Forecasting Soil Moisture Using Domain Inspired Temporal Graph Convolution Neural Networks To Guide Sustainable Crop Management.

Anonymous submission

Abstract

Climate change, population growth, and associated water scarcity present unprecedented challenges for agriculture. Consequently, climate smart agriculture demands efficient usage of resources to optimize crop production. This project aims to forecast soil moisture using domain knowledge and machine learning, which in turn can be used for crop management decisions that enable sustainable farming. Traditionally, hydrological response features, such as soil moisture, are predicted by dividing fields into small response units and solving physics-based and empirical hydrological models, which require significant computational time and oversight from domain experts. Recent work has implemented machine learning models as a tool to forecast hydrological response features. These machine learning models neglect a crucial component of traditional hydrological modeling: that spatially close units can have vastly different hydrological responses. In traditional hydrological modeling, units with similar hydrological properties are grouped together and share model parameters regardless of their spatial proximity. Inspired by this domain knowledge, we construct a novel domain-inspired temporal graph convolution neural network. Our approach consists of the following steps: (1) clustering units based on their time-varying hydrological properties, (2) constructing graph topologies for each cluster based on similarity using dynamic time warping and, (3) forecasting soil moisture for each unit using graph convolutions and a gated recurrent neural network. We train, validate and test our method on field-scale time series data consisting of 99K hydrological response units spanning 40 years in a case study in North-East of the US. Comparison with existing models for soil moisture forecast illustrates the effectiveness of using domain-inspired clustering with time series graph neural networks. The framework is currently being deployed as part of a pro bono social impact program which leverages technologies such as hybrid cloud and AI to enhance and scale non-profit and government organizations. The trained models are being deployed in a series of small-holding farms in Central Texas.

Introduction

Machine learning has changed much of our world in the past decade. However, its impact in some of the most critical areas have been negligible. A prominent example is quantifying and forecasting ground- and surface- water availability to inform agricultural practices. While water plays a fundamental role, characterizing these resources involves distinct temporal and spatial processes that must consider historical and future precipitation volumes, surface and groundwater runoff from heterogeneous sources, and evapotranspiration and other water sinks or losses. The huge landmasses, and high spatial and temporal variability makes it infeasible to collect sufficient density of observation data to implement IoT-backed decision support systems.

Traditionally, engineers have relied on physics-based models that represent hydrological processes as a set of partial differential equations that are further constrained by heuristics, empirical relationships, and expert intuition. While these allowed greater insight into spatial and temporal evolution of water over land, the associated complexity and uncertainty places a heavy burden on the expert user. Further, these models can not readily be deployed across different locations without a cumbersome calibration and validation effort. A prominent example in this regard is the Soil & Water Assessment Tool (Gassman et al. 2007) that has been widely-used to simulate the quality and quantity of both surface and ground water processes, and inform agriculture, land use, and land management practices. A corresponding body of research has developed around parameterizing and evaluating these models with prominent examples being the parameter estimation toolbox (PEST) and SWAT Calibration and Uncertainty Program (Doherty 2003; Abbaspour 2013).

The use of machine learning in hydrology has shown promising results for tasks such as, stream-flow prediction, flood forecasting, and rainfall-runoff prediction (Young, Liu, and Wu 2017; Kratzert et al. 2019a; Nearing et al. 2020; Kratzert et al. 2019b). However, these successes typically occur in regions with well developed environmental monitoring programs and relatively plentiful observation data to train on. Further, most previous research has focused on surface flow prediction, and do not resolve the more agriculturally-important subsurface flows. These subsurface and soil moisture processes dictate many critical ecosystem services such as carbon storage of soils, physical, chemical and biological soil health, and crop growth dynamics.

Precision agriculture approaches (Zhang, Wang, and Wang 2002) have developed over the past four decades by combining models, satellite, and sensor data to improve decision making. Success in precision agriculture is related to how well it can be applied to assess, manage, and evaluate the space-time continuum in crop production (Pierce and

Nowak 1999). Climate change introduces a completely different set of challenges that require drastically more granular data, and more holistic decision making, to enable an environmentally and economically sustainable response. The negative impacts of climate change are already being felt, in the form of increasing temperatures, weather variability, shifting agroecosystem boundaries, invasive crops and pests, and more frequent extreme weather events (Calzadilla et al. 2013). On farms, climate change is reducing crop yields, the nutritional quality of major cereals, and lowering livestock productivity (Bank 2016). These stressors particularly impact water constrained regions, resulting in groundwater depletion, soil erosion, and crop failures.

Adapting to these challenges require the adoption of climate-smart agriculture practices that minimize resource consumption and environmental impacts, while simultaneously ensuring food security for growing populations. Globally, large farms increasingly digitize operations to enhance sustainability, small-holding farmers lack the skills and resources to leverage AI and IoT informed decision making. This led the World Economic Forum to posit that "agriculture and farming will be redefined within a decade with the adoption of AI-driven autonomous tools" (Itzhaky 2021). However democratization of these solutions to small-holding and disadvantaged farmers requires scalable, spatially-aware machine learning models that can be informed by publicly-available datasets and sparse low-cost (less accurate) sensor data.

This paper describes a novel domain-inspired framework to forecast soil moisture. The proposed framework uses graph convolutional neural networks (GNN) to resolve complex hydrological response in a domain consisting of 3000 watersheds. While previous research has explored a GNN approach to represent spatial patterns by superimposing a graph topology over the physical streamflow network, our approach instead generates the topology based on the degree of physical and hydrological similarity between individual watersheds. This provides a more physically representative framework that is informed by the concept of group response units (GRUs), a well-established hydrological modeling technique, introduced by (Kouwen et al. 1993). GRU is composed of groups of HRUs that have similar hydrological characteristics and have more comparable hydrological response than neigboring units which might have different characteristics (e.g. crop versus livestock farming). The proposed framework is applied to forecasts of soil moisture in a case study application in NorthEast of the US.

The contributions of this paper are as follows:

- Describes a novel domain-inspired temporal graph convolution neural network. Analogous to GRUs, a clustering algorithm based on dynamic time warping (DTW) clusters together HRUs with similar features regardless of their spatial proximity. For each cluster, the graph topology is extracted from a set of similarity metrics that encompass static and dynamic hydrological catchment attributes.

- Presents experimental results that compare models using our novel GNN framework against state of the art for

time series forecasting, an LSTM model. These experiments demonstrate the increased gain from using hydrological feature information to inform prediction.

- Finally, we discuss further research opportunities to apply machine learning to improve agriculture management and environmental sustainability. In particular, the potential to use the approach to inform regions with sparse sets of monitoring datasets.

Related Work

Recent advancements in machine learning has led to widespread interest amongst hydrologists and environmental scientists as a solution to address the challenges that persist with streamflow and run-off forecasting. While previous works have approached performance levels of state-of-the-art physics-based methods (Hsu, Gupta, and Sorooshian 1995; Kratzert et al. 2019b; Nearing et al. 2020), the challenge remains whether it can generalize to finer scales and if it can perform in regions with limited training data.

Physics- or empirical-based hydrological models are well established in the literature, with research in the space receiving significant impetus with the US Clean Water Act of 1977. In these systems, the hydrological processes of water movement are represented by finite difference equations that are resolved over defined spatial and temporal ranges. Data inputs to resolve streamflow processes include meteorological forcing and a large number of parameters describing the physical characteristics of the catchment (soil properties, initial water depth, topography, topology, runoff curve number, etc.) (Devia, Ganasri, and Dwarakish 2015). Popular modelling systems include SWAT (Arnold et al. 2012), MIKE SHE (Graham and Butts 2005), WRF-Hydro (Lin et al. 2018) and the VIC framework (Gao et al. 2010). On the SWAT model alone, there are over 4,500 peer-reviewed journal articles describing its application to different hydrology studies (Srinivasan and Balmer 2021).

More recently, extensive research efforts have focused on the potential of deep learning (DL) for hydrology studies (Shen 2018; Shamshirband et al. 2020). In particular, research has focused on the potential of recurrent networks and LSTMs to resolve the complex, nonlinear, spatiotemporal relationship between meteorological forcing, soil moisture and streamflow (Kratzert et al. 2019b). In a provocative recent paper, (Nearing et al. 2021) argued that there is significantly more information in large-scale hydrological data sets than hydrologists have been able to translate into theory or models. This argument for increased scientific insight and performance from machine learning rests on the assumption that large-scale data sets are available globally (over sufficient historical periods) to condition and inform on hydrological response. While significant progress on coarse-scale hydrology dataset curation has been achieved in the US (Newman et al. 2015), and Europe (Klingler, Schulz, and Herrnegger 2021) this is not implemented for many other regions and does not approach the spatial resolutions required.

Many studies have proposed frameworks to represent the spatiotemporal properties of geophsysical systems. The most widely used framework combines convolutional neu-

ral networks (CNN) with LSTM to represent both the spatial (CNN) and temporal (LSTM) dependencies within the data. This approach has been applied to a variety of geoscientific tasks such as precipitation nowcasting from rainfall radar maps (Xingjian et al. 2015) and forecasting sea surface temperature from satellite-derived observations (Yang et al. 2017). (ElSaadani et al. 2021) use this CNN + LSTM approach to estimate soil moisture. However, this approach requires gridded input data, and relies on spatial correlations. Our proposed approach overcomes these limitations by using graphs to handle unstructured data and by connecting nodes of the graph based on hydrological similarity rather than spatial proximity.

An alternative approach aims to embed information from physics or heuristic knowledge within the network. Physics-informed DL is a novel approach for resolving information from physics. The philosophy behind it is to approximate the quantity of interest (e.g., governing equation variables) by a deep neural network (DNN) and embed the physical law to regularize the network. To this end, training the network is equivalent to minimization of a well-designed loss function that contains the PDE residuals and initial/boundary conditions (Rao, Sun, and Liu 2020).

A further stream of related work has been started by Chen et al. (2018), who presented a novel approach to approximate the discrete series of layers between the input and output state by acting on the derivative of the hidden units. At each stage, the output of the network is computed using a black-box differential equation solver which evaluates the hidden unit dynamics to determine the solution with the desired accuracy. In effect, the parameters of the hidden unit dynamics are defined as a continuous function, which may provide greater memory efficiency and balancing of model cost against problem complexity. The approach aims to achieve comparable performance to existing state-of-the-art with far fewer parameters, and suggests potential advantages for time series modeling.

Methods

Data

Leavesley et al. (1983) introduced the decomposition of watersheds into sub-areas that are assumed to be homogeneous in their hydrologic response, termed hydrologic response units (HRUs). The HRUs are characterized using topographic variables, such as elevation, slope, and geographic variables such as soil type, vegetation type and precipitation distribution. HRUs are generated by first decomposing a domain into a set of watersheds which represents the land area in which any precipitation eventually flows into the same outlet. Within sub-basins, HRUs are further delineated into smaller polygons, based on land use, soil attributes, and slope. For modelling and analysis, polygons with homogeneous hydrologic response are lumped together and resolved simultaneously. The concept of HRUs enable modelers to more effectively resolve complex issues regarding spatial variability to provide a more realistic representation of land surface processes (Prasad 2005).

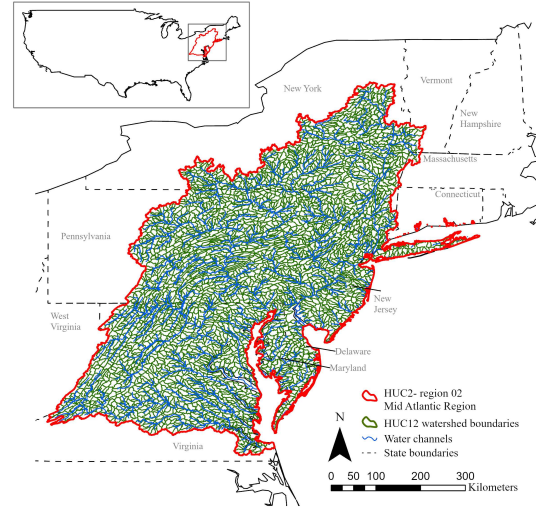We use data simulated by Soil and Water Assessment

Figure 1: Layout of the Mid-Atlantic basin along with its stream network and HUC12 watersheds.

Tool (SWAT) (Gassman et al. 2007). SWAT is the state of the art small watershed to river basin-scale model used to simulate the quality and quantity of surface and ground water and predict the environmental impact of land use, land management practices, and climate change. SWAT is widely used in developing agricultural management practices, assessing soil erosion prevention and control, non-point source pollution control and regional management in watersheds. While publicly available soil moisture re-analysis are available from institutions such as ECMWF (ERA Land-5) and NOAA (NLDAS), practical applications for agriculture management are constrained by the available resolution of $9\,km$ (Muñoz-Sabater et al. 2021) and $14\,km$ (Xia et al. 2012), respectively. Agriculture, on the other hand, requires predictions that resolve field-scale ($<500\,m$) processes. The Hydrological and Water Quality System (HAWQS) v2.0 (https://hawqs.tamu.edu/), a web-based interface of the SWAT model, was used to develop SWAT models for 3,037 watersheds at HUC12 (hydrologic unit code) resolution within HUC2- region 02, Mid-Atlantic region. The HAWQS provides a SWAT watershed model development framework with pre-loaded input data and modeling support capabilities for setting up models, running simulations, and processing outputs. To further divide delineated watersheds into HRUs, an area threshold of 0.5 km^2 was applied i.e., HRUs having area less than threshold value were not assigned a separate HRU-ID and merged with nearby HRUs. Overall, our data set consists of 3,037 watersheds divided into more than 99k HRUs. Detailed list of features associated with each HRU are included in the supplement file. Monthly data is available for each feature spanning 34 years.

Problem Formulation

Given a feature matrix $X_t \in \mathbb{R}^{n \times d}$ which is a snapshot of d feature values for n HRUs at time t, our goal is to fore-

Algorithm 1: Dynamic Time Warping Algorithm

Input: Discrete time series $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{1 \times s}$
Output: Distance between $\mathbf{x}$ and $\mathbf{y}$

```
 1:  initialize C = inf ∈ ℝ^(n×n)
 2:  C_{0,0} = 0
 3:  for i : 0 → s do
 4:      for j : 0 → s do
 5:          dist = d(x_i, y_j)^2
 6:          C_{i,j} = dist + min(C_{i-1,j}, C_{i,j-1}, C_{i-1,j-1})
 7:      end for
 8:  end for
 9:  DTW(x, y) = √(C_{s,s})
10:  return DTW(x, y)
```

cast M soil moisture values $\{Y_{t+i}\}_{i=0}^{M}$ in the future. For $M = 1$ it is called single step forecasting, for $M > 1$ it is called a multi-step forecasting. We start by solving the single step forecasting problem and extend our method to multi-step forecasting.

Single Forecast Given X_t we want to forecast the soil moisture Y_t for the next month.

Multi-step Forecast Given X_t we want to forecast the soil moisture $Y_t, .., Y_{t+12}$ for the next 12 months.

Domain Inspired Clustering

Inspired by the concept of group response units (GRUs), we build a clustering module to group HRUs that have similar hydrological characteristics. Traditionally, GRUs are constructed based on climate, land use, soil and pedotransfer properties (Poblete et al. 2020). The use of GRUs reduces the need for model calibration and allows for the transfer of model parameters among HRUs in the same group.

We propose a dynamic time warping based temporal clustering technique, which leverages the seasonality of these hydrological features to inform clustering.

First introduced in (Sakoe and Chiba 1978; Sakoe and Chiba), dynamic time warping is an algorithm for measuring the similarity between two discrete temporal signals. For the data tensor $\mathcal{X} \in \mathbb{R}^{n \times s \times d}$ containing n HRUs, s timesteps, and d features $\mathbf{x}^{i,j} := \mathcal{X}_{i,:,j}$ represents the 1D time series data for jth feature in ith HRU. The distance matrix $D \in \mathbb{R}^{n \times n}$ represents the pairwise DTW distance between all HRUs. The distance $D_{p,q}$ between two HRUs p and q is given by

$$D_{p,q} = \sum_{j=1}^{d} DTW(\mathbf{x}^{p,j}, \mathbf{x}^{q,j}) \tag{1}$$

where $DTW(.,.)$ is calculated using Algorithm 1.

Temporal Graph Convolution Neural Network (TGCN)

Graph convolution neural networks (Kipf and Welling 2016) are an extension of convolution neural networks to unstructured graph data. A graph $\mathcal{G} : (\mathcal{V}, \mathcal{E})$, has associated with it a set of nodes $\mathcal{V}$ connected by a set of edges $\mathcal{E}$. For our application each HRU represents a graph node. The adjacency matrix A is a matrix representation of the graph topology.

We use the temporal graph convolution neural network detailed in (Zhao et al. 2020) for predicting soil moisture at each node given time series features. At time t, the feature matrix X_t is updated using the graph convolution defined in (Bruna et al. 2014). The resulting *'neighbor-aware'* feature matrix Z_t is then passed on to the gated recurrent unit (GRU).

$$Z_t = Relu(AX_tW_0) \tag{2}$$
$$u_t = \sigma\left(W_u[Z_t : h_{t-1}] + b_u\right) \tag{3}$$
$$r_t = \sigma\left(W_r[Z_t : h_{t-1}] + b_r\right) \tag{4}$$
$$c_t = tanh\left(W_c\left[Z_t(r \odot h_{t-1})\right] + b_c\right) \tag{5}$$
$$h_t = (u_t \odot h_{t-1}) + (1 - u_t) \odot c_t \tag{6}$$

where u_t represents update gate, r_t represents reset gate, c_t represents cell state, h_t represents hidden state, and W_i, b_i are learnable weights and baises. The prediction $\hat{Y}_t$ is expressed as a linear transform of h_t.

We minimize the mean squared error loss during training.

$$\mathcal{L}_t = \frac{1}{n}\sum_{i=1}^{n}(Y_{t,i} - \hat{Y}_{t,i})^2 \tag{7}$$

Results

We train the LSTM model and 10 TGCN models (one for each cluster) for both the single forecast and multi-step forecasting. The number of clusters was selected based on the proportion of variance explained as described in the supplementary material.

Evaluation Metrics

We evaluate all models using mean squared error (MSE), which is a popular metric for regression. We also calculate the relative percent decrease in MSE to compare model performance. For our multi-step forecast method, we also use Kling-Gupta Efficiency (KGE) to quantify the goodness of fit. KGE is a traditional metric used in hydrology to evaluate model performances.

$$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\alpha-1)^2} \tag{8}$$

where r is the Pearson product-moment correlation coefficient, α is the ratio between the standard deviation of the predicted values and the standard deviation of the true values, and β is the ratio between the mean of the predicted values and the mean of the true values. A value of $KGE = -0.41$ corresponds to using the mean value as a benchmark predictor, therefore $KGE > -0.41$ indicates that the model improves upon the mean value benchmark (Knoben, Freer, and Woods 2019). As model becomes more accurate, $KGE \rightarrow 1$.

For model comparison, we perform a t-test to examine the statistical significance of performance improvement. Since we report test performance on independent clusters instead of k-folds, we do not violate the independence of sample assumption for the t-test. We define the following two null
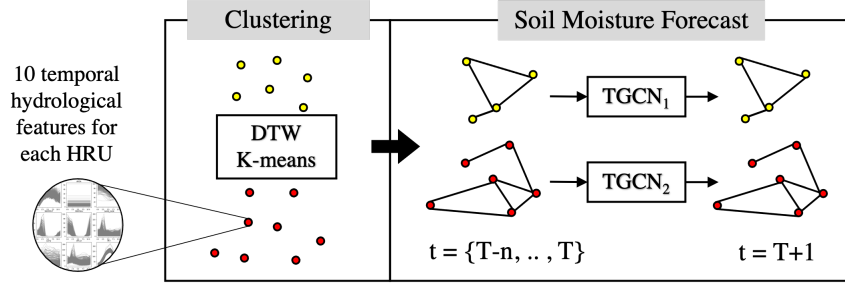
Figure 2: Schematic of our Clustering and Temporal Graph Convolution Neural network (C+TGCN) approach for soil moisture forecast.
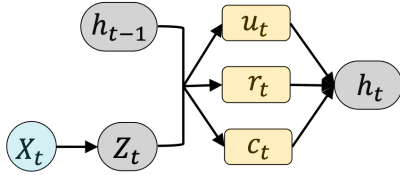

Figure 3: Schematic of a single cell of TGCN, equations (2-6).

hypotheses: $\mathcal{H}$ - model MSE has identical average values as LSTM MSE. For probability less than 0.05, we reject the null hypothesis.

Model Details

We use the first 27 years of data for training and validation and keep data from the last 7 years for testing.

Before computing the DTW distance matrix (D), we normalize the data using a custom min-max scaling. Instead of independently scaling the time series data, we normalize the time series for each feature by the minimum and maximum feature values across time series across all HRUs. Using this custom scaling we are able to preserve the relative trends in features.

We use an elbow test to estimate the number of clusters. Based on the results from the elbow test we then use K-means to split HRUs into 10 clusters. In order to avoid data leakage, we only use training data for clustering. We use functions from tslearn (Tavenard et al. 2020) to implement temporal clustering.

Once the HRUs are split into clusters, we introduce graph topology on each cluster by using the DTW distances of HRUs within the cluster. The static graphs represent disjoint subsets of HRUs and are trained independently using TGCN with same model architecture.

The model consists of a layer of graph convolution, followed by a linear transform. Output from the linear layer is then fed to the GRU which outputs the forecast Y_t. For multi-step model the GRU outputs a sequence of 12 predictions for each node. All TGCNs were trained using the Adam optimizer (Kingma and Ba 2015) with a learning rate of 1e-2 for around 100 epochs (until validation loss stopped decreasing). Weights were initialized using He initialization

(He et al. 2015). Based on the size of the graph the training took between 1.5 - 150 sec/epoch on 1 cpu core with 100G mem. Code for model setup, training and evaluation will be made available post-blind review.

Soil Moisture Forecast Results

Results from DTW + K-means clustering are shown in Figure 4, which illustrates the true normalized soil moisture values for a subset of clusters. The clusters represent distinct seasonal trends in soil moisture values.

Table 1 shows the average mean squared error of predicted soil moisture in each cluster. Using our model reduces the mean squared error across all clusters compared to the LSTM model. The relative percent reduction in mean squared error also illustrates the added benefit of using domain inspired clustering. Figure 5 shows the mean and standard deviation of KGE for all HRUs in a cluster. Since the KGE values for all clusters is greater than -0.41 this illustrates that the model improves upon a naive (mean value) model. Figure 7 and 9 show predicted values of soil moisture on a sample HRU compared to the true values. These figures illustrate the logic behind our approach, where the TGCN corresponding to every cluster is being trained to predict different trends in soil moisture, analogous to GRUs sharing model parameters in traditional hydrological modeling.

Table 2 shows a comparison of the prediction mean squared error for multi-step forecasting. The average mean squared error of LSTM across all clusters is 0.4584 with a standard deviation of 0.2179. Whereas, the average mean squared error of our method across all clusters is 0.0480 with a standard deviation of 0.0165. The p value of null-hypothesis $\mathcal{H}$ is 6.5e-6, which shows that the reduction in predicted mean squared error of our model compared to LSTM is statistically significant. Figure 6 shows that on average our method improves upon a naive (mean value) model.

Discussion

Soil moisture estimation and prediction are critical to climate-aware agriculture. Resolving these predictions requires a comprehensive assessment of heterogeneous spatial and temporal features. While well-established physics-based approaches exist, they are hindered by their high user
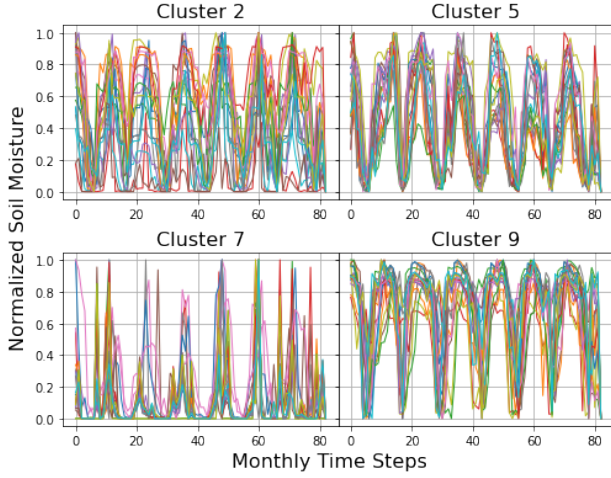
Figure 4: Plot of true soil moisture values of 20 randomly sub-sampled HRUs in cluster 2,5,7, and 9 for time steps in the test set. Soil moisture in different clusters exhibits distinct seasonal trend.

Cluster ID	LSTM MSE	C+TGCN MSE	Relative MSE Reduction
1	0.3433	0.0549	82.93%
2	0.3815	0.0573	84.93%
3	0.3588	0.0523	86.06%
4	0.3057	0.0610	79.60%
5	0.3677	0.0527	86.06%
6	0.4087	0.0543	86.19%
7	0.7326	0.0417	94.29%
8	0.4010	0.0393	91.10%
9	0.4227	0.0560	87.42%
10	0.3847	0.0591	83.43%

Table 2: Mean Squared Error (MSE) for multi-step soil moisture forecast across clusters using Clustering and TGCN (C+TGCN), compared with the LSTM model.

Cluster ID	LSTM MSE	C+TGCN MSE	Relative MSE Reduction
1	0.3433	0.0332	90.34%
2	0.3815	0.0328	91.41%
3	0.3588	0.0283	92.12%
4	0.3057	0.0399	86.95%
5	0.3677	0.0307	91.64%
6	0.4087	0.0321	92.14%
7	0.7326	0.0389	94.69%
8	0.4010	0.0217	94.58%
9	0.4227	0.0383	90.93%
10	0.3847	0.0335	91.30%

Table 1: Mean Squared Error (MSE) for single soil moisture forecast across clusters using TGCN, compared with LSTM model.

Figure 6: The plot shows the mean and standard deviation of Kling-Gupta Efficiency for each cluster for the multi-step forecast. $KGE > -0.41$ shows that the TGCN models improve upon the mean benchmark.
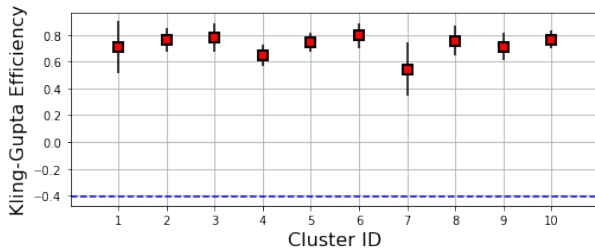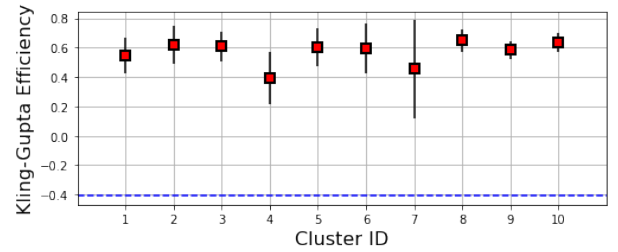

Figure 5: The plot shows the mean and standard deviation of Kling-Gupta Efficiency for each cluster for the single forecast. $KGE > -0.41$ shows that the TGCN models improve upon the mean benchmark.
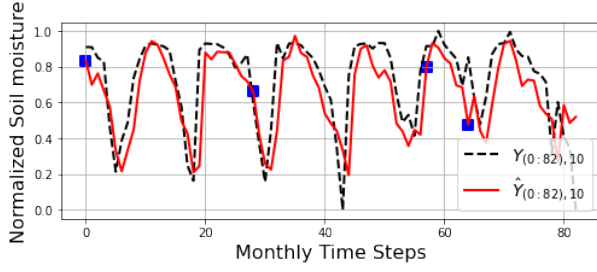
Figure 7: Predicted vs true soil moisture value for single forecast for a sample HRU(id=10) in cluster 10 from test data set. Blue boxes represent randomly sampled time steps for which multi-step results are plotted in Figure 8.
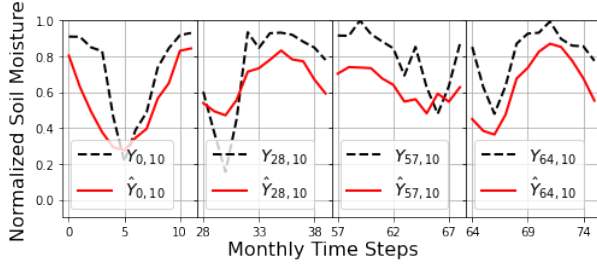

Figure 8: Predicted v.s. true soil moisture for multi-step forecast for a sample HRU(id=10) in cluster 10 from test data set.
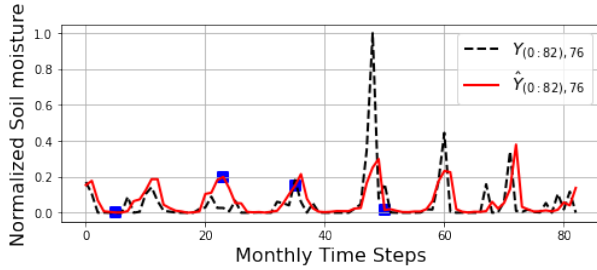

Figure 9: Predicted vs true soil moisture value for single forecast for a sample HRU(id=76) in cluster 7 from test data set. Blue boxes represent randomly sampled time steps for which multi-step results are plotted in Figure 10.
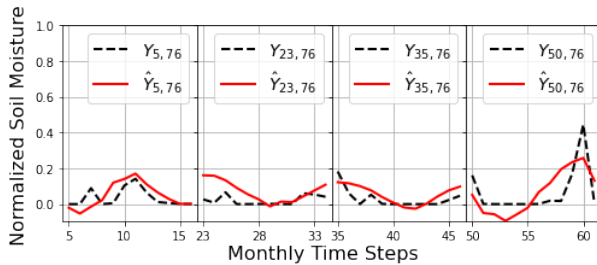

Figure 10: Predicted v.s. true soil moisture for multi-step forecast for a sample HRU(id=76) in cluster 7 from test data set.

complexity and computational expense to deploy at scale for commodity use cases. In practical terms, they are the domain of academic institutions and government organizations.

This paper describes a machine learning framework that borrows concepts from hydrological modelling to improve predictive skill and ease interpretability. There is a large volume of literature related to applying physics-informed constraints to ML which is discussed earlier. The objective in many of those studies is to augment the models with data external to the training data via methods such as modified loss functions (Daw et al. 2020), data augmentation (James, Zhang, and O'Donncha 2018), or specifying consensus filters to guide disparate models or data towards convergence (Haehnel et al. 2020).

The proposed methodology presents a natural framework to ingest information external to a time series signal positing the opportunity to enhance learning. Results demonstrate large increase in predictive skill provided by the GNN framework. Conventional techniques such as LSTM are commonly used for soil moisture prediction (Li et al. 2022). However, these approaches treat different locations independently and fail to exploit spatial dependencies.

Vyas and Bandyopadhyay (2022) described a GNN approach to forecast soil moisture based on Dynamic Graph Learning. At each timestep graph topology is updated based on a smoothness regularizer that evaluated dissimilarity for both node features and labels. Regularized dynamic graph updating have demonstrated improved model prediction in general cases (Chen et al. 2020). However, for soil moisture prediction, graph connectivity can be more effectively informed based on a systematic quantification of static and dynamic catchment attributes. Due to the high spatial and temporal heterogeneity dynamic updating can lead to spurious correlations based on synoptic similarity between features or labels. This is exacerbated by the long heterogeneous memory of soil moisture concentration. For example, the soil moisture at a point depends on weather processes together with previous moisture values over a specific window. The length of the historic window is highly dependent on local factors such as soil types, vegetation cover, and slope. For example, clay soils will have longer moisture retention than sandy soils. To accurately represent these dynamics, graph topology need to consider hydrological processes and their implications rather than individual physical descriptors.

A prominent body of literature has explored the combination of CNN and LSTM frameworks to resolve spatiotemporal processes. (e.g. (Xingjian et al. 2015; Yang et al. 2017)). These provide an intuitive and pragmatic approach to incorporate these information dimensions but are generally constrained to data on a consistent spatial grid. Applications have exclusively focused on gridded data such as satellite measurements, radar observations, and numerical model reanalysis products. Our proposed GNN framework adapts naturally to the characteristics of hydrological data. Individual polygons or hydrological response units are characterised based on their specific properties and informs a message passing between different regions based on similarity. Further, our approach provides a direct fit to modern Internet of Things (IoT) sensor networks that are typically of limited

spatial dimension but have complex (often time-lagged) dependencies between neighboring sensors. With information on the hydrological features, a graph topology can be constructed connecting different sensors.

Conclusion

This paper describes a spatiotemporal soil moisture prediction framework. Robust, high-resolution estimates are critical to most aspects of farm management, including: planting and harvesting scheduling, drought and irrigation management, and informing insurance risk and coverage. Creating a graph topology based on similarity metrics rather than the physical stream network and topography improved prediction performance by 70–90%. Further decoupling the graph topology from spatial relationships improves the generalizability of the framework. The approach can be applied on regions that share properties such as climate, soil features, and vegetation regardless of spatial proximity. This has the attractive property that data from regions with well developed monitoring programs can inform predictions in other locations or geographies. Estimating and forecasting soil moisture in ungauged basins is one of the great challenges of hydrology. This implicit form of parameters sharing enabled by the spatially decoupled graph network is a valuable contribution to this ambition.

References

Abbaspour, K. C. 2013. Swat-cup 2012. *SWAT calibration and uncertainty program—A user manual*.

Arnold, J. G.; Moriasi, D. N.; Gassman, P. W.; Abbaspour, K. C.; White, M. J.; Srinivasan, R.; Santhi, C.; Harmel, R.; Van Griensven, A.; Van Liew, M. W.; et al. 2012. SWAT: Model use, calibration, and validation. *Transactions of the ASABE*, 55(4): 1491–1508.

Bank, W. 2016. World Bank group climate change action plan.

Bruna, J.; Zaremba, W.; Szlam, A.; and LeCun, Y. 2014. Spectral Networks and Locally Connected Networks on Graphs. In Bengio, Y.; and LeCun, Y., eds., *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*.

Calzadilla, A.; Rehdanz, K.; Betts, R.; Falloon, P.; Wiltshire, A.; and Tol, R. S. 2013. Climate change impacts on global agriculture. *Climatic change*, 120(1): 357–374.

Chen, D.; Lin, Y.; Li, W.; Li, P.; Zhou, J.; and Sun, X. 2020. Measuring and relieving the over-smoothing problem for graph neural networks from the topological view. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 3438–3445.

Chen, R. T.; Rubanova, Y.; Bettencourt, J.; and Duvenaud, D. K. 2018. Neural ordinary differential equations. In *Advances in neural information processing systems*, volume 31.

Daw, A.; Thomas, R. Q.; Carey, C. C.; Read, J. S.; Appling, A. P.; and Karpatne, A. 2020. Physics-guided architecture (pga) of neural networks for quantifying uncertainty in lake temperature modeling. In *Proceedings of the 2020 siam international conference on data mining*, 532–540. SIAM.

Devia, G. K.; Ganasri, B. P.; and Dwarakish, G. S. 2015. A review on hydrological models. *Aquatic procedia*, 4: 1001–1007.

Doherty, J. 2003. *Groundwater*, 41(2): 170–177.

ElSaadani, M.; Habib, E.; Abdelhameed, A. M.; and Bayoumi, M. 2021. Assessment of a Spatiotemporal Deep Learning Approach for Soil Moisture Prediction and Filling the Gaps in Between Soil Moisture Observations. *Frontiers in Artificial Intelligence*, 4.

Gao, H.; Tang, Q.; Shi, X.; Zhu, C.; Bohn, T.; Su, F.; Pan, M.; Sheffield, J.; Lettenmaier, D.; and Wood, E. 2010. Water budget record from variable infiltration capacity (VIC) model. In *Algorithm Theoretical Basis Document for Terrestrial Water Cycle Data Records*.

Gassman, P. W.; Reyes, M. R.; Green, C. H.; and Arnold, J. G. 2007. The soil and water assessment tool: historical development, applications, and future research directions. *Transactions of the ASABE*, 50(4): 1211–1250.

Graham, D. N.; and Butts, M. B. 2005. Flexible, integrated watershed modelling with MIKE SHE. *Watershed models*, 849336090: 245–272.

Haehnel, P.; Mareček, J.; Monteil, J.; and O'Donncha, F. 2020. Using deep learning to extend the range of air pollution monitoring and forecasting. *Journal of Computational Physics*, 408: 109278.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In *2015 IEEE International Conference on Computer Vision (ICCV)*, 1026–1034.

Hsu, K.-l.; Gupta, H. V.; and Sorooshian, S. 1995. Artificial neural network modeling of the rainfall-runoff process. *Water resources research*, 31(10): 2517–2530.

Itzhaky, R. 2021. How AI will solve agriculture's water efficiency problems. https://www.weforum.org/agenda/2021/01/ai-agriculture-water-irrigation-farming/.

James, S.; Zhang, Y.; and O'Donncha, F. 2018. A machine learning framework to forecast wave conditions. *Coastal Engineering*, 137.

Kingma, D. P.; and Ba, J. 2015. Adam: A Method for Stochastic Optimization. In *ICLR (Poster)*.

Kipf, T. N.; and Welling, M. 2016. Semi-Supervised Classification with Graph Convolutional Networks. *CoRR*, abs/1609.02907.

Klingler, C.; Schulz, K.; and Herrnegger, M. 2021. LamaH— Large-Sample Data for Hydrology and Environmental Sciences for Central Europe. *Earth System Science Data Discussions*, 1–46.

Knoben, W. J. M.; Freer, J. E.; and Woods, R. A. 2019. Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, 23(10): 4323–4331.

Kouwen, N.; Soulis, E.; Pietroniro, A.; JR, D.; and Harrington, R. 1993. Grouped Response Units for Distributed Hydrologic Modeling. *Journal of Water Resources Planning*

and Management-asce - J WATER RESOUR PLAN MAN-ASCE, 119.

Kratzert, F.; Herrnegger, M.; Klotz, D.; Hochreiter, S.; and Klambauer, G. 2019a. NeuralHydrology–interpreting LSTMs in hydrology. In *Explainable AI: Interpreting, explaining and visualizing deep learning*, 347–362. Springer.

Kratzert, F.; Klotz, D.; Herrnegger, M.; Sampson, A. K.; Hochreiter, S.; and Nearing, G. S. 2019b. Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research*, 55(12): 11344–11354.

Leavesley, G.; Lichty, R.; Troutman, B.; and Saindon, L. 1983. Precipitation-runoff modeling system: User's manual. *Water-resources investigations report*, 83: 4238.

Li, Q.; Zhu, Y.; Shangguan, W.; Wang, X.; Li, L.; and Yu, F. 2022. An attention-aware LSTM model for soil moisture and soil temperature prediction. *Geoderma*, 409: 115651.

Lin, P.; Rajib, M. A.; Yang, Z.-L.; Somos-Valenzuela, M.; Merwade, V.; Maidment, D. R.; Wang, Y.; and Chen, L. 2018. Spatiotemporal evaluation of simulated evapotranspiration and streamflow over Texas using the WRF-Hydro-RAPID modeling framework. *JAWRA Journal of the American Water Resources Association*, 54(1): 40–54.

Muñoz-Sabater, J.; Dutra, E.; Agustí-Panareda, A.; Albergel, C.; Arduini, G.; Balsamo, G.; Boussetta, S.; Choulga, M.; Harrigan, S.; Hersbach, H.; et al. 2021. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13(9): 4349–4383.

Nearing, G.; Kratzert, F.; Klotz, D.; Hoedt, P.-J.; Klambauer, G.; Hochreiter, S.; Gupta, H.; Nevo, S.; and Matias, Y. 2020. A Deep Learning Architecture for Conservative Dynamical Systems: Application to Rainfall-Runoff Modeling. *AI for Earth Sciences Workshop at NeurIPS*.

Nearing, G. S.; Kratzert, F.; Sampson, A. K.; Pelissier, C. S.; Klotz, D.; Frame, J. M.; Prieto, C.; and Gupta, H. V. 2021. What role does hydrological science play in the age of machine learning? *Water Resources Research*, 57(3): e2020WR028091.

Newman, A.; Clark, M.; Sampson, K.; Wood, A.; Hay, L.; Bock, A.; Viger, R.; Blodgett, D.; Brekke, L.; Arnold, J.; et al. 2015. Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrology and Earth System Sciences*, 19(1): 209–223.

Pierce, F. J.; and Nowak, P. 1999. Aspects of precision agriculture. *Advances in agronomy*, 67: 1–85.

Poblete, D.; Arevalo, J.; Nicolis, O.; and Figueroa, F. 2020. Optimization of Hydrologic Response Units (HRUs) Using Gridded Meteorological Data and Spatially Varying Parameters. *Water*, 12(12): 3558.

Prasad, V. H. 2005. Delineation of Hydrologic Response Units (HRUs) using Remote Sensing and GIS. *Water and Energy Abstracts*, 15(1).

Rao, C.; Sun, H.; and Liu, Y. 2020. Physics informed deep learning for computational elastodynamics without labeled data. *arXiv preprint arXiv:2006.08472*.

Sakoe, H.; and Chiba, S. ???? In *Proceedings of the Seventh International Congress on Acoustics*. Budapest.

Sakoe, H.; and Chiba, S. 1978. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1): 43–49.

Shamshirband, S.; Hashemi, S.; Salimi, H.; Samadianfard, S.; Asadi, E.; Shadkani, S.; Kargar, K.; Mosavi, A.; Nabipour, N.; and Chau, K.-W. 2020. Predicting standardized streamflow index for hydrological drought using machine learning models. *Engineering Applications of Computational Fluid Mechanics*, 14(1): 339–350.

Shen, C. 2018. A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research*, 54(11): 8558–8593.

Srinivasan, R.; and Balmer, C. 2021. SWAT Literature Database for Peer-Reviewed Journal Articles.

Tavenard, R.; Faouzi, J.; Vandewiele, G.; Divo, F.; Androz, G.; Holtz, C.; Payne, M.; Yurchak, R.; Rußwurm, M.; Kolar, K.; and Woods, E. 2020. Tslearn, A Machine Learning Toolkit for Time Series Data. *Journal of Machine Learning Research*, 21(118): 1–6.

Vyas, A.; and Bandyopadhyay, S. 2022. Dynamic Structure Learning through Graph Neural Network for Forecasting Soil Moisture in Precision Agriculture. In *Proceedings of the 2022 International Joint Conference on Artificial Intelligence*, 5185–5191. IJCAI.

Xia, Y.; Mitchell, K.; Ek, M.; Sheffield, J.; Cosgrove, B.; Wood, E.; Luo, L.; Alonge, C.; Wei, H.; Meng, J.; et al. 2012. Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of Geophysical Research: Atmospheres*, 117(D3).

Xingjian, S.; Chen, Z.; Wang, H.; Yeung, D. Y.; Wong, W. K.; and Woo, W. C. 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*, 802–810.

Yang, Y.; Dong, J.; Sun, X.; Lima, E.; Mu, Q.; and Wang, X. 2017. A CFCC-LSTM model for sea surface temperature prediction. *IEEE Geoscience and Remote Sensing Letters*, 15(2): 207–211.

Young, C.-C.; Liu, W.-C.; and Wu, M.-C. 2017. A physically based and machine learning hybrid approach for accurate rainfall-runoff modeling during extreme typhoon events. *Applied Soft Computing*, 53: 205–216.

Zhang, N.; Wang, M.; and Wang, N. 2002. Precision agriculture—a worldwide overview. *Computers and electronics in agriculture*, 36(2-3): 113–132.

Zhao, L.; Song, Y.; Zhang, C.; Liu, Y.; Wang, P.; Lin, T.; Deng, M.; and Li, H. 2020. T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction. *IEEE Transactions on Intelligent Transportation Systems*, 21(9): 3848–3858.