UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

# SPATIAL-TEMPORAL ANALYSIS OF SUBSURFACE WATER CONTENT AND APPLICATIONS IN OKLAHOMA: WASTEWATER INJECTION INDUCED EARTHQUAKES AND A MULTI SENSOR SOIL MOISTURE PRODUCT

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

ZHEN HONG
Norman, Oklahoma
2021

SPATIAL-TEMPORAL ANALYSIS OF SUBSURFACE WATER CONTENT AND APPLICATIONS IN OKLAHOMA: WASTEWATER INJECTION INDUCED EARTHQUAKES AND A MULTI SENSOR SOIL MOISTURE PRODUCT

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF GEOGRAPHY AND ENVIRONMENTAL SUSTAINABILITY

BY THE COMMITTEE CONSISTING OF

Dr. John Greene, Chair

Dr. Hernan A. Moreno

Dr. Yang Hong

Dr. Renee McPherson

Dr. Thomas Neeson

**Acknowledgments**

I am extremely grateful to all the support, help and guidance I have received during my Ph.D. education. First, I would like to express my deepest appreciation and gratitude to my advisors, Dr. Aondover Tarhule, Dr. Hernan Moreno, and Dr. John Greene. In the first two years of my Ph.D. education, I received tremendous assistance and support from Dr. Aondover Tarhule. In the following years, Dr. Hernan Moreno provided great patience and endless opportunities to help me find my interests and my value in research. Under his supervision, I learned how to define a research problem, find a solution to it, and finally publish the results. The successful completion of this dissertation would not have been possible without his mentorship. Without his active guidance and support during year 2019 and 2020, when my difficult pregnancy and baby caring hindered my research progress, I would not have finished this program. At the last year of my Ph.D. education, Dr. John Greene provided thought-provoking suggestions, generous and prompt support, and continual encouragement to me.

Second, I would also like to extend my sincere thanks to my committee member and co-advisor, Dr. Yang Hong, for providing me the precious research assistant opportunities. He has provided me extensive personal and professional guidance and taught me a great deal about scientific research, class teaching and life in general.

Third, I am thankful to my members of committee, Dr. Renee McPherson, and Dr. Thomas Neeson, for all the advice and insights on my research. Also special thanks to Dr. Laurel Smith and Ms. Abby Young for helping me keep track of my

PhD progress. I would also like to thank all my colleagues in the HyDROS lab and my friends, without whom I could not have gone so far in my field of research.

Finally, I want to thank my family and my friends for the vast amounts of love, laughs and relentless support they gave me.

**Abstract**

Subsurface water is liquid water found below the ground surface, including soil water above the water table and ground water below the water table, but does not include water chemically bound to minerals or organic matter. Two important contents of subsurface water in Oklahoma have aroused the interest of more and more scientists: the wastewater injected into the ground during the oil and gas production and the surface soil moisture. This dissertation aims to develop contributions to two important topics for the sustainability of Oklahoma that are related to earthquakes and water resources: (1) the effects of deep underground waste-water injection on triggering regional seismicity and (2) the quantification of state-wide shallow-soil water content as a new tool for multiple applications in reservoir management, water resources, agriculture, natural hazards, and water management. The results of this study could help in setting sustainable limits for the oil and gas extraction industry in order to minimize the expected number and magnitude of induced quakes, thus avoiding future human and property losses. The results of this study also provide a new perspective for comparatively assessing multi-source soil moisture products, as well as a basis for objective data merging to capitalize on the strengths of multi-sensor multiplatform soil moisture products. Moreover, the new merged soil moisture product will be beneficial for multiple applications in water resources management, agriculture, and natural hazards.

# Table of Contents

**List of Figures**

**List of Tables**

# Chapter 1 Introduction

## 1.1 Background

Subsurface water is liquid water found below the ground surface, including soil water above the water table and ground water below the water table, but does not include water chemically bound to minerals or organic matter. Two important contents of subsurface water in Oklahoma have aroused the interest of more and more scientists: the wastewater injected into the ground during the oil and gas production and the surface soil moisture.

### *1.1.1 Underground Injection Control (UIC) Wells*

Petroleum has been produced in Oklahoma for more than 100 years (Murray & Holland, 2014). In 2020, Oklahoma was the fourth-largest producer of crude oil and the fourth-largest producer of marketed natural gas among the states (U.S. Energy Information Administration, 2021). The high oil and gas production caused a rapid increase in construction of underground injection wells especially Class II wells in Oklahoma.

An injection well is a devise used for placing fluids including water, wastewater, brine, or water mixed with chemicals, into various underground formations such as sandstone, limestone, or the soil layer (US EPA, 2015c). Underground fluid injection has been widely used in enhancing the recovery of oil and disposing industrial waste since 1930s (National Research Council, 2013). In 1974, the Safe Drinking Water Act (SDWA) was established to protect public drinking water quality and supplies throughout the United States (US EPA, 2015a). The SDWA requires the U. S. Environmental Protection Agency (EPA) to develop minimum requirements for injection practices in order to protect underground sources of drinking water (USDW) from endangerment.

In accordance with the act, EPA established a national program- the Underground Injection Control (UIC) program- to regulate the permitting, construction, operation, maintaining, converting, plugging, or abandonment of injection wells (US EPA, 2015b).

The Underground Injection Control program categorizes injection wells into six classes (Class I, II, III, IV, V, and VI) based on the type, depth and potential endangerment on USDW of the injection activity. Injection wells used exclusively to inject fluids in the U.S. petroleum industry are defined as UIC Class II wells (US EPA, 2015b). The petroleum industry in the United States operates approximately 180,000 Class II wells that are mostly located in Texas, California, Oklahoma and Kansas (National Research Council, 2013). Class II wells are classified in three types: disposal wells (wells used for injecting wastewater from hydraulic fracturing activities), enhanced recovery wells (wells used for injecting fluids consisting of brine, freshwater, steam, polymers, or carbon dioxide into oil-bearing formations to recover residual oil) and hydrocarbon storage wells (wells used for injecting liquid hydrocarbons into underground formations) (US EPA, 2015b). Disposal wells are designed to dispose of brines or wastewater produced in oil and gas extraction activity. About 30,000 class II disposal wells accounted for 20% of the total number of Class II wells in the United States (National Research Council, 2013). Enhanced recovery wells are used to inject fluids consisting of brine, freshwater, or carbon dioxide into a reservoir to recover great amounts of the original oil and gas. Approximately 110,000 of UIC Class II wells are enhanced oil recovery wells in the United States (Ellsworth, 2013).

### *1.1.2 Potential for Induced Seismicity from Fluid Injection*

Shallow earthquakes result from a fault slip. The conditions leading to a seismic slip are quantified by the Coulomb criterion in terms of $\tau=\mu\ (\sigma-\rho)$, where the shear stress ($\tau$) equals the

frictional strength, which is the product of the coefficient of friction (μ) and the effective stress (σ-ρ). According to the Coulomb criterion, decreasing the normal stress (σ), increasing the pore pressure (ρ), and/or increasing the shear stress (π) can trigger a seismic slip on a fault (National Research Council, 2013). Fluid injection causes earthquakes through changing the stress on faults to facilitate failure. Factors of injection processes and reservoir conditions are relevant when accessing the risk and magnitude of induced seismicity. Factors of injection processes include injection rate, injection depth, fluid volume, fluid pressure and fluid temperature. Reservoir parameters include pore pressure, rock strength and conditions of pre-existing faults (Rubinstein & Mahani, 2015; Shapiro & Dinske, 2009; Zoback, 2012). Unfortunately, many of these factors are not easily or possible to quantify due to the limitation of geological knowledge and lack of basic data related to fluid injection practices.

It is not easily to evaluate the association between injection wells and earthquakes based on their spatial relation since the distance between seismicity and injection point can be more than 10 kilometers and the depth of seismicity can be much greater than injection depth (Rubinstein & Mahani, 2015). For example, the seismicity at the Rocky Mountain Arsenal was more than 10 kilometers away from the injection point and of greater depth than the injection depth (Healy et al., 1968; Hsieh & Bredehoeft, 1981). Some recent studies have argued that the distance between induced seismicity and injection point could be more than 20 kilometers (Keranen et al., 2014). Moreover, some induced seismicity may occur long after the injection activity begins. Therefore, it is not easily or possible to determine whether the seismicity is caused by fluid injection or not.

Most injection wells cause microscale earthquakes. Only a few of these wells are suspected of inducing felt earthquakes. For example, The $M_w$ 3.9 earthquake on 31 December 2011 in Youngstown, Ohio was concluded to have been induced by the fluid injection at a deep injection

well close to the pre-existing faults (Kim, 2013). The $M_w$ 4.7 earthquake on 27 February 2011 in central Arkansas occurred within 6 km of three waste disposal wells in use (Horton, 2012). The $M_w$ 5.7 earthquake on 6 November 2011 in Oklahoma appears to be relevant to wastewater injection (Keranen et al., 2013). The thirteen earthquakes with $M_w$ 4 in February 1996, at the Rocky Mountain Arsenal northeast of Denver, Colorado were induced by a deep injection well (Hsieh & Bredehoeft, 1981). More earthquakes may be induced by fluid injection, but only earthquakes with large magnitude and sufficient geological and industrial data are included in current studies.

### *1.1.3 Earthquakes Induced by Fluid Injection in Oklahoma*

Since 2008, seismicity in the central United States has rapidly increased, predominately within regions of high unconventional oil and gas production states such as Oklahoma (Ellsworth, 2013). Seismic swarms within Oklahoma have contributed a large portion of the recent central United States seismicity. On 6 November 2011, an $M_w$ 5.7 earthquake occurred in the Wilzetta oil field near Prague, Oklahoma with complex Pennsylvanian-age Wilzetta fault system and two active fluid injection wells. It is the second largest recorded earthquake in Oklahoma. The earthquake destroyed 14 homes, injured 2 people, and caused damage to buildings and roads in the epicentral region. On 3 September 2016, an earthquake with a Mw 5.8 occurred near the northern Oklahoma town of Pawnee. It was the strongest earthquake on record to date.

A growing body of scientific research increasingly connects this upsurge in seismic activity in Oklahoma with the recent boom in oil and gas production, specifically with the wastewater injection volumes (IW) and their depth (Barbour et al., 2017; Chen et al., 2017; Hincks et al., 2018; Holland, 2013; Hough & Page, 2015; Keranen et al., 2013, 2014; Walsh & Zoback, 2015). For example, Norbeck and Rubinstein (2018) calibrated a reservoir model to calculate the hydrologic

conditions associated with the activity of 902 saltwater disposal wells into the Aurbuckle aquifer using multiple geologic parameters within Oklahoma. Their results demonstrate that the recent earthquake sequence in Oklahoma can be explained as a hydromechanical earthquake nucleation response driven primarily by the injection history of saltwater disposal wells despite of the heterogeneity and uncertainty existing in the model due to the wide spatial coverage of the study area (Norbeck & Rubinstein, 2018). Langenbruch and Zoback (2016) calibrated a statistical model that relates seismicity and injection by applying the Guttenberg–Richter Law depending on two varying-in-time and space seismogenic parameters. Their results predict that earthquakes with magnitude equal or greater than 3.0 in north-central Oklahoma should significantly decrease by the end of 2016 and approach historic levels within a few years as a respond to the mandated injection rate reduction. The authors also discussed that it will be interesting to analyze how the model and their results fluctuate in different subregions of Oklahoma (Langenbruch & Zoback, 2016). In summary, despite recent contributions of robust, but parametrically uncertain, physically based and hybrid (statistics and physics-based) models (Langenbruch et al., 2018; Langenbruch & Zoback, 2016; Norbeck & Rubinstein, 2018; Pollyea et al., 2018), more work is needed to assess the cumulative effects of underground injected water on triggering subsequent seismic activity and to identify bivariate regional migration patterns that allow for establishing clear relations between water injection and earthquake magnitude and number.

### *1.1.4 Significance of Soil Moisture within a Hydrological Cycle*

Soil moisture is a critical component of the land surface hydrological model, as the unsaturated zone partitions precipitation into soil moisture or overland flow, depending on soil texture, saturation levels, vegetation and topography. Water that is captured in the unsaturated zone is then

available for evapotranspiration processes, which are accountable for returning up to 60% of precipitation over land surfaces back into the atmosphere (Seneviratne et al., 2010). In addition, soil moisture regulates the partitioning of incoming solar energy over land surfaces, as evapotranspiration processes utilize more than half of the incoming solar energy (Trenberth et al., 2009). The evapotranspiration of soil moisture then drives the interaction between soil moisture and the atmosphere, resulting in the development of clouds, precipitation patterns and energy flux (Findell & Eltahir, 2003; Pielke, 2001). At the field scale, accurate measurement of soil moisture could benefit precision agriculture, especially precise irrigation and fertilization (Mohd Kassim et al., 2014; Woodley, 2017). At larger spatial scales, accurate soil moisture measurements could enhance capabilities in runoff and flood forecasting (Brocca et al., 2010; Crow et al., 2005), drought monitoring and prediction (Gu et al., 2008; X. Zhang, Tang, et al., 2017), numerical weather prediction (Capecchi & Brocca, 2014; Scipal et al., 2008), rainfall estimation (Brocca et al., 2012; Crow et al., 2011) and landslide prediction (Crow et al., 2012; Ray et al., 2010).

### *1.1.5 In Situ Soil Moisture Measurements*

Currently, soil moisture estimates can be obtained through three primary approaches: (1) in situ measurements, (2) remote sensing observations, and (3) Land Surface Models (LSM). A number of techniques have been developed to measure soil moisture with ground instruments, which can be categorized into classical and modern techniques (S.u. et al., 2014). For example, thermo-gravimetric (oven-drying) is a widely used classical soil moisture measurement technique (Robock et al., 2000). The modern techniques include neutron scattering technique (Hollinger & Isard, 1994), dielectric techniques (Selig & Mansukhani, 1975), soil resistivity sensor technique (Samouëlian et al., 2005), heat flux sensor technique (Valente et al., 2006), and optical techniques

(Sayde et al., 2010). With these techniques, spatially and temporally highly resolved measurements of soil moisture can be obtained at the point scale. In-situ soil moisture measurement techniques have the advantage of easy installation, high spatial and temporal resolutions, and the ability to measure soil moisture at different depths (Peng et al., 2017). Therefore, these measurements are normally recognized as the "ground truth" in validating and calibrating remote sensing and land surface model-based soil moisture retrievals. However, since soil moisture variability generally increases with extent scale (Famiglietti et al., 2008), these point measurements are not able to provide spatial representativeness of neighboring areas over a range of scales (Crow et al., 2012; Ochsner et al., 2019; Peng et al., 2017). Although previous studies have used geostatistical techniques to extrapolate in-situ soil moisture measurements to larger scales, these methods are usually complex, time-consuming, and depend on the availability of high-resolution auxiliary variables, especially over land surfaces with high spatial heterogeneity (Kang et al., 2021; Ochsner et al., 2019; Qin et al., 2013; Wang et al., 2015; Zhang et al., 2017).

### *1.1.6 Model-simulated Soil Moisture*

Model-simulated soil moisture is another source of spatially continuous soil moisture. Land surface models such as the Global Land Data Assimilation System (GLDAS; Rodell et al., 2004) and North American Land Data Assimilation System (NLDAS; Mitchell et al., 2004) can provide soil moisture estimates at various depths with high spatio-temporal resolution. Compared with in-situ measurements, Chen et al. (2013) found that all four GLDAS LSMs systematically underestimate the surface soil moisture in the Tibetan Plateau. Soil moisture simulations from NLDAS phase 2 (NLDAS-2) are also found to have large biases when compared to in-situ observations (Xia et al., 2014). Therefore, the reliability of model-simulated soil moisture varies

significantly from model to model, and over time and space. Models generally perform well in representing the variations in soil moisture and soil moisture anomalies (Albergel et al., 2012; Downer & Ogden, 2003; Meng & Quiring, 2008), but they tend to have large biases in simulating the absolute volumetric water content of the soil (Bi et al., 2016; Xia et al., 2014). Moreover, uncertainties due to model forcing, parameters, structure, and calibration affect the reliability of these soil moisture estimates (Brocca et al., 2017).

### *1.1.7 Remotely Sensed Soil Moisture Estimations*

Remote sensing soil moisture products from various sensors (e.g., microwave, optical and thermal sensors) provide global-scale soil moisture measurements but with limited spatial and temporal resolutions (Brocca et al., 2017). Microwave remote sensing techniques have gained momentum over the past 20 years with their advantages in the retrieval of soil moisture (Mohanty et al., 2017). These techniques can be categorized into two groups: active and passive microwave remote sensing. Active microwave sensors generate a series of microwave pulses which are sent to the object they observe and then they receive radiation that is reflected or backscattered from that object. Passive microwave sensors measure the thermal radiation emission from the land surface instead of sending and receiving microwave pulses (Jackson et al., 1996). Active microwave remote sensing measurements monitor soil moisture at spatial resolutions that range from tens of meters to a few kilometers. However, their low temporal resolution due to narrow operation swaths and long revisit time suppress their application in global soil moisture measurements. Passive microwave remote sensing measurements have higher temporal resolution, but also have much more coarse satellite footprints than that of active remote sensing systems. Therefore, it is difficult to use in-situ networks to validate satellite-based estimates due to the scale

mismatch between the two. In addition, all microwave remote sensing soil moisture measurements using C, X and L bands only measure soil moisture in the top five cm (or less) of the soil under low to moderate vegetation cover condition (Entekhabi et al., 2014).

*1.1.8 Soil Moisture Measurements in Oklahoma*

There are several large-scale soil moisture networks designed for monitoring soil moisture at large spatio-temporal scales operating within the United States, including the various state Mesonets, the Atmospheric Radiation Measurement Southern Great Plains site (ARM-SGP), and the Soil Climate Analysis Network (Crow et al., 2012). Established in January 1994, the Oklahoma Mesonet is a multipurpose network operating more than 110 automated stations with at least one station in each of Oklahoma's 77 counties (Brock et al., 1995; McPherson et al., 2007). Quality-assured data including temperature, humidity, solar radiation, wind speed and direction, and soil moisture are available through an operations center located at the Oklahoma Climatological Survey (OCS). Soil moisture data are collected every 30 min by soil moisture sensors at each site at four different depths (5, 25, 60, and 75 cm below the surface) (Illston et al., 2008). The Oklahoma Mesonet has been widely used for validating and calibrating remote sensing and land surface model-based soil moisture retrievals (Ford & Quiring, 2019; Gao et al., 2006; Swenson et al., 2008; Xia et al., 2014; Xu et al., 2021). However, since soil moisture variability generally increases with extent scale (Famiglietti et al., 2008), these point measurements are not able to provide spatial representativeness of neighboring areas over a range of scales (Crow et al., 2012; Ochsner et al., 2019; Peng et al., 2017). Land surface models and remote sensing techniques can provide soil moisture estimates at large scales. For example, the Noah model of NLDAS-2 provides hourly soil moisture fields at 1/8° resolution from 1979 to present at four soil layers: 0–10, 10–40, 40–100, and 100–200 cm (Xia et al., 2014). SMOS and SMAP, which have spatial resolution of roughly

25-40 km, can provide large-scale spatial patterns in surface soil moisture with extents >100 km (Mohanty et al., 2017; Ochsner et al., 2019). However, the reliability of these soil moisture products varies significantly from one to one, and over time and space (Brocca et al., 2017). In summary, each source of soil moisture observations has its strengths and weaknesses. However, none of them, at least by themselves, are adequate for providing accurate soil moisture data at high temporal and spatial resolutions. Therefore, it is vital and useful to combine these three independent data sources to capitalize on the strengths of each and to generate an optimal soil moisture product to facilitate real-world applications after comprehensive evaluation of their accuracy and error characteristics.

### 1.2 Research Goal and Objectives

This dissertation aims to develop contributions to two important topics for the sustainability of Oklahoma that are related to earthquakes and water resources: (1) the effects of deep underground waste-water injection on triggering regional seismicity and (2) the quantification of state-wide shallow-soil water content as a new tool for multiple applications in reservoir management, water resources, agriculture, natural hazards, and water management. To achieve this goal, three main tasks will be undertaken:

1. Performing a spatiotemporal analysis of the recent seismicity and industry-related wastewater injection activity in Oklahoma and develop a parsimonious predictive tool to estimate the lagged effect of previous month's injection volumes on subsequent regional seismic activity.

2. Conducting a cross-evaluation of three independently soil moisture products, as estimated from in-situ, satellite, and modeling methods over Oklahoma and over various land cover types.
3. Proposing an operational method for large-scale soil moisture mapping by blending in-situ, modeled and remote sensing data.

Objective 1 will provide an assessment of the regional collocation of wastewater injection activity and number/magnitude of earthquakes and their spatial association. Then, it explores the temporal correlations between wastewater injection volumes and number of earthquakes to develop a predictive model which will be evaluated in terms of observations and two recently published models. The results of this effort could help setting sustainable limits for the oil and gas extraction industry in order to minimize the expected number and magnitude of induced quakes, thus avoiding future human and property losses.

Objective 2 will consist of a comprehensive assessment of the Satellite SMAP_L3 (SMAP), Land Surface NOAH Model (Noah), and the interpolated Mesonet soil moisture (Mesonet) products across the State of Oklahoma at daily and seasonal timescales using the triple collocation method evaluated over different land cover types. The results of this objective are expected to provide not only a new perspective for comparatively assessing multi-source soil moisture products but also a basis for objective data merging to capitalize the strengths of multi-sensor multiplatform soil moisture products.

The effort of objective 3 adopts the TC based least square weighting method to merge the Satellite SMAP_L3, Land Surface NOAH Model, and the interpolated Mesonet soil moisture products across the State of Oklahoma and compare the merged product with equal weights merging product and SMAP L4 soil moisture product using the Automated Soil Moisture Mapping

System. The results of this objective will be beneficial for multiple applications in water resources management, agriculture and natural hazards.

### 1.3 Organization of Dissertation

This dissertation is organized into five chapters. Chapter 1 gives an introduction and lays out the research objectives. Chapters 2, 3 and 4 are dedicated to the three objectives outlined in section 1.3. These three chapters are written and formatted as stand-alone journal manuscripts, and each has its own introduction, data, methods, discussion and conclusion sections. Chapter 5 summarizes the results from all objectives and provides conclusions and a discussion of future research.

### References


Albergel, C., de Rosnay, P., Gruhier, C., Muñoz-Sabater, J., Hasenauer, S., Isaksen, L., Kerr, Y., & Wagner, W. (2012). Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations. *Remote Sensing of Environment*, *118*, 215–226. https://doi.org/10.1016/j.rse.2011.11.017

Barbour, A. J., Norbeck, J. H., & Rubinstein, J. L. (2017). The effects of varying injection rates in Osage County, Oklahoma, on the 2016 M w 5.8 Pawnee earthquake. *Seismological Research Letters*, *88*(4), 1040–1053.

Bi, H., Ma, J., Zheng, W., & Zeng, J. (2016). Comparison of soil moisture in GLDAS model simulations and in situ observations over the Tibetan Plateau. *Journal of Geophysical Research: Atmospheres*, *121*(6), 2658–2678. https://doi.org/10.1002/2015JD024131

Brocca, L., Crow, W. T., Ciabatta, L., Massari, C., de Rosnay, P., Enenkel, M., Hahn, S., Amarnath, G., Camici, S., Tarpanelli, A., & Wagner, W. (2017). A Review of the Applications of ASCAT Soil Moisture Products. *IEEE Journal of Selected Topics in Applied*

*Earth Observations and Remote Sensing*, *10*(5), 2285–2306. https://doi.org/10.1109/JSTARS.2017.2651140

Brock, F. V., Crawford, K. C., Elliott, R. L., Cuperus, G. W., Stadler, S. J., Johnson, H. L., & Eilts, M. D. (1995). The Oklahoma Mesonet: A Technical Overview. *Journal of Atmospheric and Oceanic Technology*, *12*(1), 5–19. https://doi.org/10.1175/1520-0426(1995)012<0005:TOMATO>2.0.CO;2

Chen, X., Nakata, N., Pennington, C., Haffener, J., Chang, J. C., He, X., Zhan, Z., Ni, S., & Walter, J. I. (2017). The Pawnee earthquake as a result of the interplay among injection, faults and foreshocks. *Scientific Reports*, *7*(1), 1–18.

Chen, Y., Yang, K., Qin, J., Zhao, L., Tang, W., & Han, M. (2013). Evaluation of AMSR-E retrievals and GLDAS simulations against observations of a soil moisture network on the central Tibetan Plateau. *Journal of Geophysical Research: Atmospheres*, *118*(10), 4466–4475. https://doi.org/10.1002/jgrd.50301

*Comparison of NLDAS-2 Simulated and NASMD Observed Daily Soil Moisture. Part I: Comparison and Analysis in: Journal of Hydrometeorology Volume 16 Issue 5 (2015)*. (n.d.). Retrieved June 29, 2021, from https://journals.ametsoc.org/view/journals/hydr/16/5/jhm-d-14-0096_1.xml

Crow, W. T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., Rosnay, P. de, Ryu, D., & 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). https://doi.org/10.1029/2011RG000372

Dean, T. J., Bell, J. P., & Baty, A. J. B. (1987). Soil moisture measurement by an improved capacitance technique, Part I. Sensor design and performance. *Journal of Hydrology*, *93*(1), 67–78. https://doi.org/10.1016/0022-1694(87)90194-6

Downer, C. W., & Ogden, F. L. (2003). Prediction of runoff and soil moistures at the watershed scale: Effects of model complexity and parameter assignment. *Water Resources Research*, *39*(3). https://doi.org/10.1029/2002WR001439

Ellsworth, W. L. (2013). Injection-induced earthquakes. *Science*, *341*(6142).

Entekhabi, D., Yueh, S., & De Lannoy, G. (2014). *SMAP handbook*.

Famiglietti, J. S., Ryu, D., Berg, A. A., Rodell, M., & Jackson, T. J. (2008). Field observations of soil moisture variability across scales. *Water Resources Research*, *44*(1). https://doi.org/10.1029/2006WR005804

Findell, K. L., & Eltahir, E. A. B. (2003). Atmospheric controls on soil moisture-boundary layer interactions: Three-dimensional wind effects. *Journal of Geophysical Research: Atmospheres*, *108*(D8). https://doi.org/10.1029/2001JD001515

Ford, T. W., & Quiring, S. M. (2019). Comparison of Contemporary In Situ, Model, and Satellite Remote Sensing Soil Moisture With a Focus on Drought Monitoring. *Water Resources Research*, *55*(2), 1565–1582. https://doi.org/10.1029/2018WR024039

Gao, H., Wood, E. F., Jackson, T. J., Drusch, M., & Bindlish, R. (2006). Using TRMM/TMI to Retrieve Surface Soil Moisture over the Southern United States from 1998 to 2002. *Journal of Hydrometeorology*, *7*(1), 23–38. https://doi.org/10.1175/JHM473.1

Healy, J. H., Rubey, W. W., Griggs, D. T., & Raleigh, C. B. (1968). The denver earthquakes. *Science*, *161*(3848), 1301–1310.

Hincks, T., Aspinall, W., Cooke, R., & Gernon, T. (2018). Oklahoma's induced seismicity strongly linked to wastewater injection depth. *Science*, *359*(6381), 1251–1255.

Holland, A. A. (2013). Earthquakes triggered by hydraulic fracturing in south-central Oklahoma. *Bulletin of the Seismological Society of America*, *103*(3), 1784–1792.

Hollinger, S. E., & Isard, S. A. (1994). A Soil Moisture Climatology of Illinois. *Journal of Climate*, *7*(5), 822–833. https://doi.org/10.1175/1520-0442(1994)007<0822:ASMCOI>2.0.CO;2

Horton, S. (2012). Disposal of hydrofracking waste fluid by injection into subsurface aquifers triggers earthquake swarm in central Arkansas with potential for damaging earthquake. *Seismological Research Letters*, *83*(2), 250–260.

Hough, S. E., & Page, M. (2015). A century of induced earthquakes in Oklahoma? *Bulletin of the Seismological Society of America*, *105*(6), 2863–2870.

Hsieh, P. A., & Bredehoeft, J. D. (1981). A reservoir analysis of the Denver earthquakes: A case of induced seismicity. *Journal of Geophysical Research: Solid Earth*, *86*(B2), 903–920.

Illston, B. G., Basara, J. B., Fiebrich, C. A., Crawford, K. C., Hunt, E., Fisher, D. K., Elliott, R., & Humes, K. (2008). Mesoscale Monitoring of Soil Moisture across a Statewide Network. *Journal of Atmospheric and Oceanic Technology*, *25*(2), 167–182. https://doi.org/10.1175/2007JTECHA993.1

Jackson, T. J., Schmugge, J., & Engman, E. T. (1996). Remote sensing applications to hydrology: Soil moisture. *Hydrological Sciences Journal*, *41*(4), 517–530.

Kang, J., Jin, R., Li, X., & Zhang, Y. (2021). Mapping High Spatiotemporal-Resolution Soil Moisture by Upscaling Sparse Ground-Based Observations Using a Bayesian Linear Regression Method for Comparison with Microwave Remotely Sensed Soil Moisture Products. *Remote Sensing*, *13*(2), 228. https://doi.org/10.3390/rs13020228

Keranen, K. M., Savage, H. M., Abers, G. A., & Cochran, E. S. (2013). Potentially induced earthquakes in Oklahoma, USA: Links between wastewater injection and the 2011 Mw 5.7 earthquake sequence. *Geology*, *41*(6), 699–702.

Keranen, K. M., Weingarten, M., Abers, G. A., Bekins, B. A., & Ge, S. (2014). Sharp increase in central Oklahoma seismicity since 2008 induced by massive wastewater injection. *Science*, *345*(6195), 448–451.

Kim, W.-Y. (2013). Induced seismicity associated with fluid injection into a deep well in Youngstown, Ohio. *Journal of Geophysical Research: Solid Earth*, *118*(7), 3506–3518.

McPherson, R. A., Fiebrich, C. A., Crawford, K. C., Kilby, J. R., Grimsley, D. L., Martinez, J. E., Basara, J. B., Illston, B. G., Morris, D. A., Kloesel, K. A., Melvin, A. D., Shrivastava, H., Wolfinbarger, J. M., Bostic, J. P., Demko, D. B., Elliott, R. L., Stadler, S. J., Carlson, J. D., & Sutherland, A. J. (2007). Statewide Monitoring of the Mesoscale Environment: A Technical Update on the Oklahoma Mesonet. *Journal of Atmospheric and Oceanic Technology*, *24*(3), 301–321. https://doi.org/10.1175/JTECH1976.1

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. https://doi.org/10.1175/2008JHM916.1

Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Cosgrove, B. A., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., Tarpley, J. D., Lettenmaier, D. P., Marshall, C. H., Entin, J. K., Pan, M., Shi, W., Koren, V., … Bailey, A. A. (2004). The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological

modeling system. *Journal of Geophysical Research: Atmospheres*, *109*(D7). https://doi.org/10.1029/2003JD003823

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

Murray, K. E., & Holland, A. A. (2014). *Inventory of class II underground injection control volumes in the midcontinent*.

National Research Council. (2013). *Induced seismicity potential in energy technologies*. National Academies Press.

Norbeck, J. H., & Rubinstein, J. L. (2018). Hydromechanical earthquake nucleation model forecasts onset, peak, and falling rates of induced seismicity in Oklahoma and Kansas. *Geophysical Research Letters*, *45*(7), 2963–2975.

Ochsner, T. E., Linde, E., Haffner, M., & Dong, J. (2019). Mesoscale Soil Moisture Patterns Revealed Using a Sparse In Situ Network and Regression Kriging. *Water Resources Research*, *55*(6), 4785–4800. https://doi.org/10.1029/2018WR024535

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

Pielke, R. A. (2001). Influence of the spatial distribution of vegetation and soils on the prediction of cumulus Convective rainfall. *Reviews of Geophysics*, *39*(2), 151–177. https://doi.org/10.1029/1999RG000072

Qin, J., Yang, K., Lu, N., Chen, Y., Zhao, L., & Han, M. (2013). Spatial upscaling of in-situ soil moisture measurements based on MODIS-derived apparent thermal inertia. *Remote Sensing of Environment*, *138*, 1–9. https://doi.org/10.1016/j.rse.2013.07.003

Robinson, D. A., Jones, S. B., Wraith, J. M., Or, D., & Friedman, S. P. (2003). A Review of Advances in Dielectric and Electrical Conductivity Measurement in Soils Using Time Domain Reflectometry. *Vadose Zone Journal*, *2*(4), 444–475. https://doi.org/10.2136/vzj2003.4440

Robock, A., Vinnikov, K. Y., Srinivasan, G., Entin, J. K., Hollinger, S. E., Speranskaya, N. A., Liu, S., & Namkhai, A. (2000). The Global Soil Moisture Data Bank. *Bulletin of the American Meteorological Society*, *81*(6), 1281–1300. https://doi.org/10.1175/1520-0477(2000)081<1281:TGSMDB>2.3.CO;2

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., & Toll, D. (2004). The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society*, *85*(3), 381–394. https://doi.org/10.1175/BAMS-85-3-381

Rubinstein, J. L., & Mahani, A. B. (2015). Myths and facts on wastewater injection, hydraulic fracturing, enhanced oil recovery, and induced seismicity. *Seismological Research Letters*, *86*(4), 1060–1067.

Samouëlian, A., Cousin, I., Tabbagh, A., Bruand, A., & Richard, G. (2005). Electrical resistivity survey in soil science: A review. *Soil and Tillage Research*, *83*(2), 173–193. https://doi.org/10.1016/j.still.2004.10.004

Sayde, C., Gregory, C., Gil-Rodriguez, M., Tufillaro, N., Tyler, S., Giesen, N. van de, English, M., Cuenca, R., & Selker, J. S. (2010). Feasibility of soil moisture monitoring with heated fiber optics. *Water Resources Research*, *46*(6). https://doi.org/10.1029/2009WR007846

Selig, E. T., & Mansukhani, S. (1975). Relationship of Soil Moisture to the Dielectric Property. *Journal of the Geotechnical Engineering Division*, *101*(8), 755–770. https://doi.org/10.1061/AJGEB6.0000184

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., & Teuling, A. J. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, *99*(3), 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004

Shapiro, S. A., & Dinske, C. (2009). Fluid-induced seismicity: Pressure diffusion and hydraulic fracturing. *Geophysical Prospecting*, *57*(2), 301–310.

S.u., S. L., Singh, D. N., & Shojaei Baghini, M. (2014). A critical review of soil moisture measurement. *Measurement*, *54*, 92–105. https://doi.org/10.1016/j.measurement.2014.04.007

Swenson, S., Famiglietti, J., Basara, J., & Wahr, J. (2008). Estimating profile soil moisture and groundwater variations using GRACE and Oklahoma Mesonet soil moisture data. *Water Resources Research*, *44*(1). https://doi.org/10.1029/2007WR006057

Trenberth, K. E., Fasullo, J. T., & Kiehl, J. (2009). Earth's Global Energy Budget. *Bulletin of the American Meteorological Society*, *90*(3), 311–324. https://doi.org/10.1175/2008BAMS2634.1

U.S. Energy Information Administration. (2021, April 15). *Oklahoma State Energy Profile*. https://www.eia.gov/state/print.php?sid=OK

US EPA, O. (2015a, March 25). *Safe Drinking Water Act (SDWA)* [Collections and Lists]. US EPA. https://www.epa.gov/sdwa

US EPA, O. (2015b, March 27). *Class II Oil and Gas Related Injection Wells* [Overviews and Factsheets]. US EPA. https://www.epa.gov/uic/class-ii-oil-and-gas-related-injection-wells

US EPA, O. (2015c, April 1). *General Information About Injection Wells* [Overviews and Factsheets]. US EPA. https://www.epa.gov/uic/general-information-about-injection-wells

US EPA, O. (2015d, June 11). *Underground Injection Control Well Classes* [Overviews and Factsheets]. US EPA. https://www.epa.gov/uic/underground-injection-control-well-classes

Valente, A., Morais, R., Tuli, A., Hopmans, J. W., & Kluitenberg, G. J. (2006). Multi-functional probe for small-scale simultaneous measurements of soil thermal properties, water content, and electrical conductivity. *Sensors and Actuators A: Physical*, *132*(1), 70–77. https://doi.org/10.1016/j.sna.2006.05.010

Walsh, F. R., & Zoback, M. D. (2015). Oklahoma's recent earthquakes and saltwater disposal. *Science Advances*, *1*(5), e1500195.

Wang, J., Ge, Y., Heuvelink, G. B. M., & Zhou, C. (2015). Upscaling In Situ Soil Moisture Observations to Pixel Averages with Spatio-Temporal Geostatistics. *Remote Sensing*, *7*(9), 11372–11388. https://doi.org/10.3390/rs70911372

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

Zhang, X., Zhang, T., Zhou, P., Shao, Y., & Gao, S. (2017). Validation Analysis of SMAP and AMSR2 Soil Moisture Products over the United States Using Ground-Based Measurements. *Remote Sensing*, *9*(2), 104. https://doi.org/10.3390/rs9020104

Zoback, M. D. (2012). Managing the seismic risk posed by wastewater disposal. *Earth*, *57*(4), 38.

# Chapter 2 Spatiotemporal Assessment of Induced Seismicity in Oklahoma: Foreseeable Fewer Earthquakes for Sustainable Oil and Gas Extraction?


## Abstract

In this study we present a spatiotemporal analysis of the recent seismicity and industry-related wastewater injection activity in Oklahoma. A predictive tool was developed to estimate the lagged effect of previous month's injection volumes on subsequent regional seismic activity. Results support the hypothesis that the recent boom in unconventional oil and gas production and either the mitigation policies or the drop in oil prices (or both) are potentially responsible for the upsurge and reduction in the state's seismic activity between 2006–2015 and 2016–2017, respectively. A cluster analysis reveals a synchronous migration pattern between earthquake occurrences and saltwater injection with a predominant northwest direction during 2006 through 2017. A lagged cross-correlation analysis allows extracting a power law between expected number of quakes and weighted average monthly injection volumes with a coefficient of determination of $R^2 = 0.77$. Such a relation could be used to establish sustainable water injection limits aiming to minimize seismicity to values comparable with several historically representative averages. Results from these analyses coincide with previously found sustainable limits of 5 to 6 million $\mathrm{m}^3/\mathrm{month}$ but expand to operations that could attain the same number through differential monthly planning. Findings could potentially be used for model intercomparison and regulation policies.


## 2.1 Introduction

Prior to the year 2000, the United States had an average of 21 earthquakes each year with magnitude 3.0 (i.e., $M_w$ 3.0) or greater; however, since the start of 2010, more than 300 earthquakes of equal or greater magnitude occurred in three years (Ellsworth, 2013). In the U.S.

Great Plains region, the rate of increase has been mostly attributed to excessive volumes of wastewater injection due to the unprecedented activity of the oil and gas industry (Ellsworth, 2013; Frohlich et al., 2011; Horton, 2012; Keranen et al., 2014; Kim, 2013; Llenos & Michael, 2013; Van der Elst et al., 2013; Weingarten et al., 2015). For example, the $M_w$ 3.9 earthquake on 31 December 2011 in Youngstown, Ohio was concluded to be induced by the fluid injection at a deep-injection well close to pre-existing faults (Kim, 2013). The $M_w$ 4.7 earthquake on 27 February 2011 in central Arkansas occurred within 6 km of three wastewater disposal wells in use (Horton, 2012). The 2008–2009 sequence of earthquakes with $M_w$ smaller than 3.3 at the Dallas/Fort Worth Airport area were potentially induced by brine disposal associated with the production of natural gas (Frohlich et al., 2011). The $M_w$ 5.7 and 5.8 earthquakes in 2011 and 2016 in Oklahoma appear to be relevant to wastewater injection (Barbour et al., 2017; Keranen et al., 2013). However, to conclusively determine the degree of association between wastewater injection and earthquakes remains a challenging task doe to the research limitations in data availability and regionally appropriate seismic models.

Both earthquakes' regional number and magnitude have increased during the previous decade and seismic events have become more common within the state of Oklahoma, including recorded earthquakes with $M_w$ 3.0 or greater. Keranen et al. (2014) noted that the total number of earthquakes in Oklahoma between 2008 and 2013 (i.e., 6 years) was four times those occurred from 1976 to 2007 (i.e., 31 years). Additionally, between 1974 and 2008, Oklahoma had an average of one earthquake with $M_w \geq 3.0$ each year. Comparatively, during 2013 and 2014, the state had more than 100 $M_w \geq 3.0$ quake events per year (Walsh & Zoback, 2015). On 3 September 2016, an earthquake with a $M_w$ 5.8 occurred near the northern Oklahoma town of Pawnee. It was the strongest earthquake on record to date in that location. As an immediate

response, the Oklahoma Corporation Commission (OCC) ordered the shutdown of 37 disposal wells to within an area of 1878 $km^2$ around the epicenter. OCC has also taken many other actions in response to recent earthquakes, including a disposal volume reduction plan (Oklahoma Corporation Commission, 2017).

A growing body of scientific research increasingly connects this upsurge in seismic activity in Oklahoma with the recent boom in oil and gas production, specifically with the wastewater injection volumes (IW) and their depth (Barbour et al., 2017; Chen et al., 2017; Hincks et al., 2018; Holland, 2013; Hough & Page, 2015; Keranen et al., 2013, 2014; Walsh & Zoback, 2015). Norbeck and Rubinstein (2018) calibrated a reservoir model to calculate the hydrologic conditions associated with the activity of 902 saltwater disposal wells into the Aurbuckle aquifer using multiple geology parameters within Oklahoma. Langenbruch and Zoback (2016) calibrated statistical model that relates seismicity and injection by applying the Guttenberg–Richter Law depending on two varying-in-time and space seismogenic parameters. Despite recent contributions of robust, but parametrically uncertain, physically based and hybrid (statistics and physics-based) models (Langenbruch et al., 2018; Langenbruch & Zoback, 2016; Norbeck & Rubinstein, 2018; Pollyea et al., 2018), more work is needed to assess the cumulative effects of underground injected water on triggering subsequent seismic activity and to identify bivariate regional migration patterns that allow for establishing clear relations between water injection and earthquake magnitude and number. This study has gathered exhaustive datasets pertaining to underground injection control (UIC) wells from OCC and the earthquake catalogue data from Oklahoma Geological Survey (OGS) from 2006 to 2017. The main findings of this chapter could help to set sustainable limits for oil and gas extraction industry by minimizing the expected number and magnitude of induced quakes, thus avoiding future human and property losses.

This chapter first provides a description of the data sources and magnitude of completeness to then develop spatiotemporal relations of the wastewater injection and seismic activity in Oklahoma during 2006–2017. Subsequently, it provides an assessment of the regional collocation of wastewater injection activity and number/magnitude of earthquakes and their spatial association. Then, it explores the temporal correlations between wastewater injection volumes and number of earthquakes to develop a two-parameter predictive power law. Model results are evaluated in terms of observations and two recently published models. A discussion section describes the potential uses and limitations of the achieved results. Lastly, conclusions summarize the main findings of this chapter.

## 2.2 Data Sources

Wastewater injection volumes (IW) and site location data were obtained from the OCC website http://www.occeweb.com/OG/ogdatafiles2.htm in September 2018 (Oklahoma Corporation Commission, 2018) for the calendar years 2006 to 2017. Class II injection and saltwater disposal (SWD) volume data sets were manually inspected to remove incomplete or duplicate records, as well as records without geolocation. IW data of SWD wells are available annually from 2006 to 2010, and monthly from 2011 to 2017. Since Osage County, in northeast Oklahoma is regulated by the Environmental Protection Agency (EPA) we could not include all active injection wells to date as these data were not publicly available. However, we obtained information from 10 active injection wells, within Osage, from Barbour et al (2017).

The Oklahoma earthquake database was downloaded from the OGS website http://www.ou.edu/content/ogs/research/earthquakes/catalogs.html (Oklahoma Geological Survey, 2018). Daily datasets, including epicenter location, depth and magnitude (i.e., $M_L$, $M_w$,

$m_b$ and $M_d$) are available between 1882 to present. In the interest of revealing spatiotemporal patterns of near-recent seismic activity in Oklahoma, only earthquakes occurred after January 2006 are studied in detail.

## 2.3 Earthquake Unit Homogenization and Data Completeness

### *2.3.1 Magnitude Unit Homogenization*

The type and accuracy of the earthquake recording devices have changed with time. For the most recent decades, despite the instruments remaining the same, the seismic magnitude units vary according to the maximum motion recorded by a seismograph in (1) local magnitude ($M_L$) also known as Richter magnitude, (2) duration magnitude ($M_d$), (3) body-wave magnitude ($m_b$) and (4) moment magnitude ($M_w$). The number of earthquakes which occurred between January 2006 and December 2017 is given in Table 2.1 showing the different magnitude units. Nonetheless, many earthquakes were simultaneously recorded in different scales which facilitate their unit homogenization. As the majority of seismic events are reported in $M_L$ scale, all other units are converted to $M_L$ to reduce data uncertainty introduced during this conversion. To do so, two empirical magnitude conversion relations are derived for those events with significant number of data pairs (i.e., $[M_L, m_b]$ and $[M_L, M_w]$). Since $[M_d, M_L]$ had zero pairs, a previously derived expression is applied (Brumbaugh, 1989) for such a conversion. Scatterplots with the event magnitude pairs, fitted, and 95% confidence envelopes are shown in Figure 2.1. The derived and used statistical regressions, sample size, cross-correlation coefficients and author are shown in Table 2.2.

**Table 2.1.** Number of earthquakes in different magnitude units during 2006–2017 in Oklahoma.

| Magnitude Type | Number of Earthquakes |
|---|---|
| Duration Magnitude ($\boldsymbol{M_d}$) | 1763 |
| Body-wave Magnitude ($\boldsymbol{m_b}$) | 364 |
| Local Magnitude ($\boldsymbol{M_L}$) | 25,956 |
| Moment Magnitude ($\boldsymbol{M_w}$) | 438 |

**Table 2.2.** Mathematical regressions adopted and derived to homogenize $\boldsymbol{M_d}$, $\boldsymbol{m_b}$ and $\boldsymbol{M_w}$ seismic magnitudes to local (Richter) magnitude, $\boldsymbol{M_L}$.

| Expression | Sample Size | $\boldsymbol{R^2}$ | Reference |
|---|---|---|---|
| $\boldsymbol{M_L = 0.936M_d - 0.16}$ | 17 | 0.95 | (Brumbaugh, 1989) |
| $\boldsymbol{M_L = 0.85m_b + 0.52}$ | 252 | 0.84 | (Hong et al., 2018) |
| $\boldsymbol{M_L = 0.96M_w + 0.35}$ | 440 | 0.81 | (Hong et al., 2018) |

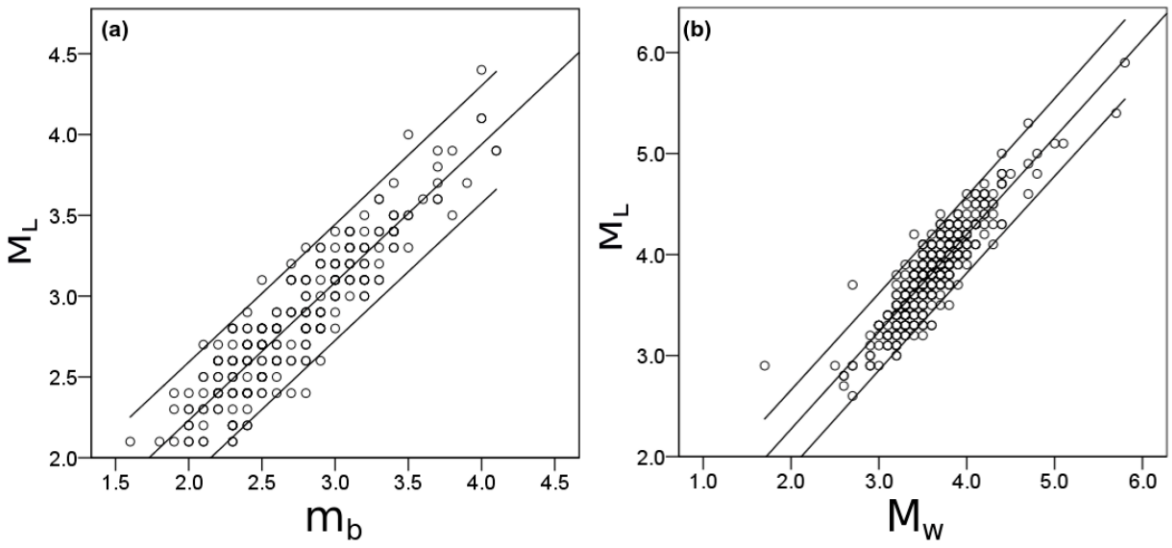


**Figure 2.1.** Linear regression plots for (**a**) $\boldsymbol{M_L}$ vs. $\boldsymbol{m_b}$ and (**b**) $\boldsymbol{M_L}$ vs. $\boldsymbol{M_w}$ for all seismic events occurred in Oklahoma between January 2006 and December 2017. Conversion equations are shown in Table 2.2.

*2.3.2 Earthquake Magnitude of Completeness*

After unit homogenization, the magnitude of all recorded earthquake events in the Oklahoma earthquake catalog for the period between 2006 and 2017 ranged from 0.1 $M_L$ to 5.9 $M_L$. The magnitude of completeness ($M_C$) is the minimum magnitude above which earthquakes within a certain region are reliably recorded. Defining $M_C$ is necessary due to the complexity, spatial and temporal heterogeneity of seismometer networks, and time series records (Habermann, 1991; Woessner & Wiemer, 2005). To assess $M_C$ for our earthquake dataset, a frequency-magnitude distribution (FMD) plot was created for the entire dataset (see Figure 2.2) based on the entire magnitude range (EMR) method proposed by Woessner and Wiemer (2005). Their method estimates the FMD based on the Gutenberg-Richter law (Gutenberg & Richter, 1944). For the data with magnitude below the assumed Mc, EMR uses a normal cumulative distribution function (Woessner & Wiemer, 2005). Woessner and Wiemer (2005) compared the EMR method with other three including maximum curvature-method (MAXC; (Wiemer & Wyss, 2000)), goodness-of-fit

test (GFT; (Wiemer & Wyss, 2000)), and Mc by b-value stability (MBS; (Cao & Gao, 2002)), and they concluded that EMR is the most favorable model to calculate Mc (Woessner & Wiemer, 2005). The FMD curve indicates a data-based suggested value of $M_C$ = 2.6 which will be used as minimum trustable $M_L$ for the subsequent analyses (Woessner & Wiemer, 2005).

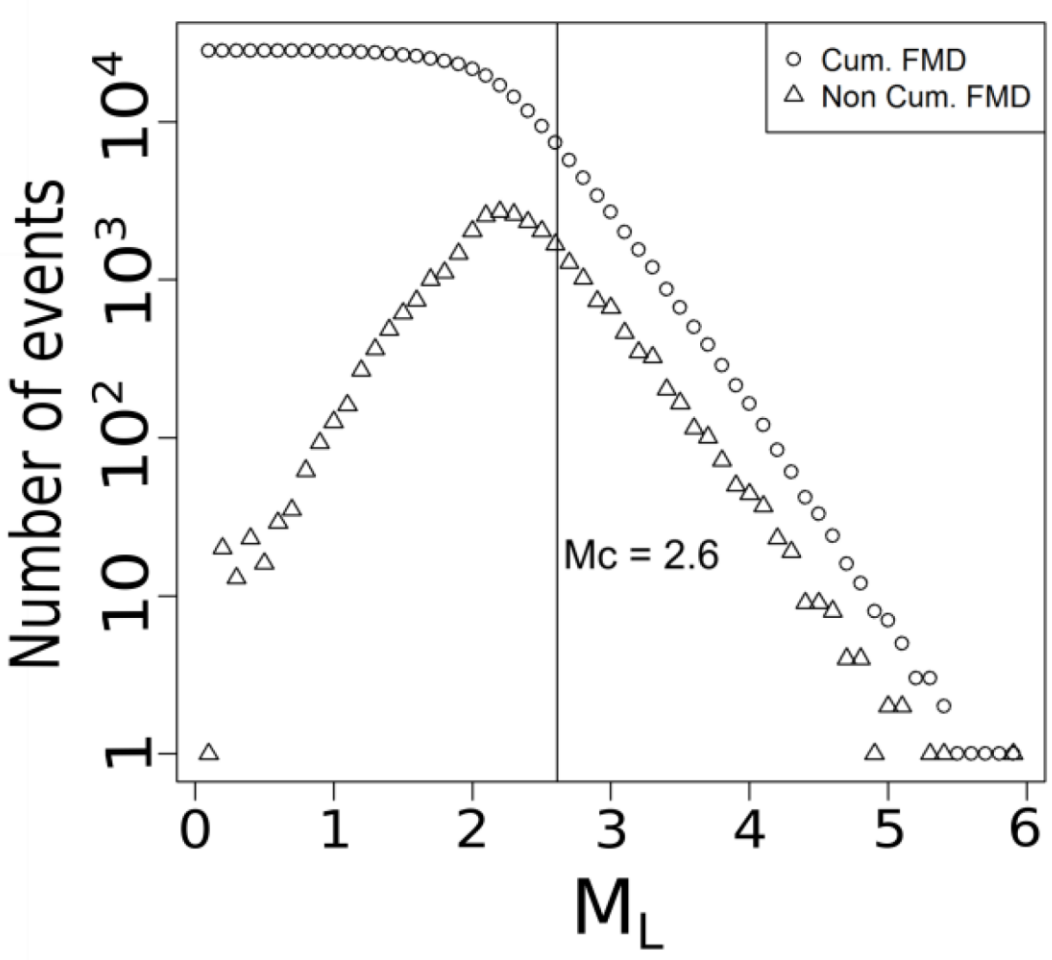


**Figure 2.2.** Cumulative and noncumulative frequency-magnitude distributions on logarithmic scale with the black line indicating magnitude of completeness ($\boldsymbol{M_C}$) for time series during 2006–2017.

### 2.4 Interannual Seismicity and Wastewater Injection Activity in Oklahoma

According to the Oklahoma Geological Survey, the earliest recorded earthquake in the state occurred on 22 October 1882 with a $M_L$ 5.0 (Oklahoma Geological Survey 2017). From 1882 to

2002, (120 years) Oklahoma had a total of 186 earthquakes with $M_L \geq M_C$, for an average of 1.55 earthquakes per year (Oklahoma Geological Survey 2017). Figures 2.3a,b shows the recent history of earthquake events (bars, N($M_L$)) in Oklahoma from 2000 to 2017 with $M_L \geq M_C$ and discretized by $M_L$ category. Comparatively to its precedent years, the number of seismic events per year with $M_L \geq M_C$ increased to 4.9 from 2003 to 2008 (39 in total; see Figure 2.3a). However, 2009 appears as a benchmark year that marks a significant increase relative to historic means (see Figure 2.3b). Between 2009 and 2017, the state had averaged 730 earthquakes per year (6570 total), which is more than four hundred times (i.e., 471) the historic average up to year 2002. Since "felt earthquakes" usually refer to those with $3 \leq M_w \leq 5$ (National Research Council 2013), Figure 3b also depicts N($M_L$) per $M_L$ category with $M_C \leq M_L < 3$, $3 \leq M_L < 4$ and $4 \leq M_L < 5$ in Oklahoma from 2000 to 2017. A peak in seismicity occurred in 2015 with 2560 events, followed by a steady decrease to 786 in 2017. This trend is replicated by each of the categories in Figure 3b (e.g., $M_C$, 3 and 4), except by the $M_L \geq 5$ whose peak occurred in 2016. The total number of "damaging earthquakes", which are those with $M_L \geq 5$, also increased after 2009 as shown in Figure 2.3b.

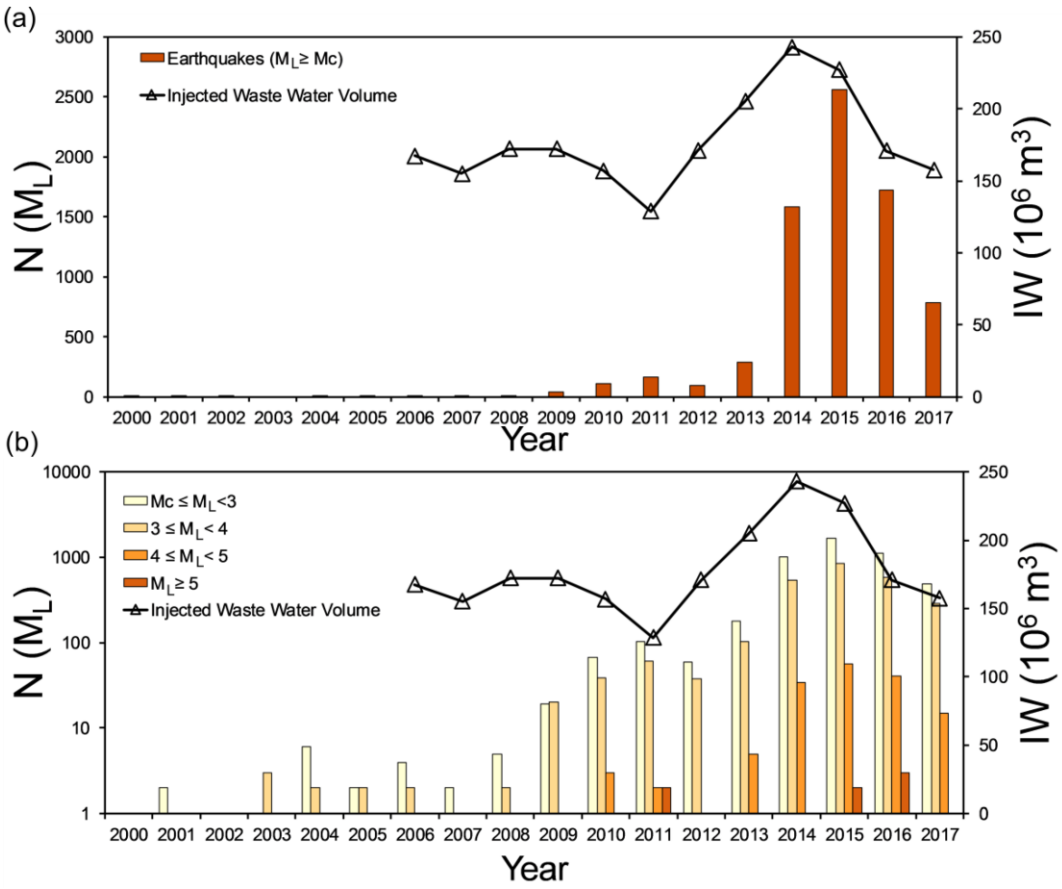


**Figure 2.3.** (**a**) Time series of total annual number of earthquakes (N($\boldsymbol{M_L}$)) with $\boldsymbol{M_L \geq M_C}$ (red bars) and oil/gas industry-related injected volumes of wastewater (IW) in million cubic meters (white triangles) in Oklahoma from 2000 to 2017. (**b**) Time series of total annual number of earthquakes (N($\boldsymbol{M_L}$)) with $\boldsymbol{M_L \geq M_C}$ per magnitude range between 2000 and 2017 and oil/gas industry-related volumes of wastewater injected (IW) between 2006 and 2017 in Oklahoma. Note the log scale for N($\boldsymbol{M_L}$) in (**b**).

Crude oil and natural gas have been extracted from Oklahoma's underground for more than 100 years (Murray & Holland, 2014). Between 2010 and 2012 Oklahoma was ranked as the 5th highest producing U.S. state (Murray & Holland, 2014). The high oil and gas production rates have caused a rapid increase in construction of underground injection Class II (UIC) wells, widely used to enhance the recovery of oil (EOR; Enhanced Oil Recovery wells) and disposing of industrial wastewater (SWD) since the 1930s (Council, 2013). Figure 2.3 shows IW (in $10^6 m^3/year$) from OCC UIC Class II wells reports that begin in 2006 to near present. From 2006 to 2012 the volume

of injected water ranged around $150 \times 10^6 m^3/year$. However, starting in 2012 a rapid increase in IW volumes is observed that peak in 2014 and 2015, followed by a sharp decline in 2016 and 2017 when the IW gets back to a number around the 2006–2012 average of $150 \times 10^6 m^3/year$. A paired time series analysis of the coupled IW and N($M_L$) reveals that both variables have shown a similar trend since the start of the unconventional use of injected water to retrieve oil and gas.

Regionally, the spatial distributions of earthquakes that occurred in Oklahoma between 2006 and 2017 with $M_L \geq M_C$ and the corresponding location of wastewater disposal wells operated during the most active year (i.e., 2014) are illustrated in Figure 2.4. In both figure panels, the different symbol sizes represent different categories of $M_L$ and IW. The spatial distribution of the two variables resembles a spatially correlated structure whose dependency functions need to be determined for different time lags. Further, during this period (2006–2017) most earthquakes occurred in central and northern Oklahoma while the largest magnitude ones occurred in the central region of the state. Historically, the largest IW volumes occurred mainly in central and northern Oklahoma. In counties like Osage, seismicity appears to be low possibly due to the dense rock bodies that reduce seismogenic potential for basement faults (Crain et al., 2017; Shah & Keller, 2017).

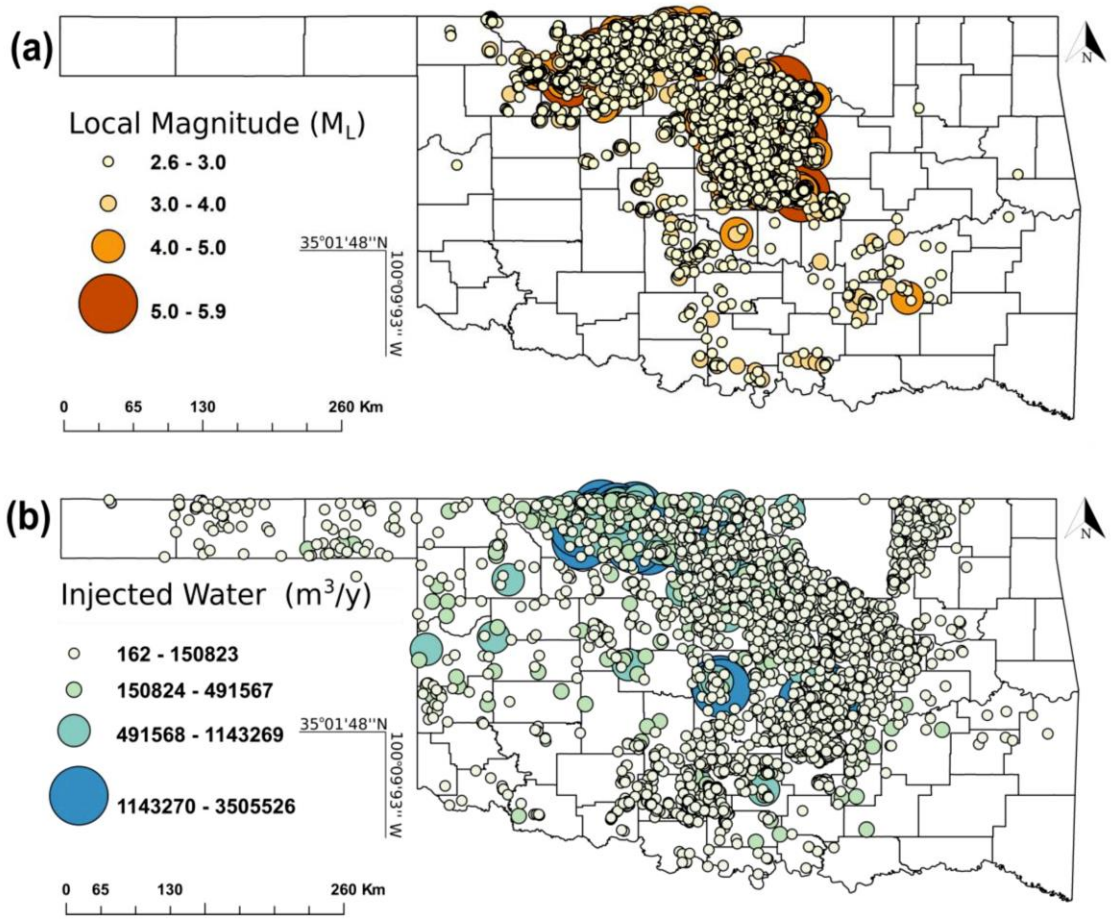


**Figure 2.4.** (**a**) Spatial distribution of earthquakes with $\boldsymbol{M_L \geq M_C}$ occurred in Oklahoma from 2006 to 2017; (**b**) Spatial distribution of wastewater disposal wells with corresponding IW volume ($\boldsymbol{m^3/year}$) operated in 2014.

## 2.5 Regional Migration Pattern of Epicenters and Wastewater Injection Activity

Since the spatial distribution of earthquakes appears to be highly conditioned by the zonal intensity of underground water injection, as shown in Figure 2.4, a cluster analysis can provide a clearer picture of the spatial co-variance between the two processes in play. Figure 2.5a shows the spatial distribution of weighted mean centers and standard deviational ellipses of all recorded Oklahoma earthquake epicenters occurred during each year from 2006 through 2017. A weighted mean center ($X_T$, $Y_T$) in any year (T) is the representative geographic location of all epicenters ($X_i$,

$Y_i$) adjusted for the local magnitude $M_L$ associated with each earthquake (i) acting as weighting factors ($w_i$) as shown in Equation (2.1) (Burt et al., 2009).

$$\overline{X_T} = \frac{\sum_{i=1}^{n} w_i X_i}{\sum_{i=1}^{n} w_i} \qquad \overline{Y_T} = \frac{\sum_{i=1}^{n} w_i Y_i}{\sum_{i=1}^{n} w_i} \qquad (2.1)$$

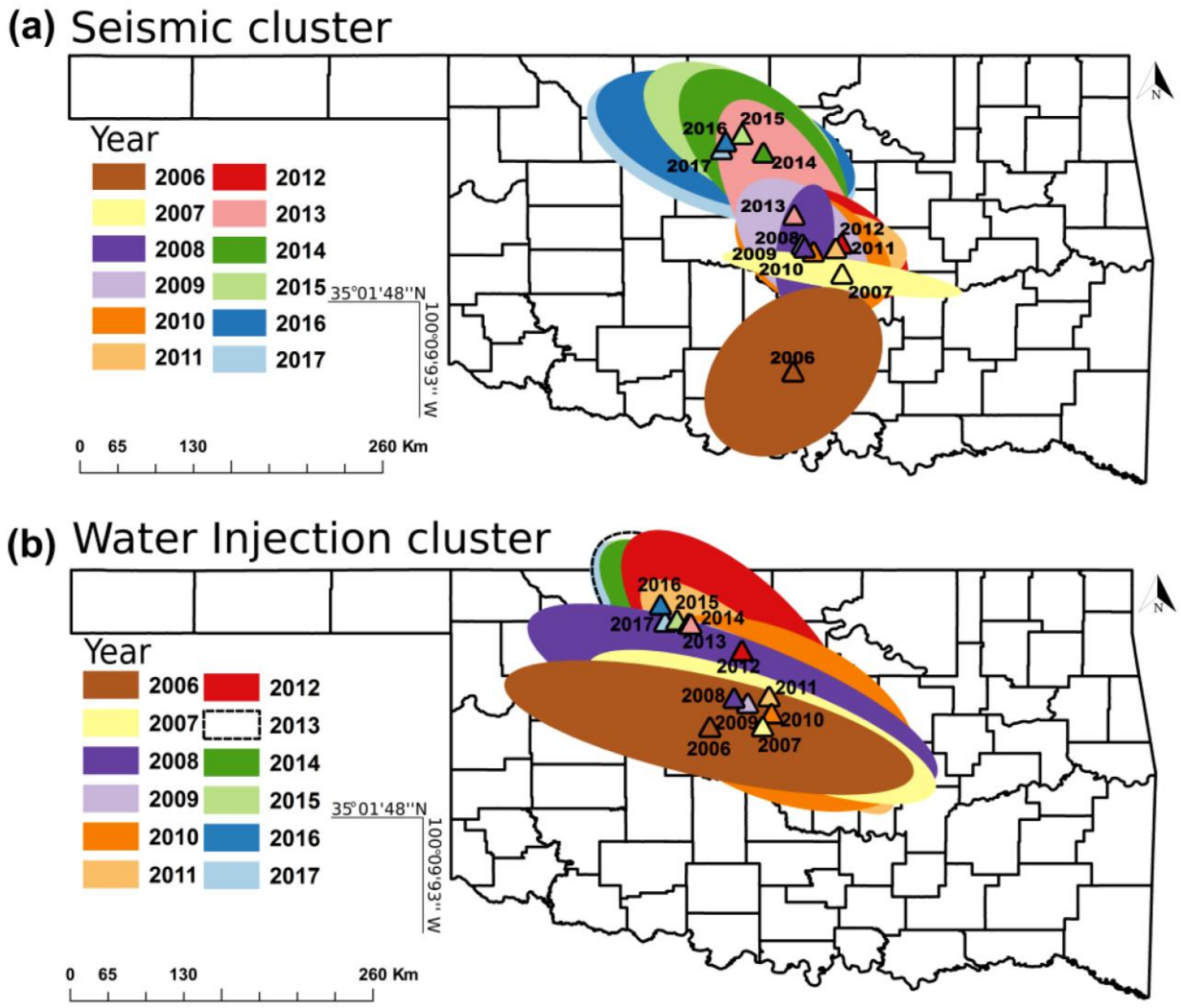


**Figure 2.5.** (**a**) Earthquake-clustering occurrence by year. Epicenter weighted mean centers (triangles) and standard deviation ellipses of all recorded earthquakes occurred in Oklahoma between 2006 and 2017; (**b**) Wastewater injection volume weighted mean centers (triangles) and standard deviation ellipses in Oklahoma between 2006 and 2017. The colors in both panels match for the same years, except by 2013 whose dashed lines are intended to improve result visualization. Coordinates of mean weighted centers are computed using Equation (2.1).

Where $w_i$ is the $M_L$ for each earthquake event (i) in a particular year T. Following equation (2.1), weighted mean centers of all earthquakes occurring in a particular year T would be closer to epicenters with the largest $M_L$ during that year. The major and minor axes of these weighted standard deviation ellipses are calculated as the second moment of the x- and y-coordinates distribution from each weighted mean center (Zoback, 2012). This approach, illustrated by Figure 2.5a, shows a generalized northwest seismic migration pattern from 2006 through 2017. Correspondingly, Figure 2.5b illustrates the weighted mean centers and standard deviation ellipses of wastewater disposal wells in each year from 2006 through 2017. Analogously to epicenters, wastewater injection locations are weighted by the volumetric magnitude of the annual injection volumes associated with each well. Thus, the weighted mean center of wastewater disposal wells in a particular year would be geographically closer to wells with larger annual injection volumes, reflecting the regional trend of well activity in that specific year. In summary, both unconventional oil and gas extraction and earthquake count show a northwest migration pattern from 2006 to 2017. To recognize year to year migration patterns, Figure 2.6 shows trends as indicated by vectors whose length is proportional to the average migration distance between consecutive years. The diagram shows some years when both processes migrated similar distances in similar directions, particularly 2007–2008 (~33 to 35 km SE), 2009–2010 (10 to 20 km NNE), 2012–2013 (~43 km SSE), 2014–2015 (~9 to 22 km SE) and 2016–2017 (7 to 15 km W). In other cases, the two vectors show an angular distance greater than 90 degrees such as 2006–2007, 2010–2011 (N-E quadrant), 2011–2012 (mostly E quadrants), 2015–2016 (mostly S quadrants). The large disparity in distances in 2006–2007, 2011–2012 and 2013, 2014 may be because injection operations moved quickly in the last months of the last year and earthquake count (since it has a lagged response) did not immediately showed the expected pattern of migration. Overall, regional migration patterns seem

to correspond to one another evidencing a zonal effect of the unconventional oil and gas industry on the number of regional earthquake count.

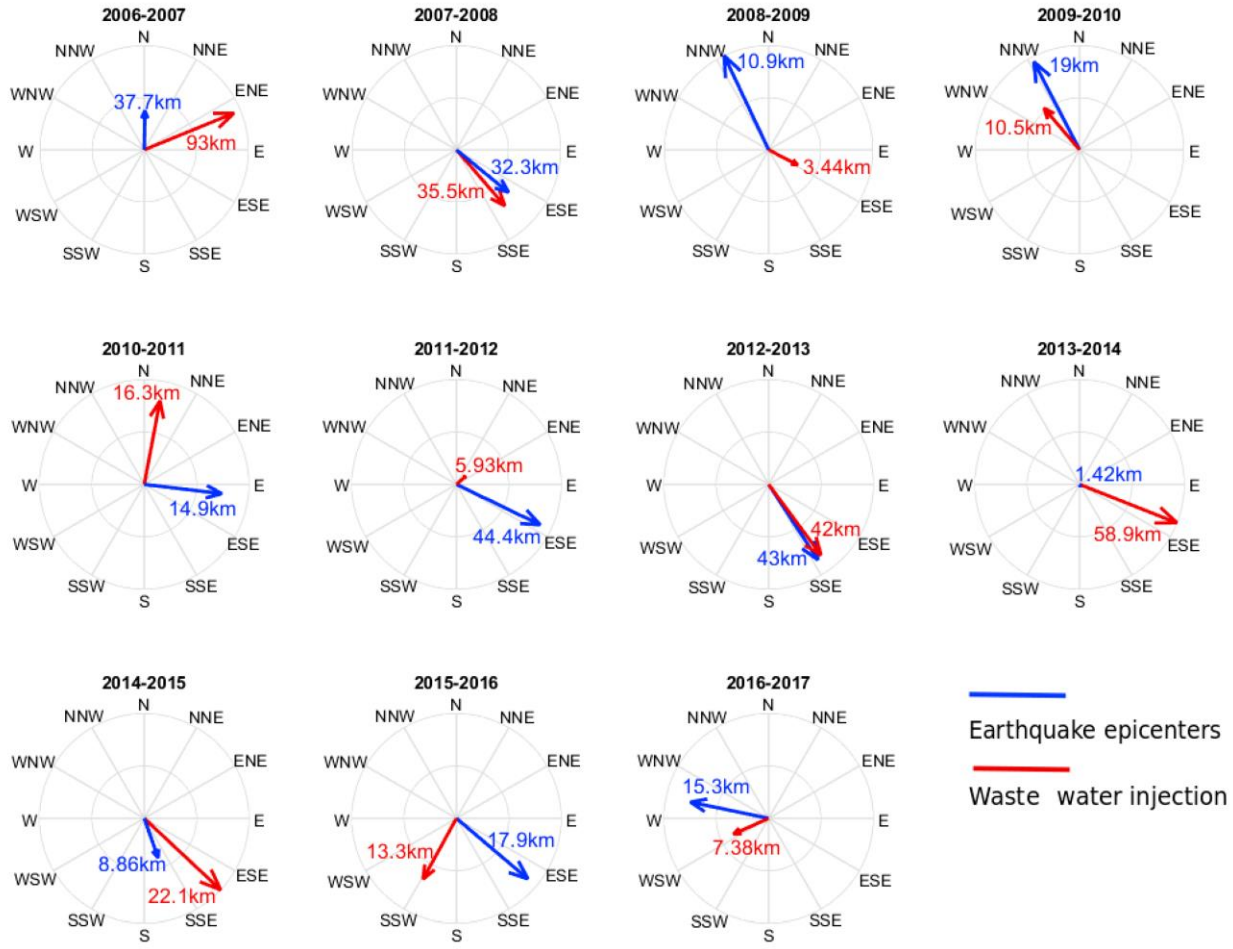


**Figure 2.6.** Yearly migration patterns between earthquakes weighted epicenters and wastewater injection activity in Oklahoma since 2006. Red and blue lines mean the average displacement of mean weighted centers of wastewater injection and earthquakes between consecutive years. The average displacement distance is also indicated within each compass diagram.

## 2.6 A Parsimonious Model of Seismicity.

With the objective of proposing a (single-variable) parsimonious regional predictive model between cumulative wastewater injection (IW) and earthquake count (N), patterns of lagged seismic responses to cumulative injected water during the 2006 to 2017 period in Oklahoma are explored. A cross-correlation analysis is carried to determine the temporal lags (i) which may be

associated with the number of expected earthquakes in a particular month t ($N_t$) as a function of $IW_{t-i}$ for i = 0, 1, 2, etc, months. This time delay can be physically expressed as the time the pressure increase takes to propagate from the injection wells to critically stressed faults in the crystalline basement (Chen et al., 2017; Langenbruch & Zoback, 2016). The cross-correlogram illustrated in Figure 2.7a reveals that lags i = 0 through −25 before the seismic events appear to mostly contribute to the bivariate co-dependence between IW (predictor) and N (predictand). Figure 2.7b quantifies the contribution of each lag i to the total correlation structure above the Pearson correlation coefficient significance threshold. According to the correlations for lags 0 to 25 months, we extract weight coefficients ($w_i$) for each lagged contribution to express $\widehat{IW}$ as a function of $IW_{t-i}$ (i = 0, 1, 2…, 25 months) as shown in Equation (2.2):

$$\widehat{IW} = \sum_0^{25} w_i IW_{t-i} = w_0 IW_t + w_1 IW_{t-1} + w_2 IW_{t-2} + \cdots w_{25} IW_{t-25} \qquad (2.2)$$

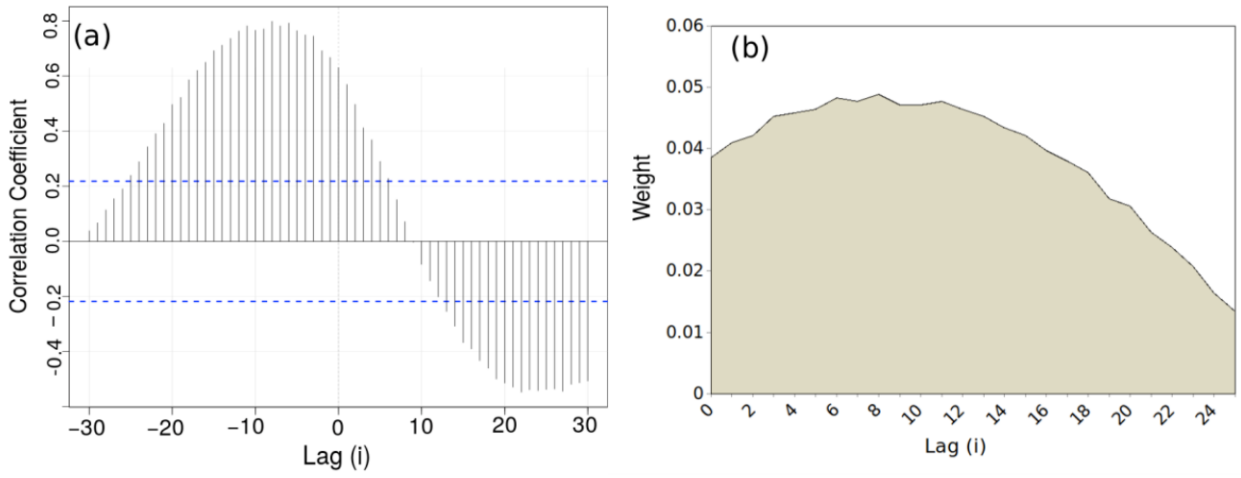


**Figure 2.7.** (**a**) Cross-correlation diagram between $\boldsymbol{IW_{t-i}}$ and $\boldsymbol{N_t}$ for different lags of IW (e.g., i = 0, 1, 2, 3…, n months). Negative numbers mean that IW precedes $\boldsymbol{N_t}$. (**b**) Contribution ($\boldsymbol{w_i}$) of each lag i to the prediction of the total of number of earthquakes in a particular month t ($\boldsymbol{N_t}$), to be applied to the predictors in Equation (2).

Figure 2.7b shows that lags 0 to 10 will be responsible for 50% of the variability in $\widehat{IW}$ in Equation (2.2), confirming that the number of earthquakes in a particular month is the result of the

pressure buildup due to previous-months water injection activity. According to the correlation structure derived, we fitted a mathematical power-law relating $\widehat{IW}$ $(m^3/month)$ to $N_t$ for all $M_L \geq M_C$ (see Equation 2.3). Figure 8 illustrates such a fitted relation applied to the logarithms of the monthly values since 2011 (complete data pairs).

$$N_t = 3.2099 \times 10^{-6} \widehat{IW}^{6.1489} \qquad (2.3)$$

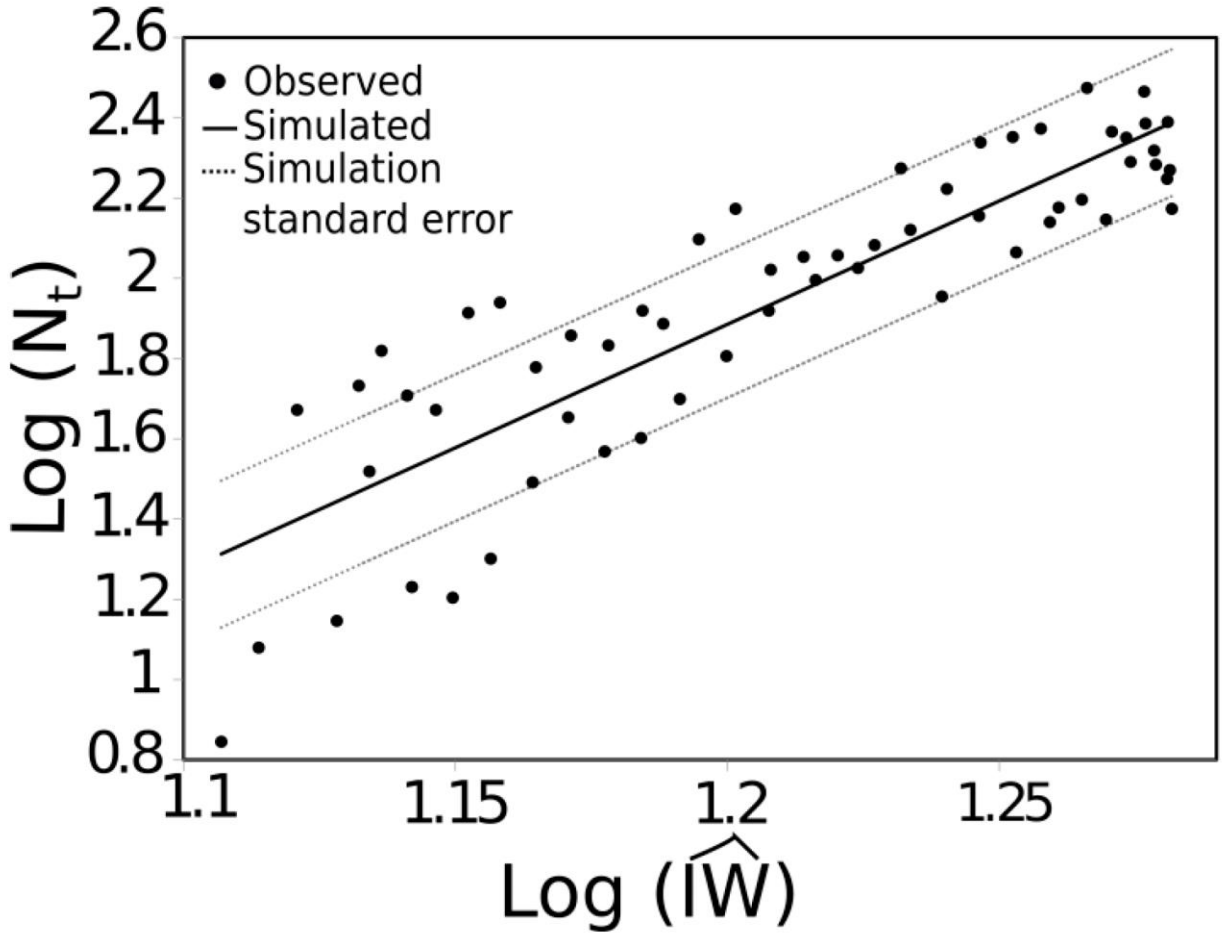


**Figure 2.8.** Regional induced-earthquake count $\boldsymbol{N_t}$ $(\boldsymbol{M_L \geq M_C})$ and $\boldsymbol{\widehat{IW}}$ estimator calibrated between years 2006 and 2017 in the state of Oklahoma. The power law explains 77% of the bivariate behavior of monthly injection and earthquakes number. Upper and lower dashed lines representing standard errors of estimates have been added to the mean predicted values.

We found that a power law of this type retrieves the highest coefficient of determination and explains 77% of the bivariate dependency between weighted cumulative regionally injected water

and seismicity in north-central Oklahoma. Standard error lines also provide a statistical estimation of the average error when using this relationship in a predictive manner. Figure 2.8 illustrates a comparison between historical benchmark periods with distinct seismic (e.g., number of events) activity in Oklahoma (see Section 2.4; Oklahoma Geological Survey, 2017). By applying this relationship to a scenario of hypothetical, constant-in-time injection rates, we can compare with historic benchmarks and study possibilities for sustainable oil and gas extraction limits (see Table 2.3) in terms of the expected number of seismic events (i.e, $N_t$). The term "sustainable limit" in column 5 of Table 2.3 refers to potential maximum injection values per month that the Oklahoma state regulation authorities (e.g., OCC and/or EPA) could consider for regulation of the oil and gas industry. This by no means accounts for other influences on environmental issues like water or energy consumption, groundwater, land or air pollution. In this table, it appears that constant and continuous rates (i.e., $IW_{t-i}$) of 5.6 million $m^3/month$ (i = 0, 1, 2, 3, etc) could reduce the number of earthquakes to pre- year 2000 conditions to 1.55 earthquakes per year with magnitude $M_L \geq M_C$. However, one could also define other limits such as the mean water injection during the period 2003–2008 (6.8 million $m^3/month$) as a sustainable limit, but at the expense of potential additional seismic occurrences (5.1 $events/year$) similar to the beginning of the 2000 decade or previous to the boom of oil and gas extraction (pre 2009–2015). An increase in $1 \times 10^6\, m^3/month$ of injected water represents different changes in seismicity across the spectrum of IW values with larger values triggering dramatic increases in seismic events, $N_t$. As an example, an increase of 1 million $m^3/month$ above 19 million $m^3/month$ (super-boom scenario) would represent more than 1000 additional earthquakes ($M_L \geq M_C$) per year.

**Table 2.3.** Predicting $\boldsymbol{N_t}$ (number of earthquakes/year) in terms of hypothetical scenarios of different weighted average ( $\widehat{\boldsymbol{IW}}$; Equation (2)) or monthly constant IW in light of historical

records and benchmark periods. Uncertainty interval estimates have been added to each predicted $\boldsymbol{N_t}$. Historical benchmark periods have been extracted from section 4 this manuscript for reasons of comparison.

| $\widehat{\boldsymbol{IW}}$ $(\times \mathbf{10^6}\, \boldsymbol{m^3/month})$ | $\boldsymbol{N_t}$ $(\boldsymbol{number/year})$ | $\boldsymbol{N_t}$ **Interval** $[\boldsymbol{min, max}]$ $(\boldsymbol{number/year})$ | **Historical Benchmark Period** | **Sustainable Limit?** |
|---|---|---|---|---|
| **1** | $3.5 \times 10^{-5}$ | $2.53 \times 10^{-5}$, $5.87 \times 10^{-5}$ | - | - |
| **3** | 0.03 | 0.02, 0.05 | - | - |
| **5** | 0.76 | 0.50, 1.17 | - | - |
| **5.6** | 1.54 | 1.01, 2.34 | 1884-2002 | Pre-2002 |
| **6.8** | 5.07 | 3.32, 7.21 | 2003-2008 | Pre oil and gas boom (2003-2008) |
| **7** | 6.05 | 3.97, 9.23 | - | - |
| **9** | 28.4 | 18.6, 43.3 | - | - |
| **11** | 97.5 | 64.0, 148.6 | - | - |
| **13** | 272 | 179, 415 | - | - |
| **15** | 657 | 431, 1001 | - | - |
| **15.2** | 712 | 467, 1086 | 2009-2017 | Peak Period |
| **15.4** | 788 | 517, 1200 | 2017 | Oil/gas price fall/OCC regulation |

| | | | | |
|---|---|---|---|---|
| **17** | 1417 | 930, 2161 | - | - |
| **18.7** | 2547 | 1671, 3882 | 2015 | Peak year |
| **19** | 2809 | 1843, 4281 | - | - |
| **20** | 3851 | 2527, 5869 | - | - |
| **21** | 5198 | 3411, 7922 | - | - |
| **23** | 9095 | 5968, 13861 | - | - |

## 2.7 Model Output Intercomparison

This section offers a model result intercomparison of our parsimonious approach with two predictive models by Norbek and Rubinstein (2018) and Langenbruch and Zoback (2016). All models use monthly injection rates as predictand and time series- derived or geology-inferred parameters. For this intercomparison, we used the database published in Norbek and Rubinstein (Norbeck & Rubinstein, 2018) on monthly injection rates, observed seismicity and model outputs from the Hydromechanical (Norbeck & Rubinstein, 2018) and Seismogenic (Langenbruch & Zoback, 2016) models. Since Norbek and Rubinstein (Norbeck & Rubinstein, 2018) used different magnitude of completeness ($M_C$ = 3.0) and removed any instances of foreshocks and aftershocks from the main events, the parsimonious model had to be re-calibrated. For an $M_C = 3$ and declustered database, Equation (2.4) represents the number of expected main shocks as a function of the antecedent 25 months of wastewater injection (using Equation 2.2). Similar to Equation (2.3), this model explains 75% of the seismic activity at a monthly time scale.

$$N_t = 7.9630 \times 10^{-3} \widehat{IW}^{3.4556} \qquad (2.4)$$

Figure 2.9 illustrates the results of such a model intercomparison considering monthly injection rates between 2008 and 2018, including observed and predicted seismicity values from

the three models. As noted by Norbeck and Rubisntein (Norbeck & Rubinstein, 2018) although the hydromechanical model outputs seem to capture the general long-term trends, after year 2016, the model overestimates seismic activity, showing some weakness in capturing sharp changes in water injection. The seismogenic model seems to capture such short-term variability more thoroughly but tends to under-predict in times of low water injection. The parsimonious model proposed in this article seems to capture both low and high seismic activity but fails at capturing short-term sharp variability in a similar fashion to the hydromechanical model.

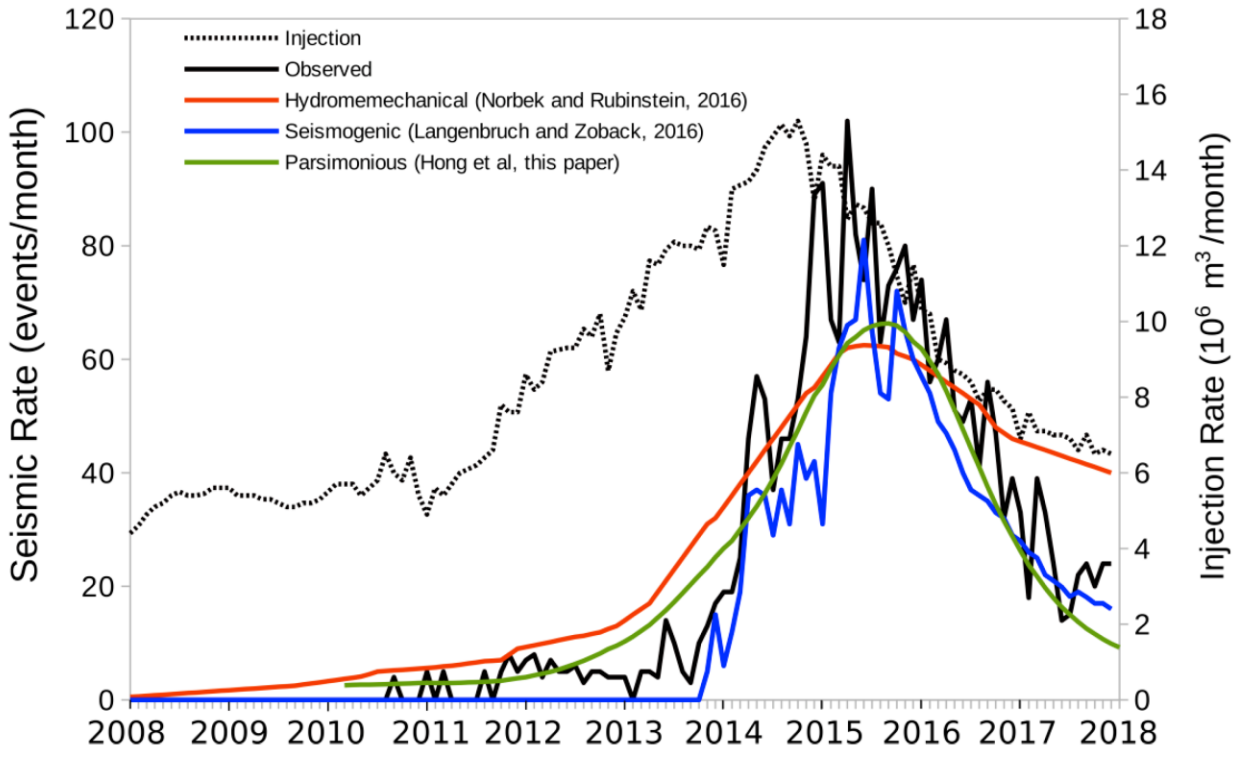


**Figure 2.9.** Model intercomparison experiment using the hydromechanical, seismogenic and parsimonious models for retrospective simulations of seismicity in Oklahoma between 2008 and 2018 in light of observed (declustered) seismic events and monthly wastewater injection rates.

## 2.8 Discussion

### *2.8.1 Acknowledging Methological Limitations*

The results from this study need to be understood considering the data and methodology limitations of our analyses. First, results mainly focus on statistical spatiotemporal relationships between wastewater injection volumes and earthquakes number and magnitude. Second, since the magnitude unit conversion (e.g., $M_w$ to $M_L$) procedure introduces a maximum uncertainty of 19% (see Section 2.3.1 and Table 2.2) the location and size of the weighted mean centers and standard deviation ellipses will have a maximum inherited error of 9.5%. Third, the analyses did not consider other influences on earthquakes' induction or generation mechanisms such as regional rock fracturing or geologic structures that propagate or moderate seismic waves. Moreover, due to the limited number of years with data (2006–2017), we do not know how the panoramic would look like in the future in views of higher (or lower) levels of wastewater injection. Further we recommend caution when planning to use the statistical relationships found here for future years as the rock systems might not behave in a linear fashion since the increasing rock-fracturing processes might propagate across larger regions, thus becoming a network of interconnected faulted systems that might translate to widespread earthquakes swarms. Finally, the results achieved in this study need to be further explored within different subregions to consider particular geological heterogeneities that could result in potentially different behaviors than the ones shown here.

### *2.8.2 Contributions to State-of-the-Art*

The conducted spatiotemporal analyses and proposed parsimonious model represent a novel contribution for prediction, model intercomparison and decision making. In terms of process

understanding, the results from this chapter are clear to relate the geographic scope and lagged dependency between wastewater injection volumes and earthquake count. Second, if used as stated, they can help predict the number of earthquakes in a particular month in terms of the antecedent monthly injection volumes. What we can define as sustainable extraction limit (conditions pre year 2000) could be 5.6 million $m^3/month$ or any combination of values of IW during the antecedent 25 months that allow obtaining around 1.5 earthquakes per year with $M_L \geq M_C$. A similar number of $5 \times 10^6\, m^3/month$ was also found by Langenbruch and Zoback (Langenbruch & Zoback, 2016). However, these authors propose a steady injection condition per month, but according to Equation (2) there could be other combinations of differential (seasonal) injection that could lead to the same result of minimum earthquakes. Third, possibly the best utility of the results of this chapter are their use as a tool for model intercomparison with current and future models. For example, Pollyea et al. (Pollyea et al., 2018) developed a geospatial analysis of the bivariate occurrence of earthquakes with the location of salt-water disposal wells. The results shown here are similar to Pollyea et al. (2018) in that there is a general pattern of north-west migration of both processes. However, one difference between these studies is that we obtained the two-axis variability ellipses and weighted the well and epicenter locations by magnitude and injection volumes, providing a more accurate description of their spatially correlated distribution. We also provide year by year direction of migration and distance patterns. Results from the model output intercomparison experiment show comparable capabilities of the Parsimonious (Hong et al., 2018) and Hydromechanical (Norbeck & Rubinstein, 2018) models in the long-term with the Parsimonious model representing better the recent decline in seismicity conditions. However, both these models seem to have a weak performance at detecting rapid changes better captured by the seismogenic model (Langenbruch & Zoback, 2016).

*2.8.3 Contributions to Sustainable Extraction and Decision Making: What are Sustainable Limits?*

Due to the earthquake upsurge since 2009, the Oklahoma Corporation Commission adopted a "traffic light" system since 2013 in response to the concerns over underground fluid injection induced earthquakes. In a "traffic light" system, if no underground fluid injection induced earthquakes occur, operators could continue their injection activities at regulated rates under a green light condition. Once an earthquake occurs, operators are under yellow light condition. They should investigate the relationship between the earthquake and injection activities and reduce injection rates. If an earthquake event induced by underground injection occurs and the triggered seismicity cannot be stopped by reducing injection rates, operators are under a red light condition and should be prepared to terminate injection activities (Oklahoma Corporation Commission, 2015; Pollyea et al., 2018; Zoback, 2012). The OCC yellow light permitting system requires operators to monitor for background seismicity and shut down wells to record bottom hole pressure every 60 days. The Oklahoma Corporation Commission has been evolving the "traffic light" system applications based on updated research results and new data (Oklahoma Corporation Commission, 2017). The slight decrease in earthquake occurrence in 2016 and 2017 (Figure 2.3) has been attributed to these mitigation efforts. However, as noted by Pollyea et al. (2018), these decreases could have also been the result of the dramatic drop in oil prices. The results presented in this manuscript could be used as a cause-effect method whose results could be used to potentially improve the current "traffic light" policy, and to inform legislators and decision makers by providing sustainable limits for oil and gas extraction in order to minimize the expected number and magnitude of subsequent quakes, thus avoiding future human and property losses. The availability of more information for upcoming years will serve to provide robustness, not only to

this, but to other current methods with the main purpose to raise conscience of the potential of human-induced seismic activity and balance out gains for economy, environment and society.

## 2.9 Conclusions

This study has gathered comprehensive datasets of oil and gas industry-related wastewater injection volumes and earthquake number with associated event magnitudes from 2006 to 2017 over Oklahoma. Data were analyzed to remove those seismic events below the threshold of magnitude completeness. First, we explored the spatiotemporal variability of both processes and conclude that a high correspondence between the two supports the hypothesis that the boom in oil and gas production through unconventional methods with wastewater injection was potentially responsible for the upsurge in the state's seismic activity during 2006 through 2015. Also, a reduction in the number of earthquakes per year in years 2016 and 2017, reflect either the mitigation policies dictated by OCC or the drop in oil and gas prices, or both. Second, a cluster analysis reveals a correlated migration pattern between earthquake occurrences and saltwater injection activity. Following the migration of the weighted wastewater injection ellipses, weighed epicenters show a predominant northwest direction pattern during the 2007–2017 period. Third, a lagged cross-correlation analysis shows that the number of induced earthquakes in a subsequent month is strongly associated with the previous 25-month cumulative wastewater injection volume and a power law can be fitted between number of quakes and weighted average monthly injection volumes as predictive tool with a coefficient of determination of $R^2 = 0.77$. Using such a relation, several sustainable extraction limits are explored and compared with historic means. Results from these analyses expand on previously sustainable limits of 5 to 6 million $m^3/month$ to potential combinations that could result in the same number of earthquakes within the 25 previous months.

A model intercomparison of our parsimonious model, a hydromechanical model, and a seismogenic model reveals a satisfactory performance of the proposed approach and similitude to the hydromechanical model outputs. Nonetheless, sharp changes in seismicity could only be more appropriately represented by the seismogenic model. The approach proposed in this chapter could potentially be regionalized according to the geology of each zone and results could potentially be used as a tool for further model intercomparison experiments and decision making on spatially varied permission distribution and regional industry development to minimize negative consequences of induced earthquakes.

**References**


Brumbaugh, D. S. (1989). A comparison of duration magnitude to local magnitude for seismic events recorded in northern Arizona. *Journal of the Arizona-Nevada Academy of Science*, 29–31.

Burt, J. E., Barber, G. M., & Rigby, D. L. (2009). *Elementary statistics for geographers*. Guilford Press.

Council, N. R. (2013). *Induced seismicity potential in energy technologies*. National Academies Press.

Crain, K., Chang, J. C., & Walter, J. I. (2017). Geophysical anomalies of Osage County and its relationship to Oklahoma seismicity. *AGU Fall Meeting Abstracts*, *2017*, S23C-0816.

Ellsworth, W. L. (2013). Injection-induced earthquakes. *Science*, *341*(6142).

Gutenberg, B., & Richter, C. F. (1944). Frequency of earthquakes in California. *Bulletin of the Seismological Society of America*, *34*(4), 185–188.

Habermann, R. E. (1991). Seismicity rate variations and systematic changes in magnitudes in teleseismic catalogs. *Tectonophysics*, *193*(4), 277–289.

Hincks, T., Aspinall, W., Cooke, R., & Gernon, T. (2018). Oklahoma's induced seismicity strongly linked to wastewater injection depth. *Science*, *359*(6381), 1251–1255.

Holland, A. A. (2013). Earthquakes triggered by hydraulic fracturing in south-central Oklahoma. *Bulletin of the Seismological Society of America*, *103*(3), 1784–1792.

Horton, S. (2012). Disposal of hydrofracking waste fluid by injection into subsurface aquifers triggers earthquake swarm in central Arkansas with potential for damaging earthquake. *Seismological Research Letters*, *83*(2), 250–260.

Hough, S. E., & Page, M. (2015). A century of induced earthquakes in Oklahoma? *Bulletin of the Seismological Society of America*, *105*(6), 2863–2870.

Keranen, K. M., Savage, H. M., Abers, G. A., & Cochran, E. S. (2013). Potentially induced earthquakes in Oklahoma, USA: Links between wastewater injection and the 2011 Mw 5.7 earthquake sequence. *Geology*, *41*(6), 699–702.

Keranen, K. M., Weingarten, M., Abers, G. A., Bekins, B. A., & Ge, S. (2014). Sharp increase in central Oklahoma seismicity since 2008 induced by massive wastewater injection. *Science*, *345*(6195), 448–451.

Kim, W.-Y. (2013). Induced seismicity associated with fluid injection into a deep well in Youngstown, Ohio. *Journal of Geophysical Research: Solid Earth*, *118*(7), 3506–3518.

Langenbruch, C., Weingarten, M., & Zoback, M. D. (2018). Physics-based forecasting of man-made earthquake hazards in Oklahoma and Kansas. *Nature Communications*, *9*(1), 1–10.

Langenbruch, C., & Zoback, M. D. (2016). How will induced seismicity in Oklahoma respond to decreased saltwater injection rates? *Science Advances*, *2*(11), e1601542.

Llenos, A. L., & Michael, A. J. (2013). Modeling earthquake rate changes in Oklahoma and Arkansas: Possible signatures of induced seismicity. *Bulletin of the Seismological Society of America*, *103*(5), 2850–2861.

Murray, K. E., & Holland, A. A. (2014). *Inventory of class II underground injection control volumes in the midcontinent*.

Norbeck, J. H., & Rubinstein, J. L. (2018). Hydromechanical earthquake nucleation model forecasts onset, peak, and falling rates of induced seismicity in Oklahoma and Kansas. *Geophysical Research Letters*, *45*(7), 2963–2975.

Oklahoma Corporation Commission. (2015). *Media Advisory—Ongoing OCC Earthquake Response*. http://www.occeweb.com/News/2015/03-25-15%20Media%20Advisory%20-%20TL%20and%20related%20documents.pdf

Oklahoma Corporation Commission. (2017). *Earthquake Response Summary*. http://www.occeweb.com/News/2017/02-24-17EARTHQUAKE%20ACTION%20SUMMARY.pdf

Oklahoma Corporation Commission. (2018). *Oil and Gas Data Files*. http://www.occeweb.com/OG/ogdatafiles2.htm

Oklahoma Geological Survey. (2018). *Earthquake Catalogs*. https://www.ou.edu/content/ogs/research/earthquakes/catalogs.html

Pollyea, R. M., Mohammadi, N., Taylor, J. E., & Chapman, M. C. (2018). Geospatial analysis of Oklahoma (USA) earthquakes (2011–2016): Quantifying the limits of regional-scale earthquake mitigation measures. *Geology*, *46*(3), 215–218.

Shah, A. K., & Keller, G. R. (2017). Geologic influence on induced seismicity: Constraints from potential field data in Oklahoma. *Geophysical Research Letters*, *44*(1), 152–161.

Van der Elst, N. J., Savage, H. M., Keranen, K. M., & Abers, G. A. (2013). Enhanced remote earthquake triggering at fluid-injection sites in the midwestern United States. *Science, 341*(6142), 164–167.

Walsh, F. R., & Zoback, M. D. (2015). Oklahoma's recent earthquakes and saltwater disposal. *Science Advances*, *1*(5), e1500195.

Weingarten, M., Ge, S., Godt, J. W., Bekins, B. A., & Rubinstein, J. L. (2015). High-rate injection is associated with the increase in US mid-continent seismicity. *Science, 348*(6241), 1336–1340.

Wiemer, S., & Wyss, M. (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the western United States, and Japan. *Bulletin of the Seismological Society of America*, *90*(4), 859–869.

Woessner, J., & Wiemer, S. (2005). Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty. *Bulletin of the Seismological Society of America*, *95*(2), 684–698.

Zoback, M. D. (2012). Managing the seismic risk posed by wastewater disposal. *Earth*, *57*(4), 38.

# Chapter 3 Cross-evaluation of Ground-based, Satellite and Land Surface Model Soil Moisture Products through the Triple Collocation Method across Oklahoma


## Abstract

Improvements in the field of soil moisture observations and modeling play a vital role in drought, water resources, flooding, and landslide management and predictability. Up to date soil moisture estimates can be obtained through three primary approaches: (1) in situ measurements, (2) remote sensing observations and (3) Land Surface Models (LSM). Each source of soil moisture data has its strengths and weaknesses. However, all three main soil moisture measurements are subject to representativeness inadequacies over various land cover types and correct interpretation and application of their products requires an in-depth understanding of their accuracy. In this chapter, we apply Triple Collocation (TC) analysis to three independent soil moisture products in order to characterize their uncertainty structures. Oklahoma is an ideal domain to test the hypotheses of this work due to the presence of marked west to east gradients in climate, vegetation, and soils. The comparison and evaluation are conducted with daily data from 01 April 2015 to 01 July 2019 over seven land cover types. The three soil moisture products evaluated include: microwave remotely sensed Soil Moisture Active Passive (SMAP) L3_SM_P_E (9 km, daily) measurements, physically based land surface modeling soil moisture estimates NLDAS_NOAH0125_H (1/8°, hourly; Noah), and the in-situ soil moisture network of the Oklahoma Mesonet (point, 30 minutes). Results indicate that in general, Mesonet is the most reliable point product, reflecting the main spatiotemporal characteristics of soil moisture, while SMAP has the lowest accuracy. The spatio-temporally integrated mean square errors (RMSE), with respect to the unknown true soil moisture, of Mesonet, Noah, and SMAP are 0.054, 0.026,

and 0.107 $m^3/m^3$, respectively, while the overall Pearson correlation coefficients (CC) are 0.805, 0.747, and 0.314. Mesonet has the best performance in shrub/scrub, herbaceuous, hay/pasture, and cultivated crops with an average correlation coefficient of 0.785. Noah achieves the best performance in evergreen, mixed and deciduous forest, with an average correlation coefficient of 0.74. SMAP has the lowest correlation coefficient values on all seven land cover types, with an average CC of 0.29. The study concludes that the TCA method provides not only a new perspective for comparatively assessing multi-source soil moisture products but also a basis for objective data merging to capitalize on the strengths of multi-sensor multiplatform soil moisture products.

### 3.1 Introduction

Soil moisture (SM) refers to the water held in the space between soil particles and is often quantified as soil water content, the ratio of water to soil in volume or weight. This variable plays a fundamental role in controlling the water, energy and biogeochemical cycles and is a fundamental factor in sustaining and ending droughts (Seneviratne et al., 2010) but also for triggering or enhancing floods and mass movements. While soil moisture only accounts for a very small amount (approximately 0.05%) of the total quantity of water within the global hydrological cycle, it plays a critical role in the climate system and the hydrological cycle (Dingman, 2015; Seneviratne et al., 2010). At the field scale, accurate measurement of soil moisture could benefit precision agriculture, especially precise irrigation and fertilization (Mohd Kassim et al., 2014; Woodley, 2017). At larger spatial scales, accurate soil moisture measurements could enhance capabilities in runoff and flood forecasting (Brocca et al., 2010; Crow et al., 2005), drought monitoring and prediction (Gu et al., 2008; Zhang, et al., 2017), numerical weather prediction

(Capecchi & Brocca, 2014; Scipal et al., 2008), rainfall estimation (Brocca et al., 2012; Crow et al., 2011) and landslide prediction (Crow et al., 2012; Ray et al., 2010).

Currently, soil moisture estimates can be obtained through three primary approaches: (1) in situ measurements, (2) remote sensing observations, and (3) Land Surface Models (LSM). In-situ soil moisture monitoring networks are particularly important for validating and calibrating remote sensing and land surface model-based soil moisture retrievals (Crow et al., 2012). For example, soil moisture data from the Oklahoma Mesonet are widely used in land surface model validation (Xia et al., 2012, 2014; Xu et al., 2021), and soil moisture remote sensing validation (Crow et al., 2012). Land surface models such as the Global Land Data Assimilation System (GLDAS) (Rodell et al., 2004) and North American Land Data Assimilation System (NLDAS) (Mitchell et al., 2004) can provide soil moisture estimates at various depths with high spatio-temporal resolution. Xia et al., (2014) assessed daily and monthly simulation skills of four NLDAS-2 land surface models and utilized 6-years (1 January 1997–31 December 2002) of daily soil moisture observed from 72 sites over the Oklahoma Mesonet network to assess daily and monthly simulation skill and found that all models are able to capture wet and dry events and show high skill. Finally, remote sensing soil moisture products from various sensors (e.g., microwave, optical and thermal sensors) provide global-scale soil moisture measurements but with limited spatial and temporal resolutions (Brocca et al., 2017). For example, The SMAP Level-3 products provides a daily composite of global soil moisture retrieved by both the Soil Moisture Active Passive radar and radiometer at 9 km spatial resolution(ONeill et al., 2019).

However, due to the characteristics of three main sources of soil moisture measurement (in situ, land surface model, and satellite), their data quality or representativeness of their values might vary over different land cover types. For example, the Oklahoma Mesonet site standards minimize

the influence of urban landscapes, irrigation, forest, bare soil, fast growing vegetation, and large bodies of water (McPherson et al., 2007). It is suggested that vegetation at the site should be uniform and low growing such as short grasses (Brock et al., 1995). Therefore, soil moisture measurements at the Oklahoma Mesonet sites may not well represent SM variations over bare soil, crops, forest, and other fast growing vegetations. The vegetation classification of NLDAS land surface models was derived from the global, 1-km, AVHRR-based, 13-class vegetation database of UMD (Noah; (Hansen et al., 2000)). For each 1/8° grid cell, Noah uses the most predominant vegetation class (Mitchell et al., 2004). Xia et al. (2014) evaluated 20-years (January 1985–December 2004) of NLDAS-2 model-simulated soil moisture with in situ measurements over the continental United States and concluded that the performance for all models is higher in the Southeast, Great Plains, Midwest, and Northwest, and lower in the Southwest and the Northeast with their dominant vegetation cover as forest, grassland, a mixture of cropland and grasslands, grassland, open shrubland, and forest respectively. Zhang et al. (2019) conducted a comprehensive validation of the SMAP Level 3 SM product with ground measurements over varied climates and landscapes from April 1, 2015 to March 31, 2018. Results showed that SMAP level 3 SM product had better performance over grassland than over cropland. In summary, all three main soil moisture measurements are subject to representativeness inadequacies over various land cover types and correct interpretation and application of their products requires an in-depth understanding of their accuracy.

Therefore, the overarching goal of this chapter is to cross-evaluate three widely used soil moisture products independently across Oklahoma and over various land cover types. Specifically, this chapter shows the results of a comprehensive assessment of the Satellite SMAP_L3 (SMAP), Land Surface NOAH Model (Noah), and the interpolated Mesonet soil moisture (Mesonet)

products across Oklahoma at daily and seasonal timescales using the TC method evaluated over different land cover types. This chapter is the first to adopt triple collocation method in intercomparison between in-situ, model-based, and satellite soil moisture products over various land cover types in Oklahoma and the results are expected to provide not only a new perspective for comparatively assessing multi-source soil moisture products over different land cover types but also a basis for objective data merging to capitalize the strengths of multi-sensor multiplatform soil moisture products. The rest of the chapter is organized as follows: Section 3.2 details the data and study area; Section 3.3 describes the methods and data processing; Section 3.4 presents the results and analysis; Section 3.5 provides a discussion; and the Section 3.6 provides some conclusions for this study.

## 3.2 Data Sources

### *3.2.1 Satellite soil moisture product: SMAP L3_SM_P_E*

Launched in January 2015, SMAP is an orbiting observatory that estimates the amount of water in the top 0 – 4 inches (0 – 10 centimeters) of soil everywhere on Earth's land surface every two to three days. SMAP was designed to provide high-resolution soil moisture information with radar (active) and radiometer (passive) sensors that operate at L-band frequencies. However, the radar instrument terminated its operation due to failure of its power supply after three months of data collection. The SMAP radiometer has been operating flawlessly and in extended operation phase since 2018 (Entekhabi et al., 2014).

In total, the SMAP mission has generated 23 distributable data products representing four levels of data processing. Level 1 products are instrument-related data sectioned into surface radar backscatter cross-section and brightness temperatures. Level 2 products are geophysical retrievals

in half orbit granules resulting from instrument data. Level 3 products are daily global composites of Level 2 data for an entire UTC day. Level 4 products are outputs from geophysical models utilizing SMAP data (Chan, 2016). The reasons why the SMAP Level 3 product is used in this chapter are: (1) Even though both Level 2 and Level 3 products are geophysical retrievals, Level 3 values are daily global composites of Level 2; and (2) while both Level 3 and Level 4 products are daily global retrievals, only Level 3 can satisfy the independency assumption of triple collocation analysis (Scipal et al., 2008). Since Level 4 products are outputs from geophysical models, it is difficult to justify that the Level 4 product is independent from the NLDAS2 Noah model derived soil moisture product.

The SMAP Level 3 product used in this study is the Enhanced L3 Radiometer Global Daily 9 km EASEGrid Soil Moisture, Version 3 (L3_SM_P_E; (ONeill et al., 2019)). It is a daily global composite of the enhanced SMAP L2_SM_P_E product, which contains gridded data of 6:00 am (descending) and 6:00 pm (ascending) SMAP radiometer-based soil moisture retrievals, ancillary data, and quality assessment flags on the global 9-km Equal-Area Scalable Earth (EASE 2.0) grid. The main output of this dataset is 0-5 cm surface soil moisture. This product is publicly available through the National Snow and Ice Data Center. Surface soil moisture data of SMAP L3_SM_P_E product pertaining to the period from 04-01-2015 to 07-01-2019 were used in this study (ONeill et al., 2019).

### *3.2.2 Model-based soil moisture product: NLDAS_NOAH0125_H*

The North American Land Data Assimilation System phase 2 (NLDAS-2) is an offline data assimilation system running four land surface models (Noah, SAC-SMA, VIC, and Mosaic) over the conterminous United States (CONUS), the southern part of Canada, and the northern portion

of Mexico with a 1/8° latitude-longitude resolution (Xia et al., 2012). The four land surface models represent different methodological approaches to land surface modeling. This study uses the simulated soil moisture from the NLDAS-2 Noah model. The Noah model is the land model of the NCEP (National Centers for Environmental Modeling Prediction) operational regional and global weather and climate models (Betts et al., 1997; F. Chen et al., 1997; Ek et al., 2003). It provides hourly soil moisture fields at 1/8° grid from 1979 to present. The Noah model has four soil layers: 0–10 cm, 10–40 cm, 40–100 cm, and 100–200 cm and simulates soil moisture at the middle of each soil layer (5, 25, 70, and 150 cm), but only the top layer value (i.e., 5 cm) is used in this study. Xia et al. (2014) compared soil moisture estimates of four NLDAS-2 land surface models (Noah, Mosaic, SAC, VIC) with three in-situ soil moisture observation data sets in the United States (the Illinois Climate Network, the Oklahoma Mesonet network, and the Soil Climate Analysis Network) to find that Noah had the smallest mean absolute error (MAE=0.036), root mean square error (RMSE=0.04) and bias (Bias=-0.033) in the comparison with the Oklahoma Mesonet observations for absolute daily soil moisture at the top 10 cm soil layer in a six year period (from January 1, 1997 to December 31, 2002). Moreover, Noah was much closer to the observations over six year averaged daily volumetric soil moisture (Xia et al., 2014). Therefore, the hourly soil moisture simulations of NLDAS-2 Noah model between the period of 04-01-2015 to 07-01-2019 were used in this study (*GES DISC Dataset: NLDAS Noah Land Surface Model L4 Monthly Climatology 0.125 x 0.125 Degree V002 (NLDAS_NOAH0125_MC 002)*, n.d.).

### *3.2.3 In situ soil moisture product: the Oklahoma Mesonet*

The Oklahoma Mesonet is a world-class statewide network of environmental monitoring stations that was established in January 1994. It measures atmospheric, hydrologic, and

meteorological variables including temperature, humidity, solar radiation, wind speed and direction, and soil moisture to aid in operational weather forecasting and environmental research across the state (McPherson et al., 2007). With at least one station in each of Oklahoma's 77 counties, the Mesonet consists of 120 automated stations across the state. These measurements are packaged into observations every five minutes, and then transmitted to the Oklahoma Climatological Survey (OCS) at the University of Oklahoma (OU), where the observed data are processed and verified for their quality, and then made public. Soil moisture data are collected every 30 min and recorded locally at each site including at the surface (Illston et al., 2004). Since the TC analysis requires three spatially and temporally collocated measurement systems and SMAP measurements are intermittent (6:00 am and 6:00 pm), the Mesonet soil moisture measurements at local solar time 6 am and 6 pm from 04/01/2015 to 07/01/2019 were used in this study. A total of 115 Mesonet sites were selected according to the data availability during the study time (see Figure 3.1). Since the data points are spread-out across the state, interpolation was conducted using ordinary Kriging, as previously suggested by (Lakhankar et al., 2010) and then re-gridded to 9 km.

### *3.2.4 Auxiliary Data*

To better understand the performance of the three soil moisture products over different land cover types, the national land cover dataset (NLCD) 2016 product was used in this study. The NLCD provides nationwide data on continental U.S. land cover and land cover change at a 30 m resolution with a 16-class legend based on a modified Anderson Level II classification system. There are fifteen land cover types within Oklahoma as shown in Figure 1. Oklahoma is divided into nine climate divisions (Figure 3.2). These nine divisions are determined by multiple factors,

including climatic conditions, county lines, crop districts, and drainage basins rather than strict climatic homogeneity (Guttman & Quayle, 1996; Illston et al., 2004). Therefore, analyzing the performance of three soil moisture products in the nine climate divisions provides a unique opportunity to understand better how land cover impacts soil moisture.

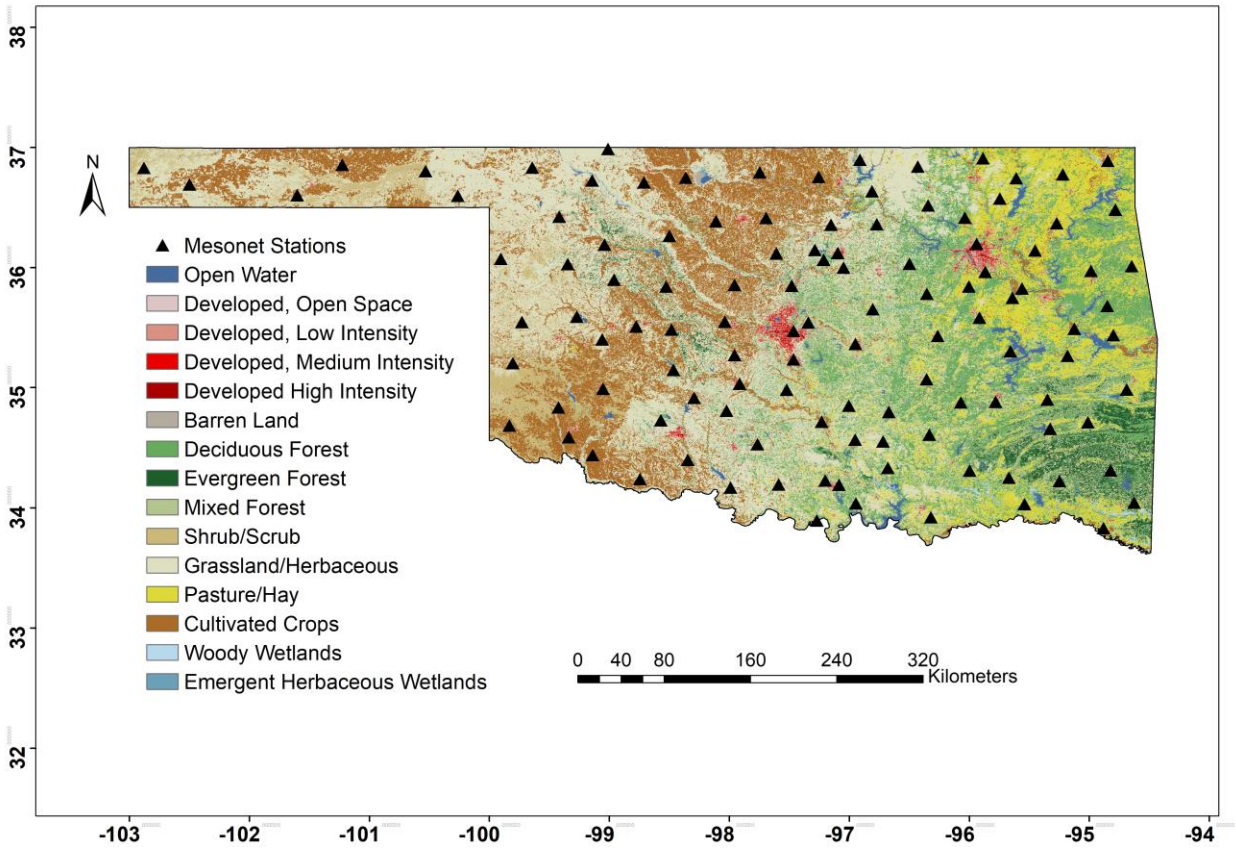


**Figure 3.1.** The distribution of in situ soil moisture stations from the Oklahoma Mesonet on a National Land Cover Dataset (NLCD) land cover type (for year 2016) map.

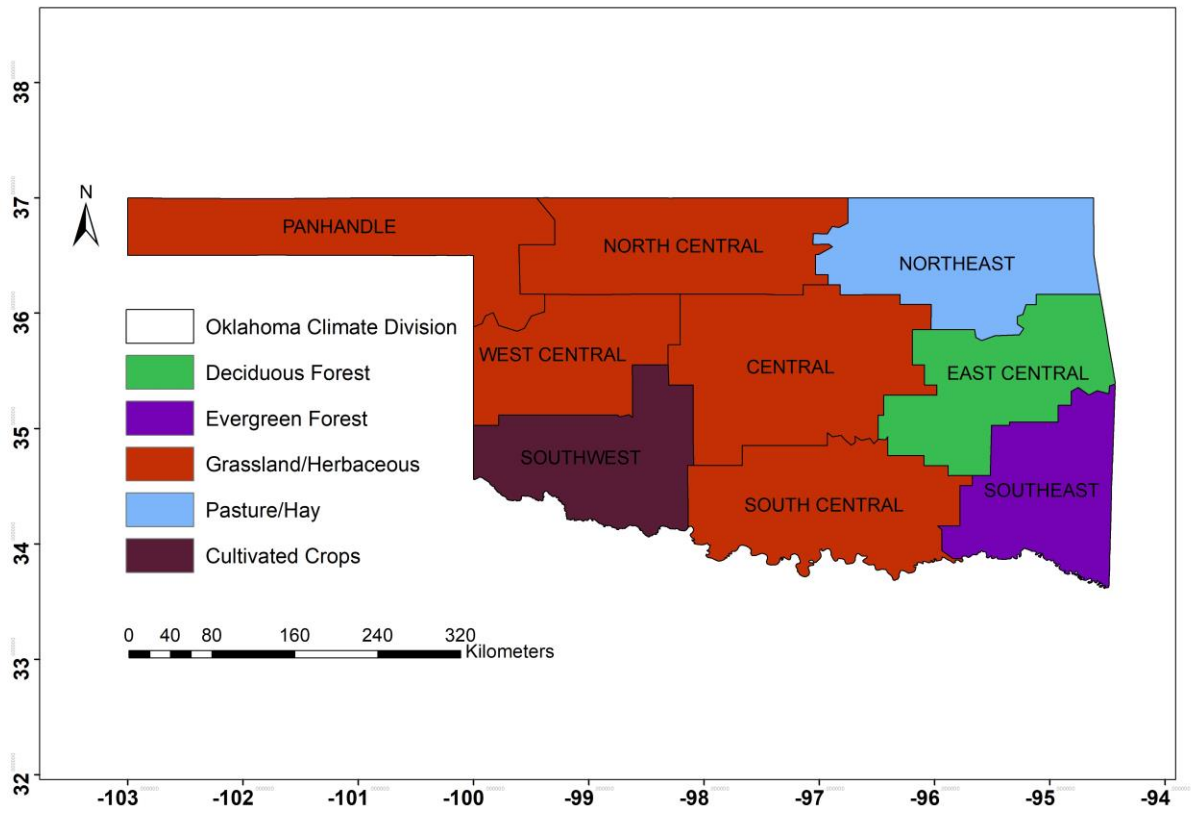


**Figure 3.2.** Oklahoma climate divisions with their major land cover types based on the 2016 NLCD map.

## 3.3 Data Processing and the Triple Collocation Method

### *3.3.1 Data processing*

Table 3.1 presents metadata of the product triplet including version, data spans, spatio-temporal resolutions and soil depth. Since the TC analysis requires three spatially and temporally collocated measurement systems, the grid of SMAP (EASE_v2) was defined as the reference for the three products. Therefore, Noah data were resampled to this grid using the area-weighted average method while Oklahoma Mesonet soil moisture measurements were matched to the EASE_v2 grid using ordinary kriging (Lakhankar et al., 2010). Figure 3.3 shows the temporal coverage of three soil moisture products during the study period (April 2015 through July 2019).

Mesonet and Noah provide continuous data but SMAP measurements are intermittent because the same swath from each orbit of SMAP is only repeated every eight days. The three collocated soil moisture products (SMAP, Mesonet, and Noah) are evaluated at daily and seasonal timescales using the TC method over different land cover types across Oklahoma. For the comparison, two time stamps are selected, one diurnal (6 AM LST) and one nocturnal (6 PM LST) according to the availability of SMAP. Table 3.2 shows the number of collocated samples for each grid point at daily and seasonal time scales. All sample sizes are larger than the suggested TC sample size (100) by Scipal et al. (2008).

**Table 3.1.** Summary of satellite (SMAP L3_SM_P_E herein called SMAP), model (NLDAS_NOAH0125_H), and in situ (Oklahoma Mesonet) soil moisture products used in this study.

| Data | Version | Data Period | Temporal Resolution | Spatial Resolution | Depth |
|---|---|---|---|---|---|
| SMAP | L3_SM_P_E | 2015 – present | daily | 9 km | 0-5 cm |
| Noah | 0125_H | 1979 - present | hourly | 0.125° | 0-10 cm |
| Mesonet | 115 sites | 1998 - present | 30 minute/daily | point | 0-5 cm |

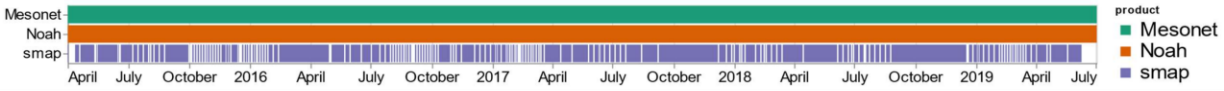


**Figure 3.3.** Temporal coverage of SMAP, Noah, and Mesonet soil moisture products used in this study.

**Table 3.2.** Sample sizes of 6 AM and 6 PM and Seasonal TC triplets.

| Sample Size | AM | PM | Spring | Summer | Autumn | Winter |
|---|---|---|---|---|---|---|
| Days | 746 | 714 | 394 | 340 | 326 | 360 |

### *3.3.2 Classical Triple Collocation*

Triple collocation (TC) analysis is a method for estimating the random error variances of three spatially and temporally collocated measurement systems of the same geophysical variable without treating any one system as perfectly observed "truth" (Stoffelen, 1998). A few assumptions are

necessary for the TC method: (1) linearity between the true soil moisture signal and the observations, (2) signal and error stationarity, i.e., their mean values and variances are assumed to remain constant over time, (3) error orthogonality, i.e., the errors are independent from the true soil moisture signal, (4) the errors of three independent products should be independent or unrelated which means they must have a zero cross-correlation, and (5) the expectation of error is treated as zero. Yilmaz and Crow (2014) conducted experiments on the TC errors due to the relevance of three products, and the results revealed that the more independent they are, the less TC induced error there will be. It is essential to consider the relevance of the inputs in order to make the TC method more reliable (Li et al., 2018). The three selected soil products selected, i.e., ground based (Mesonet), model based (Noah), and satellite-based (SMAP), all meet the above criteria.

The TC method treats all three independent products as equally important, and thus no preference or bias is introduced for any one approach. Equation (3.1) illustrates a standard form of the TC method (Zwieback et al., 2012):

$$R_i = a_i + b_i \mathrm{T} + \varepsilon_i \qquad (3.1)$$

Where $R_i$ (i $\in (X, Y, Z)$) indicates each of the three collocated soil moisture datasets X, Y and Z, T is the "relative truth," $a_i,\ b_i$ are the weights and biases to adjust, and $\varepsilon_i$ represents the error for each product $i$. Given this definition, the covariances between pairs of two different measurement systems (e.g., X and Y) would be given by

$$Cov(R_X, R_Y) = E(R_X R_Y) - E(R_X)E(R_Y)$$
$$= b_X b_Y \sigma_T^2 + b_X Cov(T, \varepsilon_Y) + b_Y Cov(T, \varepsilon_X) + Cov(\varepsilon_X, \varepsilon_Y) \qquad (3.2)$$

Where $\sigma_T^2 = Var(T)$. According to assumptions (3), (4), (5), $E(\varepsilon_X) = 0$, $(Cov(\varepsilon_X, \varepsilon_Y) = 0, X \neq Y)$, $(Cov(T, \varepsilon_X) = 0)$. Therefore, equation (3.2) reduces to

$$Q_{XY} = Cov(R_X, R_Y) = \begin{cases} b_X b_Y \sigma_T^2 & for\ X \neq Y \\ b_X^2 \sigma_T^2 + \sigma_{\varepsilon_X}^2 & for\ X = Y \end{cases} \qquad (3.3)$$

Where $\sigma_{\varepsilon_X}^2 = Var(\varepsilon_X)$. Since there are seven unknowns $(b_X, b_Y, b_Z, \sigma_{\varepsilon_X}, \sigma_{\varepsilon_Y}, \sigma_{\varepsilon_Z}, \sigma_T)$ in six equations in the 3 × 3 covariance matrix $(Q_{XX}, Q_{XY}, Q_{XZ}, Q_{YY}, Q_{YZ}, Q_{ZZ})$, there is no unique solution. However, the introduction of a new variable $\theta_X = b_X \sigma_T$, changes (3.3) to

$$Q_{XY} = Cov(R_X, R_Y) = \begin{cases} \theta_X \theta_Y & for\ X \neq Y \\ \theta_X^2 + \sigma_{\varepsilon_X}^2 & for\ X = Y \end{cases} \qquad (3.4)$$

From equation (3.4), we now have six unknowns in six equations and are able to calculate the root mean square error (RMSE) in the set of equations (3.5) that are based on the covariance of triplets (McColl et al., 2014):

$$\sigma_\varepsilon = \begin{cases} \sqrt{Q_{XX} - \frac{Q_{XY} Q_{XZ}}{Q_{YZ}}} \\ \sqrt{Q_{YY} - \frac{Q_{XY} Q_{YZ}}{Q_{XZ}}} \\ \sqrt{Q_{ZZ} - \frac{Q_{XZ} Q_{YZ}}{Q_{XY}}} \end{cases} \qquad (3.5)$$

*3.3.3 Extended Triple Collocation*

Using the same assumptions as TC, McColl et al. (2014) introduced an additional performance metric, the Pearson correlation coefficient (CC) of the measurement system with respect to the unknown target with the called "ETC" method in which CC is calculated as a set of equations (3.6).

$$\begin{cases} {CC_X}^2 = \dfrac{Q_{XY}Q_{XZ}}{Q_{XX}Q_{YZ}} \\ {CC_Y}^2 = \dfrac{Q_{XY}Q_{YZ}}{Q_{YY}Q_{XZ}} \\ {CC_Z}^2 = \dfrac{Q_{XZ}Q_{YZ}}{Q_{ZZ}Q_{XY}} \end{cases} \tag{3.6}$$

*3.3.4 Use of the Classical and Extended Triple Collocation for the Three Testing Products*

The mathematical derivations explained in equations 3.4 through 3.6 will be used to conduct the evaluation of the three study products (i.e., SMAP, Noah and Mesonet) in light of the TC method. Since both RMSE and CC are derived from covariances between the three products, they reveal the relative error as a measurement of the uncertainty. Therefore, the least uncertain product, represented by the lowest RMSE and highest CC, will have the best performance. Likewise, the most uncertain product will be associated with the highest RMSE and lowest CC. Results are presented through gridded maps of instantaneous and seasonally-discretized (i.e., spring, summer, fall, winter) RMSE and CC and boxplots for groups of pixels with the same land cover type. The instantaneous values for comparison are extracted for two hours of the day, 6:00 am and 6:00 pm, determined by the available SMAP satellite geographical overpasses during the time period of April 2015 to July 2019. The TC analysis is conducted over Oklahoma and the different land cover types are extracted from the auxiliary data (see section 3.2.4) to test the degree of dependency of each product's performance with the land cover class. According to the NLCD 2016 product, there

are fifteen (15) land cover types in the state (Figure 3.1). Table 3.3 contains the number of selected TC intercomparison pixels with a spatial resolution of nine (9) km for each land cover type in Oklahoma except open water, woody wetlands and emergent herbaceous wetlands. With the aim of achieving statistical representativeness and preserving class diversity, the developed low intensity, medium intensity and high intensity are classified as one land cover type "developed." Moreover, the triple product comparison was conducted for land cover types with more than 10 co-located pixels state-wide.

**Table 3.3.** Number of 9kmx9km grid cells in each of the land cover types in Oklahoma.

| Land Cover Type | Number of Co-located Pixels |
|---|---|
| Developed Open Space | 137 |
| Developed Low Intensity | 25 |
| Developed Medium Intensity | 7 |
| Developed High Intensity | 7 |
| Deciduous Forest | 414 |
| Evergreen Forest | 74 |
| Mixed Forest | 43 |
| Shrub/Scrub | 124 |
| Grassland/Herbaceous | 744 |
| Hay/Pasture | 229 |
| Cultivated Crops | 338 |
| Barren Land (Rock/Sand/Clay) | 7 |

## 3.4 Results

### *3.4.1 Sub-daily Product Inter-comparison of Soil Moisture Values*

Figure 3.4 shows the CC and RMSE statistics obtained for simultaneous diurnal 6:00 AM observations at a daily time scale over Oklahoma. Overall, Mesonet provides the highest spatio-temporal, integrated averaged CC ($CC_{mean}$) of 0.805, followed by Noah with an $CC_{mean}$ of 0.747. SMAP has the lowest $CC_{mean}$ of 0.314 at the state level. However, the correlation varies with different locations and products (Figure 3.4, Table 3.4). Based on the Climate Divisions from NOAA's Climate Divisional Database (Vose, et al. 2014), Mesonet provides high averaged CC values in the Southwest, West Central, South Central and Panhandle (0.92, 0.9, 0.88, 0.88 respectively) regions, while lower averaged correlation is shown in the Northeast (CC=0.58). Noah exhibits high averaged CC in the Panhandle and South Central (CC = 0.79, 0.86 respectively) regions but lower values in the Northeast and West Central divisions (CC = 0.66 and 0.67 respectively). The regional mean CC values of SMAP are generally higher in the Panhandle, Southwest, and West Central divisions (CC= 0.49, 0.47, 0.39 respectively) than those in the other six climate divisions. In terms of the RMSE, Noah provides the smallest state-wide $RMSE_{mean}$ of 0.026 $m^3/m^3$, followed by Mesonet ($RMSE_{mean} = 0.054\, m^3/m^3$) and SMAP ($RMSE_{mean} = 0.107\, m^3/m^3$) at the state level. But the error varies across different locations and products (Figure 4, Table 4). Mesonet shows low RMSE values (0.04 $m^3/m^3$) in Southwest, South Central, West Central, and Panhandle but high values in the Northeast (0.09 $m^3/m^3$). Noah exhibits small RMSE mean values in all nine divisions (equal or less than 0.04 $m^3/m^3$). SMAP exhibits low mean RMSE values (0.06 $m^3/m^3$) in the Panhandle and Southwest but higher in Northeast, East Central, Southeast (0.18, 0.15, 0.16 $m^3/m^3$respectively). In summary, among the nine climate divisions, Noah has best performance in Central (CC=0.75, RMSE=0.03 $m^3/m^3$), Northeast

(CC=0.66, RMSE=0.04 $m^3/m^3$), and East Central (CC=0.76, RMSE=0.02 $m^3/m^3$), followed by Mesonet. In the other six climate divisions, Mesonet has higher mean CC values than Noah, while Noah provides lower averaged RMSE values than Mesonet. Meanwhile, SMAP illustrates consistently lower-than-Mesonet (or Noah) averaged CC and higher-than-Mesonet (or Noah) averaged RMSE values over all nine climate divisions. Summarizing on the strengths of each product, Mesonet has better performance in the Panhandle, Southwest, West Central, and South Central and worse performance in North Central and Northeast. Noah has better performance in the Panhandle and South Central but worse in the North Central, Northeast and West Central. SMAP has better performance in the Panhandle, Southwest, West Central, South Central but less value in the Northeast, East Central, and Southeast divisions. All three products exhibit worst performances in the Oklahoma Northeast.

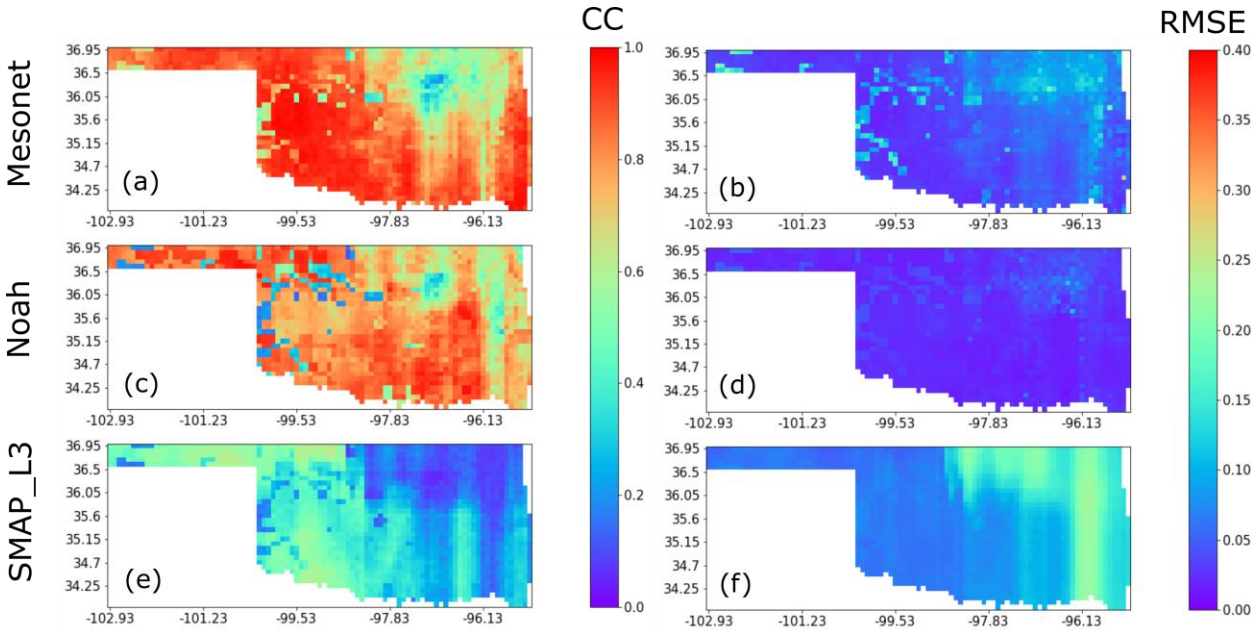


**Figure 3.4.** Product inter-comparison assessment through CC (left column) and RMSE (right column) after applying the TC method for Mesonet (first row), Noah (second row) and SMAP (third row). surface soil moisture products based on morning 6 AM soil moisture values from April 2015 through July 2019.

Figure 3.5 shows the spatial distribution of correlation and error statistics obtained from the TC triplets at local 6:00 PM standard time over Oklahoma. At the state scale, Mesonet provides the highest spatio-temporal, integrated averaged CC ($CC_{mean}$) of 0.806, followed by Noah with a $CC_{mean}$ of 0.740. SMAP has the lowest $CC_{mean}$ of 0.316. Error-wise, Noah provides the smallest $RMSE_{mean}$ of 0.026 $m^3/m^3$, followed by Mesonet ($RMSE_{mean} = 0.054\, m^3/m^3$) and SMAP ($RMSE_{mean} = 0.107\, m^3/m^3$). The geographical distributions of CC and RMSE of local 6 AM TC triplets (not shown here) resemble those at local 6 PM (Figure 3.5).

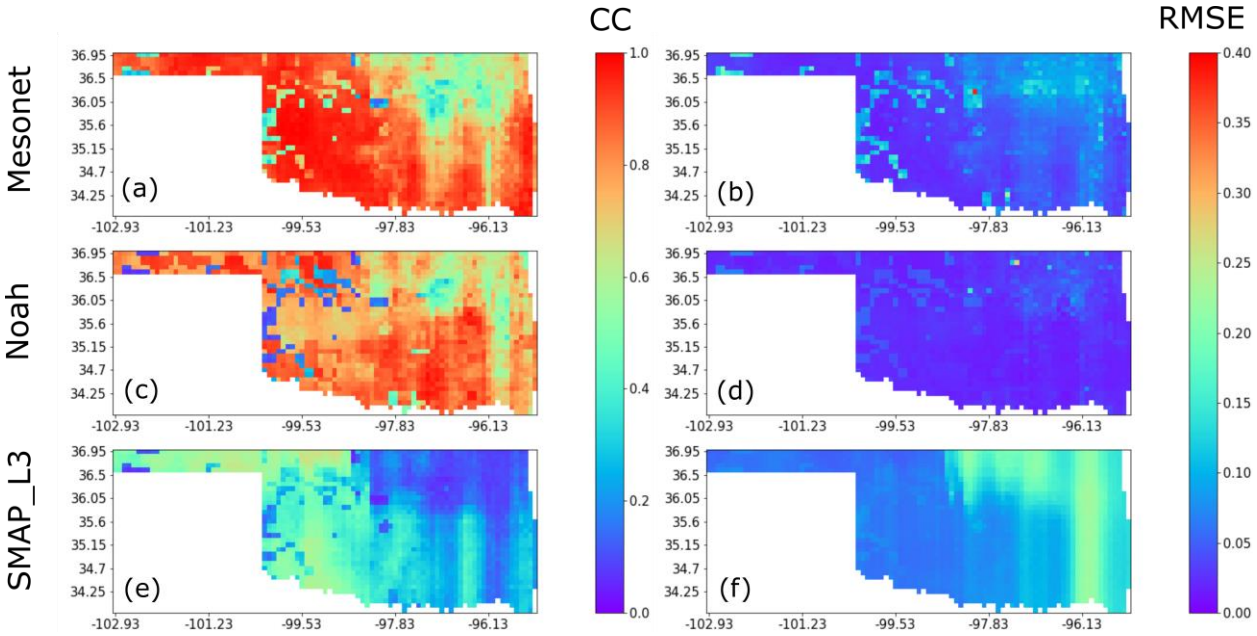


**Figure 3.5.** Product inter-comparison assessment through CC (left column) and RMSE (right column) after applying the TC method for Mesonet (first row), Noah (second row) and SMAP (third row) surface soil moisture products based on local 6 PM soil moisture values from April 2015 through July 2019.

**Table 3.4.** Average CC and RMSE values obtained from the TC triplets at local 6:00 PM for Mesonet, Noah and SMAP over nine climate divisions of Oklahoma.

| Division | Metric | Mesonet | Noah | SMAP |
|---|---|---|---|---|
| Panhandle | $CC_{mean}$ | 0.88 | 0.79 | 0.49 |
| | $RMSE_{mean}$ | 0.04 | 0.02 | 0.06 |
| West Central | $CC_{mean}$ | 0.9 | 0.67 | 0.39 |
| | $RMSE_{mean}$ | 0.04 | 0.03 | 0.07 |
| Southwest | $CC_{mean}$ | 0.92 | 0.75 | 0.47 |

| | | | | |
|---|---|---|---|---|
| | $RMSE_{mean}$ | 0.04 | 0.02 | 0.06 |
| North Central | $CC_{mean}$ | 0.8 | 0.7 | 0.32 |
| | $RMSE_{mean}$ | 0.06 | 0.03 | 0.11 |
| Central | $CC_{mean}$ | 0.75 | 0.75 | 0.28 |
| | $RMSE_{mean}$ | 0.05 | 0.03 | 0.09 |
| South Central | $CC_{mean}$ | 0.88 | 0.86 | 0.33 |
| | $RMSE_{mean}$ | 0.04 | 0.02 | 0.09 |
| Northeast | $CC_{mean}$ | 0.58 | 0.66 | 0.12 |
| | $RMSE_{mean}$ | 0.09 | 0.04 | 0.18 |
| East Central | $CC_{mean}$ | 0.76 | 0.76 | 0.23 |
| | $RMSE_{mean}$ | 0.06 | 0.02 | 0.15 |
| Southeast | $CC_{mean}$ | 0.85 | 0.76 | 0.25 |
| | $RMSE_{mean}$ | 0.05 | 0.02 | 0.16 |

*3.4.2 Seasonally-integrated Product Intercomparison of Soil moisture Values*

Figure 3.6 (spring), Figure 3.7 (summer), Figure 3.8 (fall), and Figure 3.9 (winter) show the spatial distribution of seasonally averaged CC and RMSE values obtained from the TC triplets at local 6:00 AM. The results gathered for local 6:00 PM are highly similar and consistent with those for 6:00 AM. In spring (Figure 3.6 and Table 3.5), at the state scale, Mesonet provides the highest spatio-temporal, integrated averaged CC ($CC_{mean}$) of 0.753, followed by Noah with a $CC_{mean}$ of 0.680. SMAP has the lowest (among the three products) $CC_{mean}$ of 0.463. However, the correlation values vary geographically. Mesonet provides high averaged CC values in the Panhandle, West Central, and South Central (0.89, 0.82, 0.81, respectively) divisions, while low averaged correlation is shown in the Northeast and Southeast (0.66 and 0.65). Noah exhibits high averaged CC in the Panhandle, Southwest and South Central (0.75, 0.77, 0.79 respectively) while low values in North Central, and Southeast (0.59, 0.57 respectively). The CC mean values of SMAP are generally higher in the North Central and Southwest regions (0.61 and 0.58 respectively) than those in the other six climate divisions. As for the RMSE, Noah provides the smallest $RMSE_{mean}$ of 0.0229 $m^3/m^3$, followed by Mesonet ($RMSE_{mean} = 0.0405\, m^3/m^3$) and SMAP ($RMSE_{mean} = 0.0706\, m^3/m^3$) at the state scale. Mesonet has low RMSE mean

values (ranging from 0.03 to 0.05 $m^3/m^3$) but Noah also exhibits low averaged RMSEs (ranging from 0.02 to 0.03 $m^3/m^3$) in all nine climate divisions. SMAP presents low RMSE mean values in the Northeast, East Central, and Southeast (0.09, 0.1, and 0.11 $m^3/m^3$, respectively) while relatively higher averaged RMSE values in the Panhandle and West Central (0.04 and 0.05 $m^3/m^3$). Overall, among the nine climate divisions, Mesonet has higher mean CC values than Noah, while Noah provides lower averaged RMSE values than Mesonet. SMAP values are consistently lower-than-Mesonet (or Noah) averaged CC and higher-than-Mesonet (or Noah) averaged RMSE values over all nine climate divisions. Regionally, Mesonet shows high averaged CC values in the Panhandle, West Central, and Southwest. However, the averaged RMSE values for this product in these divisions are large. Noah, on the other hand, illustrates better performance in the South Central and Southwest divisions, that worsens in the North Central and Northeast climate divisions. Finally, SMAP performs best in the Panhandle, Southwest, West Central, and North Central but presents poor results in the Northeast, East Central, and Southeast divisions.

In summer (Figure 3.7 and Table 3.6), Mesonet provides the highest spatio-temporal, integrated averaged CC ($CC_{mean}$) of 0.807, followed by Noah with a $CC_{mean}$ of 0.769. Among the three products, SMAP presents the lowest $CC_{mean}$ of 0.285. However, these correlations vary geographically. Mesonet provides high averaged CC values in the West Central, Southwest and Panhandle (0.96, 0.93, 0.87, respectively), while lower values are shown in the Northeast (0.68). Noah has the highest averaged CC in the Panhandle (0.91) but the lowest in the Northeast (0.65). The CC mean values of SMAP are generally higher in the Panhandle and Southwest (0.56 and 0.45 respectively) than those in the other six climate divisions. As for the RMSE, Noah provides the smallest $RMSE_{mean}$ of 0.0244 $m^3/m^3$, followed by Mesonet ($RMSE_{mean} = 0.0546\, m^3/m^3$) and SMAP ($RMSE_{mean} = 0.1121\, m^3/m^3$) at the state scale. Climatic division-wise, Mesonet has

low RMSE mean values, all of which are 0.03 $m^3/m^3$ in the West Central, Panhandle, and Southwest, while it presents high averaged RMSE in the Northeast (0.1 $m^3/m^3$). Noah exhibits low averaged RMSEs (ranging from 0.02 to 0.03 $m^3/m^3$) in all nine climate divisions. SMAP exhibits low RMSE mean values in the Panhandle, Southwest, and West Central (0.06, 0.07, and 0.07 $m^3/m^3$, respectively) but higher RMSE values in the Northeast, Southeast and East Central (0.19, 0.16, and 0.15 $m^3/m^3$, respectively). Overall, among the nine climate divisions, Noah performs best in the Panhandle (CC=0.91, RMSE=0.02 $m^3/m^3$), and South Central (CC=0.78, RMSE=0.03 $m^3/m^3$) divisions, followed by Mesonet. In the other six climate divisions, Mesonet has higher mean CC values than Noah, while Noah provides lower averaged RMSE values than Mesonet. SMAP values are consistently lower-than-Mesonet (or Noah) averaged CC and higher-than-Mesonet (or Noah) averaged RMSE values over all nine climate divisions. Product-wise, Mesonet has better performance in the West Central, Southwest, and Panhandle, but worse in the Northeast, East Central, and Central. Noah has better performance in the Panhandle and worse in the Northeast and SMAP has better performance in the Panhandle, Southwest, West Central, and worse in the Northeast, East Central, and Southeast. All three products show the worst performance in the Northeast Oklahoma division.

During fall (Figure 3.8 and Table 3.7), Mesonet provides the highest spatio-temporal, integrated, averaged CC ($CC_{mean}$) of 0.855, followed by Noah with a $CC_{mean}$ of 0.831. SMAP has the lowest $CC_{mean}$ of 0.335. However, the correlation varies spatially and between products. Mesonet provides high averaged CC values in the Southwest, West Central, and South Central (0.98, 0.95, 0.94, respectively). While low averaged correlation is shown in the Northeast (0.68). Noah also exhibits the highest averaged CC in the South Central, Panhandle, Southeast, and Central, all of which are 0.87 and the lowest in the Northeast (0.73). The CC mean values of SMAP

are generally higher in the Panhandle and Southwest (0.51 and 0.5 respectively) than those in the other six climate divisions. As for the RMSE, Noah provides the smallest $RMSE_{mean}$ of 0.0204 $m^3/m^3$, followed by Mesonet ($RMSE_{mean} = 0.0419\, m^3/m^3$) and SMAP ($RMSE_{mean} = 0.112\, m^3/m^3$). Noah shows consistent low RMSE values across the state (ranges from 0.02 to 0.03$m^3/m^3$). Mesonet has low RMSE mean values in the Southwest, West Central, and Panhandle (0.02, 0.03, 0.03 $m^3/m^3$, respectively), while it presents high averaged RMSE in the Northeast (0.08 $m^3/m^3$). SMAP exhibits low RMSE mean values in the Panhandle, Southwest, and West Central (0.07, 0.06, and 0.07 $m^3/m^3$, respectively) but high averaged RMSE values in the Northeast, Southeast and East Central (0.19, 0.18, and 0.15 $m^3/m^3$, respectively). Overall, among the nine climate divisions, Noah has best performance in the North Central (0.8, 0.02 $m^3/m^3$), Central (0.87, 0.02 $m^3/m^3$), Northeast (0.73, 0.03 $m^3/m^3$), East Central (0.83, 0.02 $m^3/m^3$), and Southeast (0.87, 0.02 $m^3/m^3$) followed by Mesonet. Mesonet has best performance in the Southwest (0.98, 0.02 $m^3/m^3$) followed by Noah. SMAP values are consistently lower-than-Mesonet (or Noah) averaged CC and higher-than-Mesonet (or Noah) averaged RMSE values over all nine climate divisions. Among all three, Mesonet has better performance in the West Central, Southwest, Panhandle, and South Central and worse performance in the Northeast and East Central. On the other hand, Noah shows better performance in the Panhandle but poorer in the Northeast. SMAP has better performance in the Panhandle, Southwest, West Central, and South Central, and worse in the Northeast, East Central, and Southeast. All three products provide their worst performance in Northeast.

In winter (Figure 3.9 and Table 3.8), Mesonet provides the highest spatio-temporal, integrated averaged CC ($CC_{mean}$) of 0.811, followed by Noah with an $CC_{mean}$ of 0.667. SMAP has the lowest (of the three products) $CC_{mean}$ of 0.52. Geographically, Mesonet exhibits high mean CC

values in all nine divisions (ranging from 0.78 to 0.89). Noah exhibits the highest averaged CC in the Panhandle (0.78) but the lowest values in the Central (0.55) division. The CC mean values of SMAP are generally higher in the East Central and Southwest (0.6 and 0.61 respectively) than those in the other six climate divisions. A look at the RMSE values reveals that Noah provides the smallest $RMSE_{mean}$ of 0.0217 $m^3/m^3$, followed by Mesonet ($RMSE_{mean} = 0.0507\, m^3/m^3$) and SMAP ($RMSE_{mean} = 0.0663\, m^3/m^3$). Noah shows consistent low RMSE values across the state (ranges from 0.02 to 0.03$m^3/m^3$). Mesonet has low RMSE mean values in the Southeast, Southwest and West Central (0.03, 0.04, 0.04 $m^3/m^3$, respectively), while it presents high averaged RMSE in the Central (0.08 $m^3/m^3$) climate division. SMAP exhibits low RMSE mean values in the Panhandle, Southwest, and West Central (0.04, 0.05, and 0.05 $m^3/m^3$, respectively) but high values in the Northeast, Southeast and East Central (0.09, 0.1, and 0.08 $m^3/m^3$, respectively). Overall, among the nine climate divisions, Mesonet has higher mean CC values than Noah, while Noah provides lower averaged RMSE values than Mesonet. SMAP values are consistently lower-than-Mesonet (or Noah) averaged CC and higher-than-Mesonet (or Noah) averaged RMSE values over all nine climate divisions. For each product, Mesonet shows the worst performance in the Central climate region. Noah has better performance in the Panhandle and Southwest that worsens in the Central and North Central areas. SMAP has better performance in the Southwest and Panhandle but worst in the Northeast division.

In summary, the period-integrated TC intercomparison results for Mesonet, Noah, and SMAP over nine climate divisions (Figures 3.4 and 3.5 and Table 3.4) indicate that Noah provided the best performance (as defined by the highest averaged CC and lowest averaged RMSE values) in the Central, Northeast, and East Central regions. In the other six climate divisions, Mesonet presents the highest mean CC values, while Noah provides lowest averaged RMSE values. The

seasonal TC intercomparison results for Mesonet, Noah, and SMAP over the nine climate divisions (Figures 3.6, 3.7, 3.8 and 3.9 and Tables 3.5, 3.6, 3.7 and 3.8) indicate that in spring and winter, Mesonet has higher mean CC values than Noah, while Noah provides lower averaged RMSE values than Mesonet. In summer, Noah has best performance in the Panhandle and South Central. While in the other seven climate divisions, Mesonet and Noah have comparable performances. In fall, Noah has best performance in the North Central, Central, Northeast, East Central, and Southeast. On the other hand, Mesonet best performs in the Southwest.

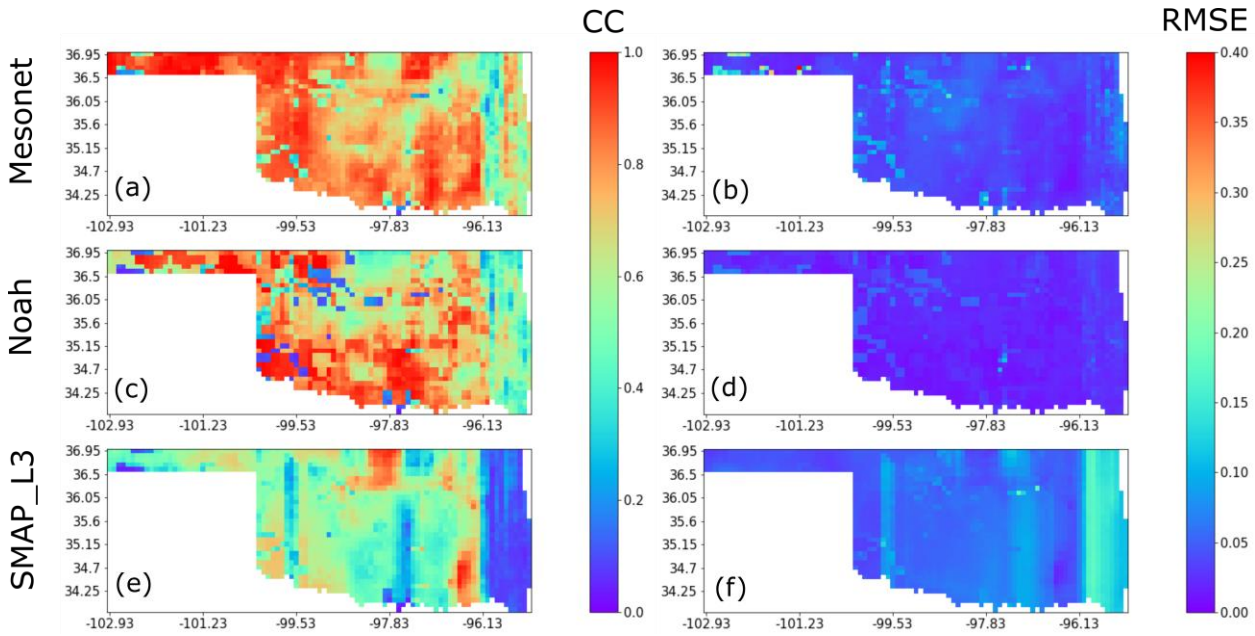


**Figure 3.6.** Product inter-comparison assessment through CC (left column) and RMSE (right column) after applying the TC method for Mesonet (first row), Noah (second row) and SMAP (third row) surface soil moisture products based on morning 6 AM values from April 2015 through July 2019, integrated (averaged) only during the Spring season months (i.e., March, April and May).

**Table 3.5.** Climate-division averaged CC and RMSE from the TC triplets at local 6:00 AM for Mesonet, Noah and SMAP during the Spring over nine climate divisions of Oklahoma.

| Division | Metric | Mesonet | Noah | SMAP |
|---|---|---|---|---|
| Panhandle | $CC_{mean}$ | 0.89 | 0.75 | 0.51 |
| | $RMSE_{mean}$ | 0.04 | 0.02 | 0.04 |
| West Central | $CC_{mean}$ | 0.82 | 0.66 | 0.54 |
| | $RMSE_{mean}$ | 0.05 | 0.03 | 0.05 |
| Southwest | $CC_{mean}$ | 0.78 | 0.77 | 0.58 |

| | | | | |
|---|---|---|---|---|
| North Central | $RMSE_{mean}$ | 0.05 | 0.02 | 0.06 |
| | $CC_{mean}$ | 0.76 | 0.59 | 0.61 |
| Central | $RMSE_{mean}$ | 0.05 | 0.03 | 0.06 |
| | $CC_{mean}$ | 0.71 | 0.7 | 0.47 |
| South Central | $RMSE_{mean}$ | 0.04 | 0.03 | 0.06 |
| | $CC_{mean}$ | 0.81 | 0.79 | 0.43 |
| Northeast | $RMSE_{mean}$ | 0.03 | 0.02 | 0.07 |
| | $CC_{mean}$ | 0.66 | 0.61 | 0.42 |
| East Central | $RMSE_{mean}$ | 0.04 | 0.04 | 0.09 |
| | $CC_{mean}$ | 0.71 | 0.66 | 0.37 |
| Southeast | $RMSE_{mean}$ | 0.03 | 0.02 | 0.1 |
| | $CC_{mean}$ | 0.65 | 0.57 | 0.22 |
| | $RMSE_{mean}$ | 0.04 | 0.02 | 0.11 |

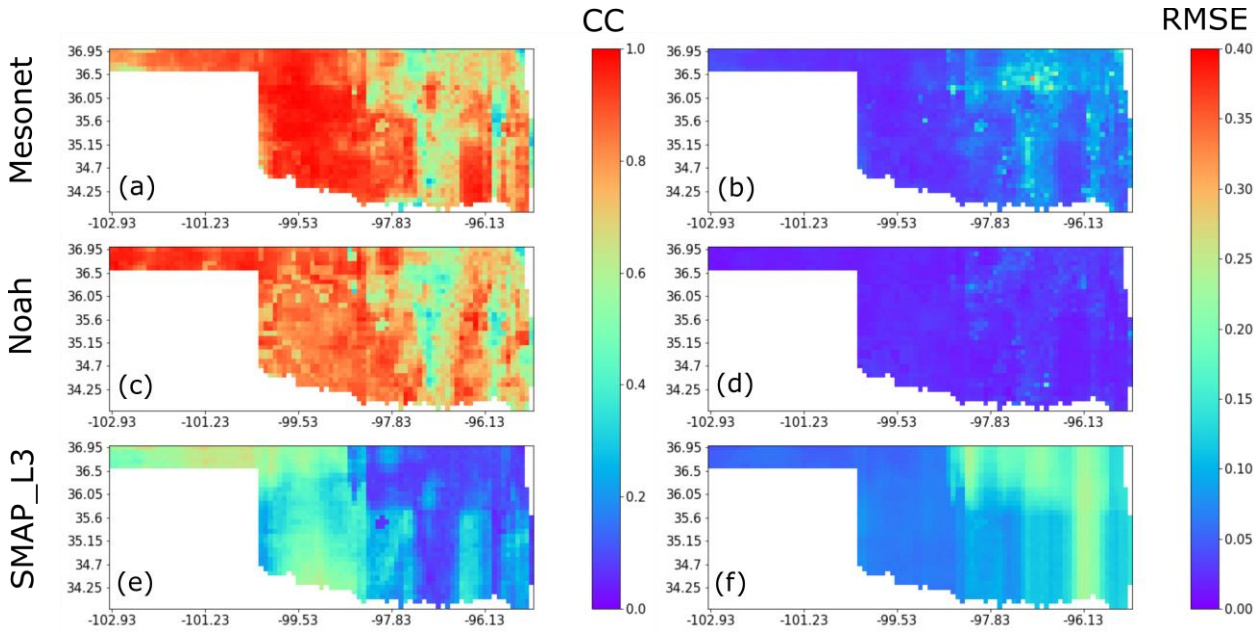


**Figure 3.7.** Product inter-comparison assessment through CC (left column) and RMSE (right column) after applying the TC method for Mesonet (first row), Noah (second row) and SMAP (third row) surface soil moisture products based on morning 6 AM values from April 2015 through July 2019, integrated (averaged)only during the Summer season months (i.e., June, July and August).

**Table 3.6.** Climate-division averaged CC and RMSE from the TC triplets at local 6:00 AM for Mesonet, Noah and SMAP during the Summer over nine climate divisions of Oklahoma.

| Division | Metric | Mesonet | Noah | SMAP |
|---|---|---|---|---|
| Panhandle | $CC_{mean}$ | 0.87 | 0.91 | 0.56 |
| | $RMSE_{mean}$ | 0.03 | 0.02 | 0.06 |
| West Central | $CC_{mean}$ | 0.96 | 0.82 | 0.36 |
| | $RMSE_{mean}$ | 0.03 | 0.02 | 0.07 |

| | | | | |
|---|---|---|---|---|
| Southwest | $CC_{mean}$ | 0.93 | 0.83 | 0.45 |
| | $RMSE_{mean}$ | 0.03 | 0.02 | 0.07 |
| North Central | $CC_{mean}$ | 0.83 | 0.82 | 0.31 |
| | $RMSE_{mean}$ | 0.05 | 0.02 | 0.11 |
| Central | $CC_{mean}$ | 0.78 | 0.72 | 0.22 |
| | $RMSE_{mean}$ | 0.06 | 0.03 | 0.1 |
| South Central | $CC_{mean}$ | 0.78 | 0.78 | 0.23 |
| | $RMSE_{mean}$ | 0.06 | 0.03 | 0.1 |
| Northeast | $CC_{mean}$ | 0.68 | 0.65 | 0.1 |
| | $RMSE_{mean}$ | 0.1 | 0.03 | 0.19 |
| East Central | $CC_{mean}$ | 0.71 | 0.7 | 0.17 |
| | $RMSE_{mean}$ | 0.07 | 0.02 | 0.15 |
| Southeast | $CC_{mean}$ | 0.8 | 0.71 | 0.22 |
| | $RMSE_{mean}$ | 0.06 | 0.02 | 0.16 |

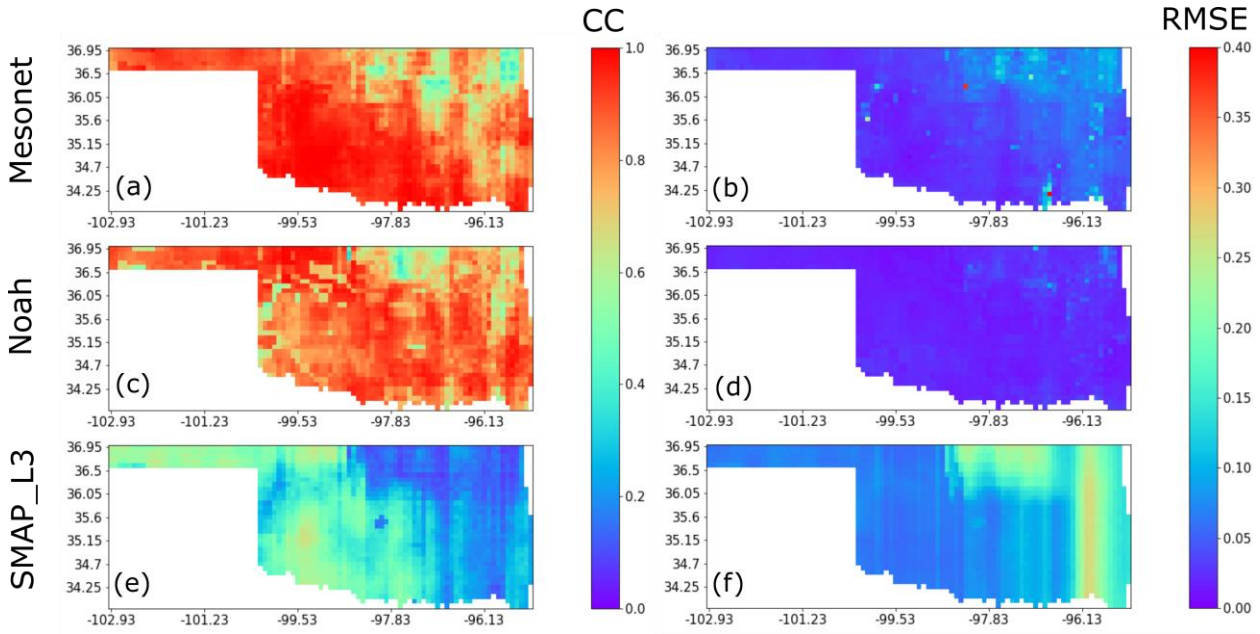


**Figure 3.8.** Product inter-comparison assessment through CC (left column) and RMSE (right column) after applying the TC method for Mesonet (first row), Noah (second row) and SMAP (third row) surface soil moisture products based on morning 6 AM values from April 2015 through July 2019, integrated (averaged)only during the Fall season months (i.e., September, October, November).

**Table 3.7.** Climate-division averaged CC and RMSE from the TC triplets at local 6:00 AM for Mesonet, Noah and SMAP during the Fall over nine climate divisions of Oklahoma.

| Division | Metric | Mesonet | Noah | SMAP |
|---|---|---|---|---|
| Panhandle | $CC_{mean}$ | 0.9 | 0.87 | 0.51 |
| | $RMSE_{mean}$ | 0.03 | 0.02 | 0.07 |
| West Central | $CC_{mean}$ | 0.95 | 0.81 | 0.43 |

| | | | | |
|---|---|---|---|---|
| Southwest | $RMSE_{mean}$ | 0.03 | 0.02 | 0.07 |
| | $CC_{mean}$ | 0.98 | 0.81 | 0.5 |
| North Central | $RMSE_{mean}$ | 0.02 | 0.02 | 0.06 |
| | $CC_{mean}$ | 0.79 | 0.8 | 0.32 |
| Central | $RMSE_{mean}$ | 0.05 | 0.02 | 0.12 |
| | $CC_{mean}$ | 0.85 | 0.87 | 0.34 |
| South Central | $RMSE_{mean}$ | 0.04 | 0.02 | 0.09 |
| | $CC_{mean}$ | 0.94 | 0.87 | 0.38 |
| Northeast | $RMSE_{mean}$ | 0.04 | 0.02 | 0.09 |
| | $CC_{mean}$ | 0.68 | 0.73 | 0.14 |
| East Central | $RMSE_{mean}$ | 0.08 | 0.03 | 0.19 |
| | $CC_{mean}$ | 0.81 | 0.83 | 0.21 |
| Southeast | $RMSE_{mean}$ | 0.06 | 0.02 | 0.15 |
| | $CC_{mean}$ | 0.82 | 0.87 | 0.23 |
| | $RMSE_{mean}$ | 0.05 | 0.02 | 0.18 |

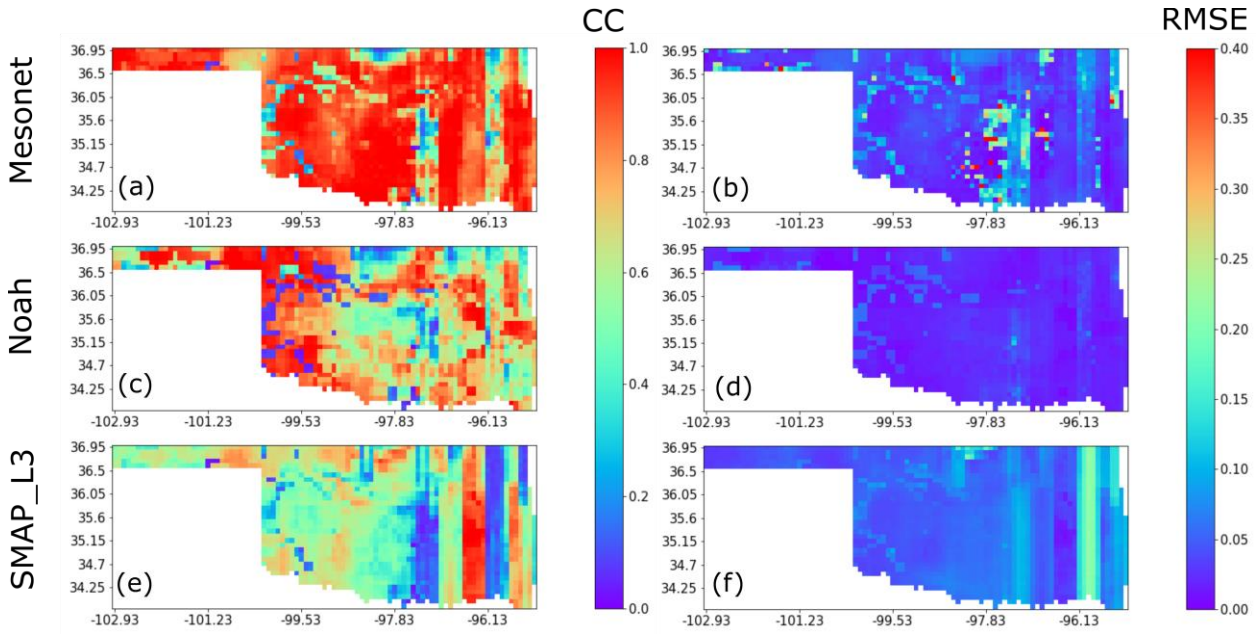


**Figure 3.9.** Product inter-comparison assessment through CC (left column) and RMSE (right column) after applying the TC method for Mesonet (first row), Noah (second row) and SMAP (third row) surface soil moisture products based on morning 6 AM values from April 2015 through July 2019, integrated (averaged)only during the Winter season months (i.e., December, January, February).

**Table 3.8.** Climate-division averaged CC and RMSE from the TC triplets at local 6:00 AM for Mesonet, Noah and SMAP during the Winter over nine climate divisions of Oklahoma.

| Division | Metric | Mesonet | Noah | SMAP |
|---|---|---|---|---|
| Panhandle | $CC_{mean}$ | 0.84 | 0.78 | 0.59 |
| | $RMSE_{mean}$ | 0.06 | 0.02 | 0.04 |

| | | | | |
|---|---|---|---|---|
| West Central | $CC_{mean}$ | 0.86 | 0.7 | 0.52 |
| | $RMSE_{mean}$ | 0.04 | 0.02 | 0.05 |
| Southwest | $CC_{mean}$ | 0.85 | 0.73 | 0.61 |
| | $RMSE_{mean}$ | 0.04 | 0.02 | 0.05 |
| North Central | $CC_{mean}$ | 0.78 | 0.58 | 0.58 |
| | $RMSE_{mean}$ | 0.05 | 0.02 | 0.06 |
| Central | $CC_{mean}$ | 0.84 | 0.55 | 0.43 |
| | $RMSE_{mean}$ | 0.08 | 0.03 | 0.06 |
| South Central | $CC_{mean}$ | 0.89 | 0.68 | 0.4 |
| | $RMSE_{mean}$ | 0.07 | 0.03 | 0.07 |
| Northeast | $CC_{mean}$ | 0.81 | 0.66 | 0.47 |
| | $RMSE_{mean}$ | 0.05 | 0.02 | 0.09 |
| East Central | $CC_{mean}$ | 0.87 | 0.72 | 0.6 |
| | $RMSE_{mean}$ | 0.05 | 0.02 | 0.08 |
| Southeast | $CC_{mean}$ | 0.85 | 0.66 | 0.55 |
| | $RMSE_{mean}$ | 0.03 | 0.02 | 0.1 |

### *3.4.3 Sub-daily Soil Moisture Product Inter-comparison by Land Cover Type*

The TC intercomparison results for Mesonet, Noah, and SMAP over different land cover types are shown in Figure 3.10 (period-integrated), Figure 3.11 (CC, across seasons) and Figure 3.12 (RMSE, across seasons). Since results for 6:00 PM are highly similar to 6:00 AM, only the morning condition is presented.

The quartile distribution and data range of the CC and RMSE for each soil moisture product over different land cover types are shown in Fig. 3.10 through box and whisker representations. Figure 3.10 illustrates that, in terms of CC, Mesonet shows the highest correlations in evergreen forest, cultivated crops, and shrub/scrub land cover types but Noah has slightly better CC values (compared to Mesonet) in deciduous forest and pasture/hay land cover types. Inter-quartile variability and ranges are similar across categories for Mesonet and Noah except for Shrub/Scrub class where Noah shows significant, below-the-mean, variability and perhaps outliers. SMAP consistently provides the lowest correlations with the unknown truth in all land cover types. As for the RMSE, Noah provides lowest mean values and interquartile ranges followed by Mesonet in all land cover types.

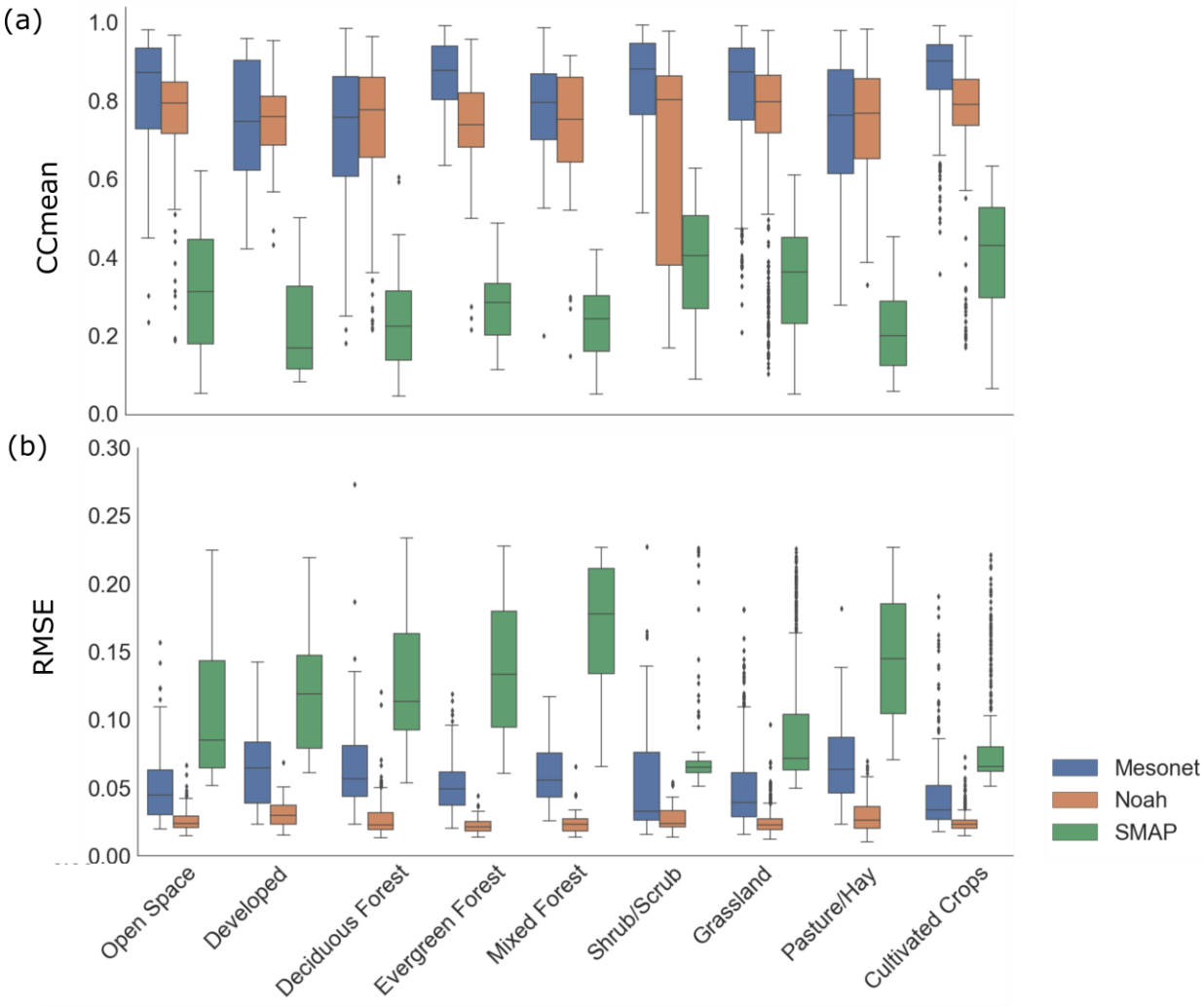


**Figure 3.10.** Soil moisture product inter-comparison results using the TC method for Mesonet, Noah and SMAP based on (a) CC and (b) RMSE for values at local 6 AM, over nine Oklahoma state representative land cover types (see Table 3) from April 2015 through July 2019.

The seasonal box and whiskers distribution of the CC for each soil moisture product over different land cover types is illustrated in Figure 3.11. During spring (Figure 3.11a), Mesonet, followed by Noah have the highest $CC_{mean}$ values over all nine land cover types. SMAP values are consistently lower-than-Mesonet (or Noah) $CC_{mean}$ values over all land cover types except for shrub/scrub. In this land cover type, the interquartile and overall data ranges are greater for Noah $CC_{mean}$ than SMAP.

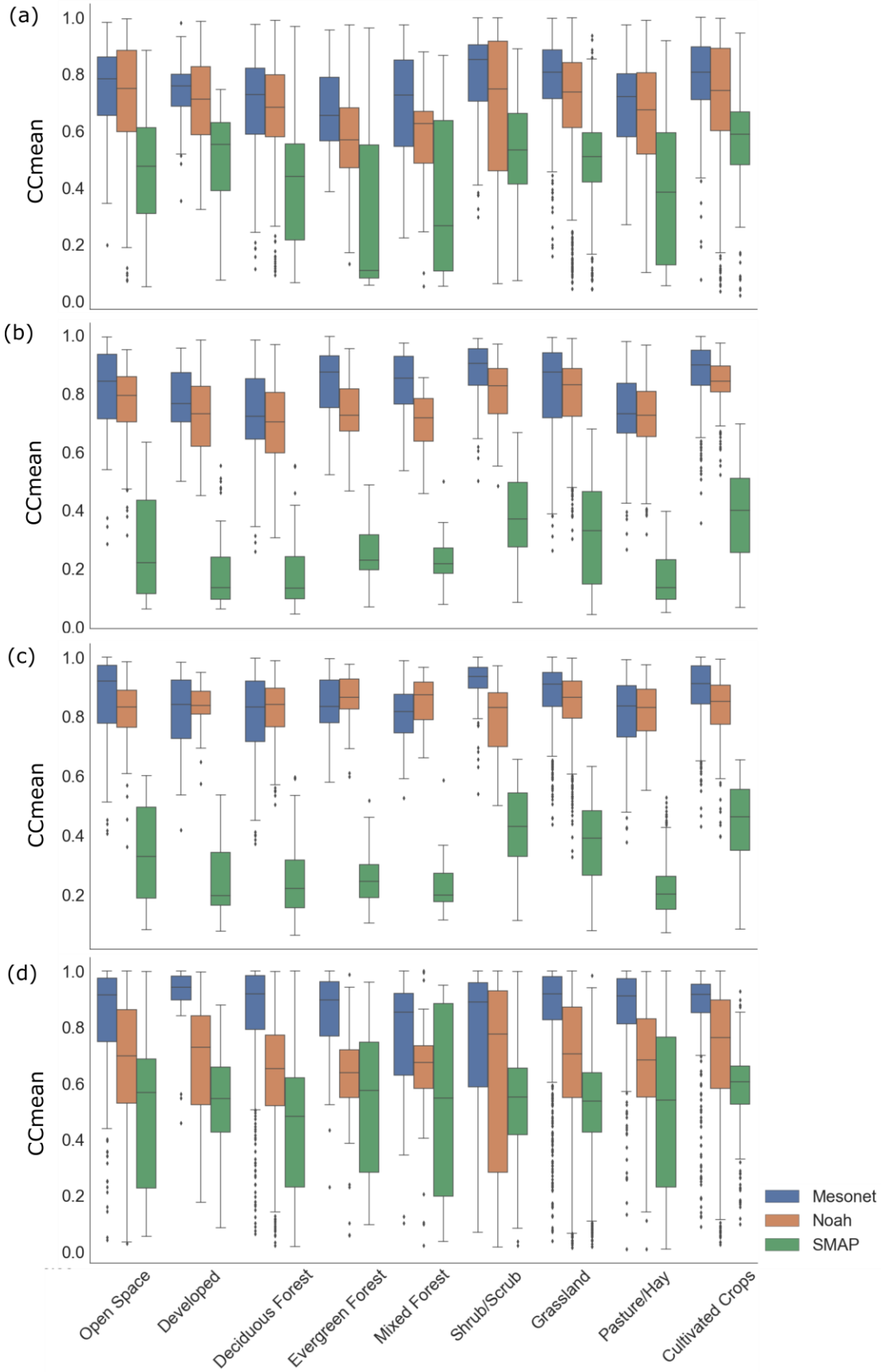
(a)
(b)
(c)
(d)
CCmean
1.0
0.8
0.6
0.4
0.2
0.0
Mesonet
Noah
SMAP
Open Space
Developed
Deciduous Forest
Evergreen Forest
Mixed Forest
Shrub/Scrub
Grassland
Pasture/Hay
Cultivated Crops

**Figure 3.11.** Mean correlation coefficient (CCmean) between each soil moisture product and the TC-derived "unknown truth" at local 6 AM, over nine Oklahoma state representative land cover types (see Table 3). Results are discretized by season: (a) spring, (b) summer, (c) fall and (d) winter for the period from April 2015 to July 2019.

During summer (Figure 3.11b), Mesonet provides the highest correlations with the TC values in evergreen forest, mixed forest, and shrub/scrub land covers. Noah provides similar correlative distributions but slightly –lower –than Mesonet $CC_{mean}$ values in other land cover types (e.g., open space, grassland, developed, deciduous forest, pasture/hay and cultivated crops). On the other hand, SMAP consistently illustrates lower –than Mesonet (or Noah) $CC_{mean}$ values over all land cover types. During Fall (Figure 3.11c), Mesonet provides the highest correlation values with the unknown TC truth over shrub/scrub, open space, grassland, pasture/hay, and cultivated crops land cover types. Noah has highest $CC_{mean}$ in developed, evergreen forest, deciduous forest, and mixed forest. The $CC_{mean}$ values of SMAP are the lowest over all land cover types. Finally, during Winter, Mesonet provides the highest $CC_{mean}$ values in all land cover types. The seasonal distributions of the RMSE of each soil moisture product over different land cover types are shown in Figure 3.12 (a through d). In spring, summer, fall and winter, Noah provides the lowest RMSE values across land cover types, but SMAP, the highest errors, except by winter in open space, shrub/scrub, and grassland when the performance is similar to that exhibited by Mesonet.

In summary, the period-integrated TC intercomparison results for Mesonet, Noah, and SMAP over different land cover types (Figure 3.10) indicate that Noah provided the best performance (with highest averaged CC and lowest averaged RMSE values) over deciduous forest, and pasture/hay land cover types. For the other land cover types, Mesonet has highest mean CC values, while Noah provides lowest averaged RMSE values. SMAP exhibits the poorest performance over all land cover types. The seasonal TC intercomparison results for Mesonet, Noah, and SMAP over

different land cover types (Figure 3.11 and 3.12) indicate that fall, Noah provided the best performance (with the highest averaged CC and lowest averaged RMSE values) in developed, evergreen forest, deciduous forest and mixed forest land cover types. In spring, summer and winter, Mesonet reaches the highest mean CC values, while Noah provides lowest averaged RMSE values over all land cover types. SMAP exhibits the least desirable performance over all land cover types across all seasons. In addition, the larger inter-quartile ranges presented by Noah for the period-integrated assessment for Shrub/Scrub (Figure 3.10) seem to be induced by large correlation variability during the Spring and Winter seasons (Figure 3.11a and d). On the other hand, SMAP illustrates lower than Mesonet (or Noah) $CC_{mean}$ values specially during spring, summer and fall that explain the low period-integrated values.

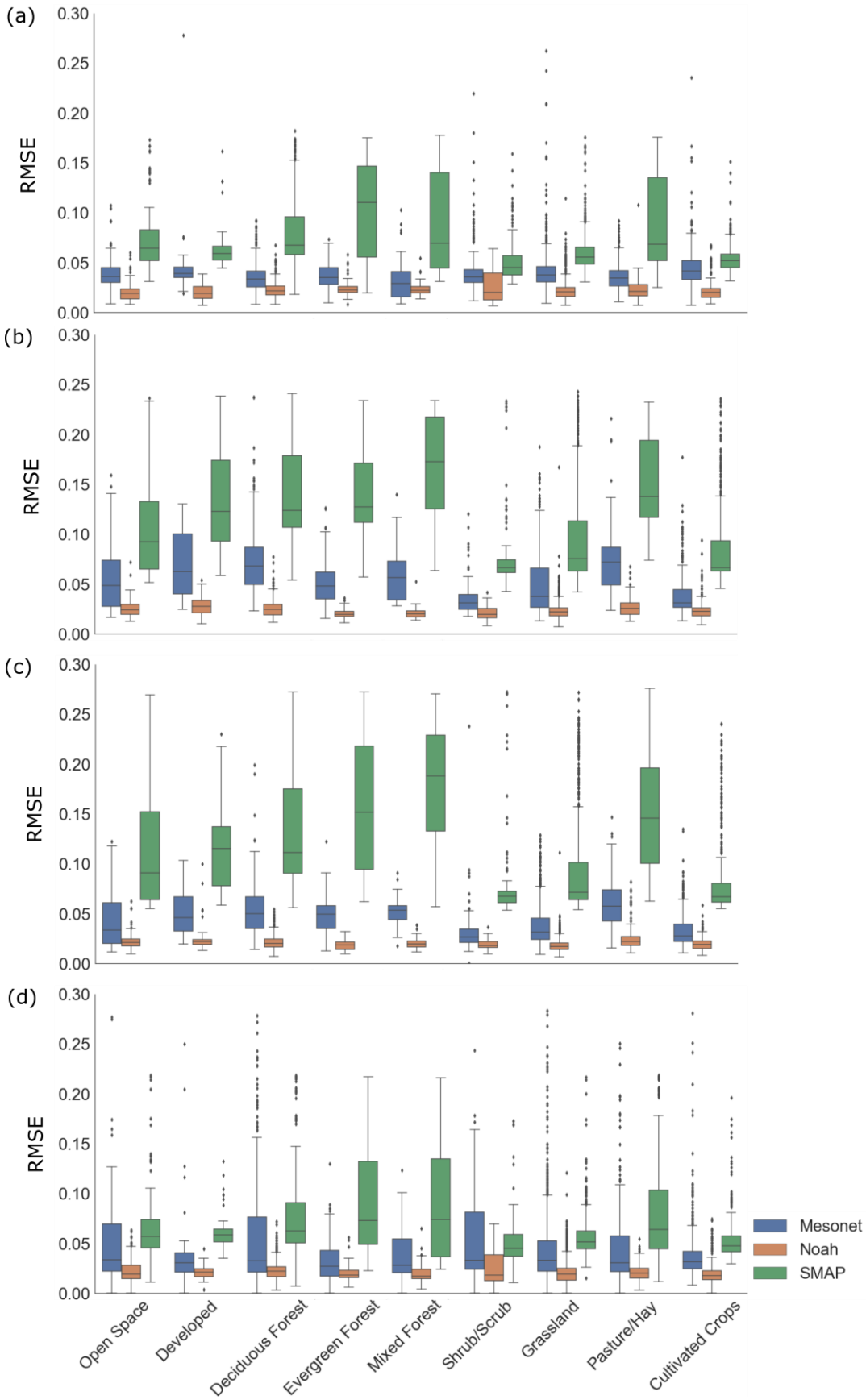
(a)
(b)
(c)
(d)
RMSE
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Open Space
Developed
Deciduous Forest
Evergreen Forest
Mixed Forest
Shrub/Scrub
Grassland
Pasture/Hay
Cultivated Crops
Mesonet
Noah
SMAP

**Figure 3.12.** Mean RMSE between each soil moisture product and the TC-derived "unknown truth" at local 6 AM, over nine Oklahoma state representative land cover types (see Table 3). Results are discretized by season: (a) spring, (b) summer, (c) fall and (d) winter for the period from April 2015 to July 2019.

### 3.5 Discussion

Soil moisture plays a critical role in the climate system and the hydrological cycle (Dingman, 2015; Seneviratne et al., 2010). However, all three main soil moisture measurements are subject to representativeness inadequacies over various land cover types, and correct interpretation and application of their products requires an in-depth understanding of their accuracy. Although triple collocation analysis has been widely used in the validation of the satellite-based and modeled soil moisture retrievals in recent years (Chen et al., 2016, 2018; Wu et al., 2020; Xu et al., 2021), here are several knowledge gaps that still exist, including: (1) of the lack of evaluation of in-situ soil moisture products as previous studies mainly focus on the evaluation of modeled and satellite-based soil moisture, rather than assessment of all three sources (modeled, remote sensing and in-situ) together; (2) There is no comprehensive evaluation of different soil moisture products at the seasonal scale and; (3) The impact of land cover on the data quality or the representativeness of each product is not often analyzed. This paper addresses all three of these knowledge gaps by conducting a comprehensive assessment of the Satellite SMAP_L3, Land Surface NOAH Model, and the interpolated Mesonet soil moisture products across Oklahoma at daily and seasonal timescales using the TC method evaluated over different land cover types during more than 4 consecutive years.

The period-integrated TC intercomparison results for Mesonet, Noah, and SMAP over nine Oklahoma state climate divisions indicate that Noah provided the best performance in the Central, Northeast, and East Central regions. The same pattern is found in the seasonal TC intercomparison

results during the Fall season. This suggests that it might be inappropriate to regard Mesonet measurements as the benchmark in the Central, Northeast, and East Central regions of Oklahoma. The reason why Noah has better quality and representativeness in these regions are: (1) the Oklahoma Mesonet site standards require the sites to be far away from urban landscapes, irrigation, forest, bare soil, fast growing vegetation, and large bodies of water to minimize those influences (Brock et al., 1995; McPherson et al., 2007). (2) The majority land cover types of these climate divisions are grassland with urban landscapes, pasture/hay, and deciduous forest (Figure 3.2). (3) The Mesonet product used in our TC intercomparison is interpolated from point-scaled to match the spatial resolution of SMAP (9 km) using the ordinary kriging method (Lakhankar et al., 2010). This interpolation method did not consider factors including soil type, land cover and topography that definitely affect the true variation of surface soil moisture. Therefore, our interpolated Mesonet product might not able to well represent the true soil moisture variations in the Central, Northeast, and East Central regions.

The seasonal TC intercomparison results for Mesonet, Noah, and SMAP over nine climate divisions indicate that in spring and winter, Mesonet has higher mean CC values than Noah, while Noah provides lower averaged RMSE values than Mesonet. Obviously, there is some disagreement between the CC and RMSE metrics, i.e., Mesonet is better than Noah in terms of CC while Noah is better than Mesonet in terms of RMSE. The reason for this is that the representativeness of errors occurred in the TC analysis can lead to biases in the TC-based error variance estimates (Gruber et al., 2016). In our study, TC is applied on one point-scale in situ data set (Mesonet) together with two coarse-scale data sets that have a comparable spatial representativeness (SMAP and Noah), even though the Mesonet product used in our TC intercomparison is interpolated from point-scale Oklahoma Mesonet and re-gridded to the same spatial resolution of SMAP using the ordinary

kriging method as suggested by Lakhankar et al. (Lakhankar et al., 2010). This interpolation method, however, did not account for large scale hydrological factors including precipitation and evapotranspiration affecting the variation of surface soil moisture. Therefore, the coarse-scale data sets (SMAP and Noah) might better capture soil moisture spatial variations that do not appear at the site location of Mesonet (e.g., localized rainfall events). In this case, the error variance estimate for the coarse-scale data sets will remain unbiased since TC will penalize the in situ site for its missing ability to capture the coarse-scale soil moisture variations (Gruber et al., 2016).

Both the period-integrated TC intercomparison and the seasonal TC intercomparison results show that SMAP exhibits the third highest performance over all climate divisions across all seasons. The reasons for this performance could be (1) microwave remote sensing is responsive to surface (~5-cm) soil moisture in regions with sparse to moderate vegetation density (Entekhabi et al., 2014) and (2) there are also challenges with retrievals in areas with complex topography, dense vegetation, near water bodies, or cities (Parinussa et al., 2011; Wagner et al., 1999). However, it is worth noting that in several areas in the Panhandle, West Central, and Southwest regions (Figure 3.4 and 3.5), where Mesonet and Noah exhibit low averaged CC and high averaged RMSE values (outliers in Figure 3.10), SMAP shows better performance than these two data sets.

In terms of their performance over different land cover types, Mesonet provides the best estimates of volumetric soil moisture over shrub/scrub, grassland, and cultivated crops, because the Oklahoma Mesonet site standards minimize the influence of urban landscapes, irrigation, forest, bare soil, fast growing vegetation, and large bodies of water (McPherson et al., 2007). It is suggested that vegetation at the Mesonet sites should be uniform and low growing such as short grasses (Brock et al., 1995). Noah provides the best estimates of volumetric soil moisture over hay/pasture, deciduous forest, mixed forest, and evergreen forest.

One limitation of this study is that the Mesonet product used in our TC intercomparison is interpolated from point-scaled Oklahoma Mesonet to spatial resolution of SMAP (9 km) using ordinary kriging. Future work could use regression kriging approaches and include independent predictors such as soil properties, land cover, topography and precipitation to increase the accuracy of the interpolated Mesonet product.

### 3.6 Conclusions

The objective of this chapter was to cross-evaluate the accuracy and error characteristics of the most commonly used, yet independent, satellite, model-based, and in situ soil moisture products. Specifically, the assessment of the SMAP L3_SM_P_E (i.e., SMAP), NLDAS_NOAH0125_H (i.e., Noah), and interpolated Oklahoma Mesonet (i.e., Mesonet) soil moisture products at daily and seasonal timescales was conducted over Oklahoma using the triple collocation method. Moreover, their performances were evaluated over different land cover types. Several conclusions are summarized as follows:

1. At the daily timescale, the interpolated Oklahoma Mesonet and Noah are found to be more reliable than SMAP for all metrics. Specifically, Mesonet provides the best estimates of volumetric soil moisture with a mean Pearson correlation coefficient of 0.805, followed by Noah with 0.747. However, Noah represents the true soil moisture variation better than our interpolated Mesonet product at meso-scale with an averaged RMSE of 0.026 $m^3/m^3$. The period-integrated TC intercomparison results for Mesonet, Noah, and SMAP over nine climate divisions indicate that Noah provided the best performance in the Central, Northeast, and East Central regions. Moreover,

the same pattern is found in the seasonal TC intercomparison results for Mesonet, Noah, and SMAP over nine climate divisions during the Fall season.

2. At dissaggregated seasonal time scales, the interpolated Oklahoma Mesonet and Noah are found to be more reliable than SMAP for all metrics in all four seasons. Specifically, Mesonet provides the best estimates of volumetric soil moisture with an averaged correlation coefficient of 0.753, 0.807, 0.855, and 0.811 in spring, summer, fall and winter respectively. However, Noah provides the best performance in representing the true soil moisture variation with an averaged RMSE of 0.0229, 0.0244, 0.0204, and 0.0217 $m^3/m^3$ in each season respectively.
3. In terms of their performance over different land cover types, Mesonet provided the best estimates of volumetric soil moisture over shrub/scrub, grassland, and cultivated crops but Noah provided the best estimates of volumetric soil moisture over hay/pasture, deciduous forest, mixed forest, and evergreen forest. This illustrates that Oklahoma Mesonet site standards minimize the influence of urban landscapes, irrigation, forest, bare soil, fast growing vegetation, and large bodies of water (McPherson et al., 2007). It is suggested that vegetation at the site should be uniform and low growing such as short grasses (Brock et al., 1995).

The TC method-based results of this study can potentially provide a new perspective for comparatively assessing multi-source soil moisture products and a basis for objective data merging to capitalize the strengths of the multi-sensor soil moisture products for the State of Oklahoma and beyond.

## References


Albergel, C., de Rosnay, P., Gruhier, C., Muñoz-Sabater, J., Hasenauer, S., Isaksen, L., Kerr, Y., & Wagner, W. (2012). Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations. *Remote Sensing of Environment*, *118*, 215–226. https://doi.org/10.1016/j.rse.2011.11.017

Betts, A. K., Chen, F., Mitchell, K. E., & Janjić, Z. I. (1997). Assessment of the land surface and boundary layer models in two operational versions of the NCEP Eta model using FIFE data. *Monthly Weather Review*, *125*(11), 2896–2916.

Bi, H., Ma, J., Zheng, W., & Zeng, J. (2016). Comparison of soil moisture in GLDAS model simulations and in situ observations over the Tibetan Plateau. *Journal of Geophysical Research: Atmospheres*, *121*(6), 2658–2678. https://doi.org/10.1002/2015JD024131

Brocca, L., Crow, W. T., Ciabatta, L., Massari, C., de Rosnay, P., Enenkel, M., Hahn, S., Amarnath, G., Camici, S., Tarpanelli, A., & Wagner, W. (2017). A Review of the Applications of ASCAT Soil Moisture Products. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(5), 2285–2306. https://doi.org/10.1109/JSTARS.2017.2651140

Brocca, L., Hasenauer, S., Lacava, T., Melone, F., Moramarco, T., Wagner, W., Dorigo, W., Matgen, P., Martínez-Fernández, J., Llorens, P., Latron, J., Martin, C., & Bittelli, M. (2011). Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe. *Remote Sensing of Environment*, *115*(12), 3390–3408. https://doi.org/10.1016/j.rse.2011.08.003

Brocca, L., Melone, F., Moramarco, T., Wagner, W., Naeimi, V., Bartalis, Z., & Hasenauer, S. (2010). Improving runoff prediction through the assimilation of the ASCAT soil moisture

product. *Hydrology and Earth System Sciences*, *14*(10), 1881–1893. https://doi.org/10.5194/hess-14-1881-2010

Brocca, L., Ponziani, F., Moramarco, T., Melone, F., Berni, N., & Wagner, W. (2012). Improving Landslide Forecasting Using ASCAT-Derived Soil Moisture Data: A Case Study of the Torgiovannetto Landslide in Central Italy. *Remote Sensing*, *4*(5), 1232–1244. https://doi.org/10.3390/rs4051232

Brock, F. V., Crawford, K. C., Elliott, R. L., Cuperus, G. W., Stadler, S. J., Johnson, H. L., & Eilts, M. D. (1995). The Oklahoma Mesonet: A Technical Overview. *Journal of Atmospheric and Oceanic Technology*, *12*(1), 5–19. https://doi.org/10.1175/1520-0426(1995)012<0005:TOMATO>2.0.CO;2

Capecchi, V., & Brocca, L. (2014). A simple assimilation method to ingest satellite soil moisture into a limited-area NWP model. *Meteorologische Zeitschrift*, *23*(2), 105–121. https://doi.org/10.1127/0941-2948/2014/0585

Chan, S. (2016). Enhanced level 3 passive soil moisture product specification document. *Jet Propulsion Lab., California Inst. Technol.: Pasadena, CA, USA*.

Chen, F., Crow, W. T., Bindlish, R., Colliander, A., Burgin, M. S., Asanuma, J., & Aida, K. (2018). Global-scale evaluation of SMAP, SMOS and ASCAT soil moisture products using triple collocation. *Remote Sensing of Environment*, *214*, 1–13. https://doi.org/10.1016/j.rse.2018.05.008

Chen, F., Crow, W. T., Colliander, A., Cosh, M. H., Jackson, T. J., Bindlish, R., Reichle, R. H., Chan, S. K., Bosch, D. D., & Starks, P. J. (2016). Application of triple collocation in ground-based validation of Soil Moisture Active/Passive (SMAP) level 2 data products.

*IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(2), 489–502.

Chen, F., Janjić, Z., & Mitchell, K. (1997). Impact of atmospheric surface-layer parameterizations in the new land-surface scheme of the NCEP mesoscale Eta model. *Boundary-Layer Meteorology*, *85*(3), 391–421.

Crow, W. T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., Rosnay, P. de, Ryu, D., & 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). https://doi.org/10.1029/2011RG000372

Crow, W. T., Berg, M. J. van den, Huffman, G. J., & Pellarin, T. (2011). Correcting rainfall using satellite-based surface soil moisture retrievals: The Soil Moisture Analysis Rainfall Tool (SMART). *Water Resources Research*, *47*(8). https://doi.org/10.1029/2011WR010576

Crow, W. T., Bindlish, R., & Jackson, T. J. (2005). The added value of spaceborne passive microwave soil moisture retrievals for forecasting rainfall-runoff partitioning. *Geophysical Research Letters*, *32*(18). https://doi.org/10.1029/2005GL023543

Crow, W. T., & Loon, E. V. (2006). Impact of Incorrect Model Error Assumptions on the Sequential Assimilation of Remotely Sensed Surface Soil Moisture. *Journal of Hydrometeorology*, *7*(3), 421–432. https://doi.org/10.1175/JHM499.1

Dingman, S. L. (2015). *Physical Hydrology: Third Edition*. Waveland Press.

Dorigo, W. A., Gruber, A., De Jeu, R. A. M., Wagner, W., Stacke, T., Loew, A., Albergel, C., Brocca, L., Chung, D., Parinussa, R. M., & Kidd, R. (2015). Evaluation of the ESA CCI soil moisture product using ground-based observations. *Remote Sensing of Environment*, *162*, 380–395. https://doi.org/10.1016/j.rse.2014.07.023

Downer, C. W., & Ogden, F. L. (2003). Prediction of runoff and soil moistures at the watershed scale: Effects of model complexity and parameter assignment. *Water Resources Research, 39*(3). https://doi.org/10.1029/2002WR001439

Draper, C., Reichle, R., de Jeu, R., Naeimi, V., Parinussa, R., & Wagner, W. (2013). Estimating root mean square errors in remotely sensed soil moisture over continental scale domains. *Remote Sensing of Environment*, *137*, 288–298.

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., & Tarpley, J. D. (2003). Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research: Atmospheres, 108*(D22).

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

Entekhabi, D., Yueh, S., & De Lannoy, G. (2014). *SMAP handbook*.

Famiglietti, J. S., Ryu, D., Berg, A. A., Rodell, M., & Jackson, T. J. (2008). Field observations of soil moisture variability across scales. *Water Resources Research, 44*(1). https://doi.org/10.1029/2006WR005804

Fan, X., Liu, Y., Gan, G., & Wu, G. (2020). SMAP underestimates soil moisture in vegetation-disturbed areas primarily as a result of biased surface temperature data. *Remote Sensing of Environment, 247*, 111914. https://doi.org/10.1016/j.rse.2020.111914

Findell, K. L., & Eltahir, E. A. B. (2003). Atmospheric controls on soil moisture-boundary layer interactions: Three-dimensional wind effects. *Journal of Geophysical Research: Atmospheres*, *108*(D8). https://doi.org/10.1029/2001JD001515

Ford, T. W., & Quiring, S. M. (2019). Comparison of Contemporary In Situ, Model, and Satellite Remote Sensing Soil Moisture With a Focus on Drought Monitoring. *Water Resources Research*, *55*(2), 1565–1582. https://doi.org/10.1029/2018WR024039

Gao, H., Wood, E. F., Jackson, T. J., Drusch, M., & Bindlish, R. (2006). Using TRMM/TMI to Retrieve Surface Soil Moisture over the Southern United States from 1998 to 2002. *Journal of Hydrometeorology*, *7*(1), 23–38. https://doi.org/10.1175/JHM473.1

*GES DISC Dataset: NLDAS Noah Land Surface Model L4 Monthly Climatology 0.125 x 0.125 degree V002 (NLDAS_NOAH0125_MC 002)*. (n.d.). https://disc.gsfc.nasa.gov/datasets/NLDAS_NOAH0125_MC_002/summary

Gruber, A., Scanlon, T., van der Schalie, R., Wagner, W., & Dorigo, W. (2019). Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology. *Earth System Science Data*, *11*(2), 717–739. https://doi.org/10.5194/essd-11-717-2019

Gruber, A., Su, C.-H., Zwieback, S., Crow, W., Dorigo, W., & Wagner, W. (2016). Recent advances in (soil moisture) triple collocation analysis. *International Journal of Applied Earth Observation and Geoinformation*, *45*, 200–211.

Gu, Y., Hunt, E., Wardlow, B., Basara, J. B., Brown, J. F., & Verdin, J. P. (2008). Evaluation of MODIS NDVI and NDWI for vegetation drought monitoring using Oklahoma Mesonet soil moisture data. *Geophysical Research Letters*, *35*(22). https://doi.org/10.1029/2008GL035772

Guttman, N. B., & Quayle, R. G. (1996). A Historical Perspective of U.S. Climate Divisions. *Bulletin of the American Meteorological Society*, *77*(2), 293–304. https://doi.org/10.1175/1520-0477(1996)077<0293:AHPOUC>2.0.CO;2

Hansen, M. C., Defries, R. S., Townshend, J. R. G., & Sohlberg, R. (2000). Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*, *21*(6–7), 1331–1364. https://doi.org/10.1080/014311600210209

Healy, J. H., Rubey, W. W., Griggs, D. T., & Raleigh, C. B. (1968). The denver earthquakes. *Science*, *161*(3848), 1301–1310.

Hollinger, S. E., & Isard, S. A. (1994). A Soil Moisture Climatology of Illinois. *Journal of Climate*, *7*(5), 822–833. https://doi.org/10.1175/1520-0442(1994)007<0822:ASMCOI>2.0.CO;2

Hsieh, P. A., & Bredehoeft, J. D. (1981). A reservoir analysis of the Denver earthquakes: A case of induced seismicity. *Journal of Geophysical Research: Solid Earth*, *86*(B2), 903–920.

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

Illston, B. G., Basara, J. B., Fiebrich, C. A., Crawford, K. C., Hunt, E., Fisher, D. K., Elliott, R., & Humes, K. (2008). Mesoscale Monitoring of Soil Moisture across a Statewide Network. *Journal of Atmospheric and Oceanic Technology*, *25*(2), 167–182. https://doi.org/10.1175/2007JTECHA993.1

Imaoka, K., Kachi, M., Kasahara, M., Ito, N., Nakagawa, K., & Oki, T. (2010). Instrument performance and calibration of AMSR-E and AMSR2. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science*, *38*(8), 13–18.

Jackson, T. J., Bindlish, R., Cosh, M. H., Zhao, T., Starks, P. J., Bosch, D. D., Seyfried, M., Moran, M. S., Goodrich, D. C., Kerr, Y. H., & Leroux, D. (2012). Validation of Soil Moisture and Ocean Salinity (SMOS) Soil Moisture Over Watershed Networks in the U.S. *IEEE*

*Transactions on Geoscience and Remote Sensing*, *50*(5), 1530–1543. https://doi.org/10.1109/TGRS.2011.2168533

Jackson, T. J., Schmugge, J., & Engman, E. T. (1996). Remote sensing applications to hydrology: Soil moisture. *Hydrological Sciences Journal*, *41*(4), 517–530.

Kang, J., Jin, R., Li, X., & Zhang, Y. (2021). Mapping High Spatiotemporal-Resolution Soil Moisture by Upscaling Sparse Ground-Based Observations Using a Bayesian Linear Regression Method for Comparison with Microwave Remotely Sensed Soil Moisture Products. *Remote Sensing*, *13*(2), 228. https://doi.org/10.3390/rs13020228

Kerr, Y. H., Waldteufel, P., Wigneron, J.-P., Martinuzzi, J., Font, J., & Berger, M. (2001). Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE Transactions on Geoscience and Remote Sensing*, *39*(8), 1729–1735.

Lakhankar, T., Jones, A. S., Combs, C. L., Sengupta, M., Vonder Haar, T. H., & Khanbilvardi, R. (2010). Analysis of large scale spatial variability of soil moisture using a geostatistical method. *Sensors*, *10*(1), 913–932.

Li, C., Tang, G., & Hong, Y. (2018). Cross-evaluation of ground-based, multi-satellite and reanalysis precipitation products: Applicability of the Triple Collocation method across Mainland China. *Journal of Hydrology*, *562*, 71–83.

Loew, A., & Schlenz, F. (2011). A dynamic approach for evaluating coarse scale satellite soil moisture products. *Hydrology and Earth System Sciences*, *15*(1), 75–90.

McColl, K. A., Vogelzang, J., Konings, A. G., Entekhabi, D., Piles, M., & Stoffelen, A. (2014). Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target. *Geophysical Research Letters*, *41*(17), 6229–6236.

McPherson, R. A., Fiebrich, C. A., Crawford, K. C., Kilby, J. R., Grimsley, D. L., Martinez, J. E., Basara, J. B., Illston, B. G., Morris, D. A., Kloesel, K. A., Melvin, A. D., Shrivastava, H., Wolfinbarger, J. M., Bostic, J. P., Demko, D. B., Elliott, R. L., Stadler, S. J., Carlson, J. D., & Sutherland, A. J. (2007). Statewide Monitoring of the Mesoscale Environment: A Technical Update on the Oklahoma Mesonet. *Journal of Atmospheric and Oceanic Technology*, *24*(3), 301–321. https://doi.org/10.1175/JTECH1976.1

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. https://doi.org/10.1175/2008JHM916.1

Miller, D. A., & White, R. A. (1998). A conterminous United States multi-layer soil characteristics data set for regional climate and hydrology modeling, Earth Interactions, 2. Web-based publication. *Res*, 711–724.

Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Cosgrove, B. A., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., Tarpley, J. D., Lettenmaier, D. P., Marshall, C. H., Entin, J. K., Pan, M., Shi, W., Koren, V., … Bailey, A. A. (2004). The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research: Atmospheres*, *109*(D7). https://doi.org/10.1029/2003JD003823

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

Mohd Kassim, M. R., Mat, I., & Harun, A. N. (2014). Wireless Sensor Network in precision agriculture application. *2014 International Conference on Computer, Information and Telecommunication Systems (CITS)*, 1–5. https://doi.org/10.1109/CITS.2014.6878963

National Research Council. (2013). *Induced seismicity potential in energy technologies*. National Academies Press.

Ochsner, T. E., Linde, E., Haffner, M., & Dong, J. (2019). Mesoscale Soil Moisture Patterns Revealed Using a Sparse In Situ Network and Regression Kriging. *Water Resources Research*, *55*(6), 4785–4800. https://doi.org/10.1029/2018WR024535

ONeill, P. E., Chan, S., Njoku, E. G., Jackson, T., & Bindlish, R. (2019). *SMAP Enhanced L3 Radiometer Global Daily 9 km EASE-Grid Soil Moisture, Version 3* [Data set]. NASA National Snow and Ice Data Center DAAC. https://doi.org/10.5067/T90W6VRLCBHI

Parinussa, R. M., Meesters, A. G., Liu, Y. Y., Dorigo, W., Wagner, W., & de Jeu, R. A. (2011). Error estimates for near-real-time satellite soil moisture as derived from the land parameter retrieval model. *IEEE Geoscience and Remote Sensing Letters*, *8*(4), 779–783.

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

Pielke, R. A. (2001). Influence of the spatial distribution of vegetation and soils on the prediction of cumulus Convective rainfall. *Reviews of Geophysics*, *39*(2), 151–177. https://doi.org/10.1029/1999RG000072

Qin, J., Yang, K., Lu, N., Chen, Y., Zhao, L., & Han, M. (2013). Spatial upscaling of in-situ soil moisture measurements based on MODIS-derived apparent thermal inertia. *Remote Sensing of Environment*, *138*, 1–9. https://doi.org/10.1016/j.rse.2013.07.003

Ray, R. L., Jacobs, J. M., & Cosh, M. H. (2010). Landslide susceptibility mapping using downscaled AMSR-E soil moisture: A case study from Cleveland Corral, California, US. *Remote Sensing of Environment*, *114*(11), 2624–2636. https://doi.org/10.1016/j.rse.2010.05.033

Reichle, R., De Lannoy, G., Koster, R., Crow, W., Kimball, J., & Liu, Q. (2020). *SMAP L4 Global 3-hourly 9 km EASE-Grid Surface and Root Zone Soil Moisture Geophysical Data, Version 5* [Data set]. NASA National Snow and Ice Data Center DAAC. https://doi.org/10.5067/9LNYIYOBNBR5

Reichle, R. H., Crow, W. T., & Keppenne, C. L. (2008). An adaptive ensemble Kalman filter for soil moisture data assimilation. *Water Resources Research*, *44*(3). https://doi.org/10.1029/2007WR006357

Reichle, R. H., Lannoy, G. J. M. D., Liu, Q., Ardizzone, J. V., Colliander, A., Conaty, A., Crow, W., Jackson, T. J., Jones, L. A., Kimball, J. S., Koster, R. D., Mahanama, S. P., Smith, E. B., Berg, A., Bircher, S., Bosch, D., Caldwell, T. G., Cosh, M., González-Zamora, Á., … Zeng, Y. (2017). Assessment of the SMAP Level-4 Surface and Root-Zone Soil Moisture Product Using In Situ Measurements. *Journal of Hydrometeorology*, *18*(10), 2621–2645. https://doi.org/10.1175/JHM-D-17-0063.1

Robock, A., Vinnikov, K. Y., Srinivasan, G., Entin, J. K., Hollinger, S. E., Speranskaya, N. A., Liu, S., & Namkhai, A. (2000). The Global Soil Moisture Data Bank. *Bulletin of the American Meteorological Society*, *81*(6), 1281–1300. https://doi.org/10.1175/1520-0477(2000)081<1281:TGSMDB>2.3.CO;2

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., &

Toll, D. (2004). The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society*, *85*(3), 381–394. https://doi.org/10.1175/BAMS-85-3-381

Rubinstein, J. L., & Mahani, A. B. (2015). Myths and facts on wastewater injection, hydraulic fracturing, enhanced oil recovery, and induced seismicity. *Seismological Research Letters*, *86*(4), 1060–1067.

Samouëlian, A., Cousin, I., Tabbagh, A., Bruand, A., & Richard, G. (2005). Electrical resistivity survey in soil science: A review. *Soil and Tillage Research*, *83*(2), 173–193. https://doi.org/10.1016/j.still.2004.10.004

Sayde, C., Gregory, C., Gil-Rodriguez, M., Tufillaro, N., Tyler, S., Giesen, N. van de, English, M., Cuenca, R., & Selker, J. S. (2010). Feasibility of soil moisture monitoring with heated fiber optics. *Water Resources Research*, *46*(6). https://doi.org/10.1029/2009WR007846

Scipal, K., Holmes, T., Jeu, R. de, Naeimi, V., & Wagner, W. (2008). A possible solution for the problem of estimating the error structure of global soil moisture data sets. *Geophysical Research Letters*, *35*(24). https://doi.org/10.1029/2008GL035599

Selig, E. T., & Mansukhani, S. (1975). Relationship of Soil Moisture to the Dielectric Property. *Journal of the Geotechnical Engineering Division*, *101*(8), 755–770. https://doi.org/10.1061/AJGEB6.0000184

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., & Teuling, A. J. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, *99*(3), 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004

Shapiro, S. A., & Dinske, C. (2009). Fluid-induced seismicity: Pressure diffusion and hydraulic fracturing. *Geophysical Prospecting*, *57*(2), 301–310.

Stoffelen, A. (1998). Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. *Journal of Geophysical Research: Oceans*, *103*(C4), 7755–7766.

S.u., S. L., Singh, D. N., & Shojaei Baghini, M. (2014). A critical review of soil moisture measurement. *Measurement*, *54*, 92–105. https://doi.org/10.1016/j.measurement.2014.04.007

Swenson, S., Famiglietti, J., Basara, J., & Wahr, J. (2008). Estimating profile soil moisture and groundwater variations using GRACE and Oklahoma Mesonet soil moisture data. *Water Resources Research*, *44*(1). https://doi.org/10.1029/2007WR006057

Tavakol, A., Rahmani, V., Quiring, S. M., & Kumar, S. V. (2019). Evaluation analysis of NASA SMAP L3 and L4 and SPoRT-LIS soil moisture data in the United States. *Remote Sensing of Environment*, *229*, 234–246. https://doi.org/10.1016/j.rse.2019.05.006

Trenberth, K. E., Fasullo, J. T., & Kiehl, J. (2009). Earth's Global Energy Budget. *Bulletin of the American Meteorological Society*, *90*(3), 311–324. https://doi.org/10.1175/2008BAMS2634.1

U.S. Energy Information Administration. (2021, April 15). *Oklahoma State Energy Profile*. https://www.eia.gov/state/print.php?sid=OK

US EPA, O. (2015a, March 25). *Safe Drinking Water Act (SDWA)* [Collections and Lists]. US EPA. https://www.epa.gov/sdwa

US EPA, O. (2015b, March 27). *Class II Oil and Gas Related Injection Wells* [Overviews and Factsheets]. US EPA. https://www.epa.gov/uic/class-ii-oil-and-gas-related-injection-wells

US EPA, O. (2015c, April 1). *General Information About Injection Wells* [Overviews and Factsheets]. US EPA. https://www.epa.gov/uic/general-information-about-injection-wells

Valente, A., Morais, R., Tuli, A., Hopmans, J. W., & Kluitenberg, G. J. (2006). Multi-functional probe for small-scale simultaneous measurements of soil thermal properties, water content, and electrical conductivity. *Sensors and Actuators A: Physical*, *132*(1), 70–77. https://doi.org/10.1016/j.sna.2006.05.010

Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., Figa-Saldana, J., de Rosnay, P., Jann, A., Schneider, S., Komma, J., Kubu, G., Brugger, K., Aubrecht, C., Züger, J., Gangkofner, U., Kienberger, S., Brocca, L., Wang, Y., … Rubel, F. (2013). The ASCAT soil moisture product: A review of its specifications, validation results, and emerging applications. *Meteorologische Zeitschrift*. https://doi.org/10.1127/0941-2948/2013/0399

Wagner, W., Lemoine, G., Borgeaud, M., & Rott, H. (1999). A study of vegetation cover effects on ERS scatterometer data. *IEEE Transactions on Geoscience and Remote Sensing*, *37*(2), 938–948.

Wang, J., Ge, Y., Heuvelink, G. B. M., & Zhou, C. (2015). Upscaling In Situ Soil Moisture Observations to Pixel Averages with Spatio-Temporal Geostatistics. *Remote Sensing*, *7*(9), 11372–11388. https://doi.org/10.3390/rs70911372

Woodley, W. (2017). *Investigating the upscaling of in situ soil moisture measurements for satellite validation*.

Wu, X., Lu, G., Wu, Z., He, H., Scanlon, T., & Dorigo, W. (2020). Triple Collocation-Based Assessment of Satellite Soil Moisture Products with In Situ Measurements in China: Understanding the Error Sources. *Remote Sensing*, *12*(14), 2275. https://doi.org/10.3390/rs12142275

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., & Meng, J. (2012). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of Geophysical Research: Atmospheres*, *117*(D3).

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

Xu, L., Chen, N., Zhang, X., Moradkhani, H., Zhang, C., & Hu, C. (2021). In-situ and triple-collocation based evaluations of eight global root zone soil moisture products. *Remote Sensing of Environment*, *254*, 112248.

Yilmaz, M. T., & Crow, W. T. (2014). Evaluation of assumptions in soil moisture triple collocation analysis. *Journal of Hydrometeorology*, *15*(3), 1293–1302.

Yilmaz, M. T., Crow, W. T., Anderson, M. C., & Hain, C. (2012). An objective methodology for merging satellite-and model-based soil moisture products. *Water Resources Research*, *48*(11).

Zeng, Y., Su, Z., Van der Velde, R., Wang, L., Xu, K., Wang, X., & Wen, J. (2016). Blending Satellite Observed, Model Simulated, and in Situ Measured Soil Moisture over Tibetan Plateau. *Remote Sensing*, *8*(3), 268. https://doi.org/10.3390/rs8030268

Zhang, N., Quiring, S. M., & Ford, T. W. (2021). Blending Noah, SMOS and In-Situ Soil Moisture Using Multiple Weighting and Sampling Schemes. *Journal of Hydrometeorology*, *1*(aop). https://doi.org/10.1175/JHM-D-20-0119.1

Zhang, R., Kim, S., & Sharma, A. (2019). A comprehensive validation of the SMAP Enhanced Level-3 Soil Moisture product using ground measurements over varied climates and landscapes. *Remote Sensing of Environment*, *223*, 82–94. https://doi.org/10.1016/j.rse.2019.01.015

Zhang, X., Tang, Q., Liu, X., Leng, G., & Li, Z. (2017). Soil Moisture Drought Monitoring and Forecasting Using Satellite and Climate Model Data over Southwestern China. *Journal of Hydrometeorology*, *18*(1), 5–23. https://doi.org/10.1175/JHM-D-16-0045.1

Zhang, X., Zhang, T., Zhou, P., Shao, Y., & Gao, S. (2017). Validation Analysis of SMAP and AMSR2 Soil Moisture Products over the United States Using Ground-Based Measurements. *Remote Sensing*, *9*(2), 104. https://doi.org/10.3390/rs9020104

Zwieback, S., Scipal, K., Dorigo, W., & Wagner, W. (2012). Structural and statistical properties of the collocation technique for error characterization. *Nonlinear Processes in Geophysics*, *19*(1), 69–80.

# Chapter 4 Triple Collocation-based Merging of Ground-based, Satellite and Land Surface Model Soil Moisture Products


## Abstract

This chapter adopted an objective methodology in blending multisource soil moisture products, identified the importance of incorporating in-situ measurements and quantified the impact of different weighting schemes on soil moisture blending. Soil moisture information from multiple sources, including satellite (SMAP_L3), land surface model (NLDAS-2 Noah), and in-situ measurements (interpolated Oklahoma Mesonet values), are used to generate blended soil moisture products at a 9-km spatial resolution and daily temporal resolution in Oklahoma. The merged product is validated against an in-situ-based soil moisture product data and shows better performance than both the equal weighting merged and SMAP Level 4 soil moisture products. Results indicate that: (1) the TC-LSW approach-based merging scheme is more appropriate than the equal weighting merging scheme; (2) in-situ measurements are valuable for improving the accuracy of blended soil moisture datasets; and (3) using multiple sources of soil moisture helps to reduce the overall uncertainty in the soil moisture estimates. The resulting combined soil moisture estimate is an improvement over currently available soil moisture products due to its reduced uncertainty and can be used as a standalone soil moisture product with available uncertainty estimates.


## 4.1 Introduction

Soil moisture is a critical component of the climate system and the hydrological cycle (Dingman, 2015; Seneviratne et al., 2010). Understanding and predicting agricultural and water resource management (Mohd Kassim et al., 2014), runoff and flooding (Brocca et al., 2010; Crow

et al., 2005), drought monitoring (Gu et al., 2008; Zhang et al., 2017), weather and climate forecasting (Capecchi & Brocca, 2014; Scipal et al., 2008), and landslides (Brocca et al., 2012; Ray et al., 2010) depend on knowledge of soil moisture variations.

Currently, soil moisture can be estimated through three primary approaches: (1) in-situ measurements, (2) remote sensing observations, and (3) Land Surface Models (LSM). In-situ soil moisture measurement techniques have the advantage of easy installation, high spatial and temporal resolutions, and the ability to measure soil moisture at different depths (Peng et al., 2017). Therefore, these measurements are normally recognized as the "ground truth" in validating and calibrating remote sensing and land surface model-based soil moisture retrievals. However, since soil moisture variability generally increases with extent scale (Famiglietti et al., 2008), these point measurements are not able to provide spatial representativeness of neighboring areas over a range of scales (Crow et al., 2012; Ochsner et al., 2019; Peng et al., 2017). Although previous studies have used geostatistical techniques to extrapolate in-situ soil moisture measurements to larger scales, these methods are usually complex, time-consuming, and depend on the availability of high-resolution auxiliary variables, especially over land surfaces with high spatial heterogeneity (Kang et al., 2021; Ochsner et al., 2019; Qin et al., 2013; Wang et al., 2015; Zhang et al., 2017). LSM can provide soil moisture estimates at various depths with high spatio-temporal resolution at larger scales. For example, the four LSMs (Noah, SAC-SMA, VIC, and Mosaic) of the North American Land Data Assimilation System phase 2 (NLDAS-2) simulate hourly soil moisture over the conterminous United States (CONUS), the southern part of Canada, and the northern portion of Mexico with a 1/8° latitude-longitude resolution at various depths (Xia et al., 2014). However, uncertainties due to model forcing, parameters, structure, and calibration affect the reliability of these soil moisture estimates (Xia et al., 2014). Finally, soil moisture retrievals of remote sensing

data, especially satellite microwave observations from active and passive sensors, have gained momentum over the past 20 years (Mohanty et al., 2017). For example, some promising satellites available in the orbit include: the Advanced Microwave Scanning Radiometer 2 (AMSR2) (Imaoka et al., 2010), Advanced Scatterometer (ASCAT) (Wagner et al., 2013), Soil Moisture and Ocean Salinity (SMOS)(Kerr et al., 2001), Soil Moisture Active and Passive (SMAP) Mission (Entekhabi et al., 2010). Various soil moisture products at regional or global scales have been retrieved from these active and passive sensors and validated against extensive field campaigns (Brocca et al., 2011; Dorigo et al., 2015; Jackson et al., 2012; Wagner et al., 2013). However, due to the limited penetration depth of microwave signals, sensors can only measure the surface soil moisture (e.g., top 5 cm of the soil column) under low to moderate vegetation cover conditions (Crow et al., 2012; Entekhabi et al., 2014). Moreover, they cannot detect soil moisture under snow or ice, or in frozen soils (Entekhabi et al., 2014). In summary, each source of soil moisture estimates is not perfect and has characteristic uncertainties. Therefore, it is desirable to merge independent products to obtain a more accurate estimate.

Data assimilation using Kalman filter-based methodologies is one of the most commonly used approaches for merging different products while taking into account their relative uncertainties (Yilmaz et al., 2012). For example, the SMAP Level-4 soil moisture product is generated by a data assimilation system that combines L-band brightness temperature measurements, precipitation observations, and the NASA Catchment land surface model (R. H. Reichle et al., 2017). However, the integration of independent data sources using Kalman filter-based methodologies in land data assimilation studies may not be optimal because these methodologies often rely on prior knowledge of product uncertainties which is arguably subjective (Crow & Loon, 2006; R. H. Reichle et al., 2008). Therefore, Yilmaz et al. (Yilmaz et al., 2012) introduced an objective

methodology which is less dependent on uncertain, user-defined error assumptions for blending satellite- and model-based SM products in the least squares framework where uncertainty estimates for each product are obtained using the triple collocation method. This chapter adopts the same methodology to blend the in-situ, remote sensing, and modeled soil moisture datasets.

Therefore, the overarching goals of this chapter are to (1) adopt an objective methodology introduced by Yilmaz et al. (Yilmaz et al., 2012) in blending multisource soil moisture products in Oklahoma as a case study and (2) evaluate the effect of incorporating in situ observations into soil moisture blending. Specifically, this chapter adopts the Triple Collocation (TC) based least square weighting method to merge the Satellite SMAP_L3, Land Surface Noah Model, and the interpolated Mesonet soil moisture products across Oklahoma and compares the merged product (OU-MSSM) with the equal weights merging product (AVE) and SMAP L4 soil moisture products using the Automated Soil Moisture Mapping System as the benchmark. The rest of the chapter is organized as follows: Section 4.2 details the data and study area; Section 4.3 describes the methods and data processing; Section 4.4 presents the results and analysis; Section 4.5 provides discussion; and the Section 4.6 provides conclusions for this study.

## 4.2 Data Sources

### *4.2.1 The Automated Soil Moisture Mapping System*

The fully automated system developed by Ochsner et al. (2019) applied regression kriging to several data sets: daily soil moisture measurements from the Oklahoma Mesonet, sand content estimates from the Natural Resource Conservation Service Soil Survey Geographic Database, and an antecedent precipitation index computed from National Weather Service multi-sensor precipitation estimates. This mapping system provides daily statewide maps of soil moisture at 5-

, 25-, and 60-cm depths at 800-m resolution from 2015-09-01 to present with a mean absolute error of ≤0.0576 $cm^3/cm^3$ across all three depths. This product is used in this study as a benchmark when comparing the blended soil moisture products and SMAP L4 product and is referred to as the OSU soil moisture measurement in the following contents.

*4.2.2Tthe Oklahoma Mesonet Soil Moisture Measurements*

Established in January 1994, the Oklahoma Mesonet is a multipurpose network operating more than 110 automated stations with at least one station in each of Oklahoma's 77 counties (Brock et al., 1995; McPherson et al., 2007). Quality-assured data including temperature, humidity, solar radiation, wind speed and direction, and soil moisture are available through an operations center located at the Oklahoma Climatological Survey (OCS). Soil moisture data are collected every 30 min by soil moisture sensors at each site at four different depths (5, 25, 60, and 75 cm below the surface; Illston et al., 2008). Since the TC analysis requires three spatially and temporally collocated measurement systems and SMAP measurements are available from 03/31/2015, the Mesonet soil moisture measurements of 115 Mesonet sites ranging from 04/01/2015 to 07/01/2019 were used in this study (see Figure 1). Since the data points are spread-out across the state, interpolation was conducted using ordinary Kriging, as previously suggested by (Lakhankar et al., 2010) and then re-gridded to 9 km. The re-gridded Oklahoma Mesonet soil moisture measurements were used as an independent data source of OU-MSSM in the objective methodology introduced by Yilmaz et al (2012).

*4.2.3 NLDAS-2 Noah Soil Moisture Estimations*

The Noah model is a land surface model (LSM) of the phase 2 of the North American Land Data Assimilation System (NLDAS-2), generating energy fluxes, water fluxes, and state variables (e.g., soil moisture) (Xia et al., 2014). It provides hourly soil moisture fields at 1/8° resolution from 1979 to present at four soil layers: 0–10, 10–40, 40–100, and 100–200 cm. Noah uses a dominant vegetation type with a varied root depth (e.g., 100 cm for grassland and 200 cm for forest and woodland). Details about the NLDAS-2 Noah LSM can be found in (Xia et al., 2012). In this study, the hourly soil moisture simulations of NLDAS-2 Noah model between the period of 04-01-2015 to 07-01-2019 (*GES DISC Dataset: NLDAS Noah Land Surface Model L4 Monthly Climatology 0.125 x 0.125 Degree V002 (NLDAS_NOAH0125_MC 002)*, n.d.) were used as an independent data source of OU-MSSM in the objective methodology introduced by Yilmaz et al (2012).

*4.2.4 SMAP L3 and L4 Soil Moisture Retrievals*

The Soil Moisture Active Passive (SMAP) mission has been providing global L-band brightness temperature observations and surface soil moisture retrievals since 31 March 2015 at about 40-km resolution from a 685-km, near-polar, sun-synchronous orbit (Entekhabi et al., 2010). In total, the SMAP mission has generated 24 distributable data products representing four levels of data processing. In this study, the SMAP Enhanced L3 Radiometer Global Daily 9 km EASEGrid Soil Moisture, Version 3 (L3_SM_P_E; [40]) product was used as an independent data source of OU-MSSM in the objective methodology introduced by Yilmaz et al. (2012) because it can satisfy the independency assumption of triple collocation analysis (Stoffelen, 1998). It is a daily global composite of gridded surface soil moisture retrievals of 6:00 am (descending) and 6:00 pm (ascending). The SMAP L3_SM_P_E product pertaining to the period from 04-01-2015

to 07-01-2019 were used in this study (ONeill et al., 2019). The SMAP Level-4 soil moisture product is generated by a data assimilation system that combines L-band brightness temperature measurements, and precipitation observations with the NASA Catchment land surface model (Reichle et al., 2017). Surface (0-5 cm) and root zone soil moisture (0-100 cm) are available every three hours on a global grid with 9-km spacing within about three days from the time of observation (Reichle et al., 2017). The SMAP Level-4 surface soil moisture product (SMAP L4) between the period of 04-01-2015 to 07-01-2019 was used in this study in the comparison with the merged product using the OSU soil moisture measurement as benchmark (Reichle et al., 2020) .

*4.2.5 Auxiliary Data*

4.2.5.1 The National Land Cover Dataset 2016

To compare the performance of the three soil moisture products (OU-MSSM, AVE and SMAP L4) over different land cover types, the national land cover dataset (NLCD) 2016 product was used in this study. The NLCD provides nationwide data on continental U.S. land cover and land cover change at a 30 m resolution with a 16-class legend based on a modified Anderson Level II classification system. There are fifteen land cover types within the state of Oklahoma as shown in Figure 4.1.

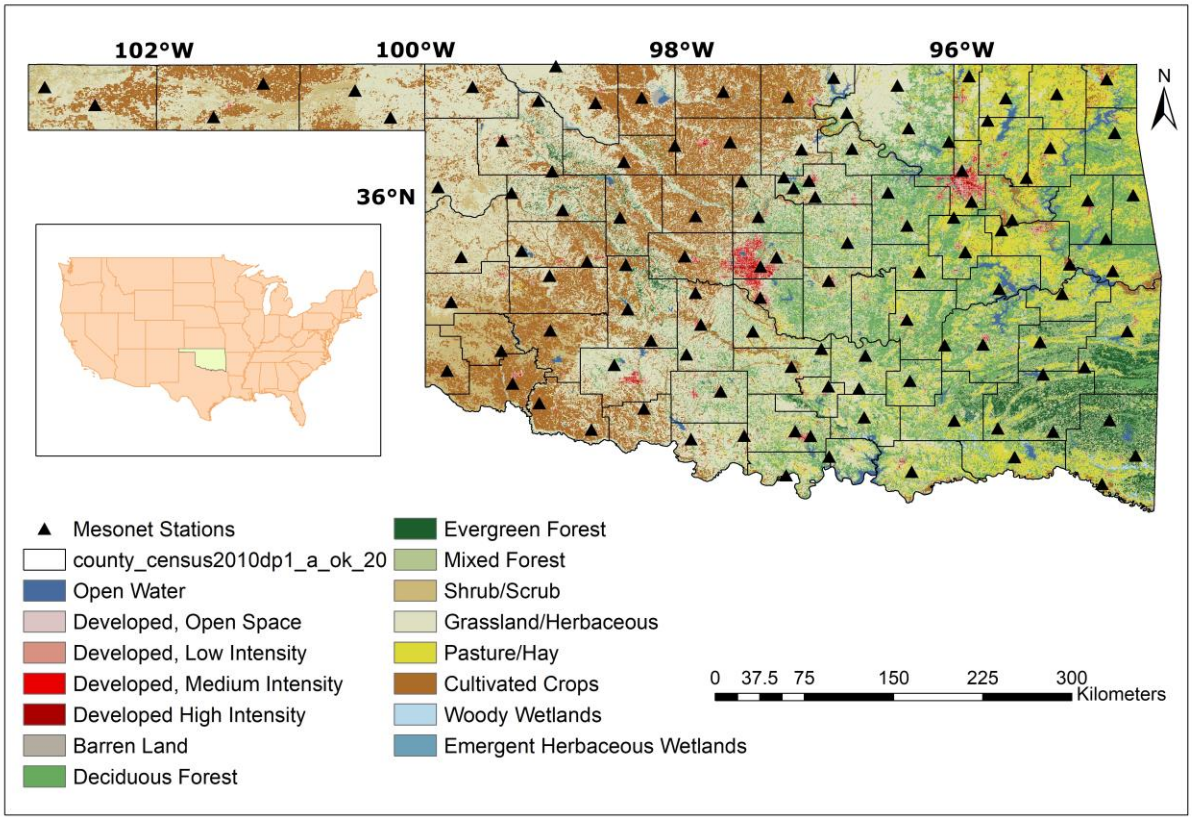


**Figure 4.1.** The distribution of in situ soil moisture stations from the Oklahoma Mesonet on a National Land Cover Dataset (NLCD) land cover type (for year 2016) map.

4.2.5.2 Oklahoma Climate Divisions

The performance of the three soil moisture products (OU-MSSM, AVE and SMAP L4) were also compared over nine climate divisions in Oklahoma (Figure 4.2). These nine divisions correspond to multiple factors such as climatic conditions, county lines, crop districts, and drainage basins rather than strict climatic homogeneity (Guttman & Quayle, 1996; Illston et al., 2004). Each climate division represents a section of the state that is considered to have homogeneous weather and climate patterns. The hottest temperatures in the state usually occur in the south, with cooler conditions towards the north (Illston et al., 2004). In addition, the division that typically receives the least amount of precipitation (504 mm/year) is the extreme northwest division (the Oklahoma

panhandle), while increased precipitation is observed toward the southeast (1200 mm/year; Illston et al., 2004). As a result, the climate of Oklahoma varies significantly across the state.

#### 4.2.5.3 STATSGO Soil Texture

The performance of the three soil moisture products (OU-MSSM, AVE and SMAP L4) were also compared over different soil textures in Oklahoma (Figure 4.2). The soil texture data (spatial resolution 9 km) was aggregated from the 1-km STATSGO database of Miller and White [1998] (Miller & White, 1998), which carries 16 texture classes by layer over 11 layers to 2-m depth. There are seven texture classes in Oklahoma including sand, sandy loam, silt loam, loam, silty clay loam, clay loam, and clay (Figure 4.2).

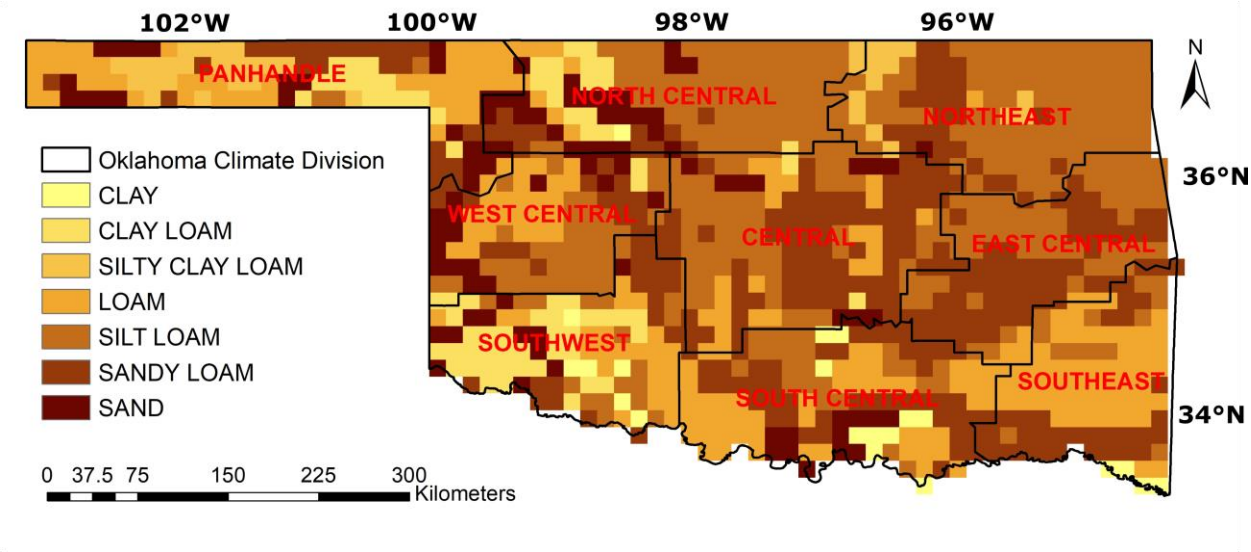


**Figure 4.2.** Oklahoma climate divisions and the 9-km STATSGO database derived soil texture classes.

## 4.3 Methodology

The soil moisture blending schemes used in this chapter were introduced by Yilmaz et al. (2012). Triple Colocation (TC) is used to estimate the error variances of parent products, including

the re-gridded Oklahoma Mesonet soil moisture measurements (Mesonet), the soil moisture simulations of NLDAS-2 Noah model (Noah), and the SMAP Level-3 soil moisture product (SMAP L3). Least square weighting (LSW) is used to merge these products. The merged product is named the University of Oklahoma Multi-Sensor Soil Moisture (OU-MSSM). We also compare the OU-MSSM with the equal weights merging product (AVE) and the SMAP L4 soil moisture product using the automated soil moisture mapping system as benchmark.

### *4.3.1 Triple Collocation*

Triple collocation (TC) analysis is a method for estimating the random error variances of three spatially and temporally collocated measurement systems of the same geophysical variable without treating any one system as perfectly observed "truth" (Stoffelen, 1998). A few assumptions are necessary for the TC method: (1) linearity between the true soil moisture signal and the observations, (2) signal and error stationarity, i.e., their mean values and variances are assumed to remain constant over time, (3) error orthogonality, i.e., the errors are independent from the true soil moisture signal, (4) the errors of three independent products should be independent or unrelated which means they must have a zero cross-correlation, and (5) the expectation of error is treated as zero. Even though some studies indicate that some of the assumptions are not always met, no proposed alternatives showed an enhanced accuracy or reliability of the error estimates (Gruber et al., 2016). For example, the difference between the climatology of the three data sets used in TC will led to the violation of signal stationarity. However, the adaptation proposed by many studies through removing the climatology of each data set individually is susceptible to estimation errors of the climatology (Crow et al., 2012; Draper et al., 2013; Gruber et al., 2016). The error stationarity of large data sets covering several years limits the representativeness of the

estimated average random error variance for subset time periods such as within different seasons. Therefore, Lowe and Schlenz (Loew & Schlenz, 2011) proposed a dynamic TC approach by applying TC analysis within 30-day windows. However, the extremely low sampling density in the time window leads to very low precision estimates. Therefore, the TC analysis in our study relies on annual error variance estimates based on a large sampling density. Yilmaz and Crow (Yilmaz & Crow, 2014) conducted experiments on the TC errors due to the relevance of three products, and the results revealed that the more independent the products are, the less TC induced error there will be. It is essential to consider the relevance of the inputs in order to make the TC method more reliable (Li et al., 2018). The three selected soil products selected, i.e., ground based (Mesonet), model based (Noah), and satellite based (SMAP L3), all meet the above criteria.

The TC method treats all three independent products as equally important, and thus no preference or bias is introduced for any one approach. Equation (4.1) illustrates a standard form of the TC method (Zwieback et al., 2012):

$$R_i = a_i + b_i \mathrm{T} + \varepsilon_i \tag{4.1}$$

Where $R_i$ (i $\in (X, Y, Z)$) indicates each of the three collocated soil moisture datasets X, Y and Z, T is the "relative truth," $a_i,\ b_i$ are the weights and biases to adjust, and $\varepsilon_i$ represents the error for each product $i$. Given this definition, the covariances between pairs of two different measurement systems (e.g., X and Y) would be given by

$$Cov(R_X, R_Y) = E(R_X R_Y) - E(R_X)E(R_Y) = b_X b_Y \sigma_T^2 + b_X Cov(T, \varepsilon_Y) + b_Y Cov(T, \varepsilon_X) + Cov(\varepsilon_X, \varepsilon_Y) \tag{4.2}$$

Where $\sigma_T^2 = Var(T)$. According to assumptions (3), (4), (5), $E(\varepsilon_X) = 0$, $(Cov(\varepsilon_X, \varepsilon_Y) = 0, X \neq Y)$, $(Cov(T, \varepsilon_X) = 0)$. Therefore, equation (4.2) reduces to

$$Q_{XY} = Cov(R_X, R_Y) = \begin{cases} b_X b_Y \sigma_T^2 & for\ X \neq Y \\ b_X^2 \sigma_T^2 + \sigma_{\varepsilon_X}^2 & for\ X = Y \end{cases} \qquad (4.3)$$

Where $\sigma_{\varepsilon_X}^2 = Var(\varepsilon_X)$. Since there are seven unknowns $(b_X, b_Y, b_Z, \sigma_{\varepsilon_X}, \sigma_{\varepsilon_Y}, \sigma_{\varepsilon_Z}, \sigma_T)$ in six equations in the 3 × 3 covariance matrix $(Q_{XX}, Q_{XY}, Q_{XZ}, Q_{YY}, Q_{YZ}, Q_{ZZ})$, there is no unique solution. However, the introduction of a new variable $\theta_X = b_X \sigma_T$, changes (4.3) to

$$Q_{XY} = Cov(R_X, R_Y) = \begin{cases} \theta_X \theta_Y & for\ X \neq Y \\ \theta_X^2 + \sigma_{\varepsilon_X}^2 & for\ X = Y \end{cases} \qquad (4.4)$$

From equation (4.4), we now have six unknowns in six equations and are able to calculate the root mean square error (RMSE) in the set of equations (4.5) that are based on the covariance of triplets (McColl et al., 2014):

$$\sigma_\varepsilon = \begin{cases} \sqrt{Q_{XX} - \frac{Q_{XY} Q_{XZ}}{Q_{YZ}}} \\ \sqrt{Q_{YY} - \frac{Q_{XY} Q_{YZ}}{Q_{XZ}}} \\ \sqrt{Q_{ZZ} - \frac{Q_{XZ} Q_{YZ}}{Q_{XY}}} \end{cases} \qquad (4.5)$$

*4.3.3 Least square weighting*

Least square weighting (LSW) is a commonly used method for blending different soil moisture data sources (Gruber et al., 2019; Yilmaz et al., 2012; Zeng et al., 2016; Zhang et al., 2021). The least square framework can be expressed as:

$$S_m = w_x S_x + w_y S_y + w_z S_z \qquad (4.6)$$

Where, $S_m$ is the merged product, $w_x$, $w_y$, and $w_z$ are the relative weights of three parent datasets $S_x$, $S_y$ and $S_z$ respectively. With $w_x + w_y + w_z = 1$, and by minimizing a cost function and the partial derivative of the cost function with respect to $w_x$ ,$w_y$, and $w_z$, the optimal estimation of the weights are obtained from:

$$w_x = \frac{\sigma_y^2 \sigma_z^2}{\sigma_x^2 \sigma_y^2 + \sigma_x^2 \sigma_z^2 + \sigma_y^2 \sigma_z^2} \qquad (4.7)$$

$$w_y = \frac{\sigma_x^2 \sigma_z^2}{\sigma_x^2 \sigma_y^2 + \sigma_x^2 \sigma_z^2 + \sigma_y^2 \sigma_z^2} \qquad (4.8)$$

$$w_z = \frac{\sigma_x^2 \sigma_y^2}{\sigma_x^2 \sigma_y^2 + \sigma_x^2 \sigma_z^2 + \sigma_y^2 \sigma_z^2} \qquad (4.9)$$

Where, $\sigma_x^2$, $\sigma_y^2$, and $\sigma_z^2$ are the TC-estimated error variance for the parent datasets.

*4.3.4 Goodness of fit and example product intercomparison*

The Pearson correlation coefficient (R), bias, and the root mean square error (RMSE) between blended products (OU-MSSM, AVE and SMAP L4) and the automated soil moisture mapping system are used for the comparison of the blended products.

*4.3.5 Daily time series product inter-comparison*

An example one-month multi-product (i.e., OU-MSSM, AVE, SMAP_L4 and OSU benchmark) performance intercomparison is also presented to provide an illustrative example of

all products performance during rainy and non-rainy days during a typical rainy summer season in Oklahoma.

## 4.4 Results

### *4.4.1 Weights from LSW*

The LSW scheme is adopted with TC-estimated error variance to calculate the weights of the triplet of the SMAP L3, Noah, and Mesonet soil moisture products (Figure. 3). According to LSW theory, the weight of the parent products is inversely proportional to the error variance. Therefore, for this triplet, Noah has the highest mean weight (0.56), followed by Mesonet (0.28), and SMAP L3 has the smallest mean weights (0.16). Spatially, the large weights (> 0.56) for Noah are clustered in east Oklahoma, while the small weights are in western Oklahoma (Figure. 4.3b). The high weights (> 0.28) for Mesonet are concentrated in the Panhandle, West Central and Southwest regions (Fig. 4.3a). The weights of SMAP L3 below 0.2 are throughout most of the study area; however, large weights (> 0.2) are shown in areas where both Noah and Mesonet has very small weights (Figure 4.3c).

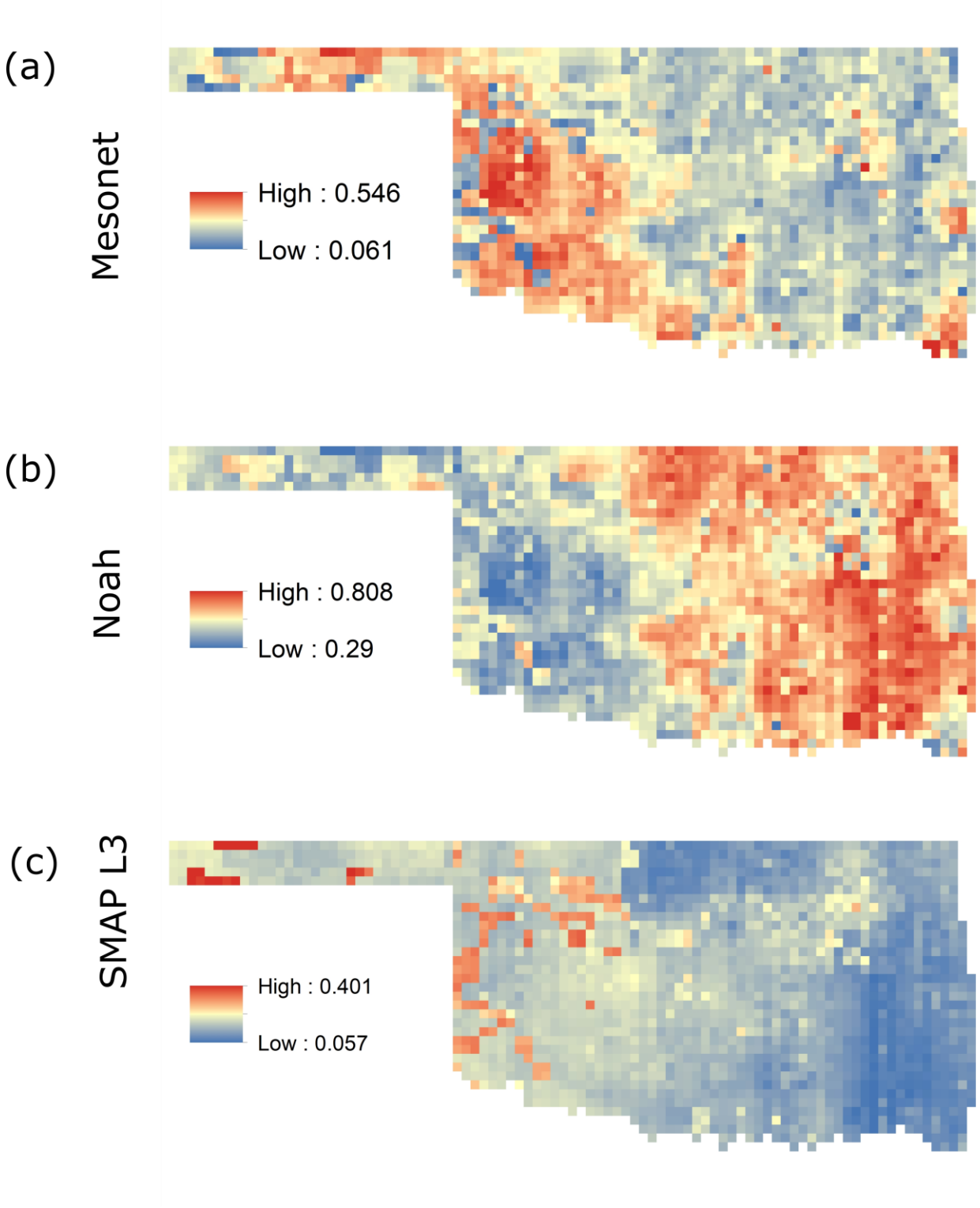
(a)
Mesonet
High : 0.546
Low : 0.061
(b)
Noah
High : 0.808
Low : 0.29
(c)
SMAP L3
High : 0.401
Low : 0.057

**Figure 4.3.** Map of the least square weighting (LSW)-derived weights based on the TC estimated error variance using soil moisture. Subfigures (a) to (c) illustrate the spatial distribution of weights for each product.

*4.4.2 Evaluation of blended products*

Figure 4.4 provides a quantitative evaluation of the blended products (OU-MSSM, AVE) and SMAP L4 using the OSU daily soil moisture measurements based on Pearson correlation coefficient (CC; first row in Figure 4.4), Root Mean Square Error (RMSE; second row), and bias (third row). The OU-MSSM presents the highest averaged CC value (0.87), followed by AVE (0.82), while the SMAP L4 has the lowest averaged CC value (0.76). Based on a one-way ANOVA test, the three products have significantly ($p<0.05$) different CC values from each other. Spatially, the high CC values (> 0.9) for OU-MSSM are concentrated in the West Central, Southwest and Northeast regions. In terms of RMSE, the OU-MSSM presents the lowest mean RMSE value (0.06), followed by AVE (0.063), while the SMAP L4 has highest mean RMSE value (0.084). Based on a one-way ANOVA test, the three products have significantly ($p<0.05$) different RMSE values from each other. Particularly, the low RMSE (< 0.05) for OU-MSSM are clustered in east Oklahoma, while the high values (> 0.07) are in west Oklahoma. As with the bias evaluation, the OU-MSSM and the AVE present positive averaged bias values (0.05 and 0.03 respectively), while the SMAP L4 has the negative averaged bias value (-0.138). A one-way ANOVA test was also performed on the bias of the three products, and the results show that the OU-MSSM has significantly ($p<0.05$) larger bias than the AVE product. Spatially, both OU-MSSM and AVE show high positive bias in Panhandle region.

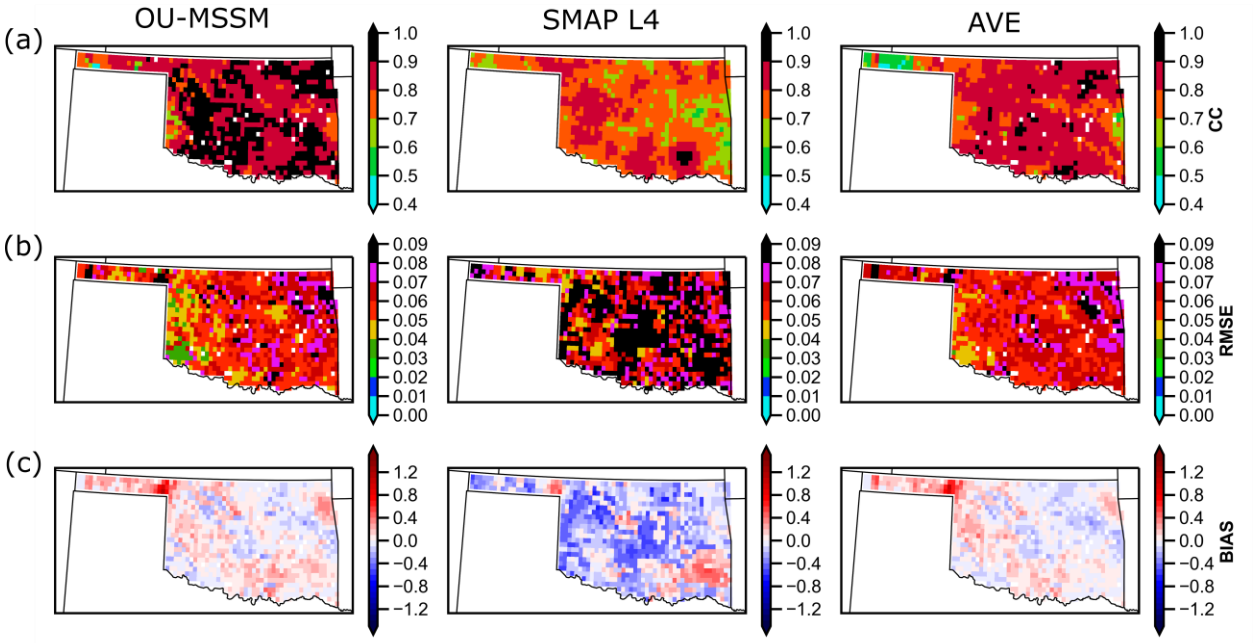


**Figure 4.4.** Evaluation of the blended products (OU-MSSM, AVE) and SMAP L4 using the OSU daily soil moisture measurements based on: (a) Pearson correlation coefficient (CC, first row), (b) Root Mean Square Error (RMSE, second row), and (c) the Bias (BIAS, third row).

Figure 4.5 provides a quantitative comparison of the blended products (OU-MSSM, AVE) and SMAP L4 with the OSU daily soil moisture measurements in terms of (a) Pearson correlation coefficient, (b) root mean square error, and (c) bias over the nine climate divisions of Oklahoma. In terms of CC, OU-MSSM shows the highest correlations with the OSU data over all nine climate divisions, followed by AVE. SMAP L4 provides the lowest correlations with the OSU product over all nine climate divisions. As for the RMSE, SMAP L4 provides the highest RMSE over all nine climate divisions. OU-MSSM shows lowest RMSE in Panhandle, West Central and Southwest. In other six climate divisions, OU-MSSM and AVE have similar RMSE values. For bias, SMAP L4 shows negative mean bias in most climate divisions except Southeast. In Southeast, its bias ranges from -0.35 to 0.56 with a mean bias of 0.117. Both OU-MSSM and AVE have positive mean bias in all nine climate divisions. OU-MSSM has larger bias than AVE in Southeast,

South Central, North Central, Northeast, Central and East Central, while it has same bias with AVE in the Panhandle, West Central, and Southwest.

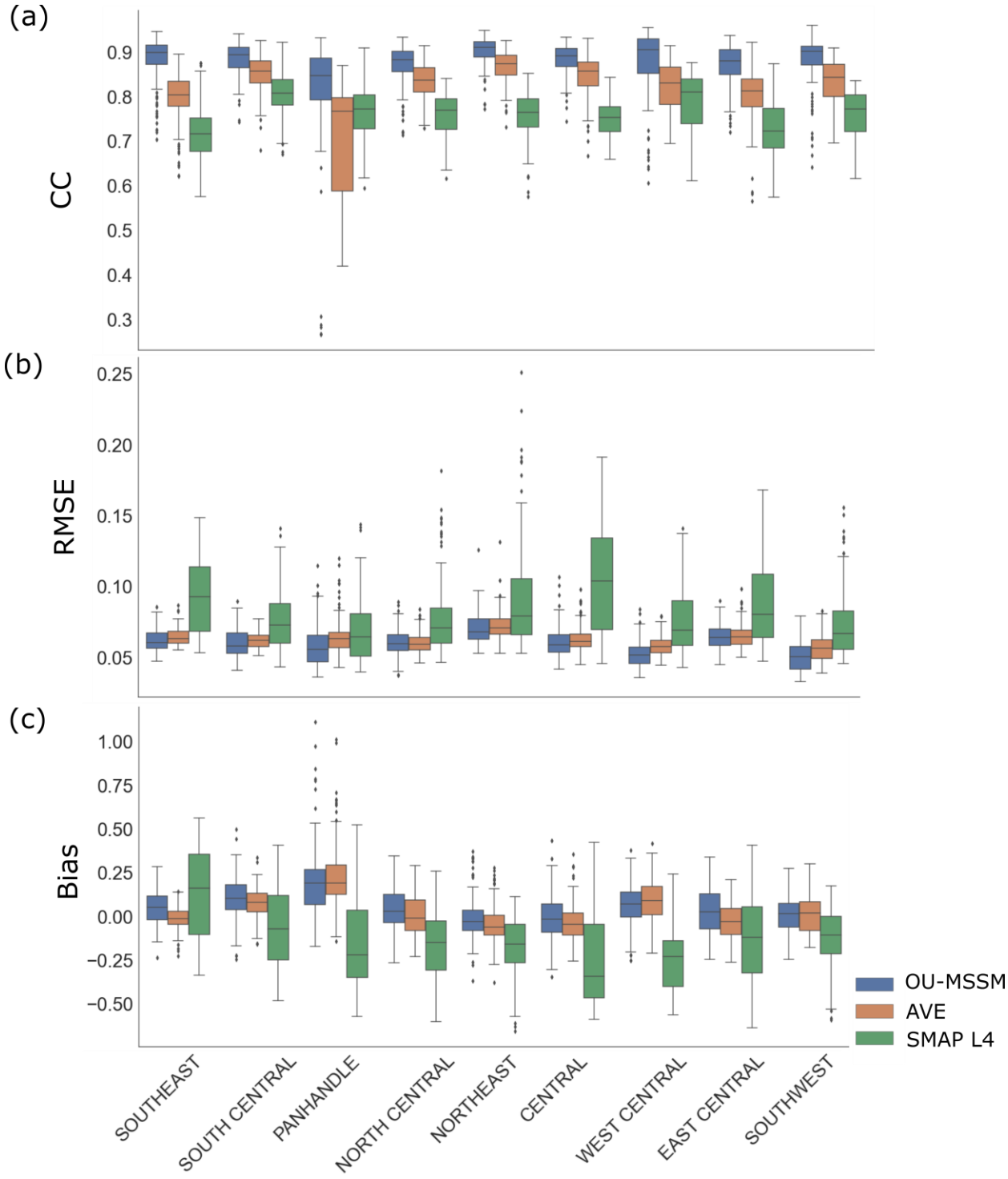


**Figure 4.5.** Comparison of the blended products (OU-MSSM, AVE) and SMAP L4 over nine climate divisions in Oklahoma: (a) Pearson correlation coefficient, (b) Root Mean Square Error, and (c) the Bias. All metrics use the benchmark (OSU) daily soil moisture product.

Figure 4.6 provides a quantitative comparison of the blended products (OU-MSSM, AVE) and SMAP L4 with the OSU daily soil moisture measurements in terms of (a) Pearson correlation coefficient, (b) root mean square error, and (c) bias over the different land cover types in Oklahoma. In terms of CC, OU-MSSM shows the highest correlations with the OSU data over all land cover types, followed by AVE. SMAP L4 provides the lowest correlations with the OSU product over all land cover types. As for the RMSE, SMAP L4 provides the highest RMSE over all land cover types. OU-MSSM shows lowest RMSE in all land cover types except mixed forest. In mixed forest, OU-MSSM and AVE have similar RMSE values. For the bias, SMAP L4 shows negative mean bias in most land cover types except mixed forest and evergreen forest. In these two land cover types, SMAP L4 has large ranges of bias (-0.56 to 0.56 and -0.49 to 0.45) which may be attributed to tree canopy and dense vegetation (Zeng et al., 2016). The OU-MSSM has smaller mean bias than AVE in shrub/scrub and developed, on the other hand, it has larger mean bias than AVE in mixed forest, evergreen forest, and deciduous forest.

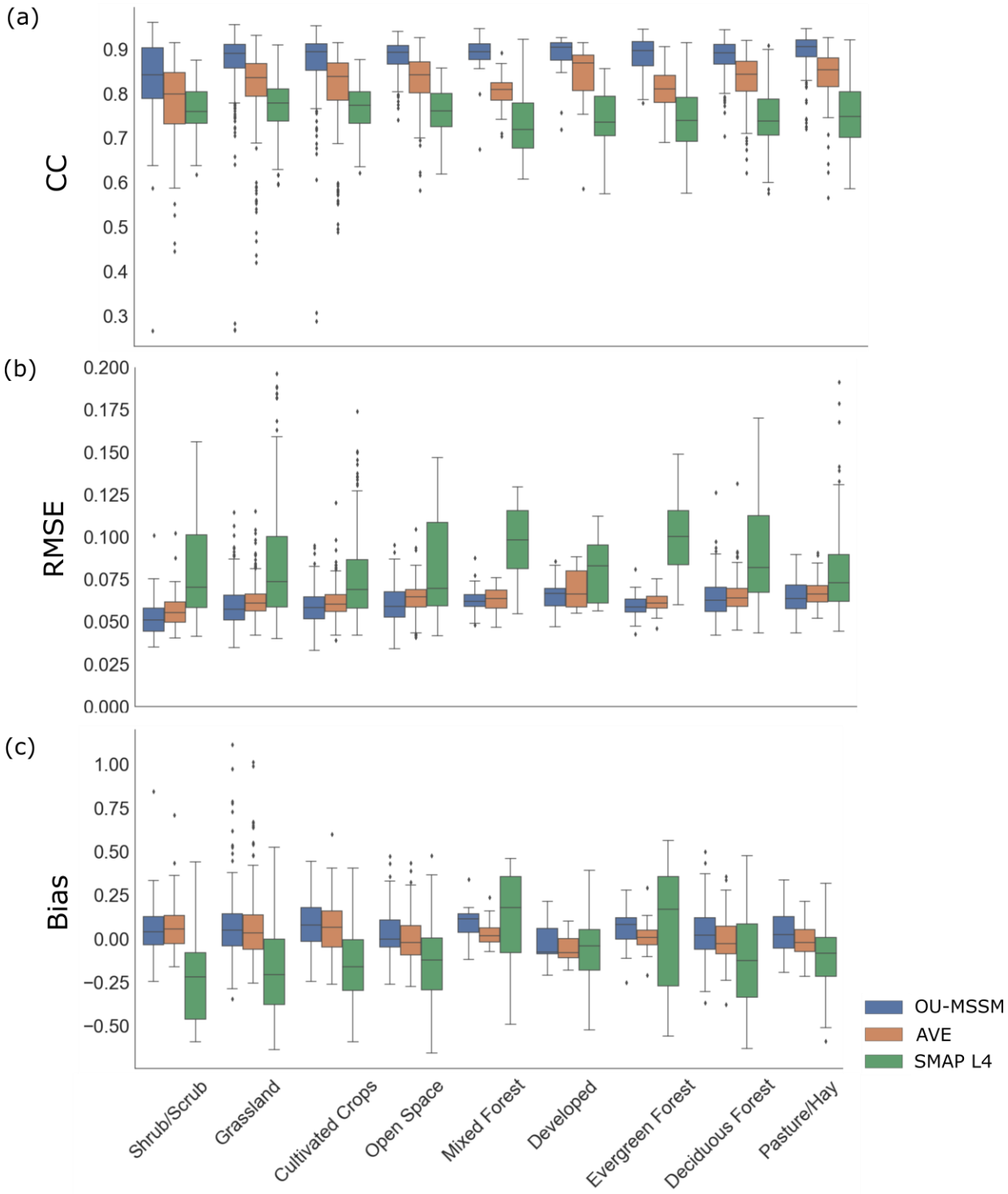


**Figure 4.6.** Comparison of the blended products (OU-MSSM, AVE) and SMAP L4 over different landcover types in Oklahoma: (a) Pearson correlation coefficient (CC), (b) Root Mean Square Error (RMSE), and (c) the Bias. All metrics use the (OSU) daily soil moisture as benchmark product.

Figure 4.7 provides a quantitative comparison of the blended products (OU-MSSM, AVE) and SMAP L4 with the OSU daily soil moisture measurements in terms of (a) Pearson correlation coefficient, (b) root mean square error, and (c) bias over the different soil texture types in Oklahoma. In terms of CC, OU-MSSM shows the highest correlations with the OSU data over loam, sandy loam, clay loam, silt loam, and clay, followed by AVE. SMAP L4 provides the lowest correlations with the OSU product over these soil texture types. For sand soil type, OU-MSSM, AVE, and SMAP-L4 have similar averaged CC values (0.746, 0.77, and 0.753 respectively), while OU-MSSM and AVE have several small outliers. For silty clay loam, OU-MSSM has the highest correlations with the OSU product (0.856), followed by SMAP-L4 (0.761) and AVE (0.728). As for the RMSE, SMAP L4 provides the highest RMSE over all soil texture types. OU-MSSM shows lowest RMSE in all soil types except sand and clay. In these two soil types, OU-MSSM and AVE have similar RMSE values (0.06 vs. 0.065, 0.063 vs. 0.0625). For bias, SMAP L4 shows negative mean bias in most soil types except clay. In clay soil type, all three products show positive mean bias. Among them, OU-MSSM has the largest positive mean bias. Both OU-MSSM and AVE have positive mean bias in all soil types except sandy loam and sand. In sandy loam, both products show similar negative mean bias (-0.01). In sand type, both products show negative mean bias and OU-MSSM has smaller negative mean bias than AVE. In loam, silty clay loam, clay loam and silt loam soil types, both OU-MSSM and AVE have positive mean bias, and OU-MSSM has larger positive mean bias than AVE.

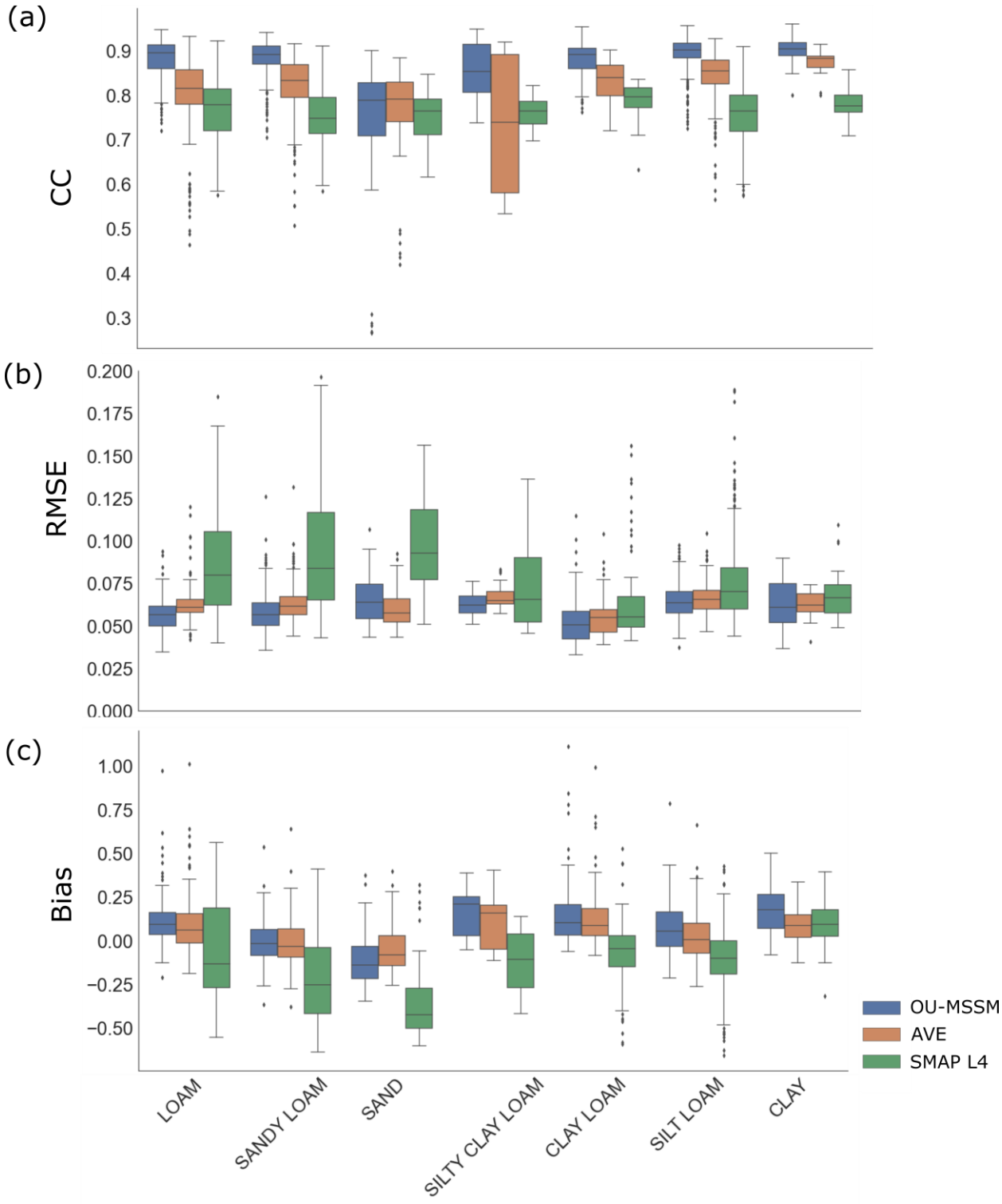


**Figure 4.7.** Comparison of the blended products (OU-MSSM, AVE) and SMAP L4 over different soil types in Oklahoma: (a) Pearson correlation coefficient, (b) Root Mean Square Error, and (c) Bias. All metrics use the (OSU) daily soil moisture as benchmark product.

Figure 4.8 displays a time series of Mesonet daily precipitation with the corresponding OU-MSSM, AVE, SMAP L4 and OSU benchmark surface soil moisture products (5 cm soil depth) at one pixel (9km * 9km) during one month summer period (August 2016) that included several rainfall events. This is in the Southwest climate division corresponding to clay loam soil and cultivated crops vegetation. The patterns of OU-MSSM and AVE are consistent with the OSU soil moisture product, which respond well to the change of rainfall. The increasing and decreasing of soil moisture follows that of the precipitation and the lags between soil moisture of the three products and the precipitation are approximately the same. Moreover, the larger the amount of precipitation, the larger the surface soil water content. The pattern of SMAP L4 is different from the other three soil moisture products. Its soil water content rises at the same time when precipitation rises. Moreover, the soil water content rises to the same level where rainfall amount differs.

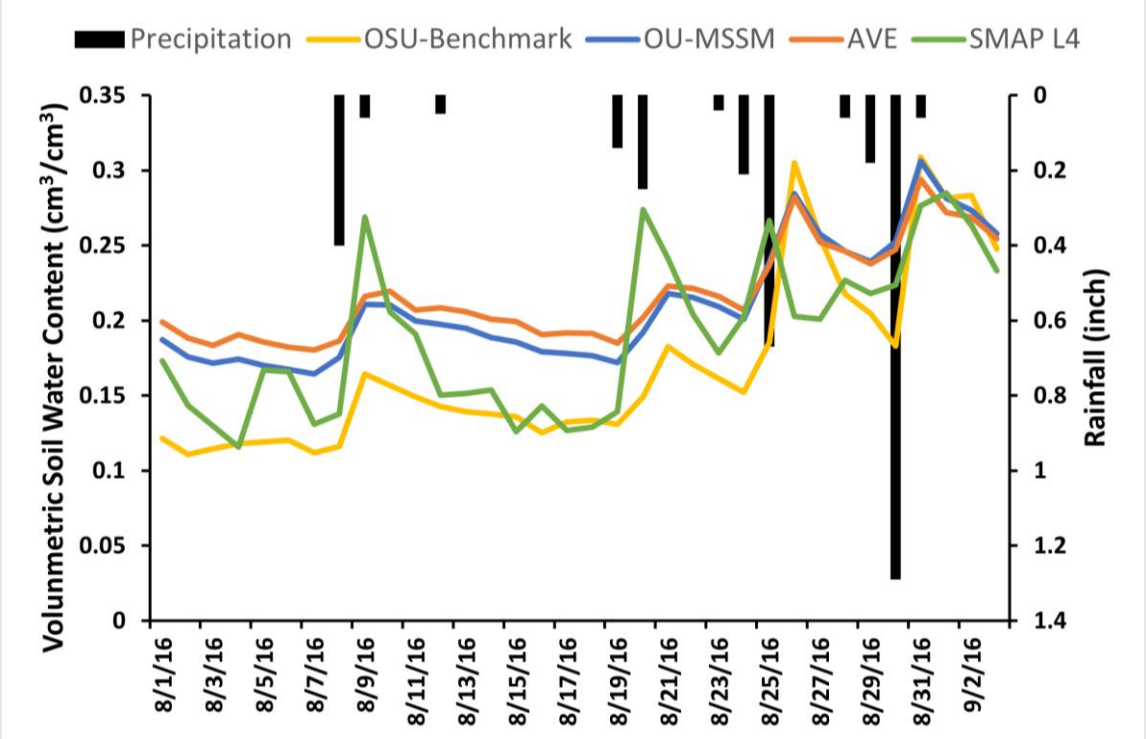


**Figure 4.8.** Precipitation (inch; vertical bars) and soil water content ($cm^3/cm^3$) time series from of OU-MSSM, AVE, SMAP L4 and OSU soil moisture products for one example location with coordinates (lat= 34.938, lon= -99.063) corresponding to clay loam soil, cultivated crops vegetation and located in the Southwest climate region of Oklahoma.

### 4.5 Discussion

This study adopted an objective methodology introduced by Yilmaz et al. (2012) in blending multisource soil moisture products in Oklahoma. The LSW scheme is adopted with TC-estimated error variance to calculate the weights of the parent products (the SMAP_L3, NLDAS Noah, and the interpolated Mesonet). The results show that the high weights of each product are clustered in different regions of Oklahoma. Specifically, the large weights (> 0.56) for NLDAS Noah are clustered in east Oklahoma, the high weights (> 0.28) for Mesonet are concentrated in the Panhandle, West Central and Southwest regions and the large weights (> 0.2) of SMAP L3 are shown in areas where both Noah and Mesonet have very small weights. This suggests that each product might not be able to represent the true soil moisture variations in Oklahoma individually.

This might be due to the characteristics of these soil moisture measurements. For example, the Oklahoma Mesonet site standards require the sites to be far away from urban landscapes, irrigation, forest, bare soil, fast growing vegetation, and large bodies of water to minimize those influences (Brock et al., 1995; McPherson et al., 2007). Therefore, it is meaningful to merge independent products from multiple sources (satellite, model, in-situ) to obtain a more accurate estimate.

In this study, the TC-LSW approach-based merging product (OU-MSSM) was also compared with an equal weighting approach-based blending product (AVE) and SMAP L4. The OU-MSSM has the highest correlation with the OSU daily soil moisture measurement with the lowest mean RMSE value, while SMAP L4 has the lowest correlation with the OSU daily soil moisture estimates with the highest mean RMSE value. Spatially, the OU-MSSM provided the best estimation of volumetric soil moisture in the Panhandle, West Central and Southwest climate divisions. It is because the weights of resampled Mesonet composites the OU-MSSM are the highest in these climate divisions and the resampled Mesonet has the highest correlation with the true soil moisture according to our previous study. While in other six divisions, the OU-MSSM has the highest averaged CC and lowest averaged RMSE it also shows larger positive bias than AVE. The SMAP L4 has the third best estimation of volumetric soil moisture in all nine climate divisions. Moreover, SMAP L4 shows negative mean bias in most climate divisions except Southeast. In Southeast, its bias ranges from -0.35 to 0.56 with a mean bias of 0.117. This large range of bias may be attributed to tree canopy and dense vegetation because higher vegetation intensity will reduce the quality of soil moisture retrieval (Zhang, et al., 2017). Fan et al.(Fan et al., 2020) also found that SMAP underestimates soil moisture in global vegetation-disturbed areas, which are related to negative biases in MERRA temperature data. However, the bias in this study (-0.138) is larger than some studies assessing SMAP L4 surface soil moisture product with in situ

measurements in the United States ((Tavakol et al., 2019) (bias = -0.006)), ((Reichle et al., 2017) (bias = -0.01)). This might be because the in-situ measurements used in this study are regression kriging-based interpolated in situ soil moisture measurements. The results indicate that (1) the TC-LSW approach-based merging scheme appears to be more optimal than the equal weighting merging scheme; (2) in-situ measurements are valuable for improving the accuracy of blended soil moisture datasets; (3) using multiple sources of soil moisture helps to reduce the overall uncertainty in the soil moisture estimates.

This study also provided a comparison of the blended products (OU-MSSM, AVE) and SMAP L4 over different land cover types and soil types in Oklahoma. The results indicate that the OU-MSSM has the best performance in all land cover types except mixed forest, evergreen forest, and deciduous forest. In these three land cover types, the OU-MSSM has larger mean positive bias than AVE. This might be because in these three land cover types, NLDAS2 Noah has more weight in the OU-MSSM calculations than in AVE, and NLDAS2 Noah displays large biases when compared to in situ observations according to Xia (Xia et al., 2014). As with soil types, the OU-MSSM has best performance in all soil types except sand. The OU-MSSM has similar performance with AVE in sand soil type.

These findings should be placed in context by acknowledging the study limitations, including: (1) the soil moisture products used in this study were all extracted from 6 a.m. observations. However, a temporal mismatch may still exist due to the different temporal resolutions of each soil moisture product. Future work can adopt methods to ensure the temporal coherence of different datasets; (2) The Mesonet product used in the TC intercomparison is interpolated from point-scaled Oklahoma Mesonet to spatial resolution of SMAP (9 km) using the ordinary kriging method. Future work could use regression kriging approaches and including independent

predictors such as soil properties, land cover, topography and precipitation to increase the accuracy of the interpolated Mesonet product; (3) Further study is required to test whether these conclusions are valid in other regions. Indeed, a primary advantage of blending in-situ soil moisture is its representation of root zone. The focus of our next step is to apply the methods developed here for blending root zone soil moisture from in-situ and model sources.

### 4.6 Conclusion

This chapter adopted an objective methodology introduced by Yilmaz et al. (Yilmaz et al., 2012) in blending multisource soil moisture products in Oklahoma, identified the importance of incorporating in-situ soil moisture into soil moisture blending and quantified the impact of different weighting schemes on soil moisture blending. Soil moisture information from multiple sources, including satellite (SMAP_L3), land surface model (NLDAS-2 Noah), and in-situ measurements (the interpolated Mesonet), are used to generate blended soil moisture products at a 9-km spatial resolution and daily temporal resolution. TC was used to estimate the error variance of the parent products, and LSW was used to generate blended soil moisture products. An equal weighting approach (AVE) and SMAP L4 were also compared with the TC-LSW approach. Several conclusions are summarized as follows:

(1) the TC-LSW approach-based merging scheme is more optimal than the equal weighting merging scheme.

(2) in-situ measurements are valuable for improving the accuracy of blended soil moisture datasets.

(3) using multiple sources of soil moisture helps to reduce the overall uncertainty in the soil moisture estimates.

(4) the OU-MSSM has best performance in all land cover types except mixed forest, evergreen forest, and deciduous forest.

(5) the OU-MSSM has best performance in all soil types except sand.

The resulting combined soil moisture estimate can be used as a standalone soil moisture product with available uncertainty estimates.

**References**

Brocca, L., Hasenauer, S., Lacava, T., Melone, F., Moramarco, T., Wagner, W., Dorigo, W., Matgen, P., Martínez-Fernández, J., Llorens, P., Latron, J., Martin, C., & Bittelli, M. (2011). Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe. *Remote Sensing of Environment*, *115*(12), 3390–3408. https://doi.org/10.1016/j.rse.2011.08.003

Brocca, L., Melone, F., Moramarco, T., Wagner, W., Naeimi, V., Bartalis, Z., & Hasenauer, S. (2010). Improving runoff prediction through the assimilation of the ASCAT soil moisture product. *Hydrology and Earth System Sciences*, *14*(10), 1881–1893. https://doi.org/10.5194/hess-14-1881-2010

Brocca, L., Ponziani, F., Moramarco, T., Melone, F., Berni, N., & Wagner, W. (2012). Improving Landslide Forecasting Using ASCAT-Derived Soil Moisture Data: A Case Study of the Torgiovannetto Landslide in Central Italy. *Remote Sensing*, *4*(5), 1232–1244. https://doi.org/10.3390/rs4051232

Brock, F. V., Crawford, K. C., Elliott, R. L., Cuperus, G. W., Stadler, S. J., Johnson, H. L., & Eilts, M. D. (1995). The Oklahoma Mesonet: A Technical Overview. *Journal of Atmospheric*

*and Oceanic Technology*, *12*(1), 5–19. https://doi.org/10.1175/1520-0426(1995)012<0005:TOMATO>2.0.CO;2

Capecchi, V., & Brocca, L. (2014). A simple assimilation method to ingest satellite soil moisture into a limited-area NWP model. *Meteorologische Zeitschrift*, *23*(2), 105–121. https://doi.org/10.1127/0941-2948/2014/0585

Chan, S. (2016). Enhanced level 3 passive soil moisture product specification document. *Jet Propulsion Lab., California Inst. Technol.: Pasadena, CA, USA*.

Chen, F., Crow, W. T., Bindlish, R., Colliander, A., Burgin, M. S., Asanuma, J., & Aida, K. (2018). Global-scale evaluation of SMAP, SMOS and ASCAT soil moisture products using triple collocation. *Remote Sensing of Environment*, *214*, 1–13. https://doi.org/10.1016/j.rse.2018.05.008

Chen, F., Crow, W. T., Colliander, A., Cosh, M. H., Jackson, T. J., Bindlish, R., Reichle, R. H., Chan, S. K., Bosch, D. D., & Starks, P. J. (2016). Application of triple collocation in ground-based validation of Soil Moisture Active/Passive (SMAP) level 2 data products. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(2), 489–502.

Chen, F., Janjić, Z., & Mitchell, K. (1997). Impact of atmospheric surface-layer parameterizations in the new land-surface scheme of the NCEP mesoscale Eta model. *Boundary-Layer Meteorology*, *85*(3), 391–421.

Crow, W. T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., Rosnay, P. de, Ryu, D., & 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). https://doi.org/10.1029/2011RG000372

Crow, W. T., Berg, M. J. van den, Huffman, G. J., & Pellarin, T. (2011). Correcting rainfall using satellite-based surface soil moisture retrievals: The Soil Moisture Analysis Rainfall Tool (SMART). *Water Resources Research*, *47*(8). https://doi.org/10.1029/2011WR010576

Crow, W. T., Bindlish, R., & Jackson, T. J. (2005). The added value of spaceborne passive microwave soil moisture retrievals for forecasting rainfall-runoff partitioning. *Geophysical Research Letters*, *32*(18). https://doi.org/10.1029/2005GL023543

Crow, W. T., & Loon, E. V. (2006). Impact of Incorrect Model Error Assumptions on the Sequential Assimilation of Remotely Sensed Surface Soil Moisture. *Journal of Hydrometeorology*, *7*(3), 421–432. https://doi.org/10.1175/JHM499.1

Dingman, S. L. (2015). *Physical Hydrology: Third Edition*. Waveland Press.

Dorigo, W. A., Gruber, A., De Jeu, R. A. M., Wagner, W., Stacke, T., Loew, A., Albergel, C., Brocca, L., Chung, D., Parinussa, R. M., & Kidd, R. (2015). Evaluation of the ESA CCI soil moisture product using ground-based observations. *Remote Sensing of Environment*, *162*, 380–395. https://doi.org/10.1016/j.rse.2014.07.023

Draper, C., Reichle, R., de Jeu, R., Naeimi, V., Parinussa, R., & Wagner, W. (2013). Estimating root mean square errors in remotely sensed soil moisture over continental scale domains. *Remote Sensing of Environment*, *137*, 288–298.

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., & Tarpley, J. D. (2003). Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research: Atmospheres*, *108*(D22).

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

Entekhabi, D., Yueh, S., & De Lannoy, G. (2014). *SMAP handbook*.

Famiglietti, J. S., Ryu, D., Berg, A. A., Rodell, M., & Jackson, T. J. (2008). Field observations of soil moisture variability across scales. *Water Resources Research*, *44*(1). https://doi.org/10.1029/2006WR005804

Fan, X., Liu, Y., Gan, G., & Wu, G. (2020). SMAP underestimates soil moisture in vegetation-disturbed areas primarily as a result of biased surface temperature data. *Remote Sensing of Environment*, *247*, 111914. https://doi.org/10.1016/j.rse.2020.111914

*GES DISC Dataset: NLDAS Noah Land Surface Model L4 Monthly Climatology 0.125 x 0.125 degree V002 (NLDAS_NOAH0125_MC 002)*. (n.d.). https://disc.gsfc.nasa.gov/datasets/NLDAS_NOAH0125_MC_002/summary

Gruber, A., Scanlon, T., van der Schalie, R., Wagner, W., & Dorigo, W. (2019). Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology. *Earth System Science Data*, *11*(2), 717–739. https://doi.org/10.5194/essd-11-717-2019

Gruber, A., Su, C.-H., Zwieback, S., Crow, W., Dorigo, W., & Wagner, W. (2016). Recent advances in (soil moisture) triple collocation analysis. *International Journal of Applied Earth Observation and Geoinformation*, *45*, 200–211.

Gu, Y., Hunt, E., Wardlow, B., Basara, J. B., Brown, J. F., & Verdin, J. P. (2008). Evaluation of MODIS NDVI and NDWI for vegetation drought monitoring using Oklahoma Mesonet soil moisture data. *Geophysical Research Letters*, *35*(22). https://doi.org/10.1029/2008GL035772

Guttman, N. B., & Quayle, R. G. (1996). A Historical Perspective of U.S. Climate Divisions. *Bulletin of the American Meteorological Society*, *77*(2), 293–304. https://doi.org/10.1175/1520-0477(1996)077<0293:AHPOUC>2.0.CO;2

Hansen, M. C., Defries, R. S., Townshend, J. R. G., & Sohlberg, R. (2000). Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*, *21*(6–7), 1331–1364. https://doi.org/10.1080/014311600210209

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

Illston, B. G., Basara, J. B., Fiebrich, C. A., Crawford, K. C., Hunt, E., Fisher, D. K., Elliott, R., & Humes, K. (2008). Mesoscale Monitoring of Soil Moisture across a Statewide Network. *Journal of Atmospheric and Oceanic Technology*, *25*(2), 167–182. https://doi.org/10.1175/2007JTECHA993.1

Imaoka, K., Kachi, M., Kasahara, M., Ito, N., Nakagawa, K., & Oki, T. (2010). Instrument performance and calibration of AMSR-E and AMSR2. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science*, *38*(8), 13–18.

Jackson, T. J., Bindlish, R., Cosh, M. H., Zhao, T., Starks, P. J., Bosch, D. D., Seyfried, M., Moran, M. S., Goodrich, D. C., Kerr, Y. H., & Leroux, D. (2012). Validation of Soil Moisture and Ocean Salinity (SMOS) Soil Moisture Over Watershed Networks in the U.S. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(5), 1530–1543. https://doi.org/10.1109/TGRS.2011.2168533

JACKSON, T. J., SCHMUGGE, J., & ENGMAN, E. T. (1996). Remote sensing applications to hydrology: Soil moisture. *Hydrological Sciences Journal*, *41*(4), 517–530. https://doi.org/10.1080/02626669609491523

Kang, J., Jin, R., Li, X., & Zhang, Y. (2021). Mapping High Spatiotemporal-Resolution Soil Moisture by Upscaling Sparse Ground-Based Observations Using a Bayesian Linear Regression Method for Comparison with Microwave Remotely Sensed Soil Moisture Products. *Remote Sensing*, *13*(2), 228. https://doi.org/10.3390/rs13020228

Kerr, Y. H., Waldteufel, P., Wigneron, J.-P., Martinuzzi, J., Font, J., & Berger, M. (2001). Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE Transactions on Geoscience and Remote Sensing*, *39*(8), 1729–1735.

Lakhankar, T., Jones, A. S., Combs, C. L., Sengupta, M., Vonder Haar, T. H., & Khanbilvardi, R. (2010). Analysis of large scale spatial variability of soil moisture using a geostatistical method. *Sensors*, *10*(1), 913–932.

Li, C., Tang, G., & Hong, Y. (2018). Cross-evaluation of ground-based, multi-satellite and reanalysis precipitation products: Applicability of the Triple Collocation method across Mainland China. *Journal of Hydrology*, *562*, 71–83.

Loew, A., & Schlenz, F. (2011). A dynamic approach for evaluating coarse scale satellite soil moisture products. *Hydrology and Earth System Sciences*, *15*(1), 75–90.

McColl, K. A., Vogelzang, J., Konings, A. G., Entekhabi, D., Piles, M., & Stoffelen, A. (2014). Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target. *Geophysical Research Letters*, *41*(17), 6229–6236.

McPherson, R. A., Fiebrich, C. A., Crawford, K. C., Kilby, J. R., Grimsley, D. L., Martinez, J. E., Basara, J. B., Illston, B. G., Morris, D. A., Kloesel, K. A., Melvin, A. D., Shrivastava, H.,

Wolfinbarger, J. M., Bostic, J. P., Demko, D. B., Elliott, R. L., Stadler, S. J., Carlson, J. D., & Sutherland, A. J. (2007). Statewide Monitoring of the Mesoscale Environment: A Technical Update on the Oklahoma Mesonet. *Journal of Atmospheric and Oceanic Technology*, *24*(3), 301–321. https://doi.org/10.1175/JTECH1976.1

Miller, D. A., & White, R. A. (1998). A conterminous United States multi-layer soil characteristics data set for regional climate and hydrology modeling, Earth Interactions, 2. Web-based publication. *Res*, 711–724.

Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Cosgrove, B. A., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., Tarpley, J. D., Lettenmaier, D. P., Marshall, C. H., Entin, J. K., Pan, M., Shi, W., Koren, V., … Bailey, A. A. (2004). The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research: Atmospheres*, *109*(D7). https://doi.org/10.1029/2003JD003823

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

Mohd Kassim, M. R., Mat, I., & Harun, A. N. (2014). Wireless Sensor Network in precision agriculture application. *2014 International Conference on Computer, Information and Telecommunication Systems (CITS)*, 1–5. https://doi.org/10.1109/CITS.2014.6878963

Ochsner, T. E., Linde, E., Haffner, M., & Dong, J. (2019). Mesoscale Soil Moisture Patterns Revealed Using a Sparse In Situ Network and Regression Kriging. *Water Resources Research*, *55*(6), 4785–4800. https://doi.org/10.1029/2018WR024535

ONeill, Peggy E., Chan, Steven, Njoku, Eni G., Jackson, Tom, & Bindlish, Rajat. (2019). *SMAP Enhanced L3 Radiometer Global Daily 9 km EASE-Grid Soil Moisture, Version 3* [Data set]. NASA National Snow and Ice Data Center DAAC. https://doi.org/10.5067/T90W6VRLCBHI

Parinussa, R. M., Meesters, A. G., Liu, Y. Y., Dorigo, W., Wagner, W., & de Jeu, R. A. (2011). Error estimates for near-real-time satellite soil moisture as derived from the land parameter retrieval model. *IEEE Geoscience and Remote Sensing Letters*, *8*(4), 779–783.

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

Qin, J., Yang, K., Lu, N., Chen, Y., Zhao, L., & Han, M. (2013). Spatial upscaling of in-situ soil moisture measurements based on MODIS-derived apparent thermal inertia. *Remote Sensing of Environment*, *138*, 1–9. https://doi.org/10.1016/j.rse.2013.07.003

Ray, R. L., Jacobs, J. M., & Cosh, M. H. (2010). Landslide susceptibility mapping using downscaled AMSR-E soil moisture: A case study from Cleveland Corral, California, US. *Remote Sensing of Environment*, *114*(11), 2624–2636. https://doi.org/10.1016/j.rse.2010.05.033

Reichle, Rolf, De Lannoy, Gabrielle, Koster, Randal, Crow, Wade, Kimball, John, & Liu, Qing. (2020). *SMAP L4 Global 3-hourly 9 km EASE-Grid Surface and Root Zone Soil Moisture Geophysical Data, Version 5* [Data set]. NASA National Snow and Ice Data Center DAAC. https://doi.org/10.5067/9LNYIYOBNBR5

Reichle, R. H., Crow, W. T., & Keppenne, C. L. (2008). An adaptive ensemble Kalman filter for soil moisture data assimilation. *Water Resources Research*, *44*(3). https://doi.org/10.1029/2007WR006357

Reichle, R. H., Lannoy, G. J. M. D., Liu, Q., Ardizzone, J. V., Colliander, A., Conaty, A., Crow, W., Jackson, T. J., Jones, L. A., Kimball, J. S., Koster, R. D., Mahanama, S. P., Smith, E. B., Berg, A., Bircher, S., Bosch, D., Caldwell, T. G., Cosh, M., González-Zamora, Á., … Zeng, Y. (2017). Assessment of the SMAP Level-4 Surface and Root-Zone Soil Moisture Product Using In Situ Measurements. *Journal of Hydrometeorology*, *18*(10), 2621–2645. https://doi.org/10.1175/JHM-D-17-0063.1

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., & Toll, D. (2004). The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society*, *85*(3), 381–394. https://doi.org/10.1175/BAMS-85-3-381

Scipal, K., Holmes, T., Jeu, R. de, Naeimi, V., & Wagner, W. (2008). A possible solution for the problem of estimating the error structure of global soil moisture data sets. *Geophysical Research Letters*, *35*(24). https://doi.org/10.1029/2008GL035599

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., & Teuling, A. J. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, *99*(3), 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004

Stoffelen, A. (1998). Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. *Journal of Geophysical Research: Oceans*, *103*(C4), 7755–7766.

Tavakol, A., Rahmani, V., Quiring, S. M., & Kumar, S. V. (2019). Evaluation analysis of NASA SMAP L3 and L4 and SPoRT-LIS soil moisture data in the United States. *Remote Sensing of Environment*, *229*, 234–246. https://doi.org/10.1016/j.rse.2019.05.006

Vogelzang, J., & Stoffelen, A. (2012). Triple collocation. *EUTMETSAT Report*.

Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., Figa-Saldana, J., de Rosnay, P., Jann, A., Schneider, S., Komma, J., Kubu, G., Brugger, K., Aubrecht, C., Züger, J., Gangkofner, U., Kienberger, S., Brocca, L., Wang, Y., … Rubel, F. (2013). The ASCAT soil moisture product: A review of its specifications, validation results, and emerging applications. *Meteorologische Zeitschrift*. https://doi.org/10.1127/0941-2948/2013/0399

Wagner, W., Lemoine, G., Borgeaud, M., & Rott, H. (1999). A study of vegetation cover effects on ERS scatterometer data. *IEEE Transactions on Geoscience and Remote Sensing*, *37*(2), 938–948.

Wang, J., Ge, Y., Heuvelink, G. B. M., & Zhou, C. (2015). Upscaling In Situ Soil Moisture Observations to Pixel Averages with Spatio-Temporal Geostatistics. *Remote Sensing*, *7*(9), 11372–11388. https://doi.org/10.3390/rs70911372

Woodley, W. (2017). *Investigating the upscaling of in situ soil moisture measurements for satellite validation*.

Wu, X., Lu, G., Wu, Z., He, H., Scanlon, T., & Dorigo, W. (2020). Triple Collocation-Based Assessment of Satellite Soil Moisture Products with In Situ Measurements in China: Understanding the Error Sources. *Remote Sensing*, *12*(14), 2275. https://doi.org/10.3390/rs12142275

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., & Meng, J. (2012). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of Geophysical Research: Atmospheres*, *117*(D3).

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

Xu, L., Chen, N., Zhang, X., Moradkhani, H., Zhang, C., & Hu, C. (2021). In-situ and triple-collocation based evaluations of eight global root zone soil moisture products. *Remote Sensing of Environment*, *254*, 112248.

Yilmaz, M. T., & Crow, W. T. (2014). Evaluation of assumptions in soil moisture triple collocation analysis. *Journal of Hydrometeorology*, *15*(3), 1293–1302.

Yilmaz, M. T., Crow, W. T., Anderson, M. C., & Hain, C. (2012). An objective methodology for merging satellite-and model-based soil moisture products. *Water Resources Research*, *48*(11).

Zeng, Y., Su, Z., Van der Velde, R., Wang, L., Xu, K., Wang, X., & Wen, J. (2016). Blending Satellite Observed, Model Simulated, and in Situ Measured Soil Moisture over Tibetan Plateau. *Remote Sensing*, *8*(3), 268. https://doi.org/10.3390/rs8030268

Zhang, N., Quiring, S. M., & Ford, T. W. (2021). Blending Noah, SMOS and In-Situ Soil Moisture Using Multiple Weighting and Sampling Schemes. *Journal of Hydrometeorology*, *1*(aop). https://doi.org/10.1175/JHM-D-20-0119.1

Zhang, R., Kim, S., & Sharma, A. (2019). A comprehensive validation of the SMAP Enhanced Level-3 Soil Moisture product using ground measurements over varied climates and landscapes. *Remote Sensing of Environment*, *223*, 82–94. https://doi.org/10.1016/j.rse.2019.01.015

Zhang, X., Tang, Q., Liu, X., Leng, G., & Li, Z. (2017). Soil Moisture Drought Monitoring and Forecasting Using Satellite and Climate Model Data over Southwestern China. *Journal of Hydrometeorology*, *18*(1), 5–23. https://doi.org/10.1175/JHM-D-16-0045.1

Zhang, X., Zhang, T., Zhou, P., Shao, Y., & Gao, S. (2017). Validation Analysis of SMAP and AMSR2 Soil Moisture Products over the United States Using Ground-Based Measurements. *Remote Sensing*, *9*(2), 104. https://doi.org/10.3390/rs9020104

Zwieback, S., Scipal, K., Dorigo, W., & Wagner, W. (2012). Structural and statistical properties of the collocation technique for error characterization. *Nonlinear Processes in Geophysics*, *19*(1), 69–80.

# Chapter 5 Summary and Conclusions

## 5.1 Overall Conclusions

Subsurface water is liquid water found below the ground surface, including soil water above the water table and ground water below the water table, but does not include water chemically bound to minerals or organic matter. Two important topics related to subsurface water in Oklahoma have aroused the interest of more and more scientists: the wastewater injected into the ground during the oil and gas production and the excess or scarcity of surface soil moisture. This dissertation aims to develop contributions to two important topics for the sustainability of Oklahoma that are related to earthquakes and water resources: (1) the effects of deep underground waste-water injection on triggering regional seismicity and (2) the quantification of state-wide shallow-soil water content as a new tool for multiple applications in reservoir management, water resources, agriculture, natural hazards, water management.

Chapter 2 gathered comprehensive datasets of oil and gas industry-related wastewater injection volumes and earthquakes number with associated event magnitudes from 2006 to 2017 over the entire state of Oklahoma. Data were analyzed to remove those seismic events below the threshold of magnitude completeness. First, we explored the spatiotemporal variability of both processes and concluded that a high correspondence between the two that supports the hypothesis that the recent boom in oil and gas production through unconventional methods with wastewater injection was potentially responsible for the upsurge in the state's seismic activity during 2006 through 2015. Also, a reduction in the number of earthquakes per year, in years 2016 and 2017, reflect either the mitigation policies dictated by OCC or the drop in oil and gas prices or both. Second, a cluster analysis reveals a correlated migration pattern between earthquake occurrences and saltwater injection activity. Following the migration of the weighted wastewater injection

ellipses, weighed epicenters show a predominant northwest direction pattern during the 2007–2017 period. Third, a lagged cross-correlation analysis shows that the number of induced earthquakes in a subsequent month is strongly associated with the previous 25-month cumulative wastewater injection volume and a power law can be fitted between number of quakes and weighted average monthly injection volumes as predictive tool with a coefficient of determination of $R^2 = 0.77$. Using such a relation, several sustainable extraction limits are explored and compared with historic means. Results from these analyses coincide and expand on previously sustainable limits of 5 to 6 million $m^3/month$ to potential combinations that could be associated with the same number of earthquakes within the 25 previous months. A model intercomparison of our parsimonious model, a hydromechanical model, and a seismogenic model reveals a satisfactory performance of the proposed approach and similitude to the hydromechanical model outputs. Nonetheless monthly sharp changes in seismicity could only be more appropriately represented by the seismogenic model. The approach proposed in this manuscript could potentially be regionalized according to the geology of each zone and results could potentially be used as a tool for further model intercomparison experiments and decision making on spatially varied permission distribution and regional industry development to minimize negative consequences of induced earthquakes.

Chapter 3 conducted a cross-evaluation of the accuracy and error characteristics of the most commonly used, yet independent, satellite, model-based, and in situ soil moisture products. Specifically, the assessment of the SMAP L3_SM_P_E (i.e., SMAP), NLDAS_NOAH0125_H (i.e. Noah), and interpolated Oklahoma Mesonet (i.e., Mesonet) soil moisture products at daily and seasonal timescales was conducted over Oklahoma using the triple collocation method. Moreover, their performances were evaluated over different land cover types. The results of this study can potentially provide a new perspective for comparatively assessing multi-source soil moisture

products and a basis for objective data merging to capitalize the strengths of the multi-sensor soil moisture products for the State of Oklahoma and beyond.

Built on the knowledge from Chapter 3, Chapter 4 adopted the TC based least square weighting method to merge soil moisture information from multiple sources, including satellite (SMAP_L3), land surface model (NLDAS-2 Noah), and in-situ measurements (interpolated Oklahoma Mesonet values), and generated blended soil moisture products at a 9-km spatial resolution and daily temporal resolution in Oklahoma. The merged product is validated against an in-situ-based soil moisture product data and shows better performance than both the equal weighting merged and SMAP Level 4 soil moisture products. The resulting combined soil moisture estimate is an improvement over currently available soil moisture products due to its reduced uncertainty and can be used as a standalone soil moisture product with available uncertainty estimates, which will be beneficial for multiple applications in water resources management, agriculture and natural hazards.

### 5.2 Future Research

To study the effects of deep underground waste-water injection on triggering regional seismicity, the study in Chapter 2 gathered comprehensive datasets of oil and gas industry-related wastewater injection volumes and earthquakes number with associated event magnitudes from 2006 to 2017 over the entire state of Oklahoma. The conducted spatiotemporal analyses and proposed parsimonious model represent a novel contribution for prediction, model intercomparison and decision making. Further attention can be devoted to the transferability of the models developed in this study beyond the state of Oklahoma. The generation and validation of the statistical relationships between industry-related wastewater injection volumes and induced

earthquakes over the other parts of the contiguous United States are worth exploring. For example, the 2008–2009 sequence of earthquakes with $M_w$smaller than 3.3 at the Dallas/Fort Worth Airport area were potentially induced by brine disposal associated with the production of natural gas (Frohlich et al., 2011). Another important area for future research would be to investigate some of the potential limitations of this research. For example, results mainly focus on statistical spatiotemporal relationships between wastewater injection volumes and earthquakes number and magnitude. However, some studies connect the induced seismic activity with not only the wastewater injection volumes (IW), but also their depth and injection rates (Keranen et al., 2013; Holland, 2013; Walsh and Zoback, 2015; Hough and Page, 2015; Barbour et al., 2017; Chen et al., 2017; Hincks, Aspinall, Cooke, and Gernon, 2018). Future research can be conducted to improve the model by including the wastewater injection depth and rates.

The second contribution of this dissertation is to provide a multi-sensor soil moisture product for Oklahoma to capitalize the strengths of three soil moisture products from different sources. Further attention can be devoted to the use this multi-sensor soil moisture product as a new tool for multiple applications in reservoir management, water resources, agriculture, natural hazards, water management. For example, the soil moisture products can be used to develop new agricultural drought indices and validate the soil moisture product either from remote sensing or hydrological models. Additional research should focus on the transferability of the TC based least square weighting method used in this dissertation beyond the State of Oklahoma especially the regions lack of high spatial temporal multi-sensor soil moisture product. Another important area for future research is to investigate some of the potential limitations of the doctoral research. For example, one limitation of this study is that the Mesonet product used in our TC intercomparison is interpolated from point-scaled Oklahoma Mesonet to spatial resolution of SMAP (9 km) using

the ordinary kriging method. Future work could use regression kriging approaches and including independent predictors such as soil properties, land cover, topography and precipitation to increase the accuracy of the interpolated Mesonet product.

In general, this dissertation paves the way for further regional adoptions of both the seismic and soil moisture models with the inclusion of new data for robustness of the relationships proposed here. In general the methodological approaches that proved to work in Oklahoma could fundamentally be applied to develop new mathematical and matricial expressions to both earthquake prediction and soil moisture tracking, specially where and when the environmental conditions are similar to those found in Oklahoma between years 2015 and 2019.