Downscaling Satellite-Based Soil Moisture Measurements with Transfer Learning and Neural Networks

Mitchell Thomas

mlt2177

EAEEE4000- Machine Learning for Environmental Engineering

**Abstract**

Satellite-based soil moisture measurements such as NASA's SMAP are valuable tools to measure soil moisture on a global scale. These measurements are essential to monitor drought vulnerability, crop health, wildfire risk, and evapotranspiration fluxes, especially in regions with poor in situ data. However, satellite soil moisture's coarse resolution limits applications past large-scale studies. In this paper, we use a transfer learning neural network approach to combine ancillary high-resolution satellite, topographic, and hydrologic data with NASA's SMAP soil moisture product in order to develop a high resolution (40 meter) soil moisture dataset over the contiguous United States at 10-day frequency.

1. **Introduction**

Soil Moisture is broadly defined as the measurement of water stored in the soil. Exact definitions vary, but generally the measurement of soil moisture refers to water stored in the unsaturated zone, or the area between the surface and the groundwater table. Often this measurement is further classified into *surface* soil moisture and *subsurface*, or *root* *zone* soil moisture, referring to the areas above and below the first few centimeters [1].

Methods of measuring soil moisture often fall into one of three categories: in situ, modeled, and satellite based. In situ soil moisture has often been seen as the gold-standard for soil moisture measurements, and when measurements are performed correctly, it can create the most accurate soil moisture profile at different depths. However, in situ data is highly vulnerable to human error and is highly susceptible to local hydrological, topographic, or soil variation. Modeled data has the potential to provide continuous soil moisture measurements at high scale. However bias and error in land surface models and significant region to region variation in modeled data limit the resolution and performance of modeled soil moisture.

Significant effort in recent years has been made to downscaling these satellite-based soil moisture products, in recent years increasingly done with machine learning. Alemohammad et al. [2] used

Meyer et al. [3] finds that Sentinel-1 backscatter in the VV and VH bands shows a strong correlation with Soil Moisture at high resolutions (20-400m)

1. **Methods**
2. **Data Preparation**

A variety of satellite, in situ, and modeled data was used in this study. Satellite and modeled data was processed and downloaded in Google Earth Engine, while in situ soil moisture was retrieved from the International Soil Moisture Network (<https://ismn.earth/>). This data is summarized in table 1.

Table 1: Satellite, Modeled, and In Situ data used in this study

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data source name | Studied Variable | Method of Retrieval | Data Type | Available range |
| In Situ Soil Moisture | Surface Soil Moisture | International Soil Moisture Network | ~100 Point based, hourly measurements across United States. | Varies by station |
| SMAP Soil Moisture | Surface Soil Moisture | Google Earth Engine | 10 km, every 3 days globally | 2015-Present |
| Sentinel-2 | Surface Reflectance at 13 wavelengths | Google Earth Engine | 10-20 m, average 5 day overpass | 2014-Present |
| Sentinel-1 | Surface Backscatter at VV and VH polarization, angle of capture. | Google Earth Engine | 10 m, average 6 day overpass | 2015-Present |
| USGS National Land Cover Database | Landcover classification over United States | Google Earth Engine | 30m raster | 2019 |
| SRTM mTPI | Multi-scale Topographic Position Index (position compared with local elevation) | Google Earth Engine | 30m raster | Derived from SRTM mission in 2000 |
| OpenLandMap Soil Water Content | Machine Learning estimated field capacity of soil | Google Earth Engine | 200m raster | N/A |

Data is processed and projected in two different ways based on the data source, the type of training done, and the spatial scale of the model.

Sentinel-2 is filtered by pixelwise cloud probability. Only pixels with a probability of cloud < 60% are included in the model. Normalized Difference Vegetation Index (NDVI), a popular remote sensing vegetation index, is computed from bands 8 and 4 and included in the study. Sentinel-2 bands are further processed by log-scaling in order to create a more gaussian distribution.

For training a model at 10km (the scale of SMAP), all other satellite and modeled data source (all bands of Sentinel-1, all bands of Sentinel-2, Landcover, mTPI, and Field Capacity) are reprojected down to 10km resolution from their native resolution. For all variables the mean over the 10km pixel is taken except landcover, where the most often classification is used. For training the final model at 40m, all satellite data sources (including SMAP) and reprojected to 40m resolution using the same logic.

Data for all sources is then selected between 2019-01-01 and 2022-01-01, and a 10-day average is computed for this three-year range to account for missing data in Sentinel-1 and Sentinel-2 and cloudy overpasses in Sentinel-2.

In situ data was downloaded from the International Soil Moisture Network, and for each site just the surface soil moisture was selected. If a site had multiple gauges, the average of these gauges was taken for the site. Significant missing data was observed, and therefore sites were only considered for study if more than 80 10-day data points over the selected three years were available. Overall, this led to \_\_\_\_ gauges across the contiguous United States with a total of \_\_\_\_\_ trainable points.

1. **Model Architecture**

Six model architectures were examined in this study and compared, three of which employed transfer learning to increase training points.

1. Model 1: Simple feed forward vanilla neural network trained on 40m satellite data (including SMAP)
2. Model 2: Simple feed forward vanilla neural network trained on 10km satellite data and SMAP
3. Model 2: Transfer learning vanilla neural network. A neural network is trained on 10km satellite and model data to predict SMAP with ancillary data. The output of this network is then fed into
4. Model 3:

Future work could explore the use of convolutional neural networks. These were not considered in this study due to the difficulty of considering spatial information at different scales – i.e. convolutional relationships would

Alternative to vanilla neural networks were considered but due to model limitations

1. **Experiment Design**
2. **Results**
3. **Discussion**

**Conclusion**

[1] Y. H. Kerr, “Soil moisture from space: Where are we?,” *Hydrogeol J*, vol. 15, no. 1, pp. 117–120, Feb. 2007, doi: 10.1007/s10040-006-0095-3.

[2] S. H. Alemohammad, J. Kolassa, C. Prigent, F. Aires, and P. Gentine, “Global Downscaling of Remotely-Sensed Soil Moisture using Neural Networks,” Global hydrology/Remote Sensing and GIS, preprint, Feb. 2018. doi: 10.5194/hess-2017-680.

[3] R. Meyer *et al.*, “Exploring the combined use of SMAP and Sentinel-1 data for downscaling soil moisture beyond the 1&thinsp;km scale,” *Hydrology and Earth System Sciences*, vol. 26, no. 13, pp. 3337–3357, Jul. 2022, doi: 10.5194/hess-26-3337-2022.