Dynamic Structure Learning through Graph Neural Network for Forecasting Soil Moisture in Precision Agriculture

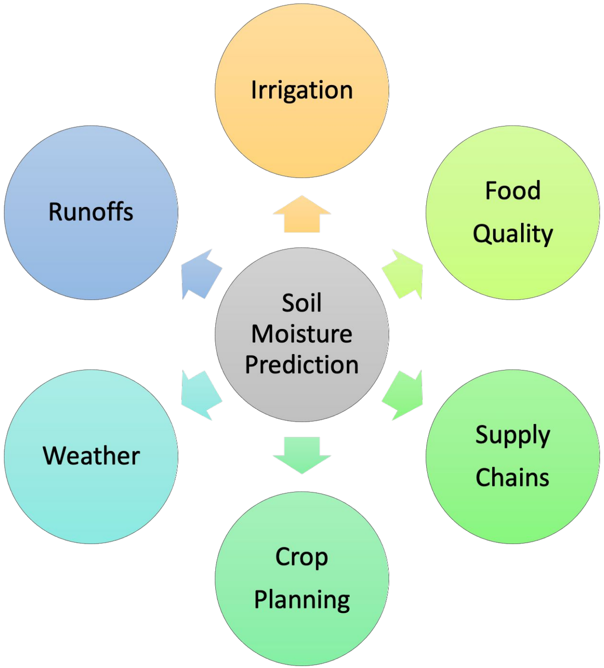
Anoushka Vyas *
Department of Computer Science
Virginia Tech, Arlington, VA, USA
anoushkav@vt.edu

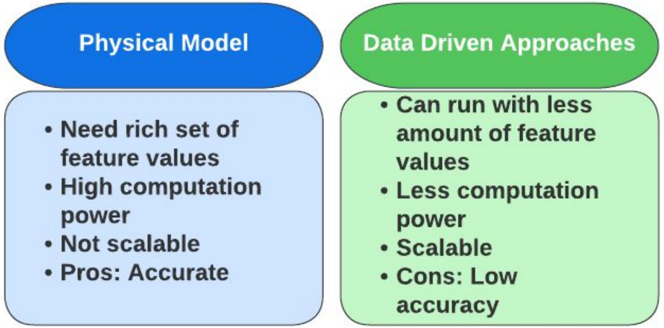
Sambaran Bandyopadhyay *
Search Science and AI
Amazon, Bangalore, India
samb.bandyo@gmail.com

The work was done when both the authors were associated to IBM Research, Bangalore

Introduction and Motivation

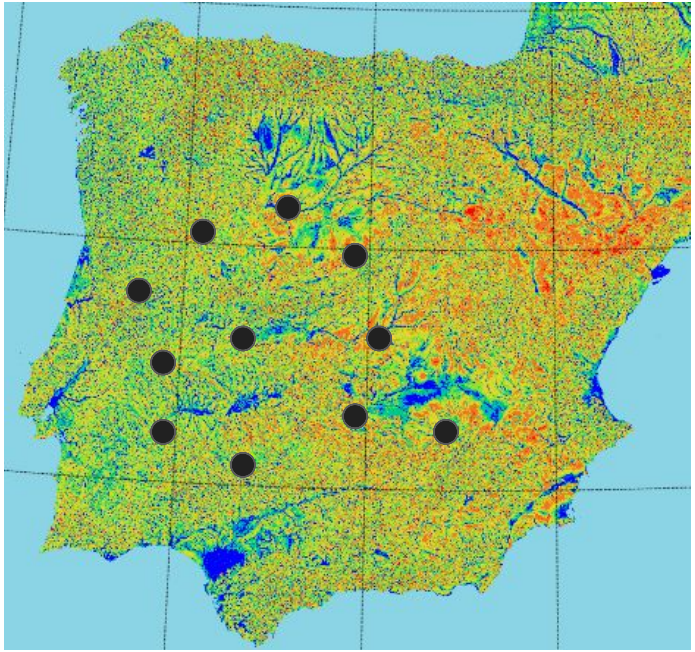
- Soil moisture measures the amount of water stored in various depths of soil.





Problem Statement

- We are given with a set of N locations over a geographic region
- There are T time steps (weekly / monthly resolution)
- For each time step, ground truth soil moisture (SM) values are given for a subset of locations
- For each location and time step, there are some input features (such as temperature, relative humidity, precipitation, NDVI, etc.) available.
- Our goal is to predict (forecast) the soil moisture at future time steps such that the forecasting model considers both past soil moisture and input feature values, and also able to learn and exploit the relation between different locations in their soil moisture values.



Background on Graph Neural Networks

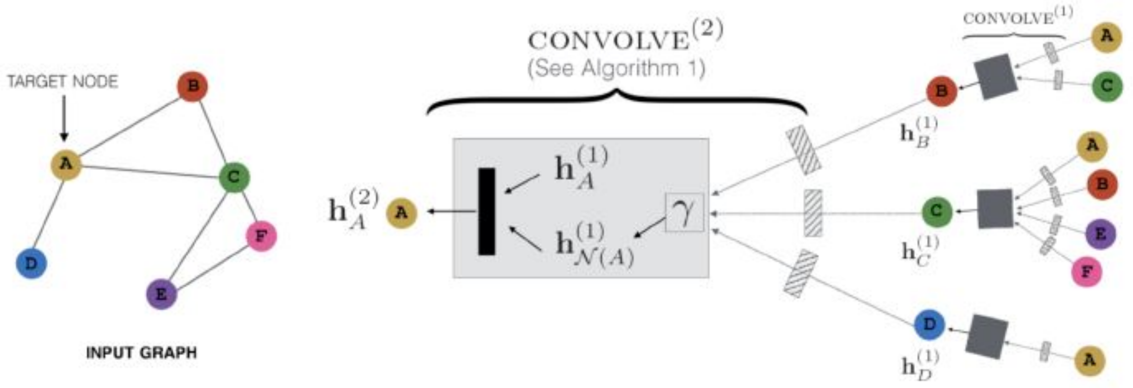


Figure: Graph Neural Networks [Ying et. al. 2018]

$$h_v^l = COMBINE^l\left(\left\{h_v^{l-1}, AGGREGATE^k\left(\{h_{v'}^{l-1} : v' \in \mathcal{N}_G(v)\}\right)\right\}\right)$$

Challenges for Our Problem

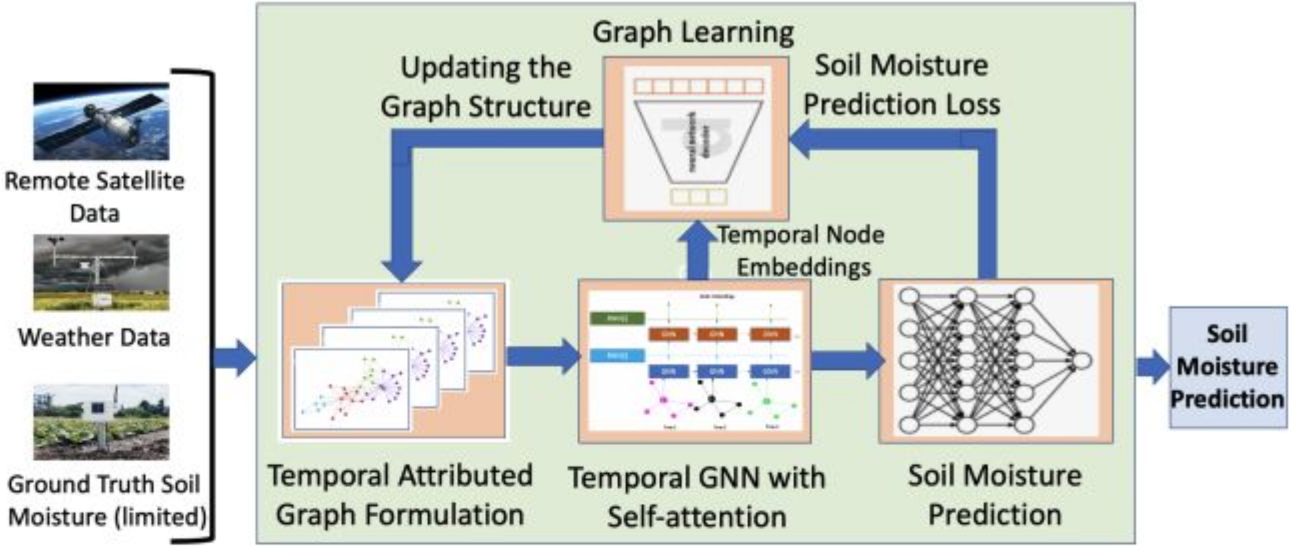
- There is spatial dependency in soil moisture values:
  - They depend on weather, crop cycles, soil quality of nearby locations.
  - It is difficult to learn in conventional ML algorithms because of huge data requirements.
  - Not all nearby locations can be correlated in soil moisture.

- Soil moisture at any location changes over time:
  - It can be due to the changes of various features of land and weather.
  - This requires a temporal (or dynamic) model for soil moisture forecasting.

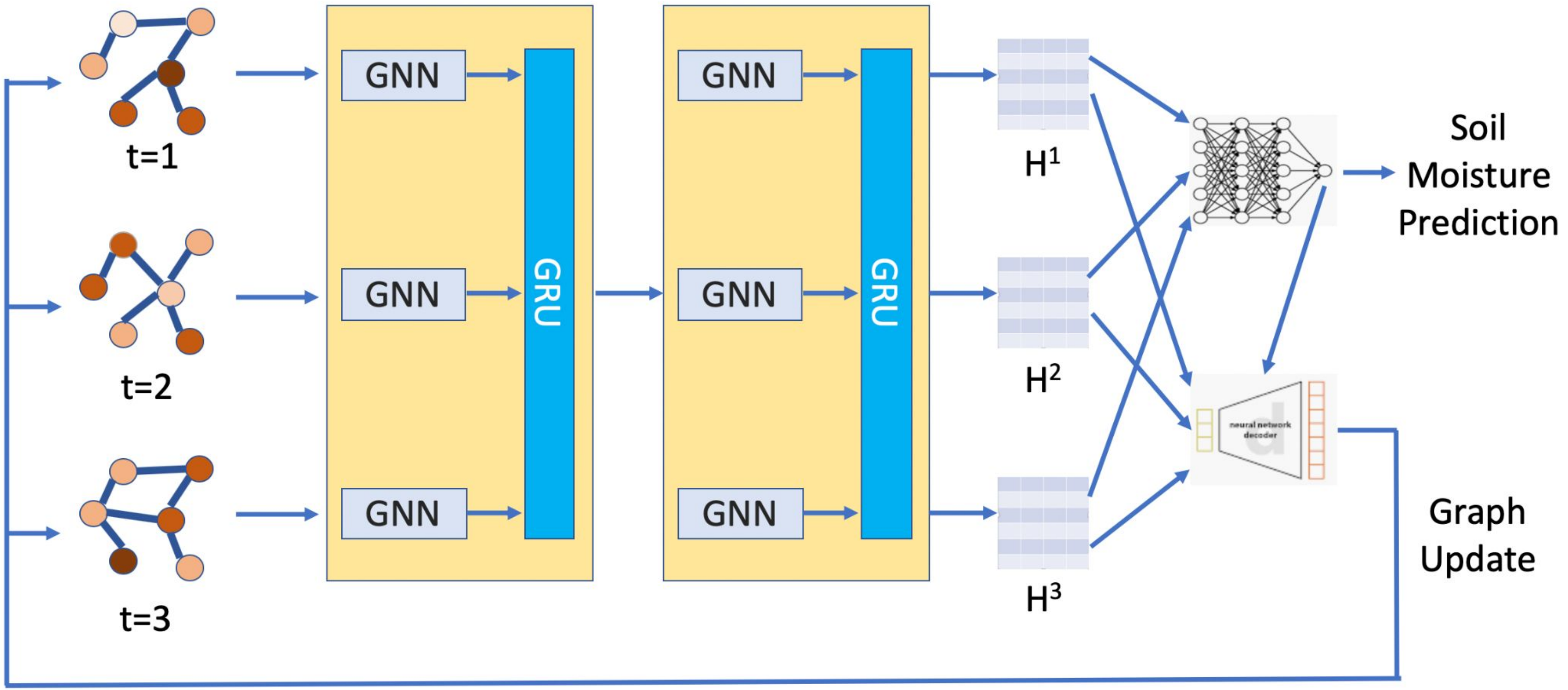
- There is a need for semi-supervised training approach:
  - There are points with available feature values but missing soil moisture values (device and communication failures).
  - Obtaining ground truth soil moisture is expensive.

- There is no ground truth graph structure given as an input for the problem of spatio-temporal data modeling required for soil moisture forecasting.

Our Solution Approach: DGLR



Our Algorithm Architecture: DGLR



Regularizers to Learn the Graph Structures

Graph Closeness Regularizer:

$$\mathcal{L}_{GC} = \sum_{t=1}^{T} \text{Distance}(A^t, \hat{A}^t)$$

Feature Smoothness Regularizer:

$$\mathcal{L}_{FS} = \sum_{t=1}^{T} \sum_{\substack{i,j \in [N] \\ i \neq j}} \hat{A}^t_{ij} ||x^t_i - x^t_j||^2_2$$

Target Smoothness Regularizer:

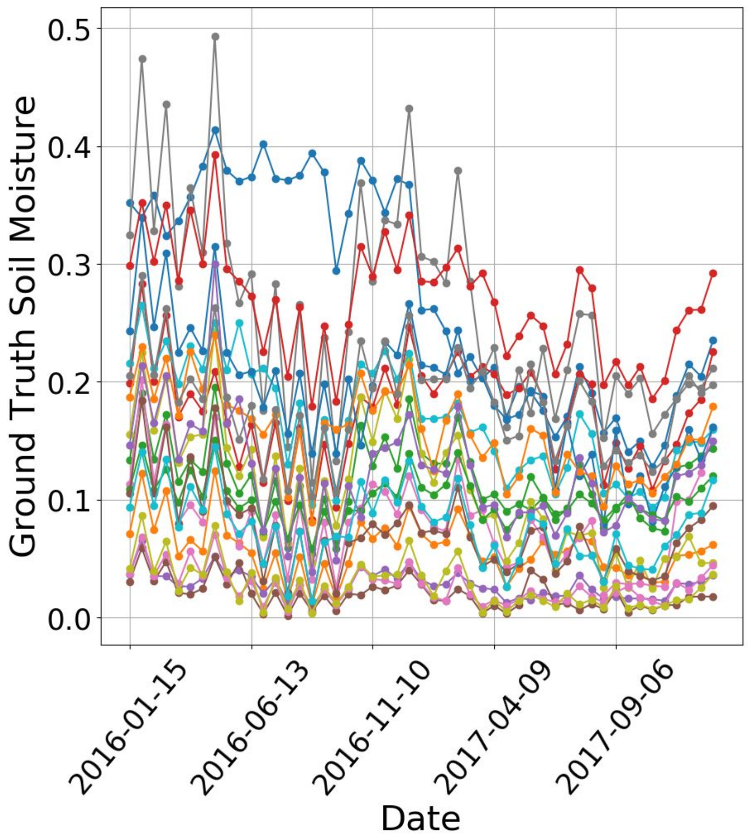
$$\mathcal{L}_{TS} = \sum_{t=1}^{T} \sum_{\substack{i,j \in N^t_{tr} \\ i \neq j}} \hat{A}^t_{ij} (s^t_i - s^t_j)^2$$

Notations	Explanations
$i, j \in \{1, 2, \cdots, N\} = [N]$	Indices over locations (nodes)
$d_{ij} \in \mathbb{R}_+$	Geographic distance between i and j
$t \in \{1, 2, \cdots, T\} = [T]$	Indices over time steps
$N^t_{tr} \in [N]$	Locations with soil moisture values at time t
$G^t = (V, E^t, X^t)$	The graph at time $t \in [T]$
$x^t_i \in \mathbb{R}^D$	Input node features for ith node at time t
$h^t_i \in \mathbb{R}^K$	Node embedding of i th node in G^t
$s^t_i \in \mathbb{R}$	Actual soil moisture for ith station at time t
$\hat{s}^t_i$	Predicted soil moisture
θ	Distance threshold for initial graph construction
$A^t \in \mathbb{R}^{N \times N}$	Initially constructed adjacency matrix
$\hat{A}^t = H^t {H^t}^{Trans}$	Reconstructed adjacency matrix
W_G	Parameter of the GNN for the graphs
$\Theta_R = \{\Theta_{R,i} : \forall i\}$	Parameters of the RNNs for all the nodes

Total loss to train the network is a weighted sum of soil moisture prediction loss (sum squared error) and the above structure learning regularizers

Experimental Evaluation

- We use two soil moisture datasets (from Spain and USA resp.) and the associated features from multiple sources.
- Some of the features used for the Spain dataset are temperature, humidity, NDVI, SAR backscattering coefficients, etc.
- For USA, the features (soil temperature, weather data, etc.) and soil moisture data are obtained from the SCAN network.
- The source code and datasets are made publicly available!!



Soil Moisture Forecasting

Algorithms	Spain			USA		
	RMSE (↓)	SMAPE % (↓)	Correlation (↑)	RMSE (↓)	SMAPE % (↓)	Correlation (↑)
SVR-Shared	0.061 ± 0.004	33.654 ± 1.236	0.020 ± 0.001	12.332 ± 0.848	39.013 ± 3.395	0.287 ± 0.031
SVR	0.052 ± 0.005	23.840 ± 2.453	0.083 ± 0.002	10.461 ± 0.718	27.377 ± 3.314	0.354 ± 0.031
Spatial-SVR	0.052 ± 0.007	23.752 ± 1.562	0.150 ± 0.013	10.245 ± 1.567	26.710 ± 4.049	0.351 ± 0.049
ARIMA	0.041 ± 0.008	19.002 ± 0.872	0.010 ± 0.001	9.290 ± 1.110	27.785 ± 1.306	0.159 ± 0.016
RNN-Shared	0.039 ± 0.003	23.443 ± 1.996	0.585 ± 0.048	7.814 ± 0.442	30.880 ± 2.355	0.191 ± 0.041
RNN	0.039 ± 0.003	21.722 ± 1.385	0.529 ± 0.049	7.338 ± 0.346	27.833 ± 1.812	0.273 ± 0.046
DCRNN	0.061 ± 0.003	31.832 ± 2.363	0.588 ± 0.062	8.161 ± 0.762	25.213 ± 1.762	0.393 ± 0.015
EvolveGCN	0.061 ± 0.004	31.789 ± 2.554	0.731 ± 0.022	8.526 ± 0.519	27.717 ± 1.919	0.415 ± 0.038
DGLR (Shared)	0.037 ± 0.003	17.533 ± 1.000	0.752 ± 0.020	7.526 ± 0.519	20.700 ± 1.210	0.493 ± 0.031
DGLR (w/o SL)	0.035 ± 0.003	15.599 ± 0.940	0.753 ± 0.017	6.970 ± 0.356	18.360 ± 0.987	0.517 ± 0.023
DGLR (w/o Sm)	0.033 ± 0.002	19.835 ± 1.005	0.759 ± 0.018	6.721 ± 0.176	17.991 ± 1.211	0.533 ± 0.013
DGLR (full model)	$\mathbf{0.031 \pm 0.001}$	$\mathbf{15.583 \pm 1.008}$	$\mathbf{0.764 \pm 0.017}$	$\mathbf{6.454 \pm 0.111}$	$\mathbf{17.002 \pm 0.987}$	$\mathbf{0.566 \pm 0.008}$

Experimental Evaluation

Forecasting with missing Soil Moisture values (semi-supervised solution)

Missing	Spain		
SM %	RMSE (↓)	SMAPE % (↓)	Correlation (↑)
0%	0.031 ± 0.001	15.583 ± 1.008	0.764 ± 0.017
10%	0.035 ± 0.004	18.121 ± 1.430	0.750 ± 0.033
20%	0.039 ± 0.002	18.864 ± 1.516	0.738 ± 0.027
30%	0.039 ± 0.006	18.962 ± 1.112	0.661 ± 0.047

Effect of threshold to construct the initial graphs

Distance	Spain		
Threshold (km)	RMSE (↓)	SMAPE % (↓)	Correlation (↑)
8	0.034 ± 0.002	16.418 ± 0.810	0.752 ± 0.042
12	0.032 ± 0.003	16.441 ± 1.106	0.752 ± 0.028
16	0.031 ± 0.001	15.583 ± 1.008	0.764 ± 0.017
20	0.032 ± 0.004	16.556 ± 1.099	0.754 ± 0.054

Conclusion

- In this work, we have addressed the problem of soil moisture modeling and prediction.
- We propose a semi-supervised dynamic graph neural network which also learns the temporal graph structures iteratively to predict soil moisture.
- It is robust in nature with respect to missing ground truth soil moisture values.
- It achieves state-of-the-art performance for data driven soil moisture prediction on real-world soil moisture datasets.

BACK UP

Our Solution Approach: DGLR

- Initial Graph Formation:
    - We create a node for each given location
    - We connect any two nodes by an undirected and unweighted edge in the graph.
    - We connect the nodes if the distance is less than some predefined threshold .
- A self attention based GNN similar to Graph Attention Networks is used to capture spatial dependency between nodes.
- The updated node embeddings from the GNN is fed to an RNN which connects graphs over different time steps to capture the temporal dependency of soil moisture.
- Based on the temporal node embeddings and predicted soil moisture values, we update the graph structure by proposing multiple intuitive regularizers.
- Our model is a semi-supervised approach and can be trained on labeled and unlabeled data points together.