A Deep Learning Approach to GNSS-R: Predicting Soil Moisture with Delay-Doppler Maps

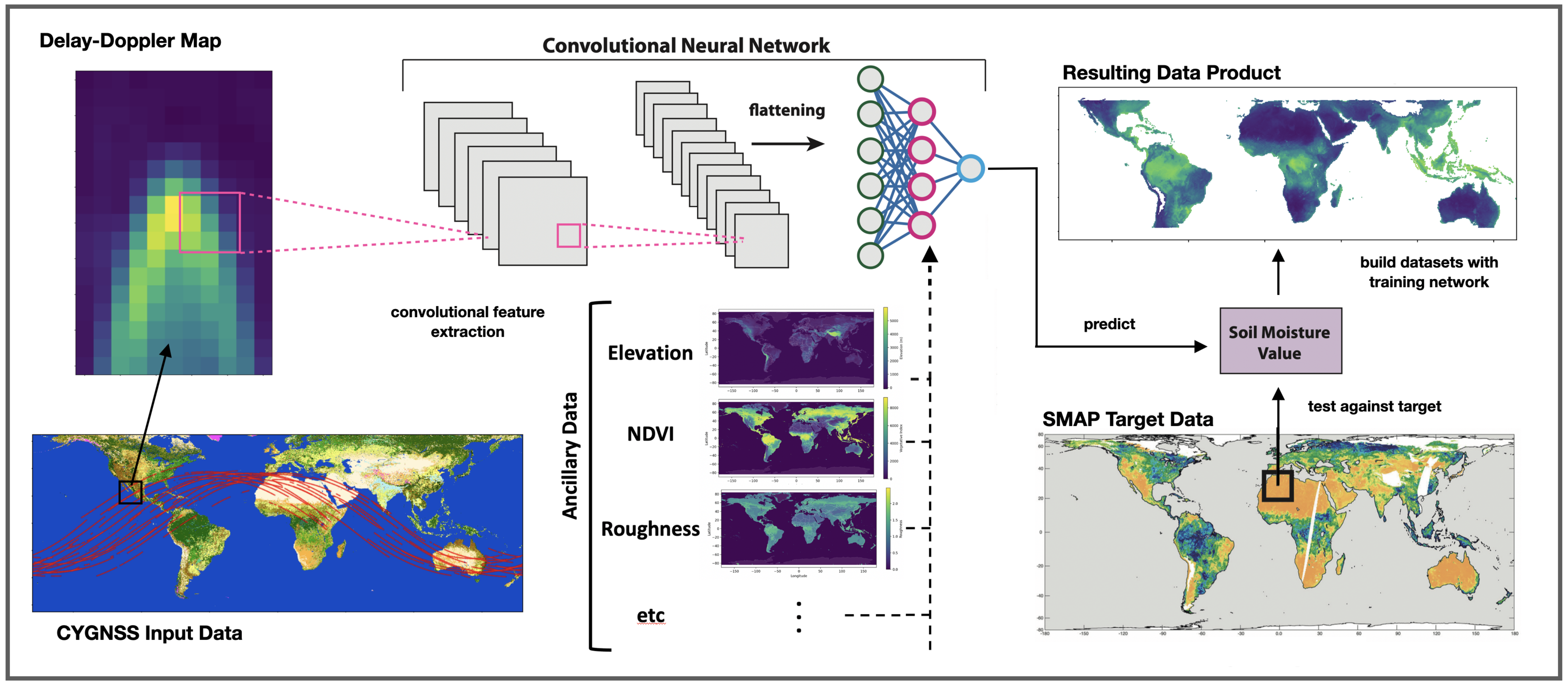
Dr. T. Maximillian Roberts†, Ian Colwell, Dr. Rashmi Shah, Dr. Stephen Lowe, Dr. Clara Chew‡

Jet Propulsion Laboratory, ‡University Corporation for Atmospheric Research

Jet Propulsion Laboratory
California Institute of Technology

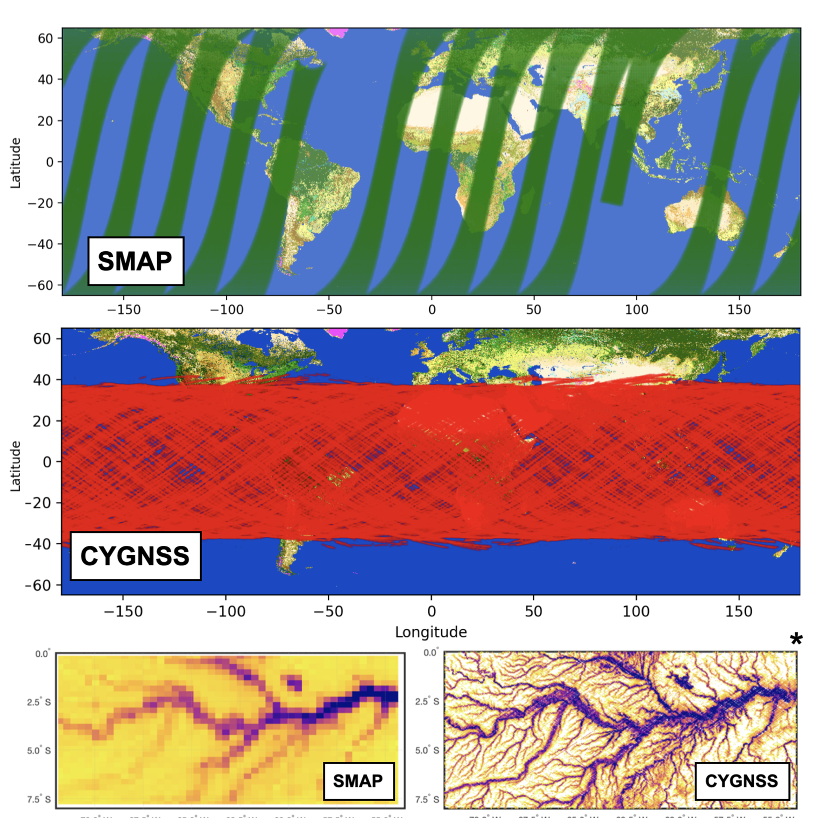
Abstract

GNSS reflection measurements can be calibrated with data from SMAP to yield estimates of soil moisture with increased spatiotemporal resolution, useful to certain hydrological/meteorological studies. Current approaches which use simple models of the relation between the DDM (delay-Doppler map) and soil moisture which can fail in certain regions. Complex information contained in the complete 2D DDM could help in these regions, and can be extracted through the application of deep learning based techniques. This approach simultaneously provides the ability to incorporate additional contextual information from external datasets. Our work explores the data-driven approach of convolutional neural networks (CNNs) to determine complex relationships between the reflection measurement and surface parameters. CYGNSS DDMs were aligned with SMAP soil moisture values and ancillary datasets, a network was developed and trained with these measurements, the results of which are analyzed and compared to existing global soil moisture products.

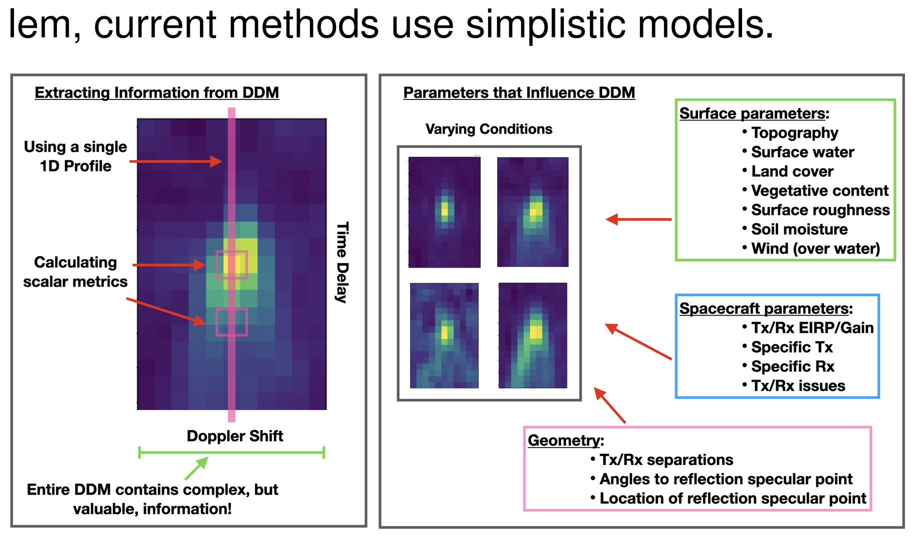


Motivation and Concept

A comparison of the spatial and temporal resolutions attainable with SMAP vs CYGNSS, globally for a single day, and composite over the Amazon river.



CYGNSS DDMs are calibrated with SMAP measurements, and are used to estimate soil moisture. The full interpretation of DDMs is a complex problem, current methods use simplistic models.



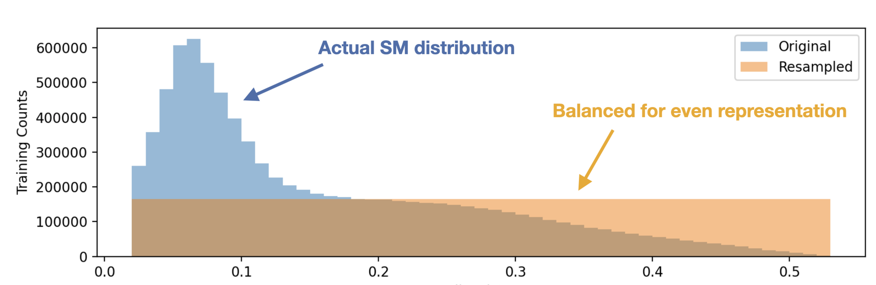
CNNs extract complex structural features from the DDMs while integrating contextual information to build data-driven models of the reflecting conditions.

Dataset Development

Parameters which influence the DDM were determined and corresponding datasets were aggregated (see below table). Each dataset was studied to understand the characteristics, outliers, and indicators of problematic values.

Data	Rate	Source	Base Resolution	Used Resolution
Primary Input: CYGNSS DDM	Per DDM	CYGNSS	0.5-7 km	0.5-7 km
Ancillary Inputs:				
CYGNSS SC Info/PRN Number	Per DDM	CYGNSS	DDM-scale	DDM-scale
Angle/Range to Reflection	Per DDM	CYGNSS	DDM-scale	DDM-scale
Gains/EIRP	Per DDM	CYGNSS	DDM-scale	DDM-scale
Latitude/Longitude	Per DDM	CYGNSS	DDM-scale	DDM-scale
Surface Elevation/Slope	Static	SMAP L1-L3 Ancillary	1 km	3 km
NDVI	Daily	SMAP L1-L3 Ancillary	1 km	3 km
Stem Factor/VWC	Static/Daily	Calculated	1 km	3 km
Land Cover	Static	GlobCover	1 km	3 km
Surface Roughness	Static	SMAP L1-L3 Ancillary	1 km	3 km
Precipitation	Daily	SMAP L1-L3 Ancillary	36 km	36 km
Surface Water	Static	Pekel	30 m	3 km
Target: SMAP SM Value	Daily	SMAP	36 km	36 km

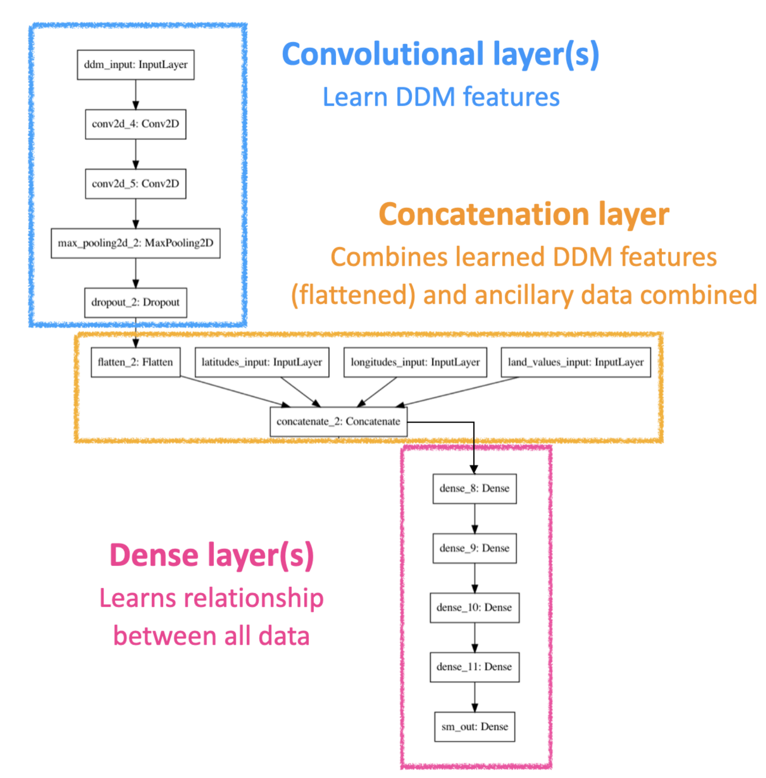
Data from different sources are “aligned", spatially and temporally, to fall within the same 3 km EASE grid cell on the same day. These training samples are compiled into a database with features designed for efficient processing.



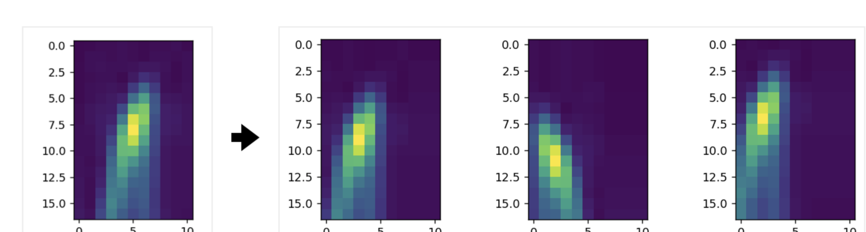
Training sets are filtered, standardized, and "balanced" to optimize for training. The above shows the balancing process, where some values are undersampled and others are oversampled to create an even distribution. Catergorical scalar data (like land type) are encoded to “one-hot" array inputs.

Neural Network Development

Development of the neural network was broken in two; developing a CNN specifically tuned for processing DDMs, and building a complete network for prediction of soil moisture.



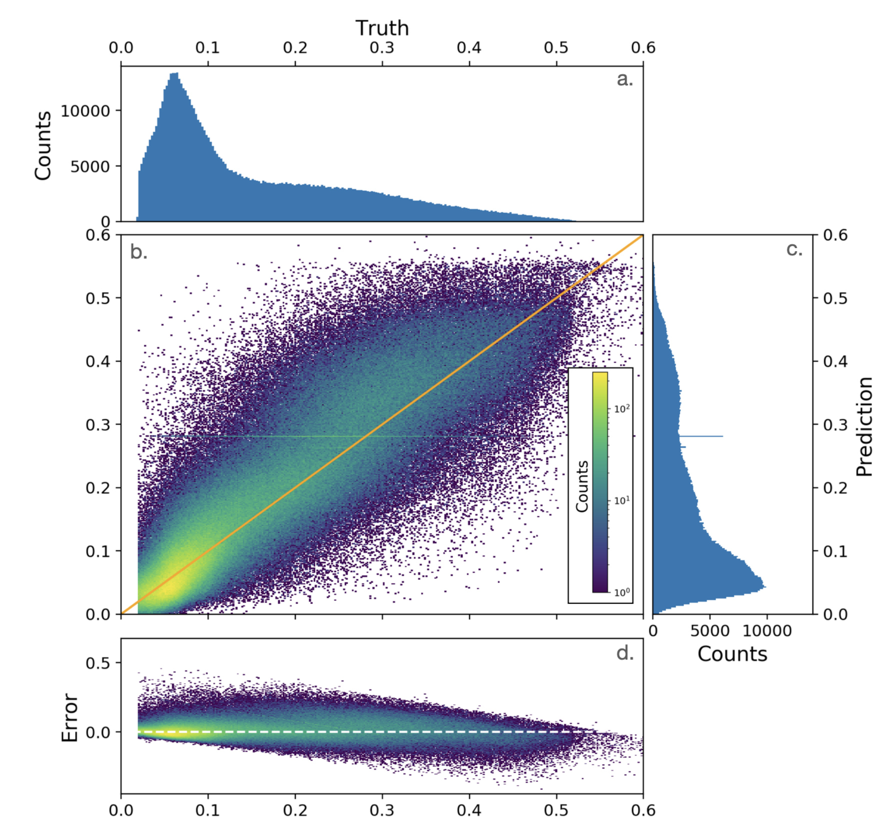
Optimization of DDM-specific CNN was studied with the toy-problem of land-type classification (blue layers only); DDM only input, land value is the target. Common CNN architectures were explored, and DDM augmentation and resolution enhancement was tested.



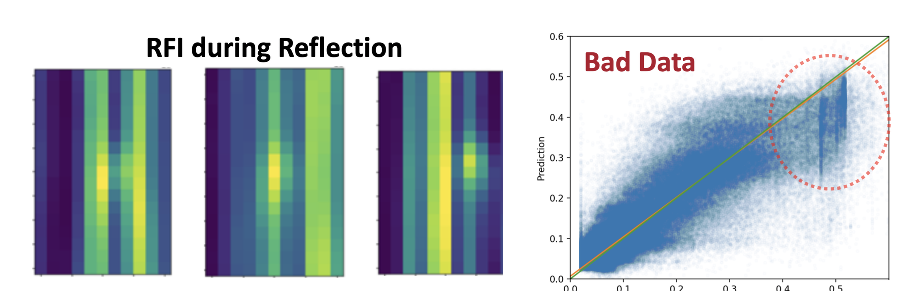
A complete network (blue, yellow, and pink layers) architecture able to accept various ancillary inputs was developed to estimate soil moisture, integrating the DDM-tuned CNN.

Analysis and Results

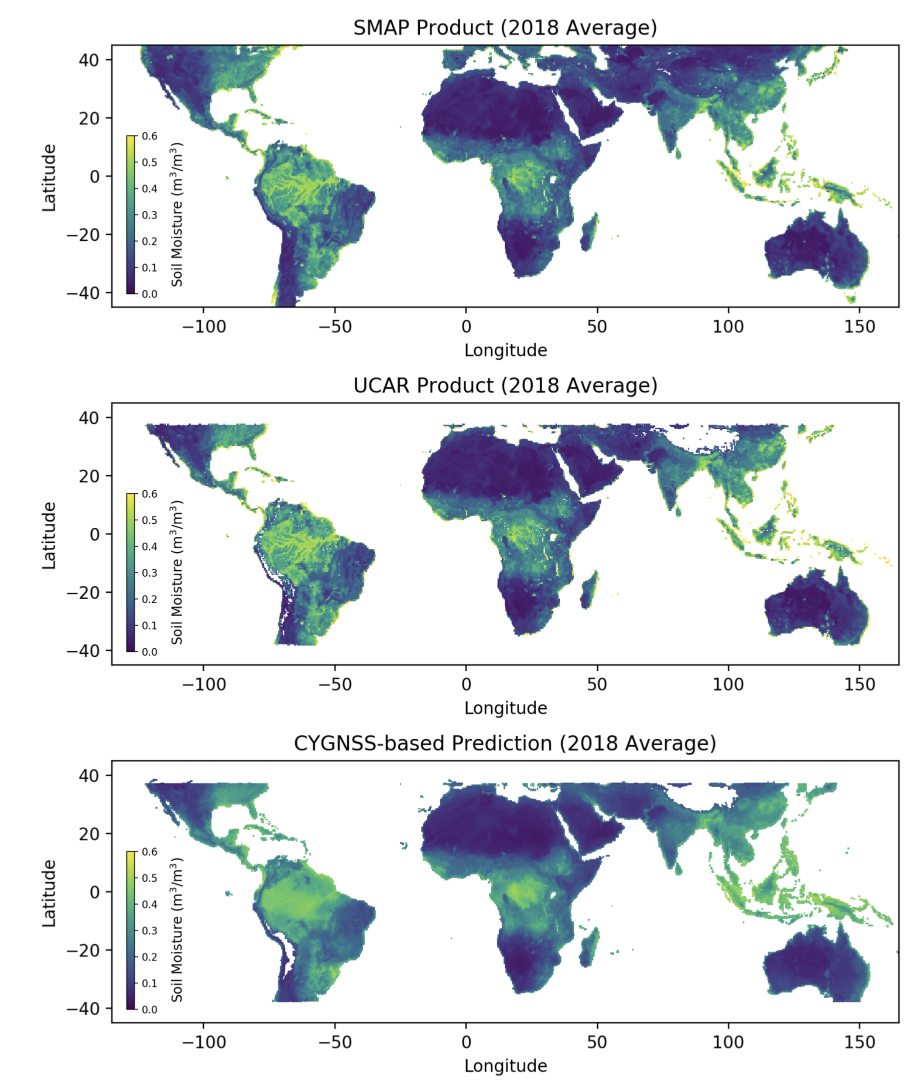
Network performance is established using unseen data split from the original dataset composed of randomly distributed points in space and time. Passing these samples through the network, a strong correlation between the predictions and targets (Pearson coefficient of 0.89) demonstrates the technique’s potential. Dataset preprocessing, training, and analysis is repeated iteratively.



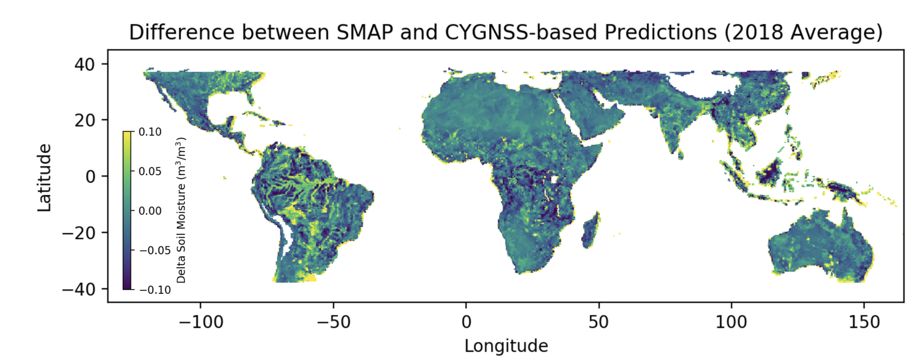
Bias toward underestimation of high SM values can be seen, likely a result of the balancing process. Other issues, such as RFI and problematic SMAP data are also revealed through analysis.



Network predictions are compared on a global scale to existing SM product of SMAP and UCAR averaged to 36 km EASE grid cells for the entire year of 2018. Qualitative comparison shows overall trends in strong agreement. However, areas with expected high SM content display less detailed structure.



Differences between predictions and the UCAR product are seen to strongly correlate with problematic SMAP quality flags removed from training data. Biases are likely created by removing samples over high surface water fraction.



Next Steps

This work has shown CNNs can be used to interpret DDMs directly with opportunity for significant improvement. Immediate follow on studies will further refine dataset filtering and spatiotemporal averaging, and add valuable, missing ancillary datasets (such as “distance-to-water"). More advanced work will create “ensembles" of networks for regional prediction, implement vector inputs for ancillary data (input region of values, not average), and include in situ measurements in training as “high value" targets. Futhermore, this concept is generalizable to other surface retrievals and product development such as “freeze-thaw”, flood/inundation, and water masks.

†Contact email: t.maximillian.roberts@jpl.nasa.gov

*Modified from Chew et al. 2019.