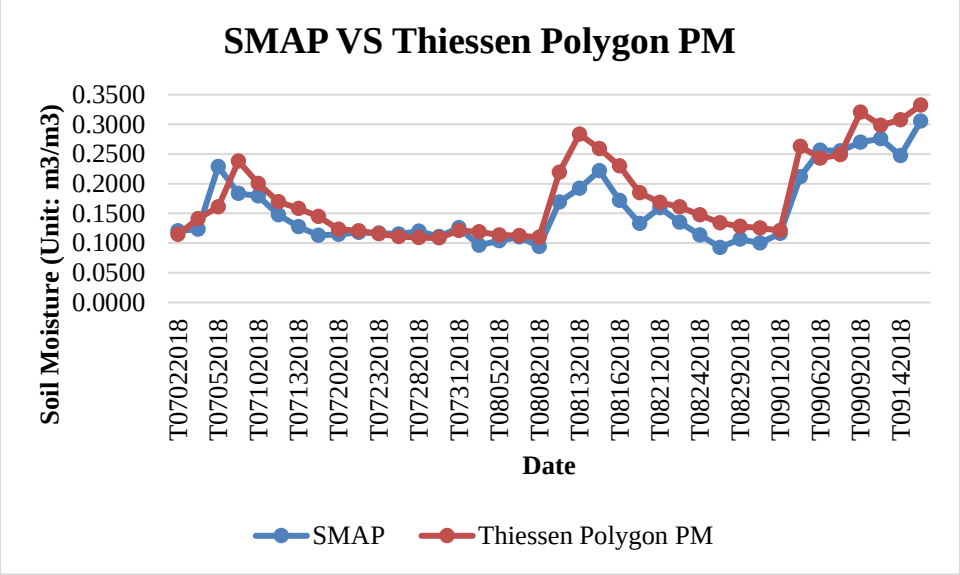


Figure 3.8. Time Plot of SMAP and Thiessen Polygon Method for the PM Data for the TxSON Stations from June to September 2018

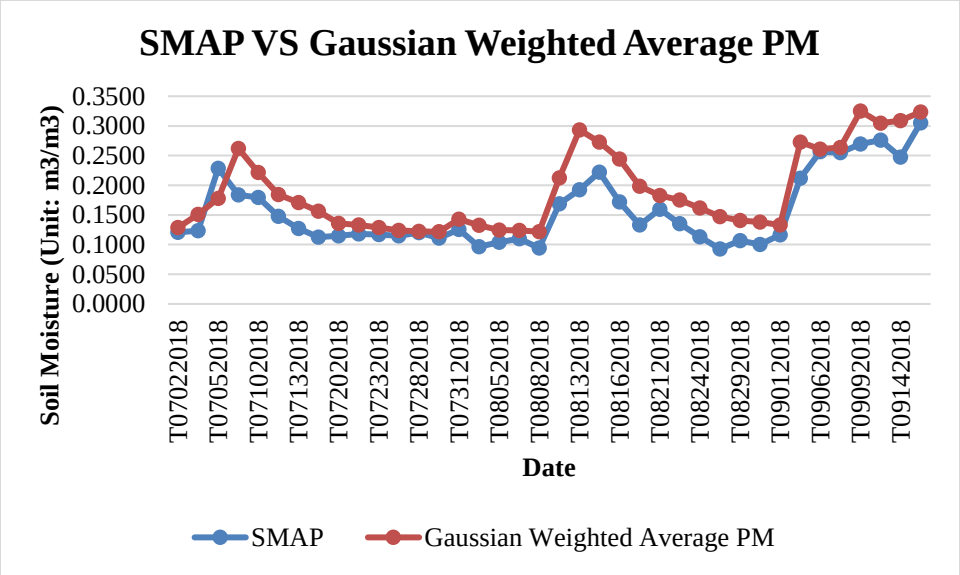


Figure 3.9. Time Plot of SMAP and Gaussian Weighted Average for the PM Data for the TxSON Stations from June to September 2018

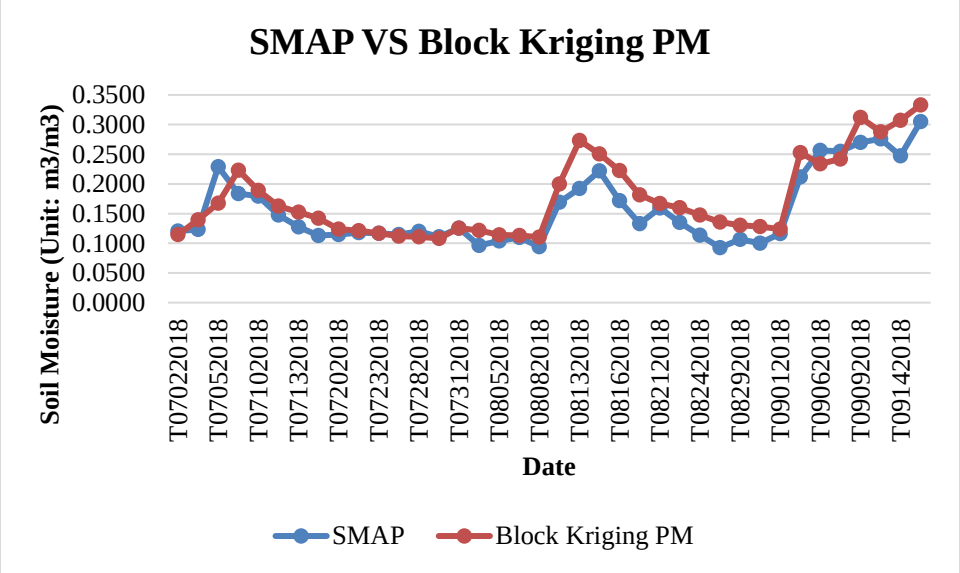


Figure 3.10. Time Plot of SMAP and Block Kriging for the PM Data for the TxSON Stations from June to September 2018

The daily soil moisture trend in the afternoon show that from July 1 to July 5 (SAMP) or July 7 (TxSON) has an increasing trend, followed by a decreasing trend until August 08. Another increase was observed from August 08 to August 13 (TxSON) or 14 (SMAP), and generally

decrease until September 1. The trend thereafter had several ups and downs. From July 1 to September 16, three soil moisture peaks occurred: July 5 or 7; August 13 or 14, and September 15.

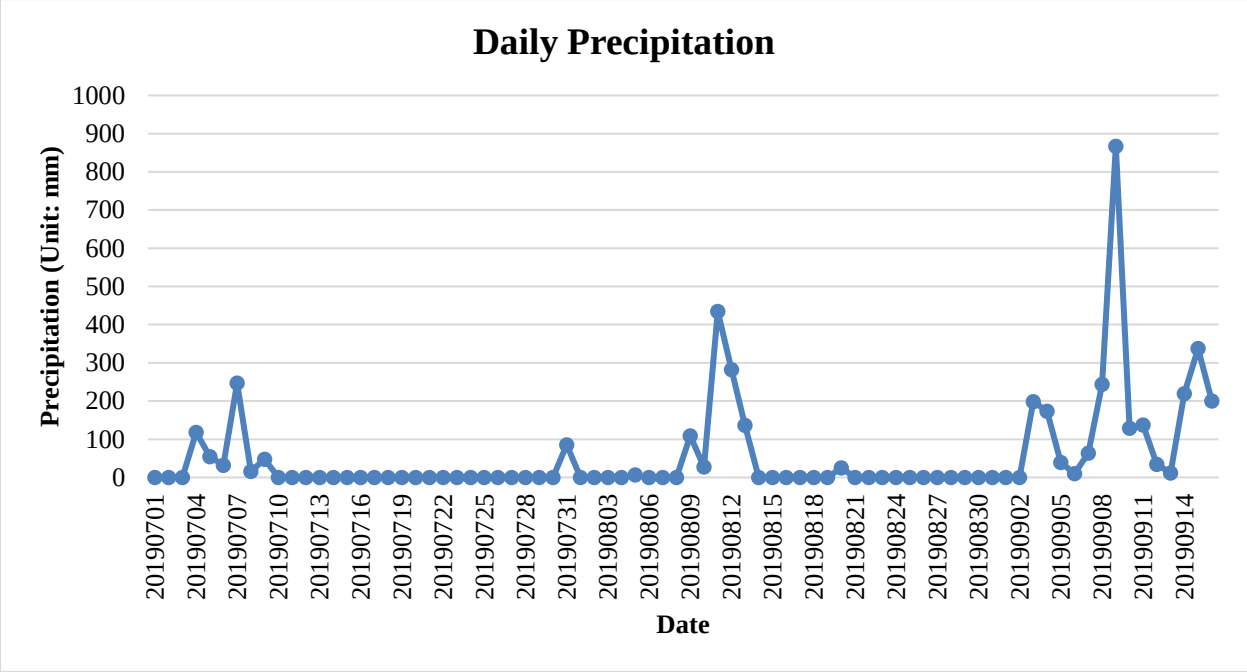


Figure 3.11. Total Precipitation for the Stations

The precipitation for the study area is shown in Figure 3.11. As shown in the figure, the precipitation is concentrated in three major periods: July 3, 2019 to July 10, 2019; August 08, 2019 to August 14, 2019; September 06, 2019 to September 15, 2019. The period of precipitation cycles correlates with that of the soil moisture.

3.4.2 Model Performance

The correlation plot of SMAP and the four methods for both the a.m. and p.m. data are shown in Figure 3.12 to Figure 3.19. Table 3.2 and 3.3 show the model performance comparison for both AM and p.m. data in term of RMSD and ubRMSD.

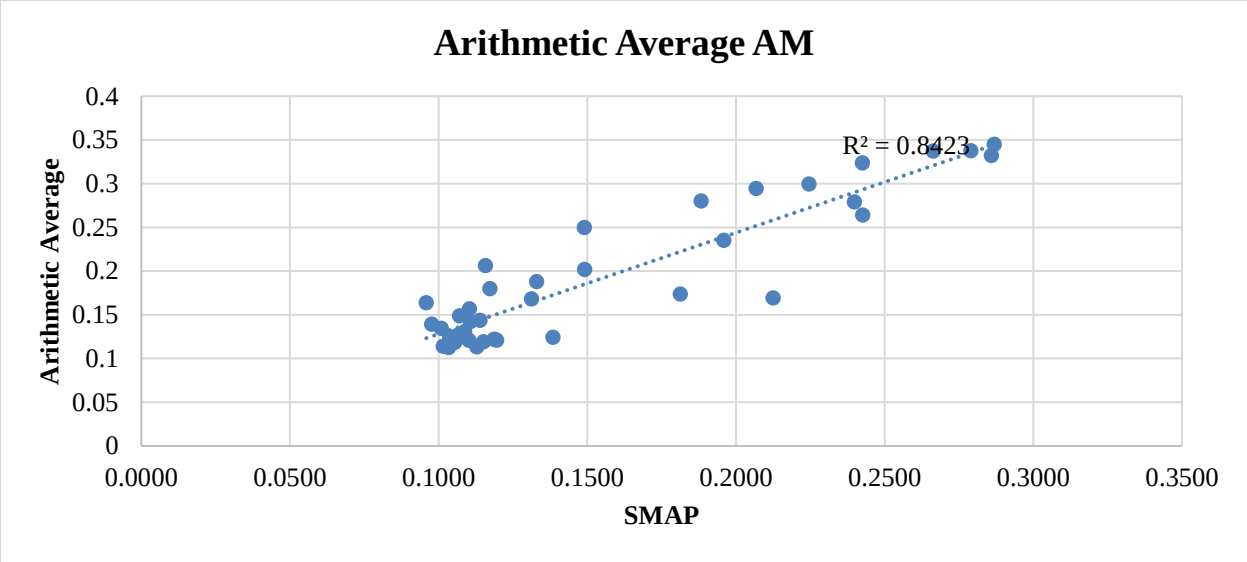


Figure 3.12. Correlation Plot of SMAP and Arithmetic Average for the AM Data

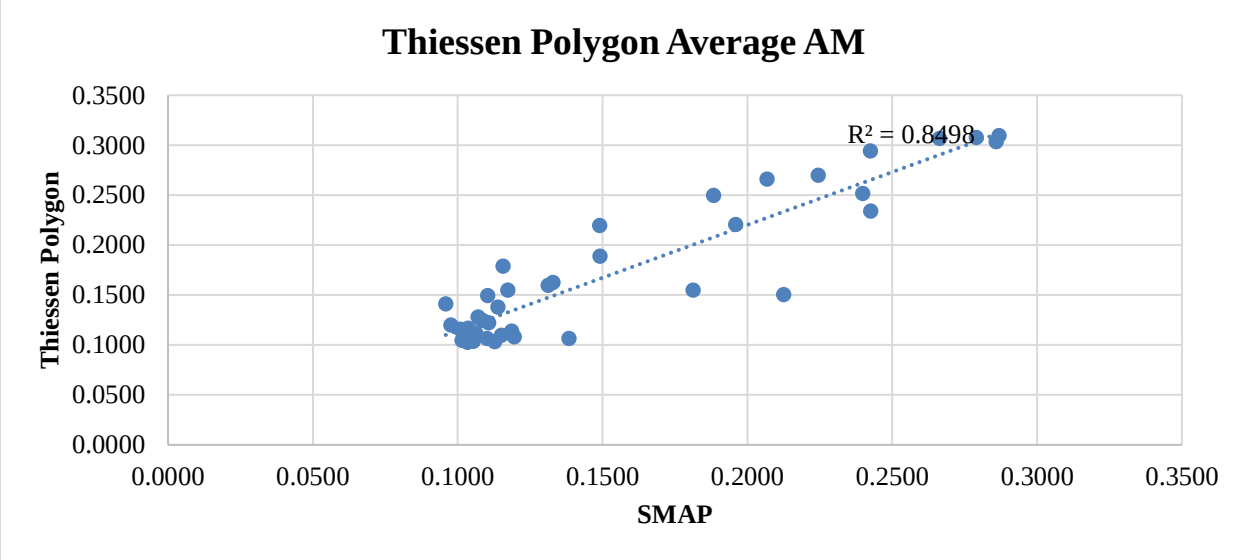


Figure 3.13. Correlation Plot of SMAP and Thiessen Polygon Method for the AM Data

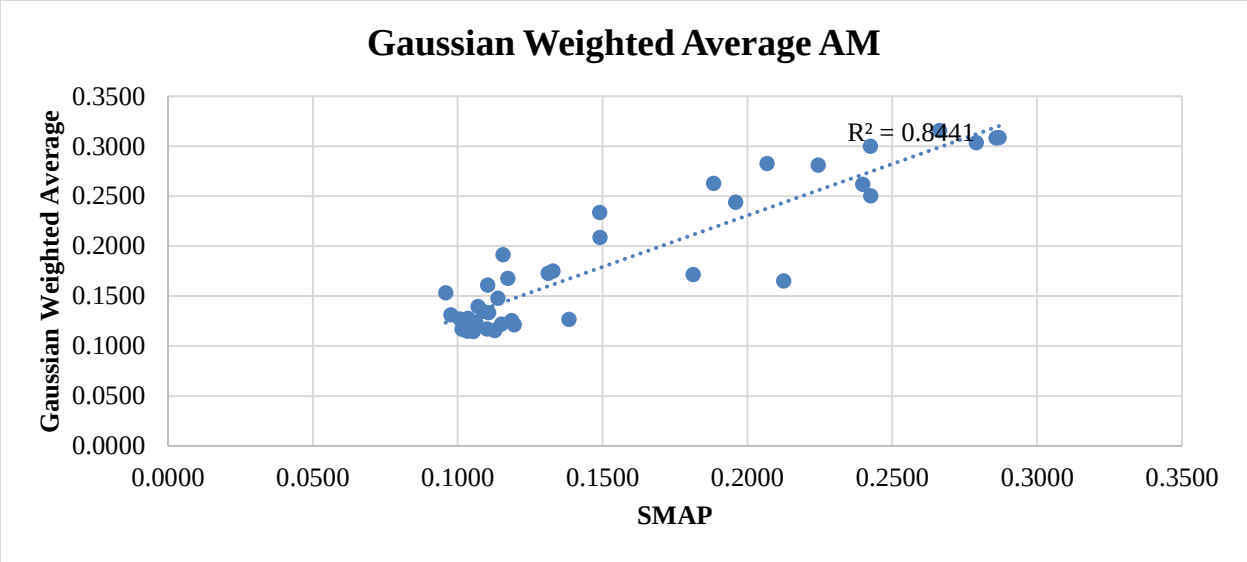


Figure 3.14. Correlation Plot of SMAP and Gaussian Weighted Average for the AM Data

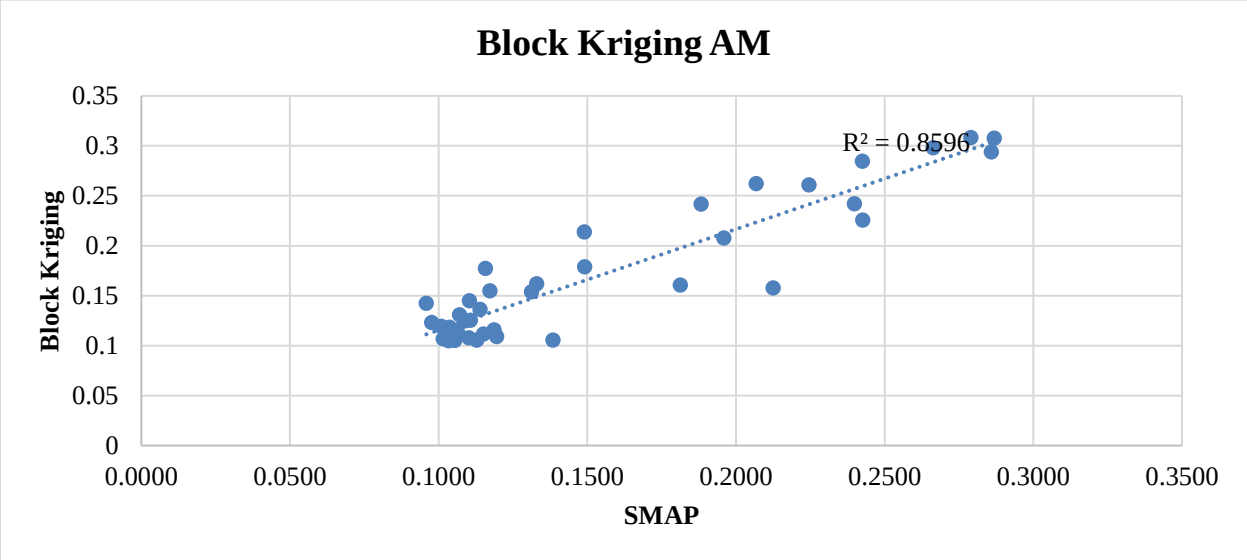


Figure 3.15. Correlation Plot of SMAP and Block Kriging for the AM Data

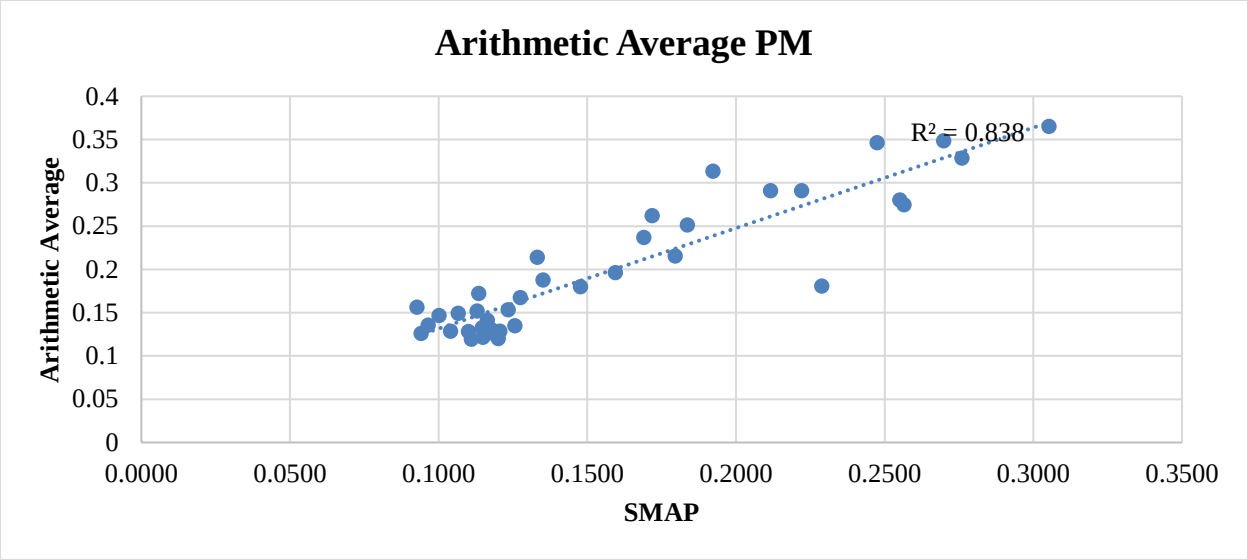


Figure 3.16. Correlation Plot of SMAP and Arithmetic Average for the PM Data

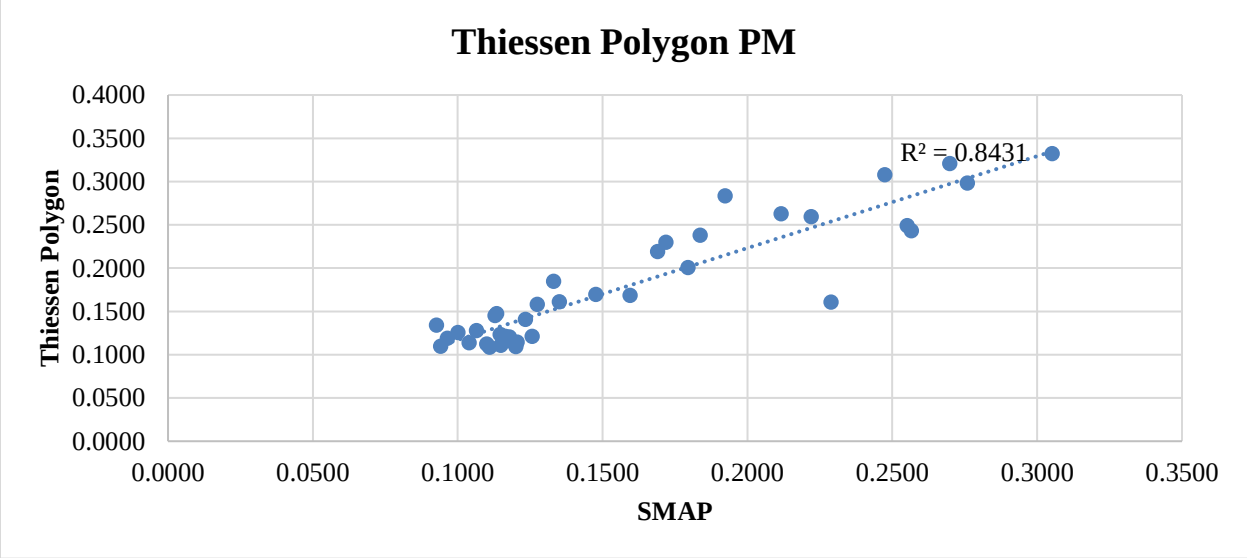


Figure 3.17. Correlation Plot of SMAP and Thiessen Polygon Method for the PM Data

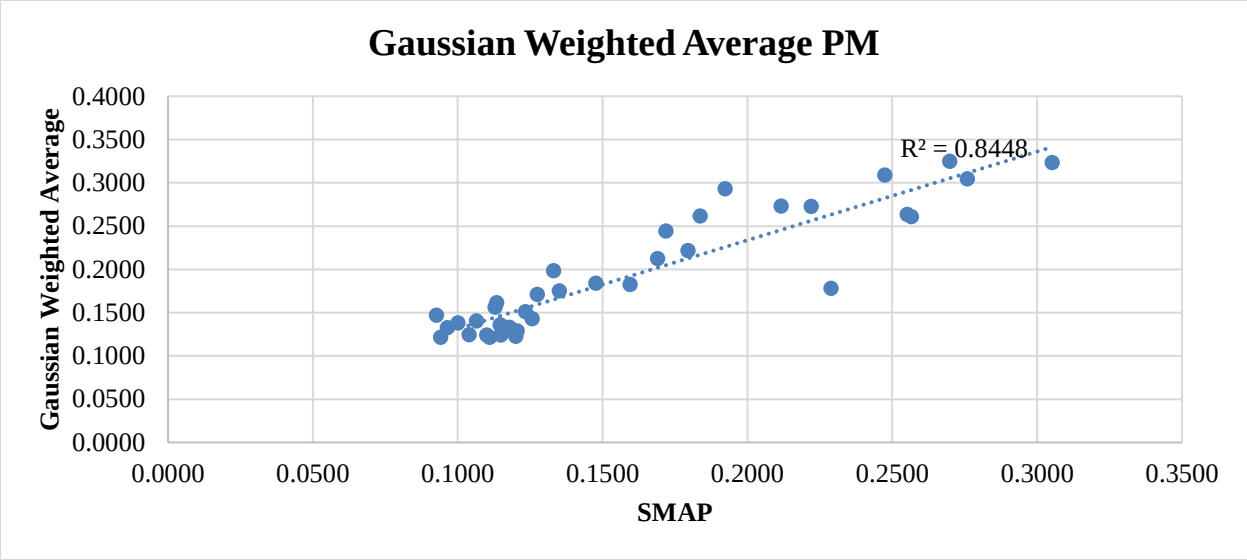


Figure 3.18. Correlation Plot of SMAP and Gaussian Weighted Average for the PM Data

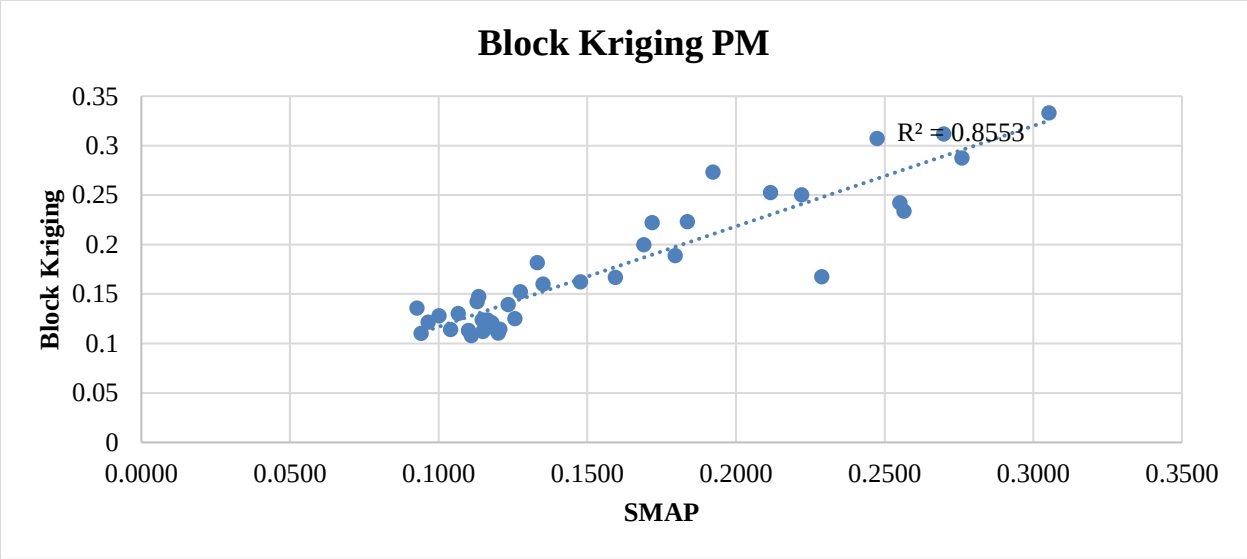


Figure 3.19. Correlation Plot of SMAP and Block Kriging for the PM Data

Table 3.2. Model Performance Comparison for the AM Data

	BK	Gaussian	Thiessen Polygon	Arithmetic Average
RMSD	0.0298	0.0400	0.0326	0.0489
ubRMSD	0.0251	0.0272	0.0275	0.0322

Table 3.3. Model Performance Comparison for the PM Data

	BK	Gaussian	Thiessen Polygon	Arithmetic Average
RMSD	0.0306	0.0419	0.0344	0.0518
ubRMSD	0.0249	0.0262	0.0276	0.0320

For both a.m. and p.m. data, Block Kriging 1 by 1 based on Thiessen polygon yielded the best accuracy. For the AM data, the best RMSD is 0.0298 (Thiessen polygon-based block kriging method), which is much smaller than all the other algorithms. The ubRMSD is 0.0251, which indicate there is not much bias between the two datasets. For the p.m. data, the best RMSD is 0.0306 (Thiessen polygon-based block kriging method), which is also much smaller than all the other algorithms. The ubRMSD is 0.0249, which indicate there is not much bias between the two datasets. Comparing the a.m. data with p.m. data, it can be found that the RMSD is a litter larger in the p.m. than in the AM, which is consistent with Chan et al. (2018) On the other hand, the ubRMSD in the afternoon is smaller than in the morning, which means the variance in the afternoon is smaller than in the morning while the bias is much larger.

3.5. Conclusions

Four methods were compared for the TxSON soil moisture upscaling: arithmetic average, Thiessen polygon; Gaussian weighted average, and Thiessen polygon-based block kriging. Thiessen polygon-based block kriging is the best overall method, which gives the best RMSD and ubRMSD.

The study confirms that the distribution of the dataset may possibly decrease the reliability of the dense network, as autocorrelation in the dataset might impact the model accuracy and increase the bias when comparing the network with satellite products. The good

performance of using Thiessen polygon-based block kriging concludes that the average function within the block kriging can smooth the result using the block of 9 km.

The study also confirmed that the a.m. soil moisture observation from SMAP performs slightly better than the p.m. soil moisture data in term of bias.

CHAPTER 4. DROUGHT APPLICATION

4.1. Introduction

Climate variability in the southeastern United States can bring regional-scale droughts. According to the National Climate Assessment for the Southeast, extreme heat and soil water deficiency are two of the four major stressors for the region (Melillo *et al.*, 2014) because a large part of the Southeast's landscape is occupied by agriculture, forests, and rangelands (McNulty *et al.*, 2015). Drought is especially a concern for agricultural and forestry management. For the agricultural sector, water deficiency during droughts has led to a reduction in crop and livestock production (Change, 2014; Parry *et al.*, 2007). For the forestry sector, water shortage could affect growth of the trees and also increase their vulnerability to wildfires (Field, 2012). A monitoring system that is able to deliver timely warnings of droughts can play a vital role in regional water resource management and economic development. The United States Department of Agriculture (USDA) and the Southeast Regional Climate Hub (SERCH) delivers science-based knowledge on climate to farmers, ranchers, and foresters to cope with climate issues such as extreme precipitation, heat stress, and drought in the southeastern United States (USDA, 2015). SERCH uses a drought mitigation tool, the Lately Identified Geospecific Heightened Threat System (LIGHTS), which is a prediction model driven by NOAA's Climate Prediction Center's Monthly Drought Outlook, Monthly Temperature and Precipitation Outlook, and Risk of Seasonal Climate Extremes in the US related to El Niño–Southern Oscillation (ENSO). Subscribers will receive a notification when

---

This chapter, previously published as "Xu, Y., Wang, L., Ross, K., Liu, C., & Berry, K. (2018). Standardized Soil Moisture Index for Drought Monitoring Based on Soil Moisture Active Passive Observations and 36 Years of North American Land Data Assimilation System Data: A Case Study in the Southeast United States. Remote Sensing, 10(2), 301."

the system predicts a drought condition in their area. With the assistance of this system, farmers and foresters can better cope with climate issues efficiently and

on a timely basis. The SERCH LIGHTS services are available in eleven States: Alabama, Arkansas, Georgia, Florida, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, and Virginia (Figure 4.1).

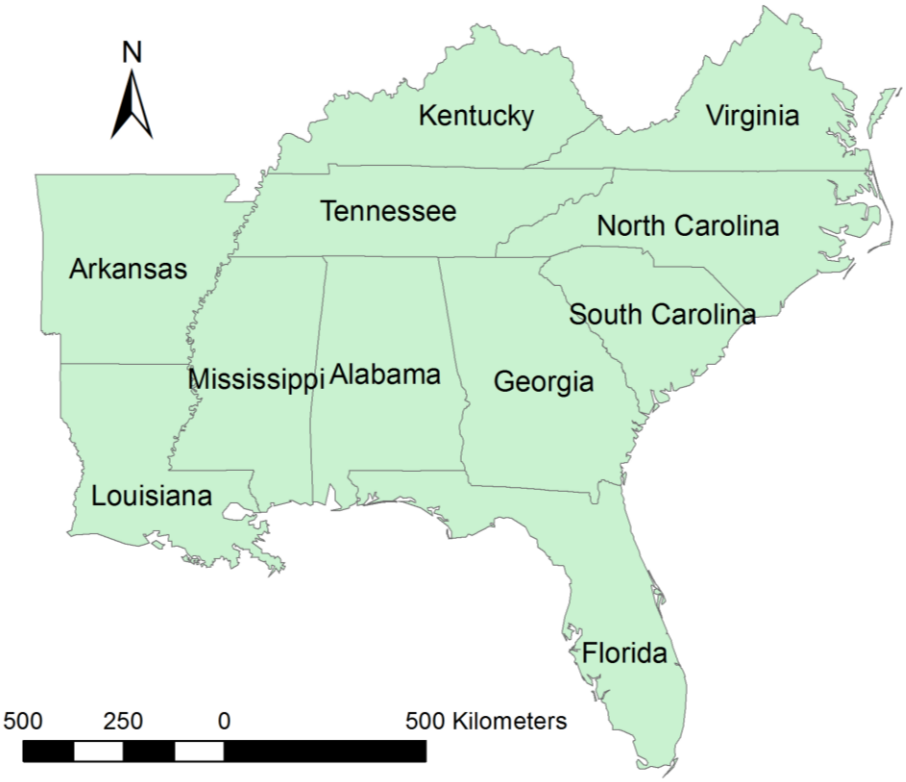


Figure 4.1. The 11 States of SERCH LIGHTS

Current implementation of LIGHTS does not include any soil moisture indices in their prediction model. Adding soil moisture data sampled by geospatial technologies can greatly improve the reliability and accuracy of the prediction model (NOAA, 2012). Several methods and indices for soil moisture retrieval were proposed in the past research (Ritchie, 1998; Thornthwaite & Mather, 1955). The Palmer Drought Severity Index (PDSI) is one of the most popular indices of drought. The PDSI measures the cumulative departure of moisture supply across space and time

(Palmer, 1965). It uses the Thornthwaite method to estimate potential evapotranspiration (PET). However, due to the complexity and uncertainty of PET estimation, the model has limitation in accuracy and application, especially in extreme climate conditions and mountainous terrain (Zargar *et al.*, 2011). The relative soil moisture index (RSMI) was designed to estimate the amount of water available in soil for crops(Sivakumar *et al.*, 2010). This model requires data of a variety of factors such as climate (rainfall rates, potential evapotranspiration), plant properties (vegetation type, leaf area, management practices, crop sensitivity to water stress, and crop water requirement for each phenological phase), and soil characteristics (soil water capacity, soil proximity to the water table) (Sheikh *et al.*, 2009; Sivakumar *et al.*, 2010). Nevertheless, it is difficult to accurately measure these variables with sufficient spatial and temporal coverage (Meng & Quiring, 2008). In addition, these variables are defined in different spatial scales and context. How to remedy the scale difference is another challenge and obstacle to use RSMI (Crow *et al.*, 2012; De Rosnay *et al.*, 2009b; Petropoulos *et al.*, 2015). Another meteorological drought index was designed as calculating the percent of precipitation from normal (Zargar *et al.*, 2011). The main advantage of this index is its simplicity and transparency (Keyantash & Dracup, 2002). However, the statistical construct has been criticized because the distributions for seasons and regions are different. For this reason, this index cannot be used to compare drought across seasons and regions (Hayes, 2006). The Palmer Z-index is a monthly standardized anomaly of available moisture (Zargar *et al.*, 2011). The Palmer Z-index was found most suitable to monitor agricultural drought in Canadian prairies (Quiring & Papakryiakou, 2003). The standardized precipitation index (SPI) (McKee *et al.*, 1993) is a popular meteorological drought index solely derived from precipitation data. SPI is expressed as deviations from the long-term mean of a normal distribution fitted on the precipitation data (Edwards & McKee, 1997). If the SPI value falls below zero for a certain period or the value

is lower than -1, a drought is said to have occurred (McKee *et al.*, 1993). The advantages of SPI include the simplicity of its definition, ability to generalize to different time scales and climate regions, as well as the ability to provide early warning of drought (Zargar *et al.*, 2011).

The abovementioned meteorological drought indices, including PDSI, SPI, and percent of precipitation, do not consider soil moisture data as an input except that the Palmer Z-index may consider soil moisture as additional input to precipitation and temperature (Illston *et al.*, 2004). Furthermore, calculating these indices usually requires at least one month of data to train the model, but such a training period does not meet SERCH LIGHTS's requirement for quick responses to drought conditions. The U.S. Drought Monitor (USDM) can provide a week's drought summary based on the abovementioned indices plus soil moisture from data assimilation systems and other models (UNL, 2014), even though the weekly drought monitoring cannot meet SERCH LIGHTS's requirement for quick responses to drought conditions. Satellite-based observation data could greatly enhance the extent and accuracy of drought prediction models.

Therefore, in this research, we make the use of Soil Moisture Active Passive (SMAP) satellite data and North American Land Data Assimilation System (NLDAS) soil moisture data to calculate a soil moisture index for drought warning called the standardized soil moisture index (SSI). SSI is based on the concept of percent of normal precipitation and Palmer Z-index, as well as the statistical construct of SPI. SSI essentially utilizes the z-score to explain how many standard deviations the soil moisture deviates from the historical mean soil moisture, and thus identifies droughts as statistical outliers in the time series.

Previous studies revealed that both the SMAP and NLDAS data are reliable soil moisture measurements. An intercomparison against SCAN in situ soil moisture measurements showed that SMAP Level 3 product outperformed the soil moisture and ocean salinity (SMOS) Level 3 product

(Al-Yaari *et al.*, 2017). The correlation between daily NLDAS data and in situ soil moisture at multiple soil depths is strong in the southeastern United States (Xia *et al.*, 2014).

The goal of this research is to incorporate SMAP data and NLDAS data for the southeastern states for updating the prediction power of the drought monitoring system, LIGHTS. It is hypothesized that integration of SMAP data into SERCH LIGHTS will increase the end-user's water management capabilities in response to drought conditions. Further introduction about SERCH LIGHTS and the project is included in the Appendix A.

4.2. Materials and Methods

4.2.1 Data Acquisition

We used the Level 3 soil moisture data from L-Band Radiometer (SMAP L3_SM_P) on board the NASA satellite Soil Moisture Active Passive (SMAP). The SMAP Level 3 product is a daily global radiometer-only soil moisture product, which provides direct soil moisture measurement at 6 a.m. local solar time in the top 5-cm layer of the soil column in units of m^3/m^3 (NASA, 2014). We obtained the data from NASA's Earth Observing System Data and Information System (EOSDIS) Reverb Echo portal on EARTHDATA, and requested to transform NetCDF files into GeoTIFFs with the WGS 1984 Geographic Coordinate System.

The second dataset is the soil moisture data from NASA NLDAS. The NLDAS Noah Land Surface Model (LSM) L4 Hourly 0.125 × 0.125 degree V002 data that measure the top 10-cm soil moisture were downloaded from Goddard Earth Sciences (GES) Data and Information Center data portal, Mirador (Xia *et al.*, 2015). The NLDAS data time zone was Coordinated Universal Time (UTC), which has an overall six-hour time difference compared with the SMAP local solar time. Therefore, 1200 UTC data were collected for each day over the 36 years.

The third dataset is from the Soil Climate Analysis Network (SCAN). SCAN stations use probes to collect soil moisture data across the United States(USDA, 2016). The probes were dielectric constant measuring devices placed at 5.08 cm depth (USDA, 2016). The USDA National Resources Conservation Service (NRCS) provides the SCAN dataset as downloadable .csv tables. Table 4.1 shows the parameters and the uses of the data.

Table 4.1. Data Description for the Agricultural Drought Monitoring

Platform & Sensor	**Parameter**	**Use**
SMAP Passive Radiometer	Soil moisture, Level-3, 36 km resolution	Daily measurement of soil moisture
NLDAS	Soil moisture, Noah model	Historical mean and standard deviation of soil moisture
USDA SCAN	Soil moisture	Validation

4.2.2 Data Processing

SMAP reached its orbit in January 2015, and the data were available since 1 April 2015. Therefore, less than two years of data have been recorded at the time of this study. A pre-processing of the SMAP data removed invalid values and outliers. The units of SMAP and NLDAS soil moisture do not match. SMAP measures volume of water per unit volume of soil. NLDAS measures soil moisture in units of kilogram per square meter of soil over variable thicknesses. Equation 4-1 converts the unit of NLDAS to the volume ratio that is similar to SMAP units.

$$\frac{\mathrm{SM_{NLDAS}}\ (\mathrm{kg/m^2})}{\mathrm{W} * \mathrm{T}} \qquad (4\text{-}1)$$

where $\mathrm{SM_{NLDAS}}$ represents the original soil moisture value and W is the density of water, or 1000 kilograms per cubic meter. T is the thickness of soil measured by NLDAS; in this case T is 0.1 m because NLDAS measures top 10-cm soil moisture.

We also notice the inconsistency of the soil depth measured by NLDAS and SMAP. The NLDAS measures the top 10 cm of the soil, 5 cm deeper than that of SMAP. Even though, according to Velpuri et al. (2016), SMAP shows a strong relationship with most soil moisture measurements at less than 20 cm depth (Colliander *et al.*, 2017b). We used a linear transformation to calibrate the two datasets. An example of the calibration on the NLDAS data is in Appendix A. Table A1 lists the calibration coefficients between NLDAS and SMAP.

4.2.3 Data Analysis

For each Julian day, there are 36 NLDAS observations from the past 36 years. Therefore, we were able to calculate the mean (μ_{NLDAS}) and standard deviation (σ_{NLDAS}) of each day. The daily SSI was calculated with Equation (2):

where x_{SMAP} is the soil moisture content from SMAP Level 3 data for a single day, μ_{NLDAS} is the

$$SSI = \frac{x_{SMAP} - \mu_{NLDAS}}{\sigma_{NLDAS}} \tag{4-2}$$

mean value of soil moisture content for the corresponding day from NLDAS, and σ_{NLDAS} is the standard deviation.

4.2.4 Validation

The SMAP mission specifies the volumetric accuracy of soil moisture to be within 0.04 (4%) m^3/m^3 in low or moderately vegetated areas in the following conditions (Colliander *et al.*, 2017b):

- Vegetation water content ≤ 5 kg/m^2
- Urban fraction ≤ 0.25
- Water fraction ≤ 0.1
- Digital Elevation Model (DEM) slope standard deviation ≤ 3 degrees

Unfortunately, the southeastern United States is not in the area where those accuracies are coherent. Therefore, we need to use other data sources to validate the soil moisture product. The validation was performed by comparing the soil moisture daily data from SMAP and NLDAS to daily soil moisture data retrieved from USDA SCAN stations. Table 4.2 lists the selected seven SCAN stations across the southeastern U.S. We selected the stations with a long-term collection of data, located in agricultural lands, plains, or grasslands, and reprehensive of diverse weather conditions. The comparison between SMAP and SCAN was on a daily basis, from 31 March 2015 to 16 July 2016. We also compared SCAN data and NLDAS data for 12 months, starting in January 2015 and ending in December 2015.

Table 4.2. Soil Climate Analysis Network (SCAN) Stations Used for Validation

Station ID	State Code	Station Name
2013	GA	Watkinsville #1
2024	MS	Goodwin Ck Pasture
2053	AL	Wtars
2039	VA	N Piedmont Arec
2005	KY	Princeton #1
2012	FL	Sellers Lake #1

SSI was validated by several soil moisture products, including PDSI and Moderate Resolution Infrared Spectroradiometer (MODIS) data. PDSI data for April 2015 were downloaded as NetCDF files in the WGS 1984 Geographic Coordinate System from the National Integrated Drought Information System on the U.S. Drought Portal. We derived a normalized difference water index (NDWI) from MODIS surface reflectance data (Gao, 1996; McFeeters, 1996; Zargar *et al.*, 2011):

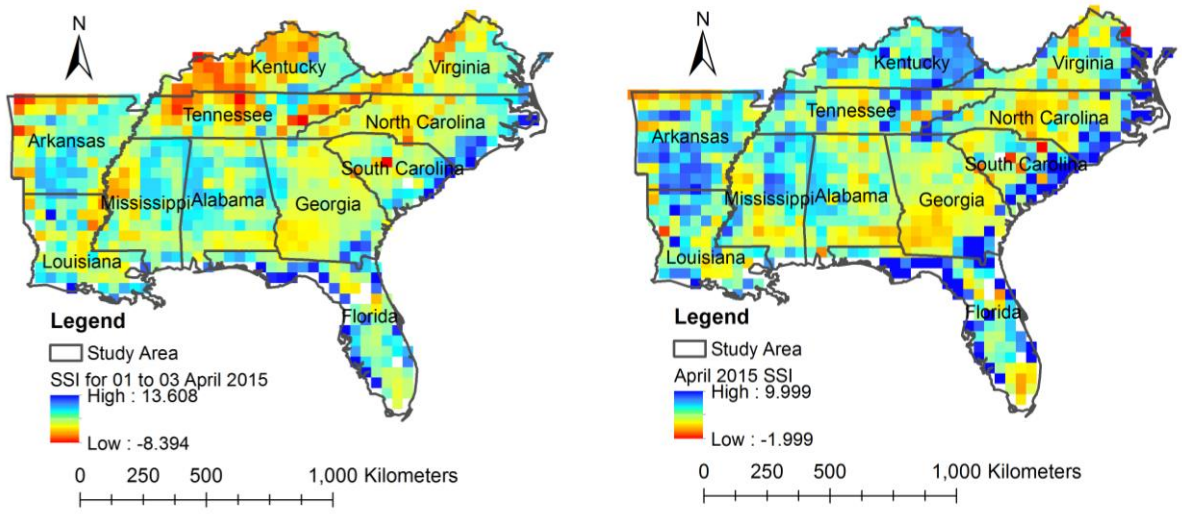
$$NDWI = \frac{NIR-SWIR}{NIR+SWIR} \quad (4\text{-}3)$$

where *NIR* is the near infrared reflectance and *SWIR* is the short-wave infrared reflectance of the MODIS data. Both PDSI and NDWI data were resampled to 36 km for the SSI validation.

4.3. Results

4.3.1 SSI Spatial Analysis

The first SMAP record was on 1 April 2015, when the SMAP radiometer began collecting routine science data. As SMAP requires a minimum of three consecutive days to cover the globe, the SSI results for each of the three consecutive days were mosaicked to cover the study area (Figure 4.2a). The standardized SSI is a z-score, indicating how many standard deviations that a SMAP value is from the historic mean. The yellow to red colors indicate negative z-scores, which means the values are lower than the historic soil moisture mean for those pixels. The green to blue colors indicate positive z-scores, which means the values are higher than the historic soil moisture average for those pixels.



(**a**) (**b**)

Figure 4.2. (a) Mosaic of the three consecutive standardized soil moisture index (SSI) maps from 1 to 3 April 2015. Areas in yellow to red represent areas that are experiencing very dry conditions, indicating drought; (b) SSI map for the whole month of April 2015.

Figure 4.2a reveals the regional climate variability for 1 to 3 April 2015 in the southeastern United States. Along the southeastern coastal area ranging from North Carolina to Florida, the soil moisture values were significantly above their historic means. This pattern diminishes as the distance inland increased. The high SSI values in the southern North Carolina and western Florida indicated a wet soil condition compared to the past 36 years. On the contrary, western Virginia and eastern Tennessee observed a below-average SSI, which indicated a dry soil condition compared to the past 36 years. Western Kentucky and Northwestern Tennessee experienced severe dry soil conditions. The remaining states, including Louisiana, Mississippi, Alabama, and most of Arkansas, Georgia, and part of Virginia, North Carolina, and South Carolina were in an average condition. Arkansas and western Louisiana generally experienced average soil moisture, with several lower values along the western border, and one pixel of higher value in northwestern Louisiana. SSI for April 2015, in contrast, shows the soil is generally wetter, except southern Mississippi, southern Alabama, Georgia, North Carolina and southern Florida (Figure 4.2b).

4.3.2 Validation of Results

SMAP Validation

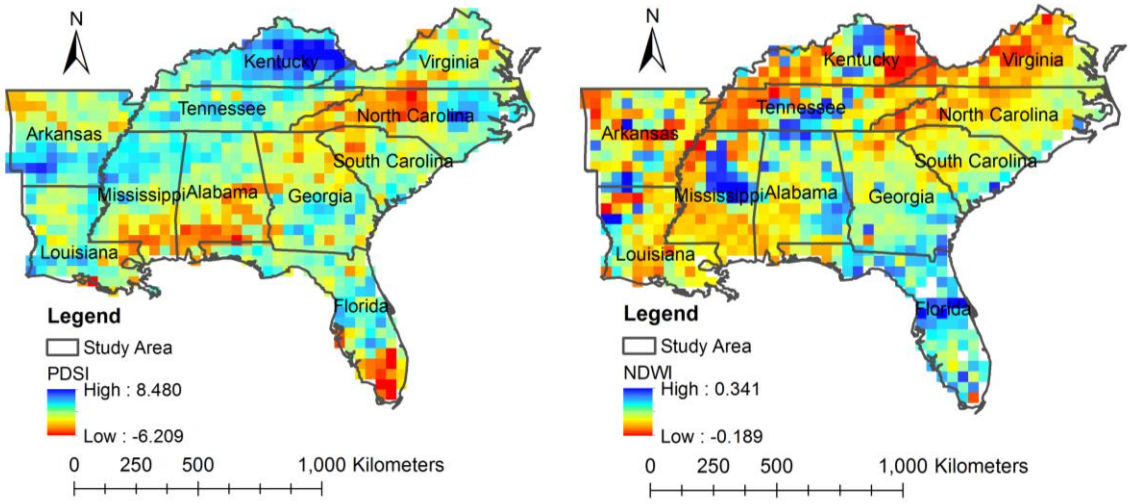
The correlations between SMAP soil moisture data and the SCAN data were between 0.1506 and 0.9177. The correlations between SCAN and NLDAS data were between 0.376 and 0.7742. The RMSEs for SMAP were between 0.0428 and 0.1379, which do not meet SMAP mission's specification (0.04 or 4% m^3/m^3) for low or moderately vegetated areas. Given that the southeastern United States are mostly covered by dense vegetation, the validation result is still acceptable for drought monitoring. Table 4.3 shows the R-squared and RMSE for SMAP in 2015 and 2016. Note that the correlation for station Uapb-Earle in 2016 was invalid due to missing SCAN data.

Table 4.3. Soil Moisture Active Passive (SMAP) Validation with SCAN Stations

Station ID	Station Name	R^2 for 2015	R^2 for 2016	RMSE for 2015	RMSE for 2016
2013	Watkinsville #1	0.6802	0.9124	0.0567	0.0791
2024	Goodwin Ck Pasture	0.7634	0.6817	0.0795	0.0591
2053	Wtars	0.4612	0.9177	0.0624	0.0428
2039	N Piedmont Arec	0.5783	0.2499	0.0712	0.0774
2005	Princeton #1	0.3115	0.5144	0.0762	0.0526
2012	Sellers Lake #1	0.2827	0.468	0.1288	0.1379
2085	Uapb-Earle	0.1506	N/A	0.0983	N/A
Average		0.4611	0.6240	0.0819	0.0748

Validation with PDSI and NDWI

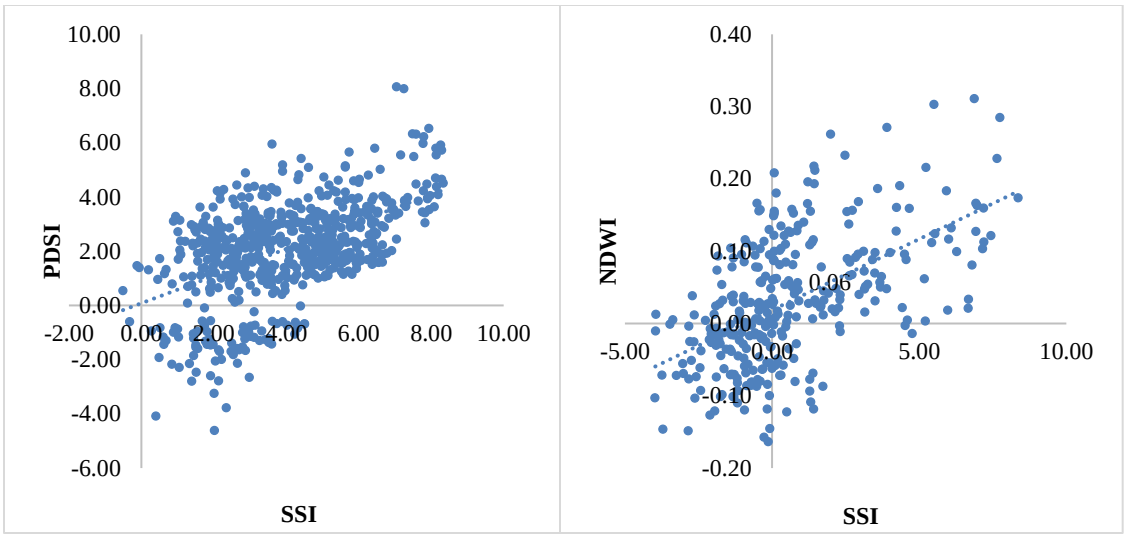
PDSI is a standardized index that spans −10 (dry) to +10 (wet) (Dai *et al.*, 2004). Figure 4.3a shows the reference image of PDSI for April 2015. Compare to the SSI result for April 2015 (Figure 4.2b), the drought patterns are generally consistent. The scatter plot shows the correlation between SSI and PDSI is moderate: the correlation coefficient (r) was 0.52 (Figure 4.4a). PDSI is effective in determining long-term drought (Dai *et al.*, 2004), but not for short time periods such as daily soil moisture deficiency. For daily comparison, MODIS NDWI was used to test the accuracy of short-term SSI.



(a) (b)

Figure 4.3. (a) Palmer drought severity index (PDSI) for April 2015. Areas in yellow and red represent areas that are experiencing dry conditions; (b) Normalized difference water index (NDWI) calculated for 01 to 03 April 2015. Likewise, areas in yellow and red represent areas that are experiencing low vegetation water content and therefore a dry condition.

NDWI is dimensionless and ranges between −1 (low vegetation water content) to +1 (high vegetation water content) (Gao, 1996). Figure 4.3b shows that the NDWI calculated for 01 to 03 April 2015 has a quite different pattern from the PDSI for April 2015 (Figure 4.3a), but has a very similar spatial pattern compared with the SSI for 1 to 3 April 2015 (Figure 4.2a). The dry condition monitored through NDWI in western Kentucky and western Tennessee matches the low-value areas by SSI. The wet condition in Florida from NDWI was also observed from SSI. This suggests that SSI is more sensitive than PDSI for short-term drought monitoring. A scatter plot shows the correlation between SSI and NDWI is strong: the correlation coefficient (r) value was 0.56 (Figure 4.4b).



(a) (b)

Figure 4.4. (a) Scatter plot for April 2015. The correlation between SSI and PDSI is moderate (r = 0.52); (b) Scatter plot for 1 to 3 April 2015. The correlation between SSI and NDWI is strong (r = 0.56).

4.4. Discussion

The SMAP validation revealed that the average correlation between SMAP data and SCAN data were 0.4611 for 2015 and 0.6240 for 2016. Four low R-squared values suggested some discrepancy between SMAP and SCAN data. The R-squared at the Uapb-Earle station in Arkansas for the year 2015 was exceptionally low (0.1506). Low R-squared values were also found at the N Piedmont Arec station in Virginia (for 2016), the Sellers Lake #1 station in Florida (for 2015), and the Princeton #1 station in Kentucky (for 2015). Figure 4.5 shows the correlations between SCAN values and SMAP values for the four stations.

To discover what caused these significantly low accuracies in the abovementioned stations, we created time-series plots to identify the outliers between SMAP and SCAN (Figure 4.5).

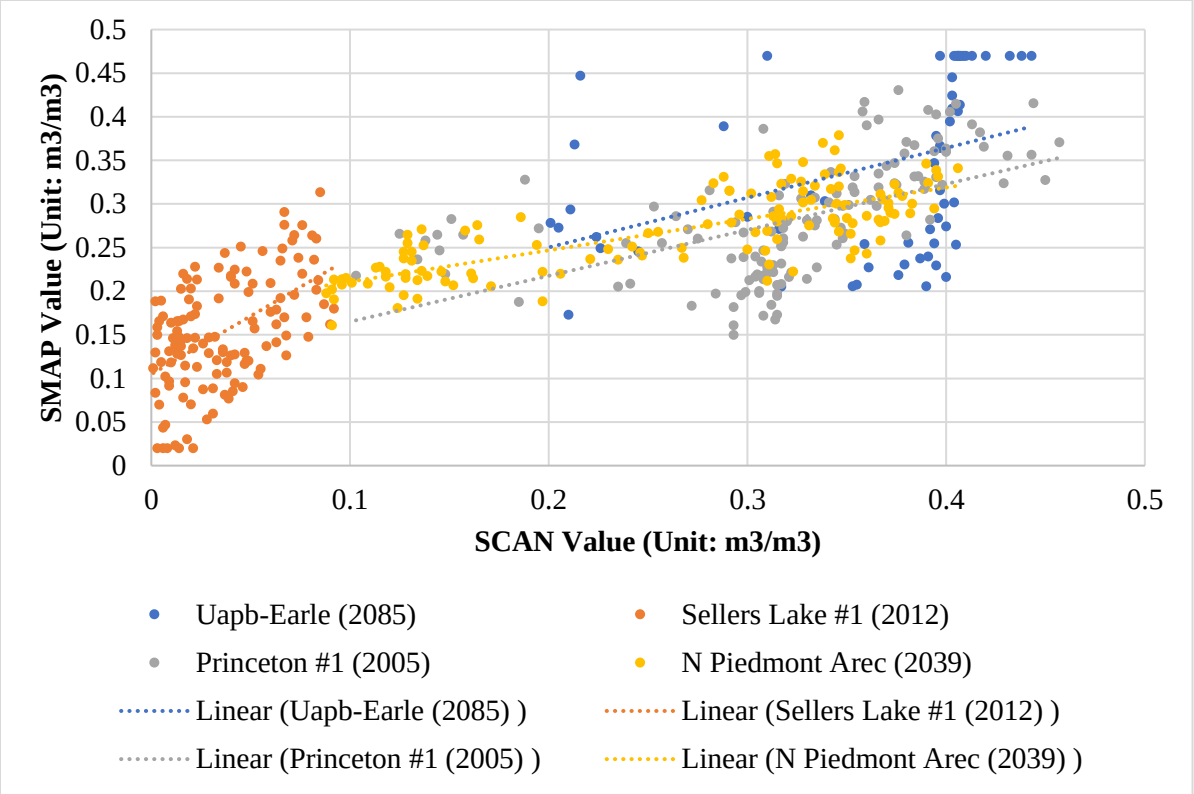
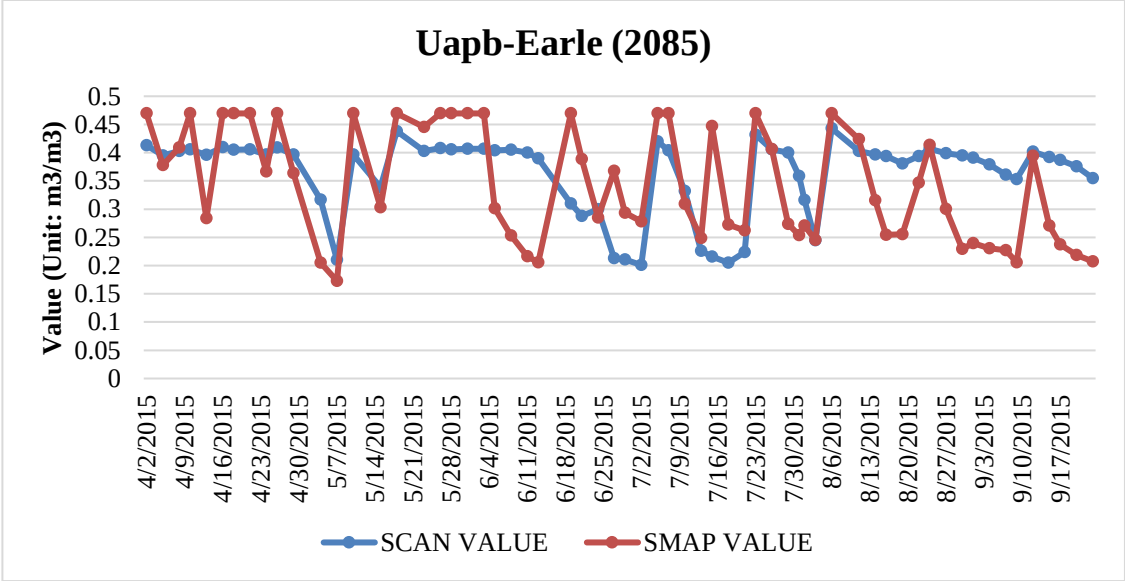
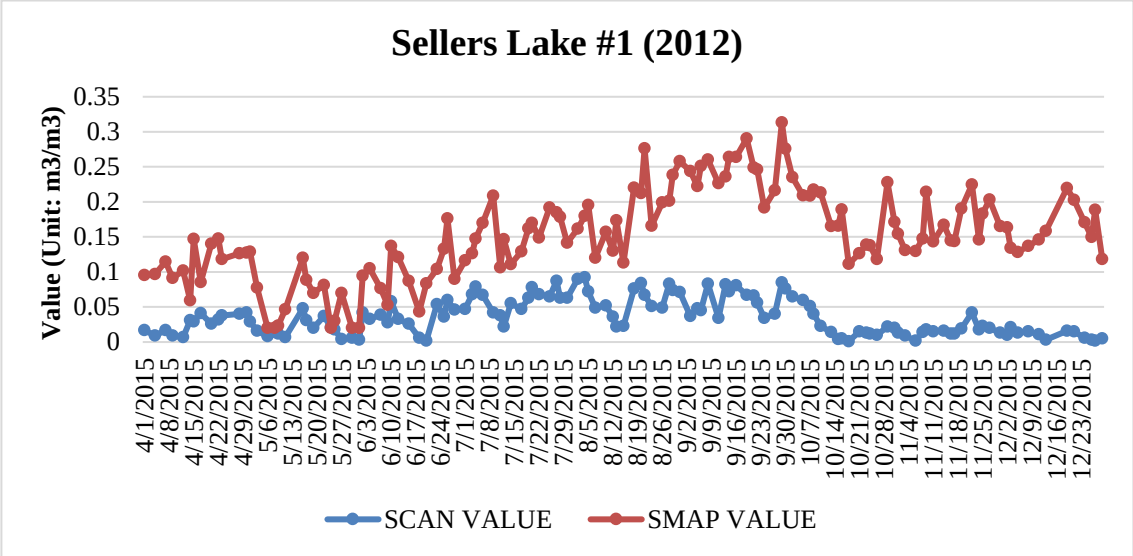


Figure 4.5. Scatter plot between SCAN values and SMAP values for the four anomaly stations with the R-squared values: Uapb-Earle station in Arkansas (for the year 2015), R-squared value was 0.1506; N Piedmont Arec station in Virginia (for 2016), R-squared value was 0.2499; the Sellers Lake #1 station in Florida (for 2015), R-squared value was 0.2827; and the Princeton #1 station in Kentucky (for 2015), R-squared value was 0.3115.

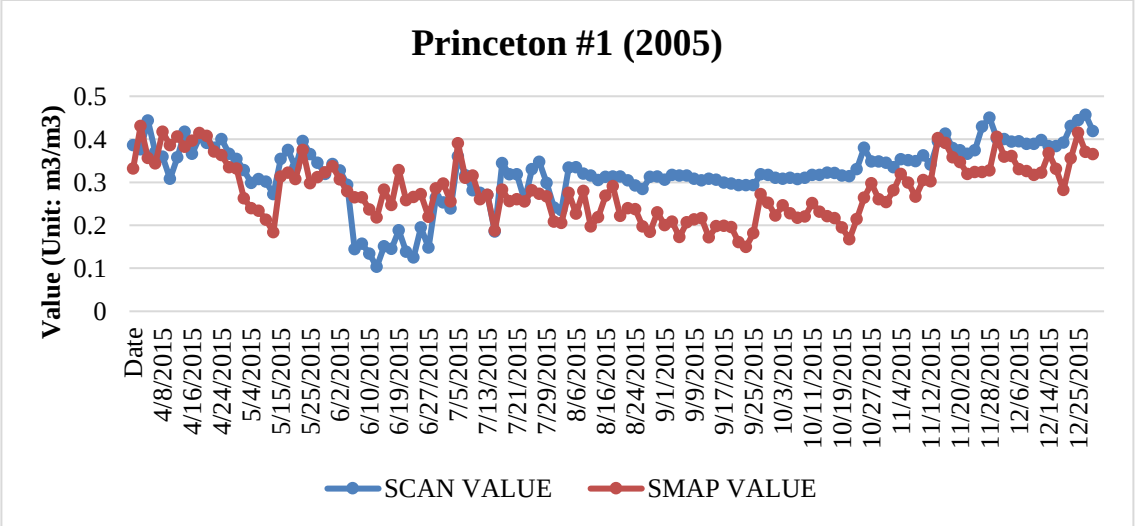


(a)

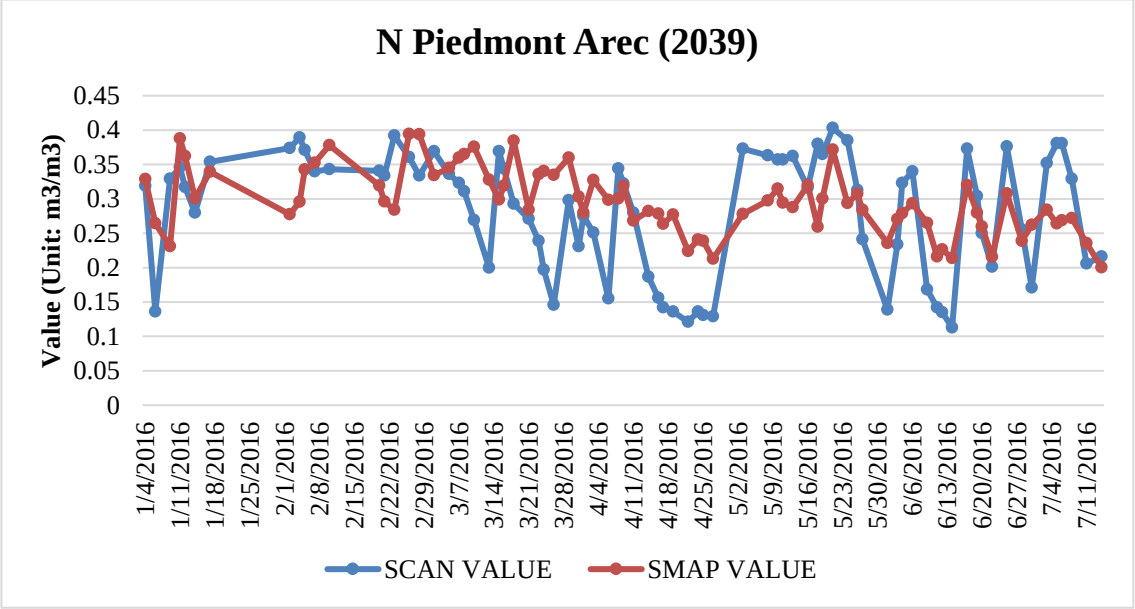


(b)

Figure 4.6. SMAP and SCAN Time Plot Comparison to Identify the Anomaly: (a) Uapb-Earle station in Arkansas (2085); (b) Sellers Lake #1 station in Florida (2012); (c) Princeton #1 station in Kentucky (2005); and (d) N Piedmont Arec station in Virginia (2039).
(figure cont'd.)



(c)



(d)

Figure 4.6a shows that at the UAPB-Earle station, the SMAP data and SCAN data are generally correlated, but the correlation drastically changed from around 4 June to 25 June, and then from 13 August to 10 September. Figure 4.6b shows the large disagreement at the Sellers Lake station. SCAN soil moisture values remained below 0.1 while SMAP abruptly jumped above 0.25 from August 21 to September 30. Figure 4.6c shows that at Princeton station, the agreement

between SMAP and SCAN was good from April 2015 to May. Nevertheless, this harmony was broken on 5 June until 27 June. In these days, SCAN observations were between 0.1 and 0.2, while SMAP data were between 0.2 to 0.3. From 6 August to 15 November, SMAP data were consistently lower than the SCAN data. Figure 4.6d shows that at N Piedmont Arec station there was a reverse trend between SMAP and SCAN from 18 January to 1 February, then 5 March to 18 March, and then from 8 April to 14 April. One of the major limitations of using SCAN data for satellite data validation is that the scales are different. The station observations are from precise sensors buried in soil, which only see a few inches of soil volume, while satellite sensors collect surface radiance from a large footprint (e.g., SMAP ~ 36 × 36 km) (Crow *et al.*, 2012). The satellite data are complex averages of the surface conditions and environments. Therefore, although direct comparison between the two datasets has been a common approach, it may not offer sufficient accuracy assessment of the satellite data.

Another limitation is the high SSI scores along the coastal areas in Figure 4.2. It has been reported in the literature that open water might lead to considerably biased soil moisture retrievals (Wu *et al.*, 2016). Making corrections of the coastal soil moisture data would require huge amount of efforts and additional data, including detailed land cover data and in-situ observations at much finer resolutions.

The last limitation of the SSI is that the calculation was based on the normality assumption of the historical data. The outliers (droughts) detected by the deviation from means are only valid if the assumption holds. Therefore, this approach is sensitive to data noise. Although we used 36 years of data to calculate the means and standard deviations, the SSI model will benefit from including longer periods if possible.

4.5. Conclusions

This research proposes a climate index called the standardized soil moisture index (SSI) to detect droughts. SSI was derived from satellite soil moisture data of SMAP and the long-term land surface model NLDAS data to facilitate short-term drought detection such as three days. By doing so, drought warnings can reach the farmers and foresters at a timely fashion.

Validation of the accuracies of the input data suggests that the SMAP soil moisture data displays good statistical correlations (R^2 = 0.4611 for 2015 and 0.6240 for 2016) with in-situ SCAN data and with acceptable RMSEs (0.0819 for 2015 and 0.0748 for 2016). However, we found large inconsistency in areas that are not friendly for satellite observations, such as vegetation, water bodies, urban, and high slope terrains.

The validation of SSI through PDSI and NDWI suggested SSI was an effective measure of soil moisture conditions. The correlation between SSI and PDSI for April 2015 is acceptable (r = 0.52), and the correlation between SSI and NDWI is slightly better (r = 0.56). Because PDSI is a monthly/weekly index. Therefore, SSI could provide shorter-term warnings than PDSI. Thus, SSI is a favorable index over PDSI for drought detection.

In summary, our SSI is a new climate index for drought detection. It is computed from daily satellite data and statistics from long-term soil moisture data, and therefore can provide short-term warning of drought conditions. Moreover, SSI is easy to interpret for farmers and foresters due to its simple and transparent statistical construct. Our research validated the SSI using multiple external sources of soil moisture data. The inconsistency of satellite observations with ground data could be solved by downscaling satellite data in future work.

CONCLUSION

This dissertation focusses on a major challenge in soil moisture mapping and spatio-temporal scaling, both of which are also major challenges in geography. The methodology includes using machine learning based approaches and spatial statistics to reconcile the scale difference between the satellite and station soil moisture observations. To this end, the following has been done in this dissertation: obtain satellite soil moisture products from passive microwave sources such as SMAP, and evaluate their accuracy and uncertainty with ground truth data; develop a soil moisture model that assimilates satellite data, probe data, and GIS spatial statistics to make appropriate predictions of soil moisture; develop a soil moisture model that upscales station observation data based on GIS-based spatial statistics methods to make reliable validation of the satellite observations; apply the soil moisture observations to an agricultural simulation for application in agricultural drought monitoring.

The conclusions, based on the research questions, have been answered. Specifically:

(1) The soil moisture product from SMAP-radiometer is not accurate when directly compared with ground truth data from the West Texas Mesonet;

(2) In-situ network data do not match the satellite pixels, but applying the spatial statistical methods is a feasible way to reconcile the scale difference between the two types of data sources;

(3) By using multi-sensor data fusion/assimilation to improve the SMAP Level 3 product, the downscaled satellite data can achieve improved spatial resolution, and smaller bias;

(4) Soil moisture prediction models can be applied to agricultural drought studies that are associated with climate change scenarios.

The present work has its novelty in using spatial statistic method to reconcile the scale difference from satellite data and ground observations, and therefore proposes new theories and solutions for dealing with the modifiable areal unit problem (MAUP) incurred in soil moisture mapping from satellite and ground stations.

One future work is on sample size effect for the stations in your upscaling work. This is to answer what is the minimal number of stations required for a reliable validation. Eliminate some stations and see its effect on RMSE. Use both upscaling and downscaling data sets to check its impact on prediction accuracy.

The second future work is spatio-temporal modeling. The spatio-temporal analysis of soil moisture is very important for flooding prediction or crop yield estimation. This future research will focus on the representation of spatio-temporal pattern of soil moisture based on multiscale and multisource data. The datasets will include satellite data from SMAP, ground station data collected from Texas Soil Observation Network (TxSON), West Texas Mesonet, Oklahoma Mesonet, and field data collection from soil moisture probes, as well as ancillary datasets from remote sensing observations. The methodology will be based on land surface models (e.g. Noah-MP) and hydrological models (e.g. WRF-Hydro) as well as soil moisture spatio-temporal modeling with deep leaning (neural network), data assimilation system (HRLDAS), and geospatial analysis (spatio-temporal regressing kriging). Triple collocation will be used to compare the performance of each methodology, and thus create an optimal model for the spatio-temporal soil moisture patterns at various regional scales.

Another future research is crop yield modeling for drought areas based on soil moisture and ancillary data. It is estimated that the 2012 drought in the Midwest led to harvest failure costing an estimated $30 billion. The 1994 drought in China caused an economic loss of almost 13.8 billion

U.S. dollars. Agricultural drought is defined by a deficit in the amount of moisture in the soil. My hypothesis is, having continuous soil moisture measurements will lead to improving crop yield forecasting, and irrigation planning. Using the rice yield and sugar cane yield data from 1980 to 2016, the research aims to build a model with long term soil moisture data (e.g. 36 years of soil moisture from North America Land Data Assimilation System (NLDAS)) and environmental variables from satellite products. Deep learning algorithms and Noah-MP-Crop land surface model will be adapted to test the hypothesis. With the proposed research, I expect to discover the climate impacts on the food-energy-water nexus and contribute to a stable water resource and food availability at regional scales.

APPENDIX: CALIBRATION OF NLDAS DATA

The calibration of NLDAS data for calculation SSI.

Table A1 shows the statistic information obtained from ArcGIS.

Using April 01, 2015 as an example,

Mean (NLDAS) = 23.705

STD (NLDAS) = 5.5943

Mean (SMAP) = 0.35731

STD (SMAP) = 0.11158

Therefore,

Mean (NLDAS) = 66.343 × Mean(SMAP)

STD (NLDAS) = 50.137 × STD(SMAP)

in which NLDAS Value = 100 times of SMAP Value because of the unit difference.

Therefore, in the equation of SSI, we should divide the mean of NLDAS by 66.343, and divide the STD by 50.137 to get a predicted SMAP value.

Table A1. NLDAS Calibration

Date	NLDAS Mean	SMAP Mean	N/S	NLDAS Std	SMAP Std	N/S
April 1, 2015	23.705	0.35731	66.34	5.5943	0.11158	50.13
April 2, 2015	23.413	0.35734	65.52	5.7149	0.08671	65.91
April 3, 2015	23.912	0.40322	59.30	6.6633	0.09153	72.80
April 1, 2016	26.860	0.38171	70.37	6.5042	0.08346	77.93
April 2, 2016	26.087	0.44331	58.84	5.4096	0.07346	73.64
April 3, 2016	24.792	0.41021	60.44	5.4099	0.08887	60.88
April 1, 2017	23.829	0.36315	65.62	6.6062	0.10416	63.43
April 2, 2017	23.036	N/A[1]	N/A [1]	6.4982	N/A [1]	N/A [1]
April 3, 2017	25.442	0.36161	70.36	8.3495	0.12412	67.27

[1] 2 April 2017 calibration was not available due to the missing data for SMAP.

REFERENCES

Ågren, A. M., Lidberg, W., Strömgren, M., Ogilvie, J., & Arp, P. A. (2014). Evaluating digital terrain indices for soil wetness mapping – a Swedish case study. *Hydrology and Earth System Sciences, 18*(9), 3623-3634. doi:10.5194/hess-18-3623-2014

Akbar, R., & Moghaddam, M. (2015). A Combined Active–Passive Soil Moisture Estimation Algorithm With Adaptive Regularization in Support of SMAP. *IEEE Transactions on Geoscience and Remote Sensing, 53*(6), 3312-3324. doi:10.1109/TGRS.2014.2373972

Al-Yaari, A., Wigneron, J. P., Kerr, Y., Rodriguez-Fernandez, N., O'Neill, P. E., Jackson, T. J., . . . Yueh, S. (2017). Evaluating soil moisture retrievals from ESA's SMOS and NASA's SMAP brightness temperature datasets. *Remote Sensing of Environment, 193*(Supplement C), 257-273. doi:https://doi.org/10.1016/j.rse.2017.03.010

Albertson, J., & Kiely, G. (2001). On the structure of soil moisture time series in the context of land surface models. *Journal of Hydrology, 243*(1), 101-119.

Atkinson, P. M. (2013). Downscaling in remote sensing. *International Journal of Applied Earth Observation and Geoinformation, 22*(Supplement C), 106-114. doi:https://doi.org/10.1016/j.jag.2012.04.012

Babaud, J., Witkin, A. P., Baudin, M., & Duda, R. O. (1986). Uniqueness of the Gaussian kernel for scale-space filtering. *IEEE Transactions on Pattern Analysis & Machine Intelligence*(1), 26-33.

Basharinov, A. Y., & Shutko, A. (1975). Simulation studies of the SHF radiation characteristics of soils under moist conditions. *Simulation studies of the SHF radiation characteristics of soils under moist conditions NASA Transl. into ENGLISH of"" Modelnyye Issledovaniya SVCh Radiatsionnykh Kharacteristik Pochvo-Gruntov v Usloviyakh Uvlazhneniya"(unpublished report) Moscow, Acad. of Sci. of the USSR, Inst. of Radio Eng. and Electronics, 1975 p 1-81*.

Beven, K. J., & Kirkby, M. J. (1979a). A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Bulletin, 24*(1), 43-69. doi:10.1080/02626667909491834

Beven, K. J., & Kirkby, M. J. (1979b). A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrol. Sci. J., 24*(1), 43-69.

Bhuiyan, H. A., McNairn, H., Powers, J., Friesen, M., Pacheco, A., Jackson, T. J., . . . Rowlandson, T. (2018). Assessing SMAP soil moisture scaling and retrieval in the Carman (Canada) study site. *Vadose Zone Journal, 17*(1).

Brassel, K. E., & Reif, D. (1979). A Procedure to Generate Thiessen Polygons. *Geographical Analysis, 11*(3), 289-303. doi:10.1111/j.1538-4632.1979.tb00695.x

Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5-32.

Brocca, L., Morbidelli, R., Melone, F., & Moramarco, T. (2007). Soil moisture spatial variability in experimental areas of central Italy. *Journal of Hydrology, 333*(2), 356-373. doi:https://doi.org/10.1016/j.jhydrol.2006.09.004

Burgess, T., & Webster, R. (1980a). Optimal interpolation and isarithmic mapping of soil properties. *European Journal of Soil Science, 31*(2), 333-341.

Burgess, T. M., & Webster, R. (1980b). OPTIMAL INTERPOLATION AND ISARITHMIC MAPPING OF SOIL PROPERTIES. *Journal of Soil Science, 31*(2), 333-341. doi:10.1111/j.1365-2389.1980.tb02085.x

Chan, S. K., Bindlish, R., O'Neill, P., Jackson, T., Njoku, E., Dunbar, S., . . . Kerr, Y. (2018). Development and assessment of the SMAP enhanced passive soil moisture product. *Remote Sensing of Environment, 204*, 931-941. doi:https://doi.org/10.1016/j.rse.2017.08.025

Chan, S. K., Bindlish, R., O'Neill, P. E., Njoku, E., Jackson, T., Colliander, A., . . . Piepmeier, J. (2016). Assessment of the SMAP passive soil moisture product. *IEEE Transactions on Geoscience and Remote Sensing, 54*(8), 4994-5007.

Change, I. P. o. C. (2014). *Climate Change 2014–Impacts, Adaptation and Vulnerability: Regional Aspects*: Cambridge University Press.

Chung, M. K. (2012). Gaussian kernel smoothing. *Lecture notes*, 1-10.

Clewley, D., Whitcomb, J. B., Akbar, R., Silva, A. R., Berg, A., Adams, J. R., . . . Moghaddam, M. (2017). A Method for Upscaling In Situ Soil Moisture Measurements to Satellite Footprint Scale Using Random Forests. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10*(6), 2663-2673. doi:10.1109/JSTARS.2017.2690220

Colliander, A., Fisher, J. B., Halverson, G., Merlin, O., Misra, S., Bindlish, R., . . . Yueh, S. (2017a). Spatial Downscaling of SMAP Soil Moisture Using MODIS Land Surface Temperature and NDVI During SMAPVEX15. *IEEE Geoscience and Remote Sensing Letters, 14*(11), 2107-2111. doi:10.1109/LGRS.2017.2753203

Colliander, A., Jackson, T., Bindlish, R., Chan, S., Das, N., Kim, S., . . . Pashaian, L. (2017b). Validation of SMAP surface soil moisture products with core validation sites. *Remote Sensing of Environment, 191*, 215-231.

Colliander, A., Jackson, T. J., Bindlish, R., Chan, S., Das, N., Kim, S. B., . . . Yueh, S. (2017c). Validation of SMAP surface soil moisture products with core validation sites. *Remote Sensing of Environment, 191*, 215-231. doi:https://doi.org/10.1016/j.rse.2017.01.021

Colliander, A., Jackson, T. J., Chan, S. K., O'Neill, P., Bindlish, R., Cosh, M. H., . . . Yueh, S. H. (2018a). An assessment of the differences between spatial resolution and grid size for the SMAP enhanced soil moisture product over homogeneous sites. *Remote Sens. Environ., 207*, 65-70. doi:10.1016/j.rse.2018.02.006

Colliander, A., Jackson, T. J., Chan, S. K., O'Neill, P., Bindlish, R., Cosh, M. H., . . . Yueh, S. H. (2018b). An assessment of the differences between spatial resolution and grid size for the SMAP enhanced soil moisture product over homogeneous sites. *Remote Sensing of Environment, 207*, 65-70. doi:https://doi.org/10.1016/j.rse.2018.02.006

Cosh, M. H., Jackson, T. J., Bindlish, R., & Prueger, J. H. (2004). Watershed scale temporal and spatial stability of soil moisture and its role in validating satellite estimates. *Remote Sensing of Environment, 92*(4), 427-435.

Cosh, M. H., Jackson, T. J., Starks, P., & Heathman, G. (2006). Temporal stability of surface soil moisture in the Little Washita River watershed and its applications in satellite soil moisture product validation. *Journal of Hydrology, 323*(1), 168-177. doi:https://doi.org/10.1016/j.jhydrol.2005.08.020

Crow, W. T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., . . . Walker, J. P. (2012). Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. *Reviews of Geophysics, 50*(2).

Crow, W. T., Ryu, D., & Famiglietti, J. S. (2005). Upscaling of field-scale soil moisture measurements using distributed land surface modeling. *Advances in Water Resources, 28*(1), 1-14. doi:https://doi.org/10.1016/j.advwatres.2004.10.004

Crow, W. T., & Yilmaz, M. T. (2014). The auto-tuned land data assimilation system (ATLAS). *Water Resources Research, 50*(1), 371-385.

Dai, A., Trenberth, K. E., & Qian, T. (2004). A global dataset of Palmer Drought Severity Index for 1870–2002: relationship with soil moisture and effects of surface warming. *Journal of Hydrometeorology, 5*(6), 1117-1130.

Das, N. N., Entekhabi, D., & Njoku, E. G. (2011). An algorithm for merging SMAP radiometer and radar data for high-resolution soil-moisture retrieval. *IEEE Transactions on Geoscience and Remote Sensing, 49*(5), 1504-1512.

De Rosnay, P., Drusch, M., Boone, A., Balsamo, G., Decharme, B., Harris, P., . . . Wigneron, J.-P. (2009a). AMMA Land Surface Model Intercomparison Experiment coupled to the Community Microwave Emission Model: ALMIP-MEM. *Journal of Geophysical Research: Atmospheres, 114*(D5). doi:10.1029/2008jd010724

De Rosnay, P., Drusch, M., Boone, A., Balsamo, G., Decharme, B., Harris, P., . . . Wigneron, J. P. (2009b). AMMA Land Surface Model Intercomparison Experiment coupled to the Community Microwave Emission Model: ALMIP-MEM. *Journal of Geophysical Research: Atmospheres, 114*(D5).

Edwards, D., & McKee, T. (1997). Characteristics of 20th century drought in the United States at multiple time scales. Fort Collins, Colorado. *Colorado State University, Dept. of Atmospheric Science*.

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., . . . Johnson, J. (2010). The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE, 98*(5), 704-716.

Entekhabi, D., Yueh, S., O'Neill, P. E., Kellogg, K. H., Allen, A., Bindlish, R., . . . Crow, W. T. (2014). SMAP handbook–soil moisture active passive: Mapping soil moisture and freeze/thaw from space.

Famiglietti, J. S., Devereaux, J. A., Laymon, C. A., Tsegaye, T., Houser, P. R., Jackson, T. J., . . . van Oevelen, P. J. (1999). Ground-based investigation of soil moisture variability within remote sensing footprints During the Southern Great Plains 1997 (SGP97) Hydrology Experiment. *Water Resources Research, 35*(6), 1839-1851. doi:10.1029/1999wr900047

Fang, B., & Lakshmi, V. (2014). Soil moisture at watershed scale: Remote sensing techniques. *Journal of Hydrology, 516*, 258-272. doi:https://doi.org/10.1016/j.jhydrol.2013.12.008

Fang, B., Lakshmi, V., Bindlish, R., & Jackson, T. J. (2018). Downscaling of SMAP Soil Moisture Using Land Surface Temperature and Vegetation Data. *Vadose Zone Journal, 17*. doi:10.2136/vzj2017.11.0198

Field, C. B. (2012). *Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change*: Cambridge University Press.

Fotheringham, A. S., & Oshan, T. M. (2016). Geographically weighted regression and multicollinearity: dispelling the myth. *Journal of Geographical Systems, 18*(4), 303-329. doi:10.1007/s10109-016-0239-5

Gao, B.-c. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment, 58*(3), 257-266. doi:http://dx.doi.org/10.1016/S0034-4257(96)00067-3

Grayson, R. B., Western, A. W., Chiew, F. H. S., & Blöschl, G. (1997). Preferred states in spatial soil moisture patterns: Local and nonlocal controls. *Water Resources Research, 33*(12), 2897-2908. doi:10.1029/97wr02174

Griffith, D. A. (2008). Spatial-Filtering-Based Contributions to a Critique of Geographically Weighted Regression (GWR). *Environment and Planning A: Economy and Space, 40*(11), 2751-2769. doi:10.1068/a38218

Hassan-Esfahani, L., Torres-Rua, A., Jensen, A., & McKee, M. (2015). Assessment of surface soil moisture using high-resolution multi-spectral imagery and artificial neural networks. *Remote Sensing, 7*(3), 2627-2646.

Hayes, M. J. (2006). *Drought indices*: Wiley Online Library.

Hong, Y., Zhang, Y., & Khan, S. I. (2016). *Hydrologic Remote Sensing: Capacity Building for Sustainability and Resilience*: CRC Press.

Illston, B. G., Basara, J. B., & Crawford, K. C. (2004). Seasonal to interannual variations of soil moisture measured in Oklahoma. *International Journal of Climatology, 24*(15), 1883-1896.

Im, J., Park, S., Rhee, J., Baik, J., & Choi, M. (2016). Downscaling of AMSR-E soil moisture with MODIS products using machine learning approaches. *Environmental Earth Sciences, 75*(15), 1120. doi:10.1007/s12665-016-5917-6

Jin, Y., Ge, Y., Wang, J., Chen, Y., Heuvelink, G. B. M., & Atkinson, P. M. (2018). Downscaling AMSR-2 Soil Moisture Data With Geographically Weighted Area-to-Area Regression Kriging. *IEEE Transactions on Geoscience and Remote Sensing, 56*(4), 2362-2376. doi:10.1109/TGRS.2017.2778420

Kang, J., Jin, R., & Li, X. (2015). Regression kriging-based upscaling of soil moisture measurements from a wireless sensor network and multiresource remote sensing information over heterogeneous cropland. *IEEE Geoscience and Remote Sensing Letters, 12*(1), 92-96.

Kerry, R., Goovaerts, P., Rawlins, B. G., & Marchant, B. P. (2012). Disaggregation of legacy soil data using area to point kriging for mapping soil organic carbon at the regional scale. *Geoderma, 170*(Supplement C), 347-358. doi:https://doi.org/10.1016/j.geoderma.2011.10.007

Keyantash, J., & Dracup, J. A. (2002). The quantification of drought: an evaluation of drought indices. *Bulletin of the American Meteorological Society, 83*(8), 1167-1180.

Knipper, K. R., Hogue, T. S., Franz, K. J., & Scott, R. L. (2017). *Downscaling SMAP and SMOS soil moisture with moderate-resolution imaging spectroradiometer visible and infrared products over southern Arizona*.

Koster, R. D., & Suarez, M. J. (2001). Soil moisture memory in climate models. *Journal of Hydrometeorology, 2*(6), 558-570.

Krivoruchko, K. (2011). *Spatial statistical data analysis for GIS users*: Esri Press Redlands.

Kwon, M., Kwon, H.-H., & Han, D. (2018). A spatial downscaling of soil moisture from rainfall, temperature, and AMSR2 using a Gaussian-mixture nonstationary hidden Markov model. *Journal of Hydrology, 564*, 1194-1207. doi:https://doi.org/10.1016/j.jhydrol.2017.12.015

Kyriakidis, P. C. (2004). A Geostatistical Framework for Area-to-Point Spatial Interpolation. *Geographical Analysis, 36*(3), 259-289. doi:doi:10.1111/j.1538-4632.2004.tb01135.x

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news, 2*(3), 18-22.

Liu, Y., Yang, Y., Jing, W., & Yue, X. (2018). Comparison of Different Machine Learning Approaches for Monthly Satellite-Based Soil Moisture Downscaling over Northeast China. *Remote Sensing, 10*(1), 31.

Loew, A., & Mauser, W. (2008). On the Disaggregation of Passive Microwave Soil Moisture Data Using <emphasis emphasistype="italic">A Priori </emphasis> Knowledge of Temporally Persistent Soil Moisture Fields. *IEEE Transactions on Geoscience and Remote Sensing, 46*(3), 819-834. doi:10.1109/TGRS.2007.914800

Lopez-Molina, C., De Baets, B., Bustince, H., Sanz, J., & Barrenechea, E. (2013). Multiscale edge detection based on Gaussian smoothing and edge tracking. *Knowledge-Based Systems, 44*, 101-111.

McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing, 17*(7), 1425-1432. doi:10.1080/01431169608948714

McKee, T. B., Doesken, N. J., & Kleist, J. (1993). *The relationship of drought frequency and duration to time scales.* Paper presented at the Proceedings of the 8th Conference on Applied Climatology.

McNulty, S., Director, S. H., Wiener, S., Treasure, S. H., Myers, J. M., Farahani, H., . . . Steele, R. F. (2015). Southeast regional climate hub assessment of climate change vulnerability and adaptation and mitigation strategies. *Agriculture Research Service, 2015*, 1-61.

Melillo, J. M., Richmond, T., & Yohe, G. (2014). Climate change impacts in the United States. *Third National Climate Assessment*.

Meng, L., & Quiring, S. M. (2008). A comparison of soil moisture models using Soil Climate Analysis Network observations. *Journal of Hydrometeorology, 9*(4), 641-659.

Merlin, O., Rudiger, C., Bitar, A. A., Richaume, P., Walker, J. P., & Kerr, Y. H. (2012). Disaggregation of SMOS Soil Moisture in Southeastern Australia. *IEEE Transactions on Geoscience and Remote Sensing, 50*(5), 1556-1571. doi:10.1109/TGRS.2011.2175000

Merlin, O., Walker, J. P., Chehbouni, A., & Kerr, Y. (2008). Towards deterministic downscaling of SMOS soil moisture using MODIS derived soil evaporative efficiency. *Remote Sensing of Environment, 112*(10), 3935-3946. doi:https://doi.org/10.1016/j.rse.2008.06.012

Mohanty, B. P., Cosh, M. H., Lakshmi, V., & Montzka, C. (2017). Soil Moisture Remote Sensing: State-of-the-Science. *Vadose Zone Journal, 16*(1). doi:10.2136/vzj2016.10.0105

Mohanty, B. P., & Skaggs, T. H. (2001). Spatio-temporal evolution and time-stable characteristics of soil moisture within remote sensing footprints with varying soil, slope, and vegetation.

Advances in Water Resources, 24(9), 1051-1067. doi:https://doi.org/10.1016/S0309-1708(01)00034-3

Montzka, C., Rötzer, K., Bogena, H. R., Sanchez, N., & Vereecken, H. (2018). A New Soil Moisture Downscaling Approach for SMAP, SMOS, and ASCAT by Predicting Sub-Grid Variability. *Remote Sensing, 10*(3), 427.

NASA. (2014). *SMAP Handbook*. In *Mapping Soil Moisture and Freeze/Thaw from Space*.

Nash, M. S., Wierenga, P. J., & Gutjahr, A. (1991). TIME SERIES ANALYSIS OF SOIL MOISTURE AND RAINFALL ALONG A LINE TRANSECT IN ARID RANGELAND. *Soil Science, 152*(3), 189-198.

Njoku, E. G., Wilson, W. J., Yueh, S. H., Dinardo, S. J., Li, F. K., Jackson, T. J., . . . Bolten, J. (2002). Observations of soil moisture using a passive and active low-frequency microwave airborne sensor during SGP99. *IEEE Transactions on Geoscience and Remote Sensing, 40*(12), 2659-2673. doi:10.1109/TGRS.2002.807008

NOAA. (2012). Fact Sheet: Drought. Retrieved from http://www.nws.noaa.gov/om/csd/graphics/content/outreach/brochures/FactSheet_Drought.pdf

Oglesby, R. J., & Erickson III, D. J. (1989). Soil moisture and the persistence of North American drought. *Journal of Climate, 2*(11), 1362-1380.

Openshow. (1979). A million or so correlation coefficients, three experiments on the modifiable areal unit problem. *Statistical Applications in the Spatial Science*, 127-144.

Owe, M., de Jeu, R., & Walker, J. (2001). A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index. *IEEE Transactions on Geoscience and Remote Sensing, 39*(8), 1643-1654.

Palmer, W. C. (1965). *Meteorological drought* (Vol. 30): US Department of Commerce, Weather Bureau Washington, DC.

Park, N.-W. (2013). Spatial Downscaling of TRMM Precipitation Using Geostatistics and Fine Scale Environmental Variables. *Advances in Meteorology, 2013*, 9. doi:10.1155/2013/237126

Parry, M., Canziani, O. F., Palutikof, J. P., van der Linden, P. J., & Hanson, C. E. (2007). *Climate change 2007: impacts, adaptation and vulnerability* (Vol. 4): Cambridge University Press Cambridge.

Peng, J., Loew, A., Merlin, O., & Verhoest, N. E. C. (2017). A review of spatial downscaling of satellite remotely sensed soil moisture. *Reviews of Geophysics, 55*(2), 341-366. doi:10.1002/2016RG000543

Perry, M., & Niemann, J. (2008). Generation of soil moisture patterns at the catchment scale by EOF interpolation. *Hydrology and Earth System Sciences Discussions, 12*(1), 39-53.

Petropoulos, G. P., Ireland, G., & Barrett, B. (2015). Surface soil moisture retrievals from remote sensing: Current status, products & future trends. *Physics and Chemistry of the Earth, Parts A/B/C, 83*, 36-56.

Piepmeier, J. (2014). SMAP calibrated, time-ordered brightness temperatures L1B_TB data product.

Piles, M., Entekhabi, D., & Camps, A. (2009). A Change Detection Algorithm for Retrieving High-Resolution Soil Moisture From SMAP Radar and Radiometer Observations. *IEEE Transactions on Geoscience and Remote Sensing, 47*(12), 4125-4131. doi:10.1109/TGRS.2009.2022088

Piles, M., Sánchez, N., Vall-llossera, M., Camps, A., Martínez-Fernández, J., Martínez, J., & González-Gambau, V. (2014). A downscaling approach for SMOS land observations: Evaluation of high-resolution soil moisture maps over the Iberian Peninsula. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7*(9), 3845-3857.

Poe, G. A. (1990). Optimum interpolation of imaging microwave radiometer data. *IEEE Transactions on Geoscience and Remote Sensing, 28*(5), 800-810. doi:10.1109/36.58966

Qin, J., Zhao, L., Chen, Y., Yang, K., Yang, Y., Chen, Z., & Lu, H. (2015). Inter-comparison of spatial upscaling methods for evaluation of satellite-based soil moisture. *Journal of Hydrology, 523*, 170-178.

Quiring, S. M., & Papakryiakou, T. N. (2003). An evaluation of agricultural drought indices for the Canadian prairies. *Agricultural and forest meteorology, 118*(1), 49-62.

Reichle, R., Crow, W., Koster, R., Sharif, H., & Mahanama, S. (2008). Contribution of soil moisture retrievals to land data assimilation products. *Geophysical Research Letters, 35*(1).

Reichle, R. H., Entekhabi, D., & McLaughlin, D. B. (2001). Downscaling of radio brightness measurements for soil moisture estimation: A four-dimensional variational data assimilation approach. *Water Resources Research, 37*(9), 2353-2364. doi:doi:10.1029/2001WR000475

Ritchie, J. (1998). Soil water balance and plant water stress. In *Understanding options for agricultural production* (pp. 41-54): Springer.

Rowlandson, T., Impera, S., Belanger, J., Berg, A. A., Toth, B., & Magagi, R. (2015). Use of in situ soil moisture network for estimating regional-scale soil moisture during high soil moisture conditions. *Canadian Water Resources Journal / Revue canadienne des ressources hydriques, 40*(4), 343-351. doi:10.1080/07011784.2015.1061948

Segal, M. R. (2004). Machine learning benchmarks and random forest regression.

Sheikh, V., Visser, S., & Stroosnijder, L. (2009). A simple model to predict soil moisture: Bridging Event and Continuous Hydrological (BEACH) modelling. *Environmental Modelling & Software, 24*(4), 542-556.

Sivakumar, M., Motha, R., Wilhite, D., & Wood, D. (2010). Agricultural Drought indices proceedings of an expert meeting. 2–4 June 2010, Murcia, Spain. *World Meteorological Organization, Geneva, Switzerland*.

Srivastava, P. K., Han, D., Ramirez, M. R., & Islam, T. (2013). Machine Learning Techniques for Downscaling SMOS Satellite Soil Moisture Using MODIS Land Surface Temperature for Hydrological Application. *Water Resources Management, 27*(8), 3127-3144. doi:10.1007/s11269-013-0337-9

Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., & Feuston, B. P. (2003). Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *Journal of Chemical Information and Computer Sciences, 43*(6), 1947-1958. doi:10.1021/ci034160g

Thornthwaite, C. W., & Mather, J. (1955). The Water Balance. Centerton, Drexel Institute of Technology-Laboratory of Climatology, 104p. *Publications in climatology, 8*(1).

Ulaby, F., Moore, R., & Fung, A. (1986). Microwave remote sensing: Active and passive. Volume 3-From theory to applications.

UNL. (2014). What is the U.S. Drought Monitor? Retrieved from http://droughtmonitor.unl.edu/data/docs/USDMbrochure.pdf

USDA. (2015). SERCH LIGHTS Alerts Help Land Managers Prepare for Drought. Retrieved from https://forestthreats.org/research/projects/2015-accomplishment-highlights/serch-lights-alerts-help-land-managers-prepare-for-drought/

USDA. (2016, 22 June 2016). Soil Climate Analysis Network (SCAN) Brochure. Retrieved from https://www.wcc.nrcs.usda.gov/scan/SCAN_brochure.pdf

Vachaud, G., Passerat de Silans, A., Balabanis, P., & Vauclin, M. (1985). Temporal Stability of Spatially Measured Soil Water Probability Density Function 1. *Soil Science Society of America Journal, 49*(4), 822-828.

Velpuri, N. M., Senay, G. B., & Morisette, J. T. (2016). Evaluating New SMAP Soil Moisture for Drought Monitoring in the Rangelands of the US High Plains. *Rangelands, 38*(4), 183-190. doi:https://doi.org/10.1016/j.rala.2016.06.002

Wang, J., Ge, Y., Heuvelink, G. B. M., & Zhou, C. (2015a). Upscaling In Situ Soil Moisture Observations to Pixel Averages with Spatio-Temporal Geostatistics. *Remote Sensing, 7*(9), 11372-11388.

Wang, Q., Shi, W., Atkinson, P. M., & Zhao, Y. (2015b). Downscaling MODIS images with area-to-point regression kriging. *Remote Sensing of Environment, 166*(Supplement C), 191-204. doi:https://doi.org/10.1016/j.rse.2015.06.003

Westenbroek, S. M., Kelson, V., Dripps, W., Hunt, R., & Bradbury, K. (2010). *SWB-A modified Thornthwaite-Mather Soil-Water-Balance code for estimating groundwater recharge* (2328-7055). Retrieved from

Wheeler, D., & Tiefelsdorf, M. (2005). Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems, 7*(2), 161-187. doi:10.1007/s10109-005-0155-6

Wilson, J., & Gallant, J. (2000). *Digital Terrain Analysis in Terrain Analysis: Principles and Applications* (Vol. 479).

Wink, A. M., & Roerdink, J. B. (2004). Denoising functional MR images: a comparison of wavelet denoising and Gaussian smoothing. *IEEE transactions on medical imaging, 23*(3), 374-387.

Wu, Q., Liu, H., Wang, L., & Deng, C. (2016). Evaluation of AMSR2 soil moisture products over the contiguous United States using in situ data from the International Soil Moisture Network. *International Journal of Applied Earth Observation and Geoinformation, 45*(Part B), 187-199. doi:https://doi.org/10.1016/j.jag.2015.10.011

Xia, Y., Ek, M. B., Wu, Y., Ford, T., & Quiring, S. M. (2015). Comparison of NLDAS-2 Simulated and NASMD Observed Daily Soil Moisture. Part I: Comparison and Analysis. *Journal of Hydrometeorology, 16*(5), 1962-1980. doi:10.1175/jhm-d-14-0096.1

Xia, Y., et al. NCEP/EMC. (2016). *NLDAS Noah Land Surface Model L4 Hourly 0.125 x 0.125 degree V002, version 002*.

Xia, Y., Sheffield, J., Ek, M. B., Dong, J., Chaney, N., Wei, H., . . . Wood, E. F. (2014). Evaluation of multi-model simulated soil moisture in NLDAS-2. *Journal of Hydrology, 512*(Supplement C), 107-125. doi:https://doi.org/10.1016/j.jhydrol.2014.02.027

Xing, C., Chen, N., Zhang, X., & Gong, J. (2017). A Machine Learning Based Reconstruction Method for Satellite Remote Sensing of Soil Moisture Images with In Situ Observations. *Remote Sensing, 9*(5), 484.

Xu, Y., Wang, L., Ross, K., Liu, C., & Berry, K. (2018). Standardized Soil Moisture Index for Drought Monitoring Based on Soil Moisture Active Passive Observations and 36 Years of North American Land Data Assimilation System Data: A Case Study in the Southeast United States. *Remote Sensing, 10*(2), 301.

Yee, M. S., Walker, J. P., Monerris, A., Rüdiger, C., & Jackson, T. J. (2016). On the identification of representative in situ soil moisture monitoring stations for the validation of SMAP soil moisture products in Australia. *Journal of Hydrology, 537*, 367-381.

Yuan, S., & Quiring, S. M. (2017). Comparison of three methods of interpolating soil moisture in Oklahoma. *International Journal of Climatology, 37*(2), 987-997.

Zargar, A., Sadiq, R., Naser, B., & Khan, F. I. (2011). A review of drought indices. *Environmental Reviews, 19*(NA), 333-349.

Zhan, X., Houser, P. R., Walker, J. P., & Crow, W. T. (2006). A method for retrieving high-resolution surface soil moisture from hydros L-band radiometer and radar observations. *IEEE Transactions on Geoscience and Remote Sensing, 44*(6), 1534-1544.

VITA

Yaping Xu was born in Weifang, Shandong province, China. He received his B.E. degree in Geographic Information System (GIS) from Guizhou University, Guizhou, China, in August 2009 and his M.S. degree in Cartography and Geographic Information System (GIS) from the University of Chinese Academy of Sciences, Beijing, in January 2013. He worked as a Research Engineer in China Spatial Technology Institute from the year 2012-2014. In August 2014, He came to the Louisiana State University (LSU) to pursue his Ph.D. degree in Geography. During his Ph.D. study, he participated in the Southeast Agriculture project sponsored by NASA and served as the team lead in 2016, and also participated in a project funded by Louisiana Transportation Research Center that measures soil moisture based on infrared and thermal sensors on UAV.

He has won several awards including the "Innovative Application of ESRI GIS Technology Poster Competition at the 2019 AAG Meeting Winner Award", "The 35th Annual Louisiana Remote Sensing & GIS Workshop Student Poster Competition, First Prize Award", "Louisiana Sea Grant Coastal Connection Competition Thesis Presentation 2nd Place", "Louisiana State University 2017-2018 Dissertation Year Fellowship", "NASA Southeast US Agriculture Project Contribution of Project Lead (Nominated)" and so on. In addition, he is a co-founder and president of the GIScience Club, an official student association at Louisiana State University.