U.S. DEPARTMENT OF ENERGY | Office of Science

DOE/SC-ARM-23-018

Real-Time, Simultaneous Soil Water Content and Meteorological Data Measurement to Support TRACER over Harris County, Texas
Field Campaign Report: Part I

MH Young
T McKinney
H Dashtian
B Young

April 2023

ARM

DISCLAIMER

DOE/SC-ARM-23-018

Real-Time, Simultaneous Soil Water Content and Meteorological Data Measurement to Support TRACER over Harris County, Texas Field Campaign Report: Part I

MH Young
H Dashtian
T McKinney
B Young
All at Bureau of Economic Geology, University of Texas at Austin

April 2023

How to cite this document:

Young, MH, H Dashtian, T McKinney, and B Young. 2023. Real-Time, Simultaneous Soil Water Content and Meteorological Data Measurement to Support TRACER over Harris County, Texas Field Campaign Report: Part I. U.S. Department of Energy, Atmospheric Radiation Measurement user facility, Richland, Washington. DOE/SC-ARM-23-018.

Work supported by the U.S. Department of Energy,
Office of Science, Office of Biological and Environmental Research

Acronyms and Abbreviations

ARM	Atmospheric Radiation Measurement
CYGNSS	Cyclone Global Navigation Satellite System
HCFCD	Harris County Flood Control District
IOP	intensive operational period
NASA	National Aeronautics and Space Administration
SMAP	Soil Moisture Active Passive
TCEQ	Texas Commission on Environmental Quality
TRACER	Tracking Aerosol Convection Interactions Experiment
TxSON	Texas Soil Observation Network

Contents

Figures

1.0 Summary

The main purpose of this project was to provide ground-truth and satellite-based soil water content data in the Houston, Texas, area to support the U.S. Department of Energy Atmospheric Radiation Measurement (ARM) user facility's 2022 field campaign, the Tracking Aerosol Convection Interactions ExpeRiment (TRACER). This involved installing four soil monitoring stations at different locations in the Houston area, two of which were part of the ARM facility and two were ancillary. The two stations associated directly with ARM were installed alongside other facilities managed and maintained by the TRACER research team during the intensive operational period (IOP), one located at the La Porte, Texas airport and the other near Guy, Texas. Data from these stations were assimilated with similar data collected by the Harris County Flood Control District (HCFCD) to improve the spatial coverage of the real-time monitoring network. The ground-truth data were compared to the satellite-based (National Aeronautics and Space Administration [NASA]'s Soil Moisture Active Passive [SMAP] mission) data in the TRACER area of interest, and analyzed further to nowcast gridded data over the Houston area. The nowcasted, gridded data product was made available to TRACER researchers for use in climate and land-atmosphere interaction modeling that can help understand formation and persistence of convective storms in dense urban areas, and predict possible environmental events such as floods.

This project was subdivided into several distinct parts: field collection of soil water content and meteorological data at select locations around Houston, and use of these data for improving the quality and timeliness of gridded soil water content data products. The monitoring stations were installed in the area east, south, and west of the city of Houston, with two stations installed each in 2021 and 2022 (the different time lines were in part due to project delays stemming from the COVID-19 pandemic). Data were collected every five minutes. Hourly averaged data were streamed to the University of Texas at Austin, quality-checked, and uploaded in bulk to password-protected folders for the Texas Commission on Environmental Quality (TCEQ) and TRACER collaborators. In general, uptime of these stations was excellent, though communications issues were noted at one station.

We tested two satellite products that could provide gridded soil water content data, including the Cyclone Global Navigation Satellite System (CYGNSS) and SMAP system, specifically the Level 4 (L4) product, both available through NASA. These systems were tested against one another and against data from a long-term soil monitoring network operated by the University of Texas at Austin, known as the Texas Soil Observation Network (TxSON). The results showed that SMAP was more stable and reflected the variability of ground conditions better than CYGNSS. For these reasons, we chose to pursue use of SMAP in the Houston area of interest.

As with nearly all satellite products available to the scientific research community, SMAP data release contains 2-5 days of latency, which is the time delay between satellite overflight and data availability. This latency limits operationalization of these data for some types of geologic hazards, especially those involving floods and convective cell generation (the subject of TRACER). In this project, we tried to address some of the challenges of streaming, assimilating, and using real-time soil water content data from various sources. The activities to accomplish this project, in addition to installing the soil monitoring stations described above, include assimilating and harmonizing soil water content data from two different monitoring networks, quality-checking the data, downloading satellite-derived, gridded soil water data for the study area, and using these data for nowcasting soil water content in real time. The nowcasting tool

reduced/eliminated the latency in the SMAP data, with a very small level of error (~0.003 m^3/m^3) (see Figure 1). SMAP gridded products through our approach are now available every six hours, and are being transmitted to researchers at the Pacific Northwest National Laboratory, where process-based regional climate modeling is being undertaken.

Significant progress in assimilating ground-based and satellite-based soil monitoring data was shown, and could be used for various environmental hazards such as flood/drought monitoring and risk mitigation. The existence of regional soil monitoring networks in Texas and across the U.S. could provide soil water content in real time, in some cases at depths up to 100 cm. Tied together, these data could be used for various operational (such as calibration and validation of satellite-derived data) and application purposes (such as real-time forecasting of flash floods). Assimilating various soil water content and meteorological data from various regional networks and satellite systems for nowcasting and forecasting regional gridded soil water products is making progress, but significant challenges related to spatiotemporal variability and data accessibility, along with real-time training and scoring (error estimation) of data-driven models and machine learning model structures, remain. Little progress has been made in the use of real-time, large-scale soil moisture observations (both ground-based and satellite-derived) within the context of land-atmosphere interactions modeling.

Opportunities abound to advance the science and practice of using large-scale soil water monitoring for the sake of improving Earth system monitoring and modeling, land-atmosphere interactions, nowcasting and forecasting. High-resolution, nowcasted and forecasted soil water data over urban areas is an important parameter that is currently not available for any regional, national, or global scale.

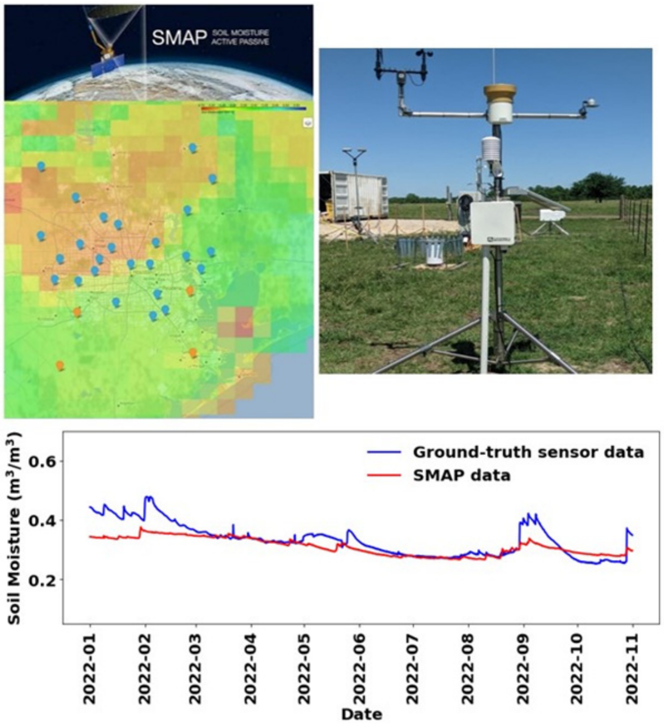

Figure 1. Houston, Texas area of interest for this project (upper left), photograph of deployed soil water and weather station at the Guy, Texas site (upper right), and example showing results of ground-truth and nowcasted SMAP water content data (bottom).

2.0 Results

2.1 Station Installation and Data Collection and Streaming

On May 7, 2021, the first ARM-specific station was installed, this one at the La Porte, Texas airport site. The facility was added immediately to the TxSON network. This station records and monitors soil water content, electrical conductivity, and soil temperature at four depths (5, 10, 20, and 50 cm), as well as several above-ground parameters, including temperature, precipitation, wind speed and direction, air temperature and relative humidity, and solar radiation. The installation and data collection in these stations follow the same procedure of TxSON. Data from each sensor was collected every five minutes. Figure 2 shows the station at La Porte.

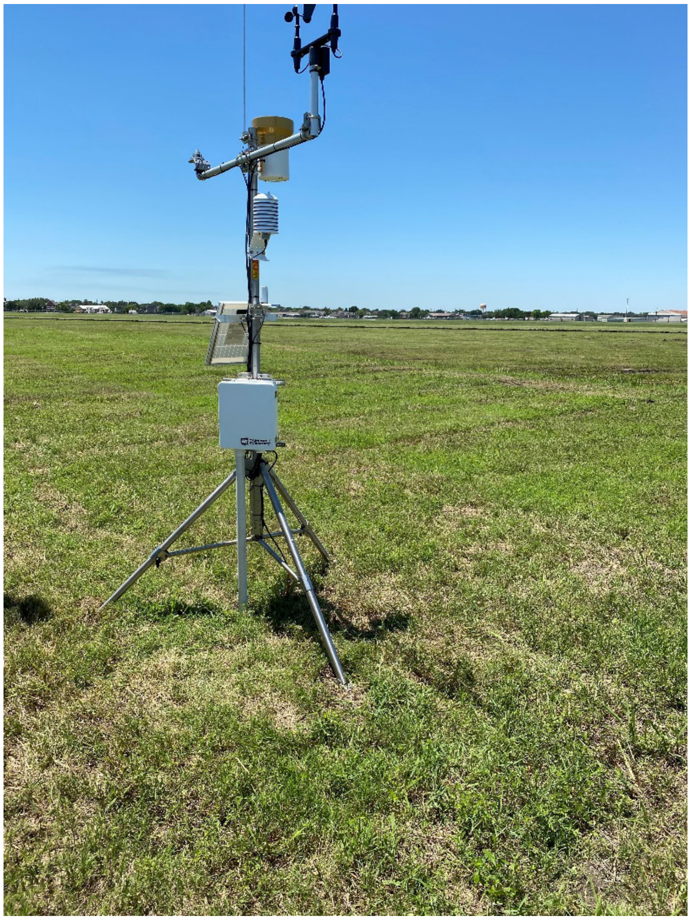

Figure 2. Soil water content monitoring station with additional meteorological monitoring equipment located in La Porte, Texas. Instruments are similar to Figure 2 in part II of this ARM field campaign report, DOE/SC-ARM-23-019.

Data were collected using the LoggerNet software (Campbell Scientific, Inc., Logan, Utah), which communicates with the Campbell Scientific dataloggers used in this project and pulls data from the loggers to the main TxSON Hub ("server") at the Bureau of Economic Geology at the University of Texas at Austin. Data were stored on the individual loggers continuously, and uploaded hourly. Data transfer was done in real time. Data were made available to TRACER researchers through a password-protected API. The final data were published and are available to the public (Dashtian and Young 2023). Data have been downloaded numerous times already by other researchers.

The interactive plots of soil water content data for La Porte, Texas are available at: https://coastal.beg.utexas.edu/soilmoisture2/data/CR300_19/SM_P_Plot_All.html

2.2 Comparison of SMAP and In Situ Data

We also compared the in situ soil water data with the SMAP L4 data. We developed source code to download and compare SMAP data for each grid to the in situ data collected from the geographically closest station. For each collection time, *t*, we use the following equation to calculate the depth-averaged water content at each station from ground surface to 75-cm depth:

$$SM_m(t) = \frac{1}{\sum_{i=1}^{i=4} h_i} \sum_{i=1}^{i=4} SM_i(t) * h_i,\ h_1 = 5,\ h_2 = 10,\ h_3 = 20,\ h_4 = 50$$

SMAP L4 water content data are expressed as the average over 100-cm depth over a 9x9-km^2 area. Figure 3 compares the ground-based, sensor-derived and the satellite-based water content in the Houston area. The data show good correspondence between ground-based and satellite-derived water content data. The response to the precipitations is more obvious (with higher fluctuations) in ground-based data.

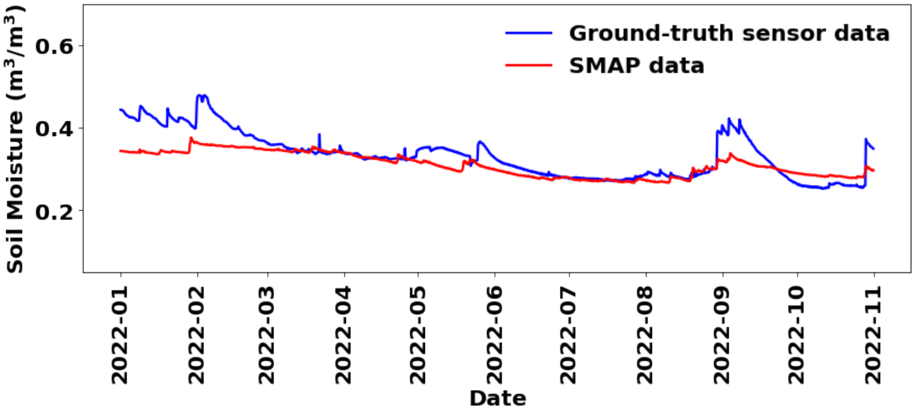

Figure 3. Example of a comparison of soil water content and the SMAP L4 grid data over an area southwest of downtown Houston.

2.3 Nowcasting SMAP Using Deep Learning Approach

We have developed a machine learning algorithm to assimilate in situ and satellite-derived soil water content data, and to train the model to nowcast soil water maps at the same scale as the satellite. Usually, satellite-derived data are not available for the current time, but rather are delayed 2-5 days between the overflight and publication. This is known as latency, and it restricts applications of satellite-derived soil water content data in the forecasting of environmental events, such as flash floods. The nowcasted data includes "n" estimated rasters over the Houston area, corresponding to the missing (not yet available) SMAP data. The numerical processing essentially brings latent SMAP data to near-real time, which allows it to be operationalized for flood risk estimates, or other environmental risks. The numerical approach and the raw data from the in situ stations were shared with TRACER collaborators from the

Pacific Northwest National Laboratory, who were tasked with process-level atmospheric modeling. The nowcasted data over Houston are being published as an online tool (https://coastal.beg.utexas.edu/soilmoisture2/TRACER_SM_P.html).

Figure 4 (a) shows the front end of the web tool, which allows the user to toggle on/off specific layers as time approaches current. Figure 4 (b) shows an example of a water content raster over the Houston area, the predicted (nowcasted) raster using our method, and the absolute error of estimation. Errors below 0.01 m^3/m^3 highlight the accuracy of the method. Figure 4 (c) shows a comparison of actual and nowcasted water content time series, throughout most of 2022, at grid #420 in our study area.

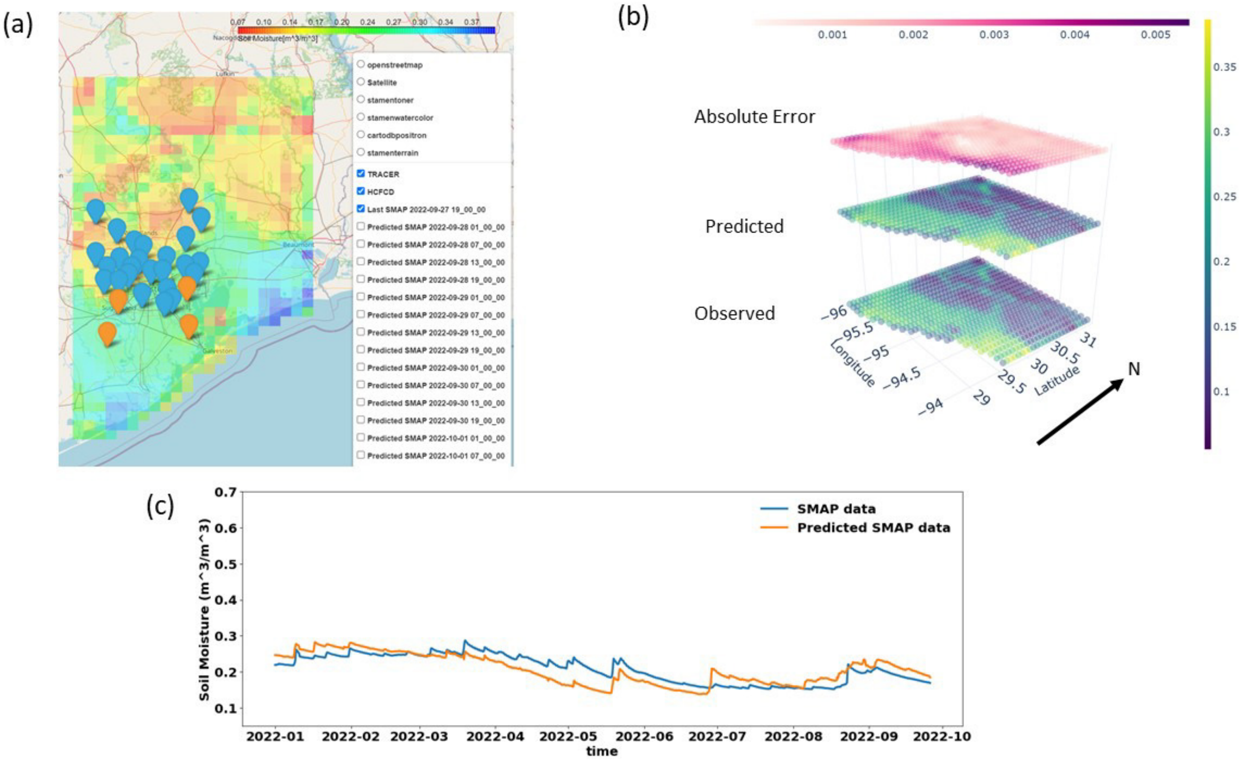

Figure 4. a) The front end of the web tool for nowcasting soil water rasters over the Houston area; b) example of observed raster from SMAP L4, predicted (nowcasted) and its associated absolute error raster over Houston; c) comparison of machine learning nowcasted soil water content data and SMAP-derived data for a specific 9-km-x-9-km grid in the study area.

3.0 Publications and References

Dashtian, H, and MH Young. 2023. Soil Water Content and Meteorological Data to Support Tracking Aerosol Convection Interactions ExpeRiment (TRACER) over Harris County, Texas. Texas Data Repository, V1, https://doi.org/10.18738/T8/FN3RWZ

Dashtian, H, MH Young, D Niyogi, BE Young, and T McKinney. 2023. "Nowcasting NASA Soil Moisture Active Passive Data Using In Situ Sensor Data and Deep Learning." To be submitted to *Geophysical Research Letters*.

4.0 Lessons Learned

One main lesson learned is that, usually, climate and atmospheric modeling groups tend to use homogenous and fixed soil water content data or statistically derived, predefined soil water content data in their studies. Ground-based monitoring stations, similar to that used in this project, are expensive to maintain and require operational experience. Satellite data, such as SMAP, can be computationally expensive and complex to download and subset for specific regions; thus, modeling groups usually use predefined soil water content data in their modeling, instead of real-time data. Providing real-time soil water content data using ground-based stations is necessary for accurately forecasting numerous land-surface related events that require soil water data as input. There is no currently available nowcast product that provides soil water content at regional or national scale; thus, data assimilation and use of numerical, statistical, and machine/deep learning methods are necessary to gap-fill soil water content data in both space and time.

ARM

www.arm.gov

U.S. DEPARTMENT OF
ENERGY

Office of Science