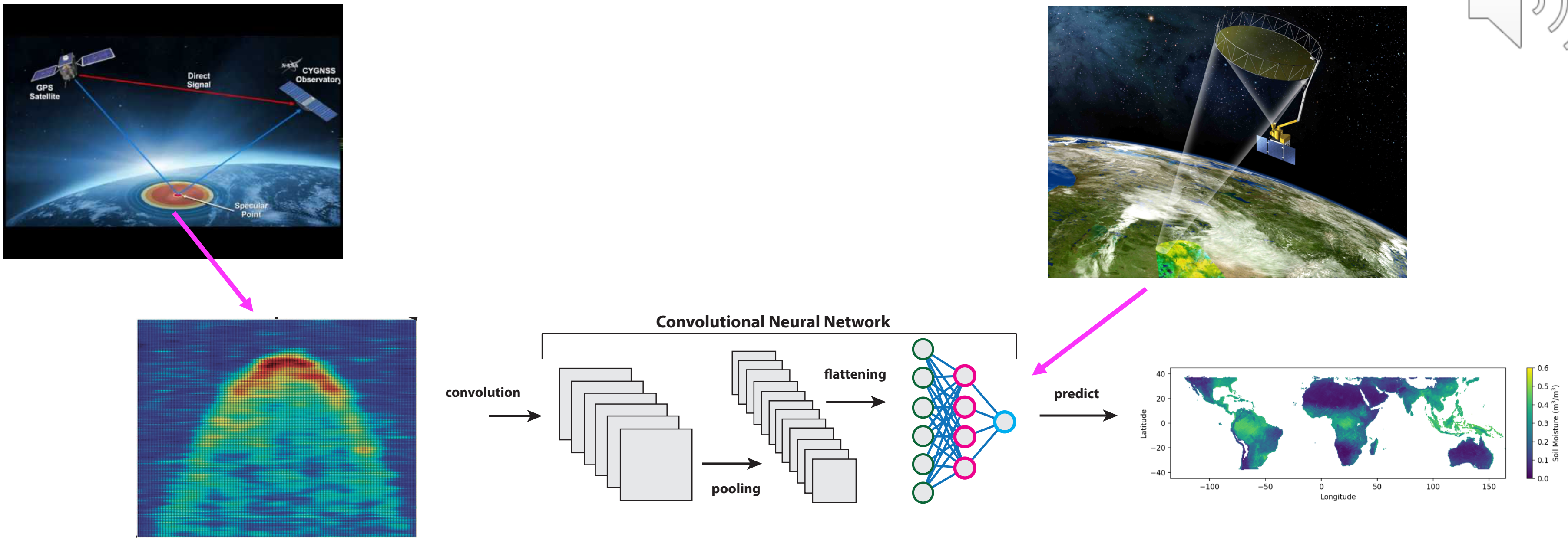


Using Convolutional Neural Networks and GPS to Measure Soil Moisture

A Deep Learning Based Approach to Soil Moisture Estimation Using GNSS-Reflectometry Data

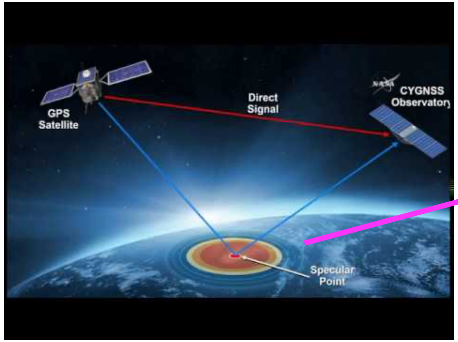
Principal Investigator: T. Maximillian Roberts 335F
Co-Is: Ian Colwell 174B, Rashmi Shah 335A, Steve Lowe 335B
Program: Spontaneous Concept

Assigned Presentation # RPC-064

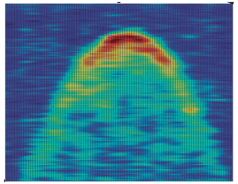
Jet Propulsion Laboratory
California Institute of Technology

Tutorial Introduction

GNSS-R



DDM



Time Delay

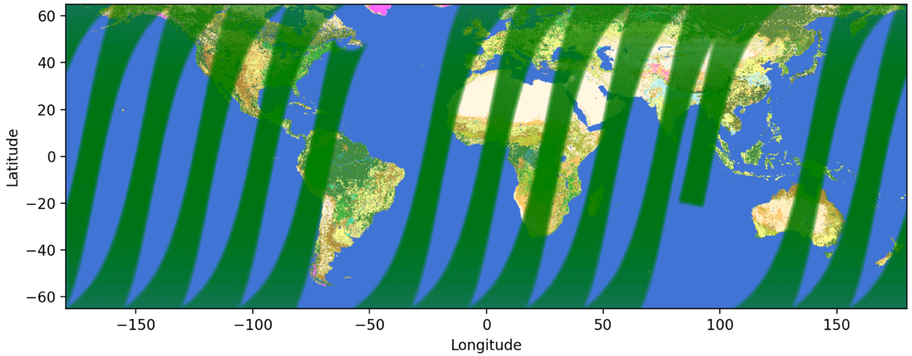
Doppler Shift

Abstract

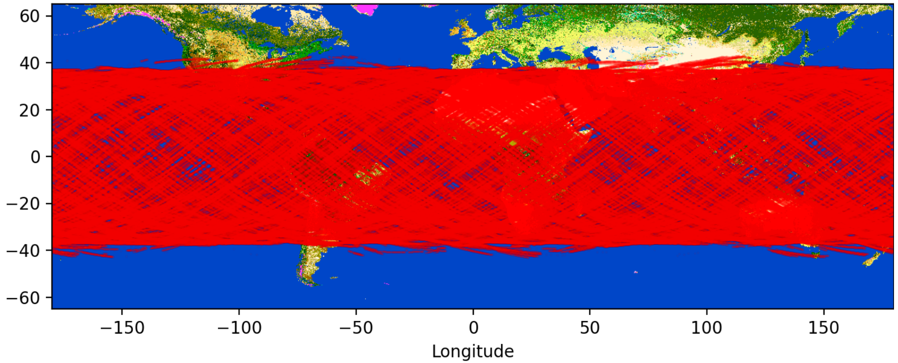
SMAP is the standard for global soil moisture estimation but has a revisit rate too slow for some hydrological/ meteorological studies. The high spatiotemporal resolution of CYGNSS GNSS reflection measurements (delay-Doppler maps, DDMs) can address this issue and have been successfully calibrated with SMAP measurements. Unfortunately, the standard approach, which uses only the peak value of the DDM, fails in regions where there is low variation in soil moisture or complex surface conditions. We hypothesize that information from the entire 2D DDM could help in these regions. The application of deep learning based techniques has the potential to extract additional information from the full DDM, while simultaneously providing the option to incorporate additional contextual information from external datasets. This work explored the data-driven approach of convolutional neural networks (CNNs) to determine complex relationships between the reflection measurement and surface parameters, providing a mechanism to achieve improved global soil moisture estimates. CYGNSS DDMs were trained on aligned SMAP soil moisture values, provided with the context of ancillary datasets. Data was aggregated into training sets, and a CNN was developed to process them. Results of this training were studied using an unbiased subset of samples, and compared to existing global soil moisture products.

Problem Description: Enhancing SM Resolution

Single Day SMAP AM (repeat rate: 3 days)



Single Day CYGNSS (repeat rate: 0.25 days)

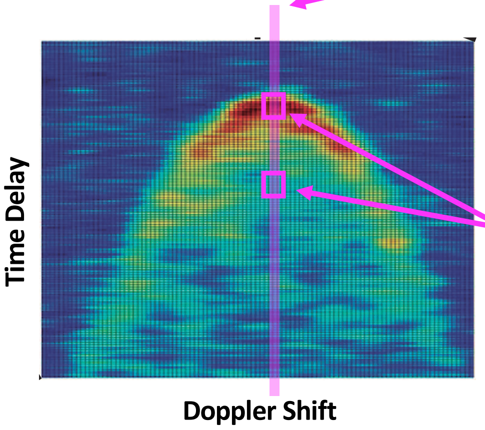


- SMAP[1] spatiotemporal resolution insufficient for some hydrological studies[2,3] (10 km, < 2 days)
- CYGNSS[4] GNSS-R data can supplement SMAP, increasing resolution
- This has been done by a group at UCAR[5] using simple linear relation (SM = b*power + a)
- Method fails in regions with stable SM, or where confounded by complex conditions

Problem Description: Utilization of Complex Data

DDM

Extract a single 1D profile



Extract key single-value metric

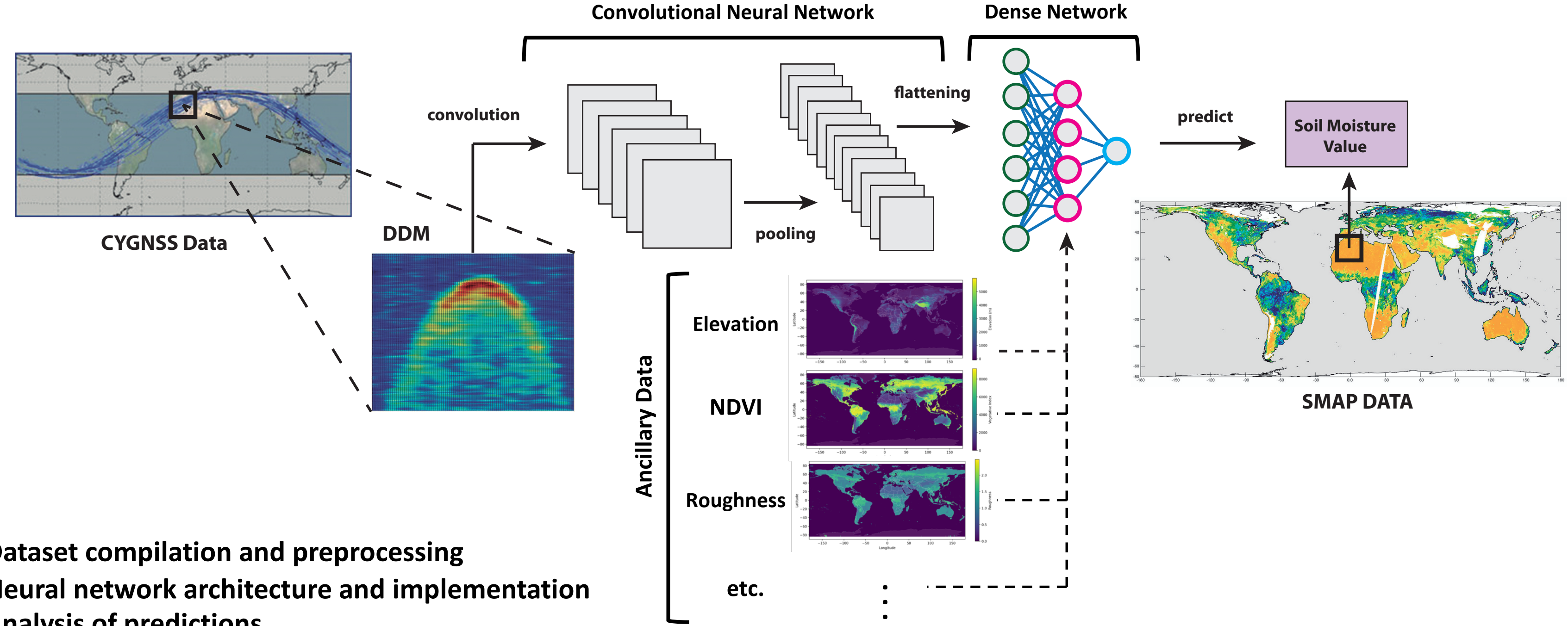
Information rich, but complex, data

Contributions from:

- Surface topography/roughness
- Water/vegetation
- Spacecraft geometry/antenna

- Delay-Doppler maps downlinked from receiver
- 2D array contains analytically complex data
- Traditional key metrics are extracted
- Majority of data is thrown out
- Discarded data gives information that could be interpreted with sufficient contextual input

Methodology: Overview



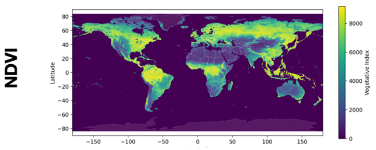
1. Dataset compilation and preprocessing
2. Neural network architecture and implementation
3. Analysis of predictions

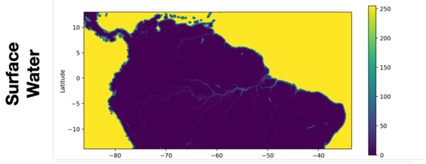
Methodology: Datasets and Preprocessing

1. Compile relevant and useful datasets

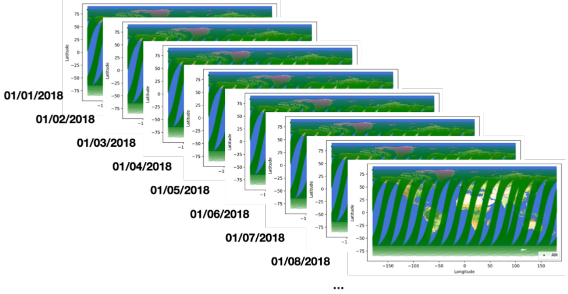
Data	Rate	Source	Base Resolution	Used Resolution
Primary Input: CYGNSS DDM	Per DDM	CYGNSS	0.5-7 km	0.5-7 km
Ancillary Inputs:				
CYGNSS SC Info/PRN Number	Per DDM	CYGNSS	DDM-scale	DDM-scale
Angle/Range to Reflection	Per DDM	CYGNSS	DDM-scale	DDM-scale
Gains/EIRP	Per DDM	CYGNSS	DDM-scale	DDM-scale
Latitude/Longitude	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/SMAP Ancillary	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

2. Study datasets to gain intuition/find numerical issues

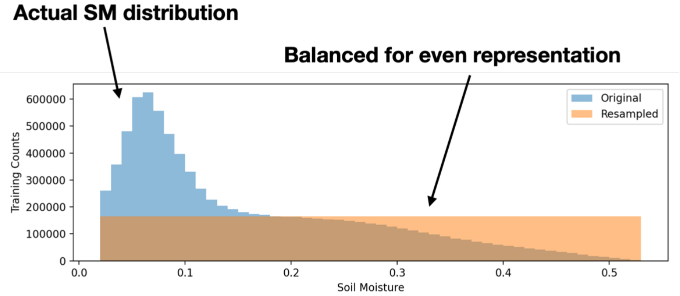




3. Build training dataset with features for processing

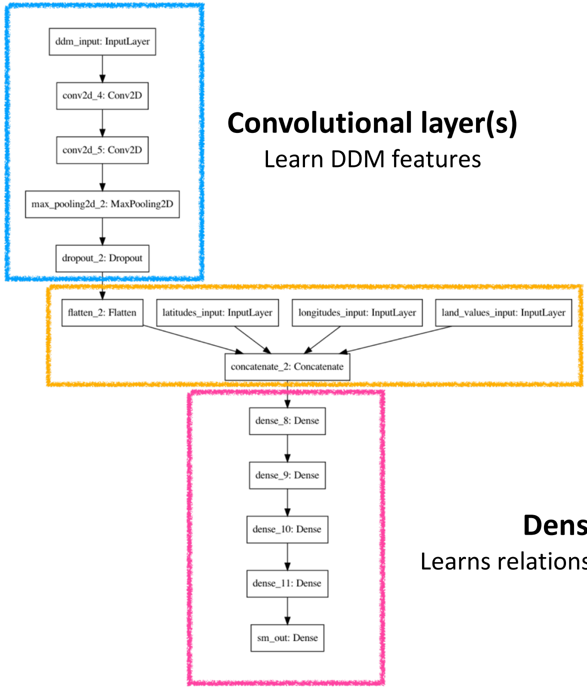


4. Filter, standardize, and balance data to optimize for training



For full CYGNSS constellation, used ~10-20 million samples for 2018

Methodology: Network Design



Convolutional layer(s)
Learn DDM features

Isolated Land Type CNN

Inputs: DDMs
Outputs: Land Type

Concatenation layer
Combines learned DDM features (flattened) and ancillary data combined

Dense layer(s)
Learns relationship between all data

Soil Moisture CNN

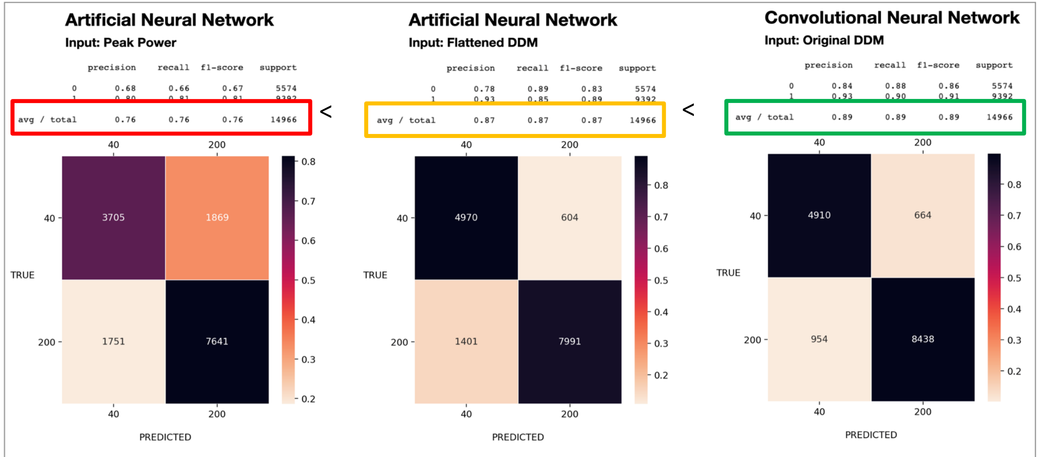
Inputs: DDMs, Elevation, NDVI, Roughness, etc.
Outputs: Soil Moisture

Implemented with TensorFlow and Keras libraries in Python

Methodology: CNNs in Isolation

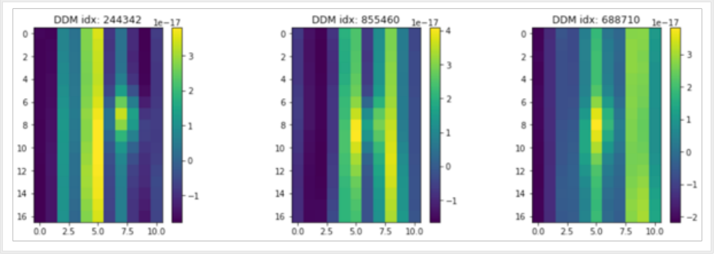
Independent study to determine best practices when using DDMs as inputs for CNN

Should CNNs even be used?



40: Closed to open (> 15%) broadleaved evergreen and/or semi-deciduous forest (>5m)
200: Bare areas

Are there issues with the DDMs?



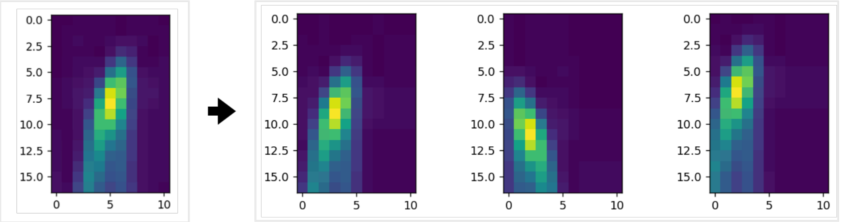
RFI in DDMs

Methodology: CNNs in Isolation

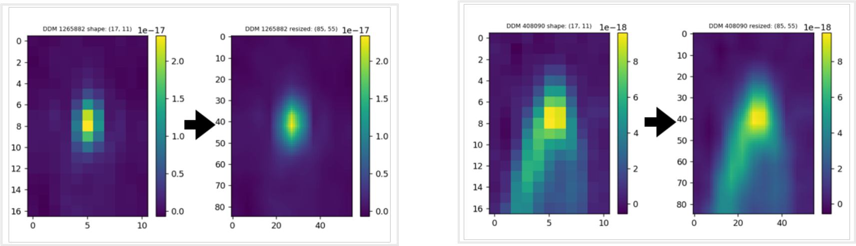
Independent study to determine best practices when using DDMs as inputs for CNN

Can DDM inputs be modified to improve performance?

DDM augmentation at train time help with network generalization



Increased resolution of DDMs improved performance

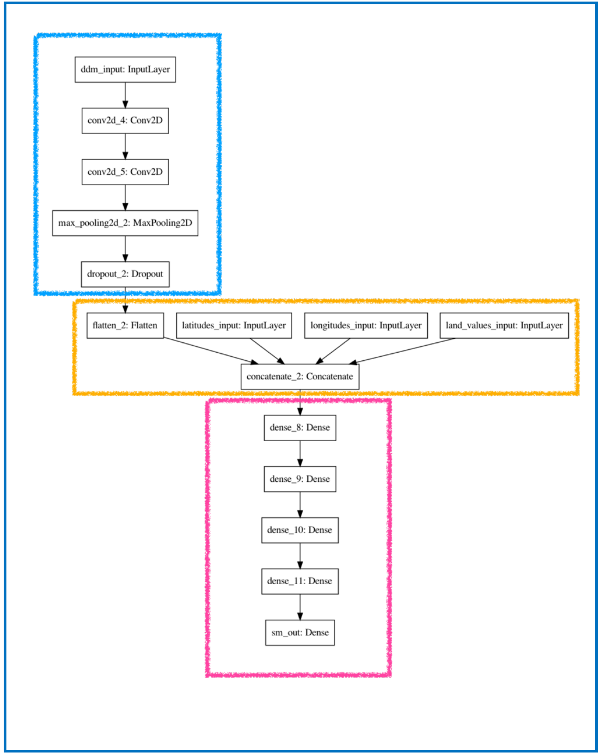


Can we improve the network architecture?

- **Borrowed elements from modern architectures.**
- **Complex models overfit and more difficult to generalize.**
- **Opted for simpler architecture.**

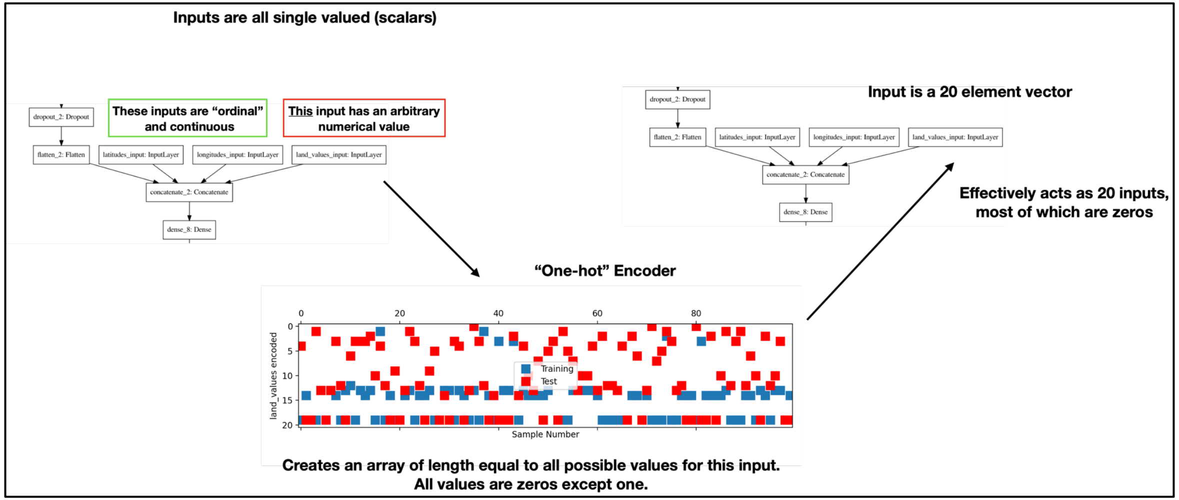
Lessons learned transferred to improve CNN used in SM network

Methodology: Complete Network

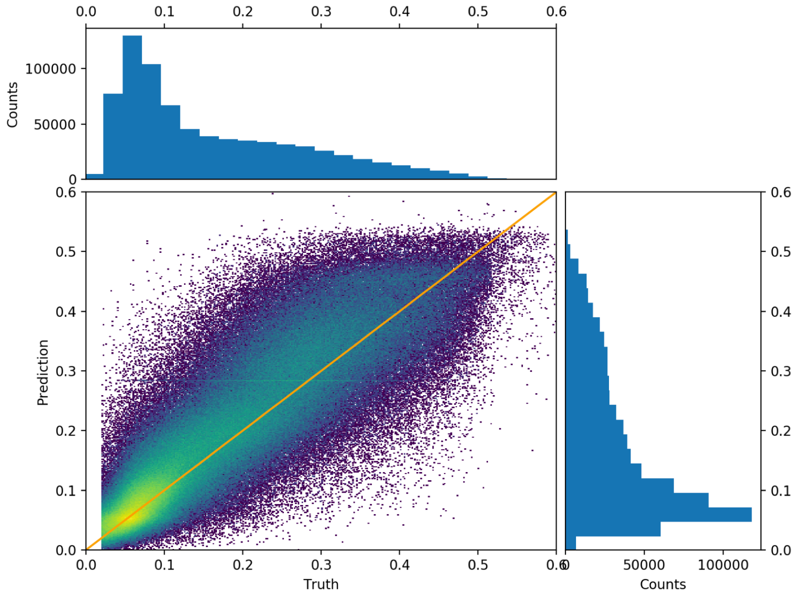


- Target SM as the output type
- Design for for flexibility in input options
- Design for variety in nature of inputs (continuous, categorical, CNN output)

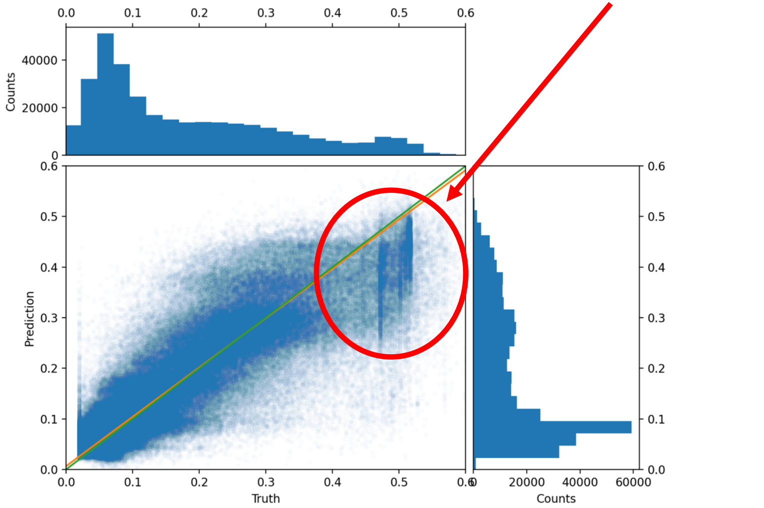
Specific Problem: Categorical Data



Methodology: Analysis (Correlation)

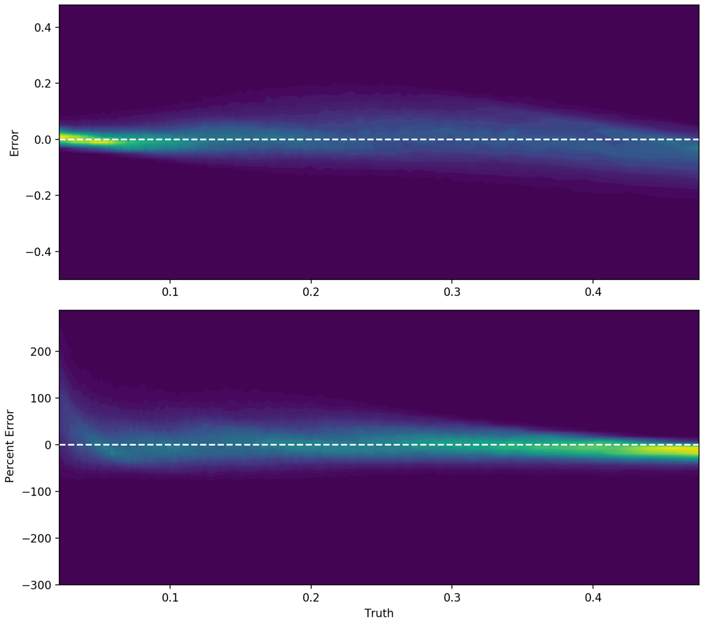


Problematic?

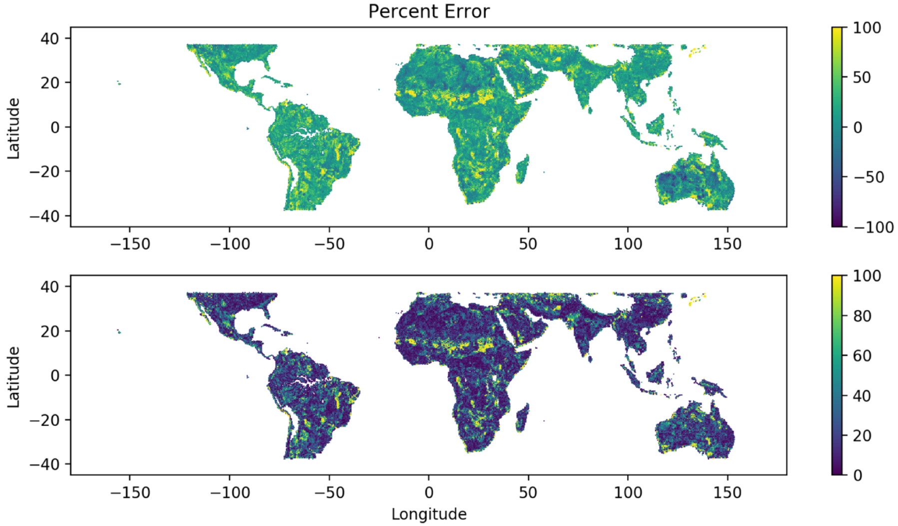


Methodology: Analysis (Error)

Error/Percent Error with SM

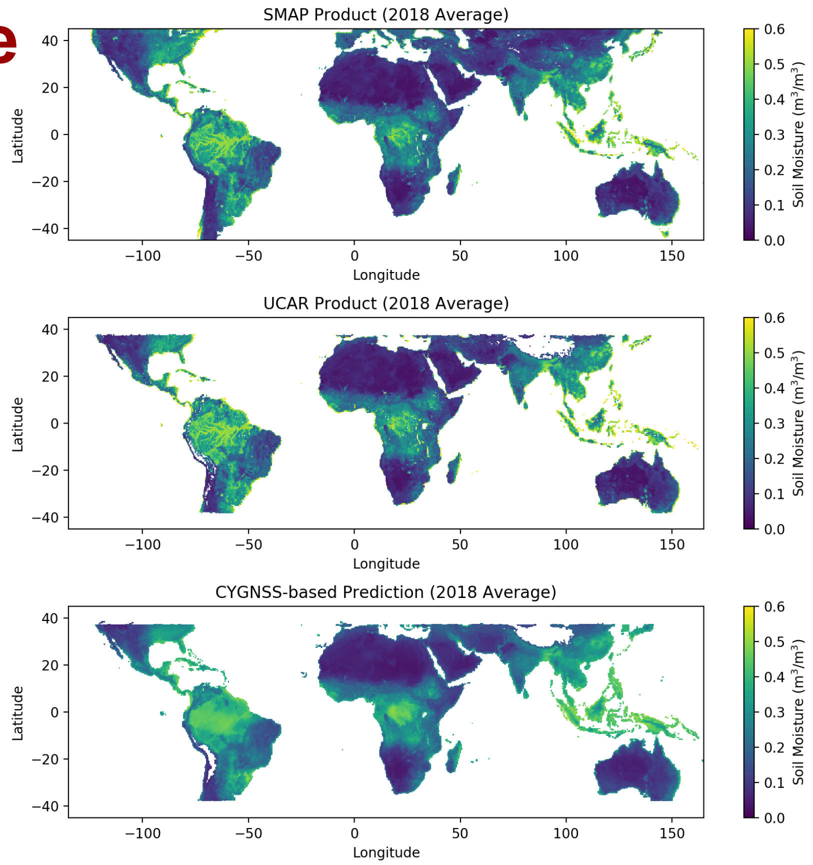


Percent Error Geographically



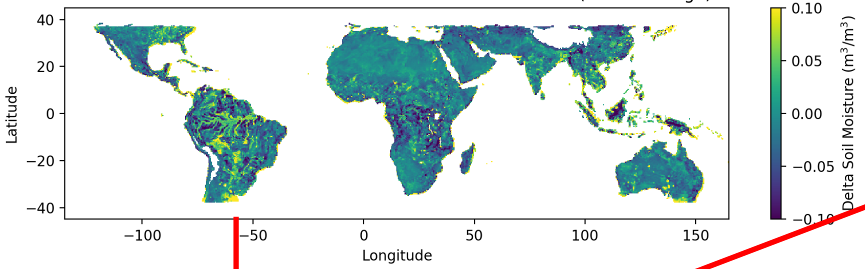
Results: Global Trends Agree

- Similar global trends/features to SMAP[4] and UCAR[3]
- Increased coverage from UCAR
- Predictions differ in high SM areas (Amazon, Congo)

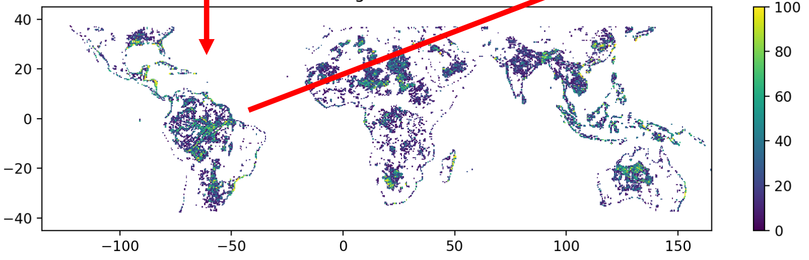


Results: Systematically Low for High SM (Two Causes)

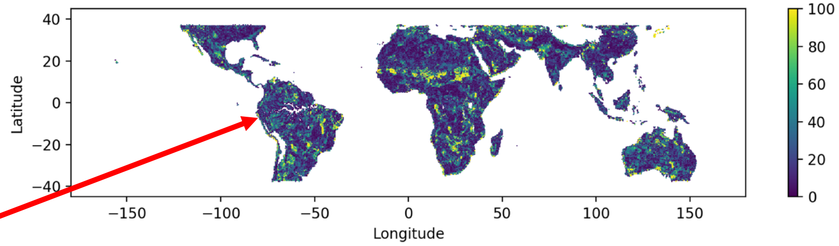
1) Differences can be correlated with SMAP flags



Somewhat…



Amazon filtered from training/test set!

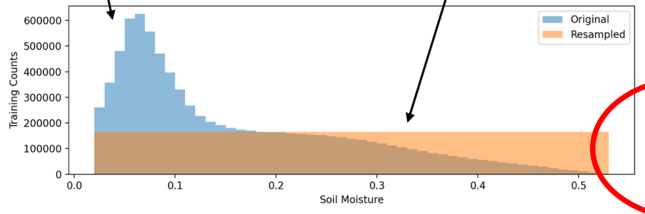


2) Balancing in training puts emphasis on low SM values

No representation in training!

Actual SM distribution

Balanced for even representation



<u>But are these predictions wrong, or are these erroneous SMAP values??</u>

Next Steps

These initial results are important in that:

- We've shown CNNs can be used to interpret DDMs directly
- Initial exploration shows strong correlation and opportunity for improvement
- Concept generalizable to other surface retrievals with DDMs

Immediate and advanced next steps would be:

- Explore influence of individual datasets extensively, refine input choices and filtering
- Study the spatial and temporal averaging to optimize performance
- Adding valuable, missing ancillary datasets (such as "distance-to-water")
- Fine tune network architecture and training parameters
- "Ensemble" of networks for regional prediction
- Use vector inputs for ancillary data (input region of values, not average)
- Include in situ measurements in training as "high value" targets
- Applications to other targets, like "freeze/thaw", flood/inundation, and water masks

References and Acknowledgements

[1] Kim, S., J. Van Zyl, R. S. Dunbar, E. G. Njoku, J. T. Johnson, M. Moghaddam, and L. Tsang. 2016. SMAP L3 Radar Global Daily 3 km EASE-Grid Soil Moisture, Version 3. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center.

[2] Ruf, C., P. Chang, M.P. Clarizia, S. Gleason, Z. Jelenak, J. Murray, M. Morris, S. Musko, D. Posselt, D. Provost, D. Starkenburg, V. Zavorotny, CYGNSS Handbook, Ann Arbor, MI, Michigan Pub., ISBN 978-1-60785-380-0, 154 pp, 1 Apr 2016.

[3] Brocca, L.; Moramarco, T.; Melone, F.; Wagner, W. A new method for rainfall estimation through soil moisture observations. Geophys. Res. Lett. 2013, 40, 853–858.

[4] Brivio, P. A., Colombo, R., Maggi, M. & Tomasoni, R. Integration of remote sensing data and GIS for accurate mapping of flooded areas. Int. J. Remote Sens. 23, 429–441 (2002).

[5] C. Chew and E. Small, "Description of the UCAR/CU Soil Moisture Product," *Remote Sensing 2017, Vol. 9, Page 362*, vol. 12, no. 10, pp. 1558–26, May 2020.

Thanks for Clara Chew at UCAR for conversations and valuable input!