**Exploring the Benefits of Downscaled Remote Sensing Soil Moisture for Drought Monitoring**

Zachary T. Leasor1\*, Chen Zhao2, Luyu Liu1, and Steven M. Quiring1

1Department of Geography, The Ohio State University, Columbus, OH, USA

2Synoptic Data PBC, Salt Lake City, UT

\**Corresponding author address:* Zachary Leasor, Department of Geography, The Ohio State University, 1145 Derby Hall, 154 N Oval Mall, Columbus, OH 43210

E-mail: [leasor.4@osu.edu](mailto:leasor.4@osu.edu)

**Target Journal:** Soil Science Society of America Journal

**Abstract**

Improved methods for visualizing soil moisture at a fine resolution can help to bridge the gap between drought monitoring and local impacts. Recent advances in remote sensing technology have provided additional tools for monitoring near-surface soil moisture across the continental US (CONUS). This research leverages remote sensing products to accurately downscale soil moisture and produce national maps of soil moisture at a fine resolution. Remote sensing soil moisture data obtained from the NASA Soil Moisture Active Passive Level 4 (SMAP L4) satellite mission are utilized in this study. To downscale these soil moisture data, ancillary variables such as precipitation, soil texture, vegetation, and other physiographic information are considered to downscale soil moisture to a 1-kilometer resolution. Random forest, a non-parametric machine learning modeling approach, is used to find the best relationship that can determine the optimal soil moisture value at any location. This method produces a downscaled soil moisture product which is significantly different from the original data product. However, when comparing the downscaled products to direct in situ measurements, nearest neighbor interpolations display higher correlations and lower error metrics. Examining the random forest variable importance also provides an opportunity to explore linkages between soil moisture and environmental variables Elevation is the most important variable to consider when downscaling soil moisture while soil texture characteristics are the least important variable to consider. Atmospheric variability influences daily fluctuations in variable importance for temperature, precipitation, elevation, and NDVI data.

**Introduction (need to bolster literature review)**

Drought is natural hazard with impacts that span a broad range of systems. This is evident when considering the many definitions of drought such as meteorological, agricultural, hydrological, ecological, and socioeconomic (Bachmair et al. 2016). The agricultural losses and reductions in water resources that accompany drought typically have negative economic impacts in the U.S. To compound this risk, drought impacts can persist longer than other costly disasters (Wilhite et al. 2007). Definitions of drought are based on indicators such as soil moisture so that the onset, termination, and severity of drought events can be quantified (Svoboda et al. 2002; Quiring 2009). Successful drought mitigation relies on guidance and cooperation among atmospheric scientists, local stakeholders, and policymakers. Improved methods for visualizing drought at a fine scale can help to bridge the gap between drought severity classification and local agricultural impacts.

Recent advances in remote sensing technology have provided additional tools for monitoring near-surface soil moisture across CONUS. However, the resolution of remote sensing soil moisture is often too coarse for some hydrological applications. Insert paragraph describing remote sensing soil moisture.

Insert paragraph describing methods for downscaling soil moisture

This research will leverage previous research in climatology, remote sensing, geographic information systems (GIS), and statistics to visualize national soil moisture characteristics at a fine resolution to improve drought monitoring at local scales. To improve the accuracy of drought monitoring, this research will find an optimal method for downscaling drought indicators by minimizing error and maintaining the spatial representation of soil moisture measurements. Random forest, a machine learning method used across many disciplines (Boulesteix et al. 2012; Gray et al. 2013; Im et al. 2016) will be used to downscale satellite-derived soil moisture data. Geographic characteristics such as precipitation, soil texture, vegetation, land cover, and elevation will also be included in the downscaling process to best relate soil moisture to local conditions. Therefore, spatiotemporal patterns of agricultural drought will be displayed across CONUS at a 1 km resolution. Therefore, our project has provided a proof of concept downscaling approach by utilizing machine learning methods in conjunction with climate and environmental data.

**Data and Methods**

*Soil Moisture Data*

Remote sensing Level 4 (L4) gridded soil moisture data is obtained from the NASA Soil Moisture Active Passive (SMAP) mission. The L4 data are available at a 9 km resolution with daily temporal coverage beginning in March, 2015. Insert paragraph describing SMAP products.

In situ measurements of soil moisture in the top 5 cm of the soil layer will be used to validate the results of the downscaling approach. The validation of SMAP soil moisture will focus on historical data during the 2016 calendar year. Insert paragraph describing the in situ data used in this study.

*Climate and Environmental Data*

High quality temperature and precipitation data from the Parameter-elevation Relationships on Independent Slopes Model (PRISM) climate group will also be used in this study (Daly et al. 2008). PRISM uses local regressions between climate variables and relevant geographic factors such as elevation to interpolate station measurements and produce grids of climatic variables. PRISM grids are provided at 4 km resolution and will be used as supplementary data in the downscaling of remote sensing soil moisture products. Add PRISM details.

Sand, silt, and clay percentages are retrieved from the US Department of Agriculture (USDA) gridded soil survey (gSSURGO) geographic database at a 1km resolution. Add gSSURGO details.

The MODIS instrument onboard the Terra and Aqua satellites provides various environmental products that are related to soil moisture. In this study, the Normalized Difference Vegetation Index (NDVI) data are used as a proxy for vegetation health. Using the most recent data (version 6), all variables are available at a daily time step and a 9 km spatial resolution. All MODIS data will be downloaded from NASA’s Earth Observing System Data and Information System (EOSDIS; <http://reverb.echo.nasa.gov/>). Add NDVI details.

The digital elevation data over CONUS will be determined using the 1 km resolution Digital Elevation Model (DEM) obtained from global 30-arc second elevation dataset from the US Geological Survey (USGS) GTOPO30 (Han et al. 2012). The chosen spatial resolution best matches the resolution of the other data products that will be used in this research. Add DEM details.

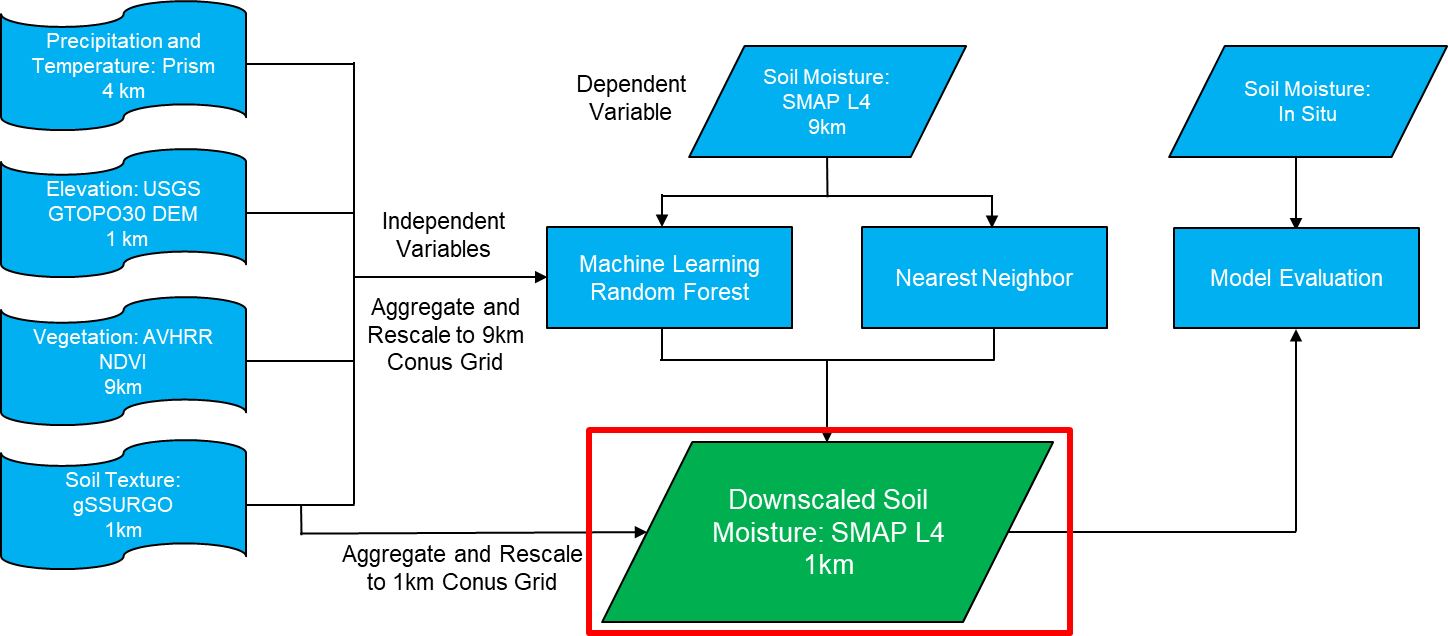
*Downscaling and Interpolation Methods*

Remote sensing has been successfully used in estimating soil moisture over large areas in recent years. Data from sources such as the SMAP mission have provided useful and spatially continuous soil moisture estimates for multiple purposes (Entekhabi et al. 2010). However, the resolution of remote sensing soil moisture is quite coarse for hydrological applications (Brocca et al. 2010; Fang et al. 2016). To overcome this, multiple downscaling methods have been developed in previous research. Using supplementary data that provide information at finer resolutions, downscaling techniques can be classified into 3 main categories: satellite-based methods, geoinformation based methods and model-based methods (Peng et al. 2017). Recent research is utilizing machine learning approaches to downscale satellite soil moisture by linking microwave derivations with optical/IR parameters (Im et al. 2016; Park et al. 2017). However, few studies have attempted to combine satellite parameters with climate and environmental data to downscale satellite soil moisture, and this research aims to fill this gap.

Random forest models are developed using the ‘scikit-learn’ (<http://scikit-learn.org>) package in Python 2.7. For each day, a random forest algorithm is developed by using a bootstrapping approach to subset the data and create 100 decision trees. Include model specifications once training/testing has been completed.

Parameter tuning

A flowchart of the methods that will be used in this project is shown in Figure 1. First, to build the downscaling model, ancillary data are aggregated to a 9 km resolution. Next, the parameters are extracted for every pixel along with SMOS soil moisture. The machine learning model will use SMAP soil moisture as the dependent variable and the 9 km data as predictors. The model developed at the native resolution will then be applied to a 1 km grid to generate a 1 km soil moisture product. This product will then be compared to downscaled soil moisture using a simple nearest neighbor reaggregation method to determine the value of a more robust downscaling method. Both soil moisture products are also be evaluated using in situ soil moisture measurements.



**Figure 1.** Schematic illustrating the data and methods used in this study.

**Results**

*Model Configurations*

Include our optimal configuration based on hyperpameter tuning/cross validation.

*Variable Importance*

The Gini importance calculation quantifies the impurity-based model performance for every independent variable. The feature importance values sum to 1 and provide an opportunity to quantify how heavily the decision tree splits rely on each variable. Elevation is typically the most important input variable. Variable importance changes by day, especially precipitation and vegetation conditions. Include more variables and discuss time series of feature importance based on best performing model configuration.

Include information regarding which daily importance values vary the most.

*Comparison with In Situ and Nearest Neighbor Aggregations*

Include maps with stations and NN/RF maps.

Include boxplots with MAE, RMSD, Correlations.

**Conclusions**

Random forest downscaling provides a 1 km gridded soil moisture product that is significantly different from the original SMAP L4 soil moisture. However, model validation demonstrates that the random forest downscaled soil moisture still performs worse than the original SMAP product when compared to station data. Nearest neighbor interpolations using only the SMAP L4 data display higher correlations and lower error metrics when compared with in situ data. There is also no clear distinction in model performance when using different random forest model configurations. In fact, reduced variable models may perform best.

Examining the random forest variable importance provides an opportunity to examine linkages between soil moisture and ancillary data. Elevation is the most important variable to consider when downscaling soil moisture. Soil texture characteristics are the least important variable to consider. Atmospheric variability influences daily fluctuations in variable importance for temperature, precipitation, elevation, and NDVI data.

**Discussion**

Future research will continue to improve model performance by exploring hyperparameter tuning in random forest models and consider other variables which may improve the downscaled product. The 1 km soil moisture data will also be considered for use in drought mapping and crop yield applications. This research also seeks to produce downscaled soil moisture maps in near real-time. Future research will explore methods for visualizing and hosting downscaled maps on the web.

**References**

Bachmair, S., and Coauthors, 2016: Drought indicators revisited: the need for a wider consideration of environment and society. *Wiley Interdisciplinary Reviews: Water*, **3**, 516–536, https://doi.org/10.1002/wat2.1154.

Boulesteix, A.-L., S. Janitza, J. Kruppa, and I. R. König, 2012: Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, **2**, 493–507, https://doi.org/10.1002/widm.1072.

Brocca, L., F. Melone, T. Moramarco, and R. Morbidelli, 2010: Spatial-temporal variability of soil moisture and its estimation across scales. *Water Resources Research*, **46**, https://doi.org/10.1029/2009WR008016.

Daly, C., M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris, 2008: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, **28**, 2031–2064.

Entekhabi, D., and Coauthors, 2010: The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE*, **98**, 704–716.

Fang, K., H. Li, Z. Wang, Y. Du, and J. Wang, 2016: Comparative analysis on spatial variability of soil moisture under different land use types in orchard. *Scientia Horticulturae*, **207**, 65–72, https://doi.org/10.1016/j.scienta.2016.05.017.

Gray, K. R., P. Aljabar, R. A. Heckemann, A. Hammers, and D. Rueckert, 2013: Random forest-based similarity measures for multi-modal classification of Alzheimer’s disease. *NeuroImage*, **65**, 167–175, https://doi.org/10.1016/j.neuroimage.2012.09.065.

Han, W., Z. Yang, L. Di, and R. Mueller, 2012: CropScape: A Web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support. *Computers and Electronics in Agriculture*, **84**, 111–123, https://doi.org/10.1016/j.compag.2012.03.005.

Im, J., S. Park, J. Rhee, J. Baik, and M. Choi, 2016: Downscaling of AMSR-E soil moisture with MODIS products using machine learning approaches. *Environ Earth Sci*, **75**, 1120, https://doi.org/10.1007/s12665-016-5917-6.

Park, S., J. Im, S. Park, and J. Rhee, 2017: Drought monitoring using high resolution soil moisture through multi-sensor satellite data fusion over the Korean peninsula. *Agricultural and Forest Meteorology*, **237–238**, 257–269, https://doi.org/10.1016/j.agrformet.2017.02.022.

Peng, J., A. Loew, O. Merlin, and N. E. C. Verhoest, 2017: A review of spatial downscaling of satellite remotely sensed soil moisture. *Reviews of Geophysics*, **55**, 341–366, https://doi.org/10.1002/2016RG000543.

Quiring, S. M., 2009: Developing objective operational definitions for monitoring drought. *Journal of Applied Meteorology and Climatology*, **48**, 1217–1229.

Svoboda, M., and Coauthors, 2002: The drought monitor. *Bulletin of the American Meteorological Society*, **83**, 1181–1190.

Wilhite, D. A., M. D. Svoboda, and M. J. Hayes, 2007: Understanding the complex impacts of drought: A key to enhancing drought mitigation and preparedness. *Water Resour Manage*, **21**, 763–774, https://doi.org/10.1007/s11269-006-9076-5.