Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23)
Special Track on AI for Good

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

Muneeza Azmat[1] , **Malvern Madondo**[1] , **Arun Bawa**[2] , **Kelsey Dipietro**[3] , **Raya Horesh**[1] , **Michael Jacobs**[4] , **Raghavan Srinivasan**[2] , **Fearghal O'Donncha**[1]
[1]IBM Research
[2]Texas A&M
[3]Sandia National Lab
[4]IBM Corporate Social Responsibility
azmatmun@msu.edu, malvern.madondo@emory.edu, arun.bawa@ag.tamu.edu, kdipiet@sandia.gov, rhoresh@us.ibm.com, Michael.Jacobs@ibm.com, r-srinivasan@tamu.edu, feardonn@ie.ibm.com

Abstract

Agriculture faces unprecedented challenges due to climate change, population growth, and water scarcity. These challenges highlight the need for efficient resource usage to optimize crop production. Conventional techniques for forecasting hydrological response features, such as soil moisture, rely on physics-based and empirical hydrological models, which necessitate significant time and domain expertise. Drawing inspiration from traditional hydrological modeling, a novel temporal graph convolution neural network has been constructed. This involves grouping units based on their time-varying hydrological properties, constructing graph topologies for each cluster based on similarity using dynamic time warping, and utilizing graph convolutions and a gated recurrent neural network to forecast soil moisture. The method has been trained, validated, and tested on field-scale time series data spanning 40 years in northeastern United States. Results show that using domain-inspired clustering with time series graph neural networks is more effective in forecasting soil moisture than existing models. This framework is being deployed as part of a pro bono social impact program that leverages hybrid cloud and AI technologies to enhance and scale non-profit and government organizations. The trained models are currently being deployed on a series of small-holding farms in central Texas.

1 Introduction

Over the past decade, machine learning (ML) has transformed various domains, but it has had little impact on crucial areas such as quantifying and predicting water availability for agricultural purposes. Ground- and surface-water availability depends on various temporal and spatial factors such as precipitation volumes, heterogeneous sources of runoff, evapotranspiration, and water losses. Due to the large geographical extents and high variability, collecting enough observation data for implementing IoT-based decision support systems is infeasible.

Traditionally, engineers have relied on physics-based models that represent hydrological processes as a set of partial differential equations constrained by heuristics, empirical relationships, and expert intuition. While these allow 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].

Precision agriculture approaches [Zhang *et al.*, 2002] have developed over the past four decades by combining simulation, satellite, and sensor data to improve decision making. Precision agriculture accomplishments are related to how well they can be applied to assess, manage, and evaluate crop production [Pierce and Nowak, 1999]. Climate change introduces a completely different set of challenges that require more granular data, and a holistic decision making approach 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 requires the adoption of climate-smart agriculture practices that minimize resource consumption and environmental impacts, while simultaneously ensuring food security for growing populations. Globally, while large farms increasingly digitize operations to enhance sustainability, small-holding farmers lack the skills and resources to leverage AI and IoT-backed 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" [Forum, 2021]. However democratization of these solutions to small-holding and disadvantaged farmers requires scalable machine learning models that can be informed by publicly-available datasets and sparse low-cost sensor data.

This paper describes a novel domain-inspired framework to forecast soil moisture. The proposed framework uses temporal graph convolutional neural networks (TGCN) 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, a well-established hydrological modeling technique, introduced by [Kouwen *et al.*, 1993]. A Group response unit is composed of groups of hydrological response units (HRUs) that have similar hydrological characteristics and consequently have more comparable hydrological response than neighboring HRUs which might have different characteristics (e.g. crop versus livestock farming). The proposed framework is applied to predict soil moisture for a case study application in Northeastern United States.

The contributions of this paper are as follows:

- We describe a novel domain-inspired temporal graph convolution neural network. Analogous to group response units, 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.
- We present 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.

2 Related Work

Recent advancements in machine learning have 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 *et al.*, 1995; Kratzert *et al.*, 2019; 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. 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 *et al.*, 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.*, 2009]. 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.*, 2019]. 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 *et al.*, 2021] other regions are still constrained by data limitations.

Many studies have proposed ML methods to represent the spatiotemporal properties of geophsysical systems. The most widely used frameworks combine convolutional neural 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 physical laws 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 *et al.*, 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.

3 Methods

3.1 Data

[Leavesley *et al.*, 1983] introduced the decomposition of watersheds into sub-areas that are assumed to share land-surface characteristics, termed hydrologic response units (HRUs). The HRUs are characterized using topographic variables, such as elevation and 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 the Soil and Water Assessment Tool (SWAT) [Gassman *et al.*, 2007]. SWAT is a 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 reanalysis 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 $9\,\mathrm{km}$ [Muñoz-Sabater *et al.*, 2021] and $14\,\mathrm{km}$ [Xia *et al.*, 2012], respectively. Agriculture, on the other hand, requires predictions that resolve field-scale ($< 500\,\mathrm{m}$) processes.

The Hydrological and Water Quality System (HAWQS) v2.0 [Chen *et al.*, 2020b] (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 model-

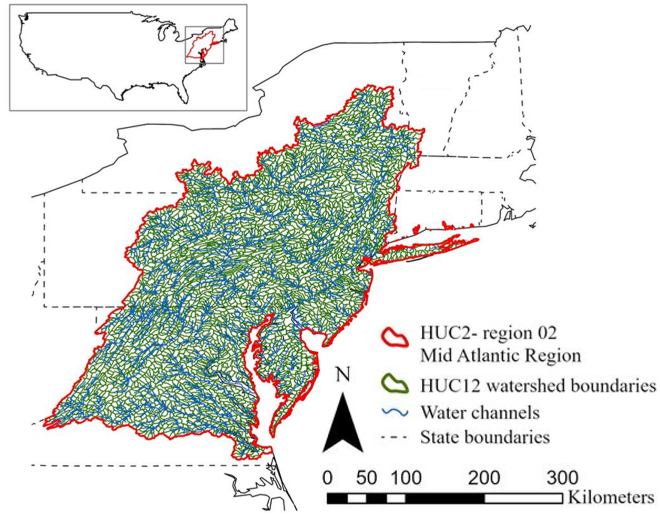

Figure 1: Layout of the Mid-Atlantic basin along with its stream network and HUC12 watersheds.

ing 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 information about the features associated with each HRU is included in the supplementary materials. Monthly data is available for each feature spanning 34 years.

3.2 Problem Formulation

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

3.3 Domain Inspired Clustering

Inspired by the concept of group response units, we build a clustering module to group HRUs that have similar hydrological characteristics. Traditionally, group response units are constructed based on climate, land use, soil and pedotransfer properties [Poblete *et al.*, 2020]. The use of group response units 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

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 $\mathbf{y}$

```
initialize C = inf ∈ R^{n×n}
C_{0,0} = 0
for i : 0 → s do
    for j : 0 → s do
        dist = d(x_i, y_j)^2
        C_{i,j} = dist + min(C_{i−1,j}, C_{i,j−1}, C_{i−1,j−1})
    end for
end for
DTW(x, y) = sqrt(C_{s,s})
return DTW(x, y)
```

hydrological features to inform clustering. First introduced in [Sakoe and Chiba, 1978], dynamic time warping (DTW) quantifies 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 $D_{\text{p,q}}$ between two HRUs p and q is given by

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

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

3.4 Temporal Graph Convolution Neural Network (TGCN)

Graph convolution neural networks are an extension of convolution neural networks to unstructured graph data [Kipf and Welling, 2016]. 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. 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. Figure 2 describes the information flow of a single cell of the TGCN.

We minimize the mean squared error loss during training.

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

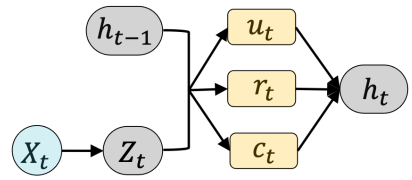

Figure 2: Schematic of a single cell of TGCN, equations (2-6).

Figure 3 summarizes the model architecture that groups similar HRUs and implements a TGCN prediction framework.

4 Results

We train 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. Results are benchmarked against an LSTM model and a distance-based TGCN model that clusters solely based on a single static measure of similarity (hydrological curve number).

4.1 Evaluation Metrics

We evaluate all models using mean squared error (MSE) and Kling-Gupta Efficiency (KGE). We also calculate the relative percent decrease in MSE to compare model performance. KGE is a traditional metric used in hydrology to evaluate model performance and is expressed as:

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

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 *et al.*, 2019]. As model becomes more accurate, $KGE \to 1$.

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. The null hypotheses $\mathcal{H}$ states that TGCN MSE has identical average values as LSTM MSE. For probability less than 0.05, we reject the null hypothesis.

4.2 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.

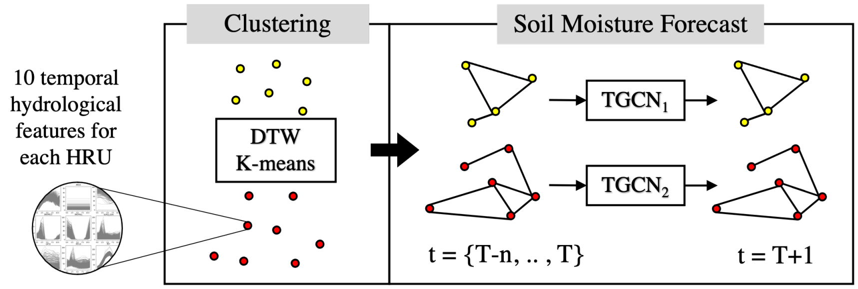

Figure 3: Schematic of our Clustering and Temporal Graph Convolution Neural network (C+TGCN) approach for soil moisture forecast.

To estimate the number of clusters, we employ an elbow test and only use training data to prevent data leakage. After dividing the HRUs into clusters, we construct graph topology by using the DTW distances of HRUs within each cluster. The resulting static graphs represent disjoint subsets of HRUs and are independently trained using TGCN with the same model architecture. The model comprises a layer of graph convolution followed by a linear transform. The output from the linear layer is fed into the GRU, which produces the forecast Y_t. For the multi-step model, the GRU generates a sequence of 12 predictions for each node. We trained all TGCNs using the Adam optimizer [Kingma and Ba, 2015] with a learning rate of 1e-2 for approximately 100 epochs (until validation loss stopped decreasing). The eights 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 memory.

In order to establish the added benefit of graph topology and DTW clustering we conduct a comparative evaluation of our model's performance against the following baseline models

$LSTM_N$

The Long Short-Term Memory (LSTM) model is trained on the entirety of the HRU data and is agnostic to any clustering information. The subscript N represents the forecast length. This constitutes the most elementary model, lacking any graph topology or clustering information.

$C_{DTW} + LSTM_N$

This model entails separate training of LSTM models for each cluster. As a result, despite the absence of explicit graph topology, the model effectively capitalizes on the hydrological attributes specific to each cluster.

$C_S + TGCN_N$

Similar to our proposed method, this model also uses TGCNs. However, the distinction lies in the manner in which graph topology is introduced. In this case, clustering is performed based on a static future (average hydrological curve number), effectively disregarding any seasonal trends.

Code for model setup, training and evaluation is available at https://github.com/IBM/tgcn-soil-moisture.

4.3 Soil Moisture Forecast Results

Table 1 compares performance of our proposed algorithm against the baseline methods. The test MSE is computed for each cluster, and the mean and standard deviation values across all clusters are reported. The addition of HRU connectivity through graphs and seasonal information via DTW clustering leads to an improvement in the performance of both single-step and multi-step forecast models. This improvement remains consistent across the models.

Table 2 shows the average mean squared error of predicted soil moisture in each cluster. MSE reduces across all clusters compared to the LSTM model. Figure 4 shows the mean and standard deviation of KGE for all HRUs in each cluster. KGE is between between 0.4–0.8, indicating an effective model. All instances report values greater than -0.41 illustrating clear improvement upon a naive model.

Time trace plots of predicted and true values of soil moisture on a sample HRU are displayed in Figures 5 – 8. One-step ahead forecasts are presented in Figures 5 and 7 while 12-step ahead forecasts are presented in Figures 6 and 8. These figures illustrate the rationale behind our approach, in which TGCNs corresponding to each cluster are trained to predict different soil moisture trends, similar to group response units that share model parameters in conventional hydrological modeling.

Table 3 compares prediction error for multi-step forecasting. Across all clusters, LSTM has an average MSE of 0.4584 with a standard deviation of 0.2179, while our method has an average MSE of 0.0480 with a standard deviation of 0.0165. The null-hypothesis $\mathcal{H}$ has a p value 6.5e-6, indicating that the decrease in MSE of our model compared to LSTM is statistically significant. Figure 4 demonstrates that our method outperforms a naïve model.

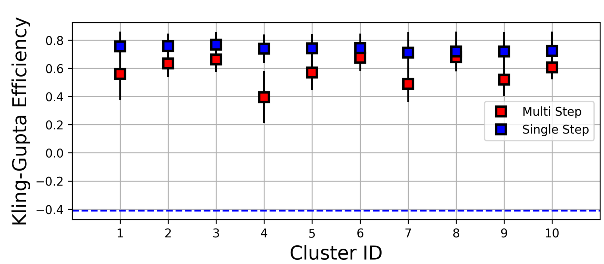

Figure 4: The plot shows the mean and standard deviation of Kling-Gupta Efficiency for each cluster for single and multistep forecast. $KGE > -0.41$ shows that the $C_{DTW} + TGCN$ models improve upon the mean benchmark.

Clustering		Graph structure	Forecast Length		Model name	Mean Test MSE	Std Test MSE
Static	DTW		1 mo	12 mo			
			✓		$LSTM_1$	0.384	0.096
	✓		✓		$C_{DTW} + LSTM_1$	0.394	0.061
✓		✓	✓		$C_S + TGCN_1$	0.086	0.059
	✓	✓	✓		$C_{DTW} + TGCN_1$	0.033	0.005
				✓	$LSTM_{12}$	0.372	0.419
	✓			✓	$C_{DTW} + LSTM_{12}$	0.458	0.218
✓		✓		✓	$C_S + TGCN_{12}$	0.079	0.040
	✓	✓		✓	$C_{DTW} + TGCN_{12}$	0.048	0.016

Table 1: Performance evaluation of our proposed approach ($C_{DTW} + TGCN$) relative to various baseline methods. The assessed architectures incorporate a fusion of graph and clustering techniques to demonstrate the advantages of utilizing dynamic time warping clustering and graph topology. The outcomes are computed by averaging the test performance for each cluster and reporting the mean and standard deviation of the MSE.

Cluster ID	$C_{DTW}+$ $LSTM_1$ **MSE**	$C_{DTW} + TGCN_1$ **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 2: Mean Squared Error (MSE) for single soil moisture forecast across DTW-clusters using TGCN, compared with LSTM model.

Cluster ID	$C_{DTW}+$ $LSTM_{12}$ **MSE**	$C_{DTW} + TGCN_{12}$ **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 3: Mean Squared Error (MSE) for multi-step soil moisture forecast across clusters using DTW-Clustering and TGCN, compared with the LSTM model.

5 Discussion

Accurate estimation and prediction of soil moisture are crucial for climate-smart agriculture and necessitate a thorough evaluation of diverse spatial and temporal features. Although physics-based approaches are established, they have limitations due to their significant user complexity and computational expense, making them unsuitable for widespread adoption in commodity use cases. As a result, these approaches are primarily employed by academic institutions and government organizations.

This paper introduces a machine learning framework that draws on hydrological modeling principles to enhance predictive accuracy and interpretability. Several studies have examined the application of physics-based constraints to machine learning, as previously discussed. These studies typically aim to incorporate external data into the models through techniques like modified loss functions [Daw *et al.*, 2020], data augmentation [James *et al.*, 2018], or specifying consensus filters to guide disparate models or data towards convergence [Haehnel *et al.*, 2020].

The proposed methodology offers a viable framework for incorporating external information into time series signals, with the potential to improve learning outcomes. The GNN framework results in a significant improvement in predictive accuracy. Conventional techniques, like LSTM, are widely used for soil moisture prediction, [Li *et al.*, 2022], but they treat distinct locations as independent and do not leverage 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.*, 2020a]. 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 hetereogeneity dynamic updating can lead to spurious correlations based on synoptic similarity between fea-

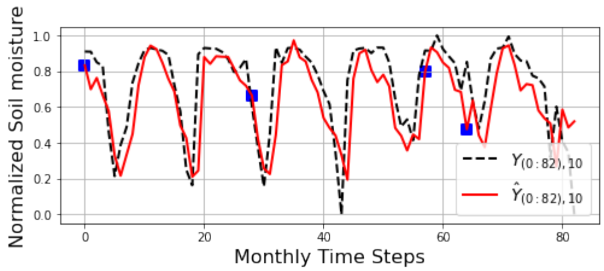

Figure 5: Predicted vs true soil moisture value for single forecast for a sample HRU(id=10) in cluster 10 from test data set. Y represents ground truth, $\hat{Y}$ represents predicted soil moisture and blue boxes represent randomly sampled time steps for which multi-step results are plotted in Figure 6.

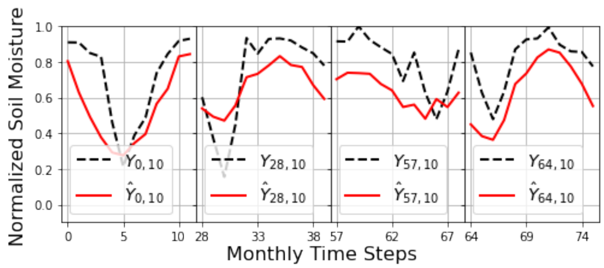

Figure 6: Predicted ($\hat{Y}$) v.s. true (Y) soil moisture for multi-step forecast for a sample HRU(id=10) in cluster 10 from test data set.

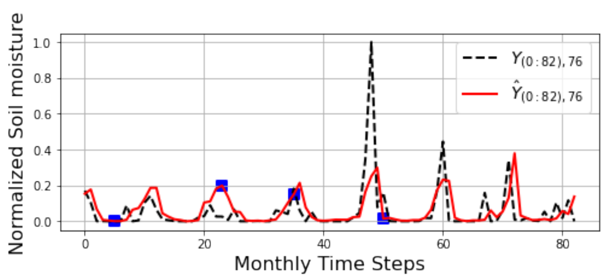

Figure 7: Predicted vs true soil moisture value for single forecast for a sample HRU(id=76) in cluster 7 from test data set. Y represents ground truth, $\hat{Y}$ represents predicted soil moisture and blue boxes represent randomly sampled time steps for which multi-step results are plotted in Figure 8.

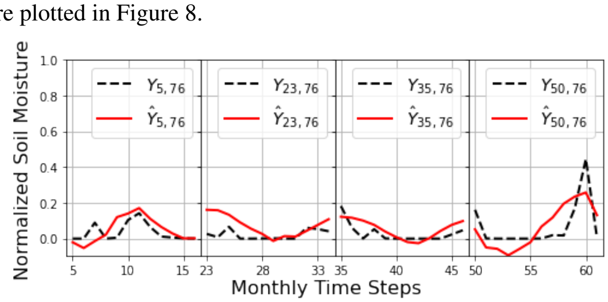

Figure 8: Predicted ($\hat{Y}$) v.s. true (Y) soil moisture for multi-step forecast for a sample HRU(id=76) in cluster 7 from test data set.

tures 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. These provide an intuitive and pragmatic approach to incorporate both 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.

Additionally, our approach provides a direct fit to modern Internet of Things (IoT) sensor networks which are generally limited in spatial scope but exhibit intricate, time-lagged interdependencies among adjacent sensors. By utilizing hydrological feature data, a graph topology can be established to connect diverse sensors based on established physics-based correlations.

6 Conclusion

Robust, high-resolution soil moisture 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 to regions that share properties such as climate, soil features, and vegetation, regardless of spatial proximity, allowing data from well-monitored regions to inform predictions in under-monitored regions.

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. Informed by well-established hydrological understanding and using a computationally efficient TGCN, the framework is particularly applicable for regions with limited computational resources or observation data. This is particularly important in hydrology where collecting high-quality data is both time consuming and expensive.

References

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

[Arnold *et al.*, 2012] Jeffrey G Arnold, Daniel N Moriasi, Philip W Gassman, Karim C Abbaspour, Michael J White, Raghavan Srinivasan, Chinnasamy Santhi, RD Harmel, Ann Van Griensven, Michael W Van Liew, et al. Swat: Model use, calibration, and validation. *Transactions of the ASABE*, 55(4):1491–1508, 2012.

[Bank, 2016] World Bank. World bank group climate change action plan 2016-2020. http://hdl.handle.net/10986/24451, 2016. World Bank, Washington, DC.

[Bruna *et al.*, 2014] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. In Yoshua Bengio and Yann LeCun, editors, *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.

[Calzadilla *et al.*, 2013] Alvaro Calzadilla, Katrin Rehdanz, Richard Betts, Pete Falloon, Andy Wiltshire, and Richard SJ Tol. Climate change impacts on global agriculture. *Climatic change*, 120(1):357–374, 2013.

[Chen *et al.*, 2018] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in neural information processing systems*, volume 31, 2018.

[Chen *et al.*, 2020a] Deli Chen, Yankai Lin, Wei Li, Peng Li, Jie Zhou, and Xu Sun. 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, pages 3438–3445, 2020.

[Chen *et al.*, 2020b] Manyu Chen, Philip W Gassman, Raghavan Srinivasan, Yuanlai Cui, and Raymond Arritt. Analysis of alternative climate datasets and evapotranspiration methods for the upper mississippi river basin using swat within hawqs. *Science of the Total Environment*, 720:137562, 2020.

[Daw *et al.*, 2020] Arka Daw, R Quinn Thomas, Cayelan C Carey, Jordan S Read, Alison P Appling, and Anuj Karpatne. 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*, pages 532–540. SIAM, 2020.

[Devia *et al.*, 2015] Gayathri K Devia, B Pa Ganasri, and G Sa Dwarakish. A review on hydrological models. *Aquatic procedia*, 4:1001–1007, 2015.

[Doherty, 2003] John Doherty. Ground water model calibration using pilot points and regularization. *Groundwater*, 41(2):170–177, 2003.

[ElSaadani *et al.*, 2021] Mohamed ElSaadani, Emad Habib, Ahmed M. Abdelhameed, and Magdy Bayoumi. 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, mar 2021.

[Forum, 2021] World Economic Forum. How ai can save water in agriculture, 2021.

[Gao *et al.*, 2009] H Gao, Q Tang, X Shi, C Zhu, T Bohn, F Su, J Sheffield, M Pan, D Lettenmaier, and EF Wood. Algorithm theoretical basis document "water budget record from variable infiltration capacity (vic) model". *Algorithm Theoretical Basis Document for Terrestrial Water Cycle Data Records*, 2009.

[Gassman *et al.*, 2007] Philip W Gassman, Manuel R Reyes, Colleen H Green, and Jeffrey G Arnold. The soil and water assessment tool: historical development, applications, and future research directions. *Transactions of the ASABE*, 50(4):1211–1250, 2007.

[Graham and Butts, 2005] Douglas N Graham and Michael B Butts. Flexible, integrated watershed modelling with mike she. *Watershed models*, 849336090:245–272, 2005.

[Haehnel *et al.*, 2020] Philipp Haehnel, Jakub Mareček, Julien Monteil, and Fearghal O'Donncha. Using deep learning to extend the range of air pollution monitoring and forecasting. *Journal of Computational Physics*, 408:109278, 2020.

[He *et al.*, 2015] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1026–1034, 2015.

[Hsu *et al.*, 1995] Kuo-lin Hsu, Hoshin Vijai Gupta, and Soroosh Sorooshian. Artificial neural network modeling of the rainfall-runoff process. *Water resources research*, 31(10):2517–2530, 1995.

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

[Kingma and Ba, 2015] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR (Poster)*, 2015.

[Kipf and Welling, 2016] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[Klingler *et al.*, 2021] Christoph Klingler, Karsten Schulz, and Mathew Herrnegger. Lamah— large-sample data for hydrology and environmental sciences for central europe. *Earth System Science Data Discussions*, pages 1–46, 2021.

[Knoben *et al.*, 2019] Wouter JM Knoben, Jim E Freer, and Ross A Woods. Inherent benchmark or not? comparing nash-sutcliffe and kling-gupta efficiency scores. *Hydrology and Earth System Sciences*, 23(10):4323–4331, 2019.

[Kouwen *et al.*, 1993] Nicholas Kouwen, Eric Soulis, Alain Pietroniro, Donald JR, and R. Harrington. Grouped response units for distributed hydrologic modeling. *Journal of Water Resources Planning and Management-asce - J WATER RESOUR PLAN MAN-ASCE*, 119, 05 1993.

[Kratzert *et al.*, 2019] Frederik Kratzert, Daniel Klotz, Mathew Herrnegger, Alden K Sampson, Sepp Hochreiter, and Grey S Nearing. Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research*, 55(12):11344–11354, 2019.

[Leavesley *et al.*, 1983] GH Leavesley, RW Lichty, BM Troutman, and LG Saindon. Precipitation-runoff modeling system: User's manual. *Water-resources investigations report*, 83:4238, 1983.

[Li *et al.*, 2022] Qingliang Li, Yuheng Zhu, Wei Shangguan, Xuezhi Wang, Lu Li, and Fanhua Yu. An attention-aware lstm model for soil moisture and soil temperature prediction. *Geoderma*, 409:115651, 2022.

[Lin *et al.*, 2018] Peirong Lin, Mohammad Adnan Rajib, Zong-Liang Yang, Marcelo Somos-Valenzuela, Venkatesh Merwade, David R Maidment, Yan Wang, and Li Chen. 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, 2018.

[Muñoz-Sabater *et al.*, 2021] Joaquín Muñoz-Sabater, Emanuel Dutra, Anna Agustí-Panareda, Clément Albergel, Gabriele Arduini, Gianpaolo Balsamo, Souhail Boussetta, Margarita Choulga, Shaun Harrigan, Hans Hersbach, et al. Era5-land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13(9):4349–4383, 2021.

[Nearing *et al.*, 2020] Grey Nearing, Frederik Kratzert, Daniel Klotz, Pieter-Jan Hoedt, Günter Klambauer, Sepp Hochreiter, Hoshin Gupta, Sella Nevo, and Yossi Matias. A deep learning architecture for conservative dynamical systems: Application to rainfall-runoff modeling. *AI for Earth Sciences Workshop at NeurIPS*, 2020.

[Nearing *et al.*, 2021] Grey S Nearing, Frederik Kratzert, Alden Keefe Sampson, Craig S Pelissier, Daniel Klotz, Jonathan M Frame, Cristina Prieto, and Hoshin V Gupta. What role does hydrological science play in the age of machine learning? *Water Resources Research*, 57(3):e2020WR028091, 2021.

[Newman *et al.*, 2015] AJ Newman, MP Clark, Kevin Sampson, Andrew Wood, LE Hay, A Bock, RJ Viger, D Blodgett, L Brekke, JR Arnold, et al. 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, 2015.

[Pierce and Nowak, 1999] Francis J Pierce and Peter Nowak. Aspects of precision agriculture. *Advances in agronomy*, 67:1–85, 1999.

[Poblete *et al.*, 2020] David Poblete, Jorge Arevalo, Orietta Nicolis, and Felipe Figueroa. Optimization of hydrologic response units (HRUs) using gridded meteorological data and spatially varying parameters. *Water*, 12(12):3558, December 2020.

[Prasad, 2005] V Hari Prasad. Delineation of hydrologic response units (hrus) using remote sensing and gis. *Water and Energy Abstracts*, 15(1), 2005.

[Rao *et al.*, 2020] Chengping Rao, Hao Sun, and Yang Liu. Physics informed deep learning for computational elastodynamics without labeled data. *arXiv preprint arXiv:2006.08472*, 2020.

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

[Shamshirband *et al.*, 2020] Shahabbodin Shamshirband, Sajjad Hashemi, Hana Salimi, Saeed Samadianfard, Esmaeil Asadi, Sadra Shadkani, Katayoun Kargar, Amir Mosavi, Narjes Nabipour, and Kwok-Wing Chau. Predicting standardized streamflow index for hydrological drought using machine learning models. *Engineering Applications of Computational Fluid Mechanics*, 14(1):339–350, 2020.

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

[Srinivasan and Balmer, 2021] Raghavan Srinivasan and Curtis Balmer. SWAT Literature Database for Peer-Reviewed Journal Articles, 2021.

[Vyas and Bandyopadhyay, 2022] Anoushka Vyas and Sambaran Bandyopadhyay. 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*, pages 5185–5191. IJCAI, 2022.

[Xia *et al.*, 2012] Youlong Xia, Kenneth Mitchell, Michael Ek, Justin Sheffield, Brian Cosgrove, Eric Wood, Lifeng Luo, Charles Alonge, Helin Wei, Jesse Meng, et al. 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), 2012.

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

[Yang *et al.*, 2017] Yuting Yang, Junyu Dong, Xin Sun, Estanislau Lima, Quanquan Mu, and Xinhua Wang. A cfcc-lstm model for sea surface temperature prediction. *IEEE Geoscience and Remote Sensing Letters*, 15(2):207–211, 2017.

[Zhang *et al.*, 2002] Naiqian Zhang, Maohua Wang, and Ning Wang. Precision agriculture—a worldwide overview. *Computers and electronics in agriculture*, 36(2-3):113–132, 2002.

[Zhao *et al.*, 2020] Ling Zhao, Yujiao Song, Chao Zhang, Yu Liu, Pu Wang, Tao Lin, Min Deng, and Haifeng Li. T-gcn: A temporal graph convolutional network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 21(9):3848–3858, 2020.