Uncertainty-guided Active Learning for Measurement Prediction in Sparsely Observed Agricultural Fields

Deepesh Agarwal[1], Aabila Tharzeen[1], Balasubramaniam Natarajan[1]

[1] Kansas State University, Manhattan, Kansas, USA

A paper from the Proceedings of the
16th International Conference on Precision Agriculture
21-24 July 2024
Manhattan, Kansas, United States

Abstract.

The sustainability of farming methods relies on the quality of soil health. Rich soil supplies vital nutrients to plants. Long-term and extensive monitoring of soil is crucial for obtaining important information regarding moisture and nutrient dynamics. Comprehending the dynamics of soil moisture and nutrients is particularly important to predict and improve agricultural productivity. Nevertheless, access to data related to soil health is frequently impeded by the absence of complete information from agricultural fields. Multiple factors, including number of available sensors, challenges in data collection, communication network disruptions, and maintenance problems contribute to the quality and quantity of data gathered. With only a few spatial locations sampled for measuring soil characteristics, data from agricultural fields tend to be spatially sparse. One of the critical challenges in precision agriculture is to determine how to exploit this spatially sparse data to obtain soil health awareness of the entire field and use that information for effective management. This work addresses this challenging problem via an innovative machine learning framework. Semi-supervised learning approaches like Active Learning (AL) are very beneficial for predicting measurements across the entire field by using spatially sparse data. This work proposes an AL-based methodology for prediction of soil information across the entire farm based on sparsely collected field measurements. The field measurement prediction problem is represented in the form of a graph. The learning process is then executed by using a combination of AL and state-of-the-art Graph Neural Network (GNN) techniques. A major advantage of the proposed methodology is that the information across both space and time domains are simultaneously captured within the learning scheme. The proposed framework also supports systematic propagation of uncertainty through the machine learning algorithm to finally obtain the uncertainty in model predictions, which is then leveraged as an AL heuristic. The proposed ML framework is applied to spatially sparse soil moisture and temperature data. Results demonstrate that even with a limited fraction of sensors and strategically selected measurement locations, it is possible to predict soil moisture/temperature over a large spatial area with over 90% accuracy.

Keywords.
Spatio-Temporal processes, Active Learning, Uncertainty, Graph Neural Network, Dilated Causal Convolutional Neural Network, Precision Agriculture.

Introduction

In modern agricultural practices, there is an environmental obligation to ensure efficient usage of water. Frequent monitoring of soil moisture and temperature can help farmers to optimize irrigation, reduce water usage and ensure that the plants receive the exact amount of water required (Gałęzewski et al., 2021). Accurately monitoring soil moisture and temperature can also be a valuable tool to detect changes in the soil environment that may indicate the presence of plant disease (Montgomery et al., 2024). Thus, frequent monitoring of soil moisture is critical for many applications and efficient agriculture management. Soil dynamics vary across both space and time, necessitating regular measurements for various applications. For instance, in regional planning, it is essential to have soil moisture data at different spatial and temporal scales, which require measurements at various depths and time intervals (Joshi & Mohanty, 2010). To effectively represent these variations, data can be well-represented in the form of a spatio-temporal graph, aiding in the quantitative monitoring of soil health.

Over the past decade, the use of machine learning (ML) algorithms such as deep learning (Saleem et al., 2021), support vector machines (Pereira et al., 2022), random forests (Basha et al., 2020) and so on has significantly increased in soil health research. This increase is attributed to their non-parametric nature and their ability to capture complex, non-linear relationships (Padarian et al., 2020). Recently, advanced ML methodologies like graph neural networks (Q. Zhang et al., 2022) and recurrent neural networks (Mythili & Al, 2021) have also been explored in order to efficiently capture spatial and temporal variations among the soil measurements, respectively. Despite the importance of soil health data, its availability from agricultural lands is often hindered by several factors, including the number of sensors and their reliability, challenges with data collection, maintenance issues, and inadequate network connectivity. Moreover, deploying large number of sensors in the field can be prohibitively expensive, resulting in spatially sparse data from the field (Sishodia et al., 2020).

Missing data creates substantial issues for ML-based analysis. Omitting missing data can lead to biased or incorrect conclusions, complicating the identification of temporal or spatial patterns (Y. Zhang & Thorburn, 2021). Therefore, developing effective approaches to handle incomplete data is crucial when designing machine learning-based decision support models. Given the constraints of sparse measurements, we propose a novel uncertainty-guided framework to strategically choose sensor locations to take measurements from rather than relying on random placement. This can be achieved using semi-supervised approaches, such as active learning, which enable the use of a limited number of in situ sensors while maintaining high-quality predictive modeling of soil health (Chen & Wang, 2023).

Related Work

Spatio-temporal graphs incorporate both spatial and temporal dimensions into their structure. They represent connections or relationships between entities not only in space but also over time and are useful in modeling complex systems such as transportation systems (Lablack & Shen, 2023), climate patterns (Laurini, 2019) and precision agriculture (San Emeterio de la Parte et al., 2023). Owing to their applications in a wide variety of complex systems, learning in spatio-temporal graphs has attracted significant research attention recently. Among these studies, there is a proposal to explicitly model the structure information into the convolutional operation and combine it with pseudo three-dimensional networks to capture temporal dependencies in the traffic data (Zhang et al., 2020). Another work proposes to learn temporal features using an attention-based Long Short-Term Memory (LSTM) network and the spatial features using Multi-

Task Learning (MTL) output layers (Lin et al., 2020). Deep Spatio-Temporal Graph Convolutional Network (DSTGCN) exploits a combination of graph convolution operations and standard convolutions to capture the dynamic variations in both spatial and temporal perspective (Yu et al., 2021).

LSTM unit and convolutional neural network are used to encode the spatial features of the target points, which are then fed to temporal attention mechanism to complete the decoding process and predict grain temperature (Duan et al., 2021). A graph-based Recurrent Neural Network (RNN) is proposed that incorporates both geographical and temporal knowledge in the model for crop yield prediction (Fan et al., 2022). More recently, Dynamic Traffic Correlation-based Spatio-Temporal Graph Convolutional Network (DTC-STGCN) is proposed to extract a dynamic adjacency matrix from different traffic characters to describe dynamic spatio-temporal correlations. Further, an attention and dynamic adjacency matrix-based GCNs framework is used to capture dynamic spatial features, while a LSTM is used to capture urban traffic temporal features, respectively (Xu et al., 2023). A multi-output and spatiotemporal model combining GCN and transformer is used to capture spatial correlations and temporal dependencies of the sensor network in a granary (Qu et al., 2023). Hierarchical Spatio-temporal Graph Neural Network (HiSTGNN) incorporates an adaptive graph learning module that constructs a self-learning hierarchical graph and leverages convolution and gated temporal convolution with a dilated inception as the backbone to capture hidden spatial dependencies and diverse long-term meteorological trends (Ma et al., 2023).

Most of the approaches proposed for learning in spatio-temporal graphs require huge amounts of training data, which might be difficult to obtain in practical scenarios like precision agriculture. Despite the advances in sensor technology, several challenges remain in collecting consistent and comprehensive soil data from the agricultural fields. Sensor reliability, data transmission issues, maintenance, and field operations often lead to sparse and uneven data distribution. The high cost and logistical challenges of widespread sensor deployment further worsen these issues (Lloret et al., 2021). Developing machine learning-based predictive frameworks to model the dynamics of these sparse soil measurements is challenging. Sparse measurements can introduce bias and are not always representative of the space that is being learned (Santana et al., 2022). There are existing works that capture spatial and temporal dynamics in sparse data setups by imputing the data at missing locations (Tharzeen et al., 2023). However, these methods may bias the results as the missingness itself might be informative (Van Ness et al., 2023). Semi-supervised learning methods such as Active Learning (AL), have addressed the sparsity in data in a machine learning setting. AL can significantly reduce the need for extensive labeled data by strategically selecting the most informative samples for labeling, resulting in higher prediction accuracy with limited data (Agarwal et al., 2021), (Agarwal & Natarajan, 2022). The integration of AL with Graph Neural Networks (GNNs) has emerged as a potent approach for spatial and temporal data prediction.

The application of AL and GNNs in environmental and agricultural data prediction has been explored in recent studies. (Y. Wu et al., 2019) demonstrated the efficacy of GNNs in handling spatial dependencies in various applications. Similarly, (Z. Wu et al., 2021) have illustrated the advantages of semi-supervised learning on graph structures, facilitating effective learning from sparsely labeled data. Though GNNs excel in capturing complex dependencies and relationships in data structured as graphs, they fail to capture temporal dependencies. This work, for the first time, presents a novel uncertainty-guided AL framework for applications involving spatio-temporal processes and graphs.

Contributions

This work addresses the problem of active learning in spatio-temporal graphs. For the first time, we propose a novel uncertainty-guided learning scheme involving a combination of Dilated Causal Convolutional Neural Network (DC-CNN) and Graph Neural Network (GNN). We systematically propagate the uncertainties associated with graph node features (representing the noise in sensor measurements) through all the layers of DC-CNN and GNN. This helps us estimate the

uncertainty associated with the model predictions, which is then leveraged to guide the informative sample selection via querying in the context of AL. We evaluate the effectiveness of the proposed methodology by conducting experiments over a field-scale sensor network data set for monitoring and modeling the spatial and temporal variation of soil moisture and temperature in a dryland agricultural field. The experimental evaluation reveals that it is possible to predict soil moisture/temperature over a large spatial area with over 90% accuracy even with a limited fraction of sensors and strategically selected measurement locations. The proposed methodology is generic and can be applied to the problem of monitoring of any soil parameter (such as nitrogen, phosphorus, potassium etc.) within any spatio-temporal graph setup, irrespective of the learning scheme and model architecture. Such an approach of utilizing uncertainty information to support AL in spatio-temporal graphs has not been reported in the literature so far and is a novel contribution of this work.

The remainder of this article is organized as follows: background on spatio-temporal graphs, DC-CNN, GNN and AL is provided in Section 2. We discuss the specific problem formulation targeted in this work and the proposed approach to implement uncertainty-guided AL on spatio-temporal graphs in Section 3. Experimental evaluation is outlined in Section 4 and the article ends with concluding remarks in Section 5.

Preliminaries

Spatio-Temporal Graphs

Spatiotemporal graphs represent data that is indexed in both spatial and temporal dimensions. These types of graphs are useful for modeling complex systems with features and relationships that change over space and time, such as transportation networks, soil parameter dynamics in agriculture, climate patterns, and biological processes (Del Mondo et al., 2021). In the context of spatiotemporal graphs, nodes/vertices represent spatial entities such as traffic intersections or location of field sensors. Edges/links represent the relationships between the nodes, such as roads between the intersection, communication links between sensing units or parametric dependencies across soil sensor locations. As a result, the spatial aspect of the spatiotemporal graph captures the geographical or logical locations of the nodes, as well as the spatial relationships represented by distance metrics, adjacency matrices, or spatial clustering techniques. On the other hand, temporal aspect captures the dynamic nature of the graph, such as the change in features or attributes over time and the appearance or disappearance of nodes or edges, thereby helping to model the evolution of the system it represents. Advances in machine learning have led to the development of novel models such as Graph Attention Networks (GAT) (Wang et al., 2022), a combination of Graph Neural Networks (GNNs) and Recurrent Neural Networks (Jain et al., 2016), and Temporal Convolution Neural Networks (Zhao et al., 2020) which can effectively learn from spatiotemporal data. Leveraging recent technological advancements, spatiotemporal graphs offer a powerful framework for modeling and analyzing complex systems where spatial and temporal dimensions interplay.

Dilated Causal Convolutional Neural Networks

Dilated Causal Convolutional Neural Networks (DC-CNN) provide an efficient mechanism for capturing temporal information in sequential data and for temporal prediction tasks (Tully et al., 2023). The major component of the dilated causal convolution is the convolutional neural network. In the context of sequential data, a 1D convolutional layer is applied to the data. Causal convolution ensures that the output at time step t only depends on inputs from time steps t and earlier. This is crucial for sequence prediction tasks where future data points should not be used to predict past or present values. In a causal convolution, the convolution operation is modified so that each filter only accesses past and present input values. Traditional convolutional layers have a limited receptive field, meaning they can only consider a small window of the input at a time. Dilated convolution introduces a spacing (dilation) between the filter elements, allowing the network to have a larger receptive field without increasing the number of parameters or

computational cost excessively (Oord et al., 2016). The dilation factor decides the number of input elements to skip which is decided by the spacing between the filter elements. For instance, with a dilation factor of 2, the filter skips every other input element.

Compared to other frameworks used for capturing temporal relationships, such as RNN, DC-CNN maintain a manageable number of parameters while being able to look back at a larger context in the input sequence (Chang et al., 2017). The hierarchical nature of the dilation factor allows the network to capture both short-term and long-term dependencies. Lower layers with smaller dilation factor capture finer details, while higher layers with larger dilation rates capture broader patterns (Oord et al., 2016).

Graph Neural Networks

The field of graph neural networks (GNNs) has experienced remarkable and swift advancements in recent years. GNNs are a class of deep learning models designed to perform inferencing on graph structured data (Scarselli et al., 2009). GNNs can capture complex relations and interactions in various domains such as sensor networks, transportation systems, social networks and so on (Gupta et al., 2021). Unlike traditional neural networks that operate on regular grids like images or sequences, GNNs are tailored to capture the underlying structure and properties of graph-based data, which consists of nodes representing entities or objects and edges representing relationships or interactions between these nodes (Zhou et al., 2020). The nodes representing entities are called node embeddings of a graph. These vectors capture the structural information and properties of the nodes, making them suitable for various downstream tasks like node classification, link prediction, and clustering.

GNNs operate by passing messages between nodes, where each node updates its representation by aggregating messages from its neighbors and combining them with its current state. Various GNN variants exist based on different aggregation and combination operations (S. Zhang et al., 2019), (Veličković et al., 2018). In this work, we use GraphSAGE, an inductive framework that can generalize to unseen nodes and subgraphs, unlike transductive models that require re-training when new nodes are added (Hamilton et al., 2017). GraphSAGE enhances scalability and adaptability to dynamic graphs by sampling a fixed-size set of neighbors for each node and aggregating their features to form the updated representation for the node of interest. This approach significantly reduces the computational complexity associated with large graphs. The equation used to update node features of a node u at i^{th} node embedding layer is as follows.

$$h_u^{(i)} = f^{(i)}\left(h_u^{(i-1)}, h_{N(u)}^{(i-1)}\right)$$
$$= g\left(\theta_C^{(i)}\, h_u^{(i-1)}, \theta_A^{(i)}\, \tilde{A}\left(h_{N(u)}^{(i-1)}\right)\right)$$

Active Learning

AL is a semi-supervised machine learning approach that selectively queries the most informative data points for labeling, aiming to build accurate models with fewer labeled instances (Agarwal & Natarajan, 2022). It allows the learning algorithm to leverage information from external sources (e.g., human experts) and train an efficient decision model using much fewer labeled instances as compared to traditional supervised ML methodologies. This approach is particularly useful in situations where labeling is expensive or time-consuming, such as in natural language processing, computer vision, medical imaging, and speech recognition. The learning algorithm in AL settings is referred to as Active Learner and the external information source is termed as the Oracle. The Active Learner is permitted to select the data it wants to learn from. Thus, it poses queries to the Oracle in the form of unlabeled data instances, and the Oracle is expected to supply corresponding labels. By employing various query strategies such as uncertainty sampling, query by committee and sampling diversity, AL helps in selecting the candidate query instance that is the most informative, enhancing model performance while operating under a labeling budget (Agarwal et al., 2021). In this work, we employ an AL strategy combined with Graph Neural Networks (GNN) and Dilated Convolutional Neural Networks (DC-CNN) to predict soil health in

agricultural fields.

Methodology

This work presents a novel framework to perform Active Learning on spatio-temporal graphs. The proposed architecture involves two modules: (i) Dilated Causal Convolutional Neural Network for learning in the temporal domain, and (ii) Graph Neural Network for the spatial counterpart. The proposed framework supports systematic propagation of uncertainty in node features through the DC-CNN and GNN layers using Assumed Density Filtering (ADF), followed by moment matching. This allows us to derive the uncertainty associated with model predictions, which is further leveraged to select the most informative nodes for AL in spatio-temporal graphs.

Problem Formulation

We consider a spatio-temporal graph $G = (A, E, X_t)$, where $A \in R^{N \times N}$ denotes the spatial adjacency matrix; N denotes the number of nodes and the element $a_{i,j} \in A$ is 1 if the nodes i and j are connected via an edge. E represents the spatial relationship between the nodes, and $X_t \in R^{N \times D}$ is the set of D dimensional dynamic features corresponding to each node at time t. The uncertainty in node features is assumed to be Gaussian with zero mean and known variance. For a node u in G, the uncertainty in the feature vector can be represented as $h_u \sim \mathcal{N}(h_u^*, \Sigma)$ where h and h^* are the measured and true values of the feature vector, respectively, and Σ is a diagonal matrix comprising of known variances of noise in individual features. It is also assumed that the noise in features of different nodes is uncorrelated.

The uncertainties in ML frameworks can be either aleatoric or epistemic. The intrinsic randomness of data caused due to noisy or faulty measurements is referred to as aleatoric uncertainty. In a graph-structed data, it can be captured via uncertainty associated with node features and probabilistic links (Munikoti et al., 2023). The uncertainty in node features takes acre of the measurement noise and probabilistic links handle the imprecise information about underlying graph structure. The inability of a model to comprehend the underlying process leads to epistemic uncertainty. In a graph-structed data, it can be captured via uncertainty in parameters of GNN and the associated activation functions. In this work, we consider the propagation of aleatoric uncertainty through the layers of DC-CNN and GNN, and propagation of epistemic uncertainty shall be considered as a part of the future work. The proposed methodology shall be evaluated over a field-scale sensor network data set for monitoring and modeling the spatial and temporal variation of soil health in a dryland agricultural field (Gasch et al., 2017). Therefore, in the context of measurements obtained from agricultural fields, every node in G corresponds to a location with two sensors, soil temperature and volumetric water content, each with a hourly time-series measurement for 1000 time instances. The node features comprise of the truncated Singular Value Decomposition (SVD) of the weighted adjacency matrix (Lingam et al., 2021). This helps to improve the learning scheme by incorporating higher order neighbor embeddings, while keeping self-embedding separate from the neighbor embeddings in aggregation function of GNN (Lingam et al., 2021).

In this work, we systematically quantify the uncertainties through different layers of DC-CNN and GNN in order to finally obtain the uncertainty in model predictions because of noise in the node features. Incorporating prior information about the data, such as sensor or measurement noise, allows for more reliable predictions by quantifying uncertainties in data through confidence intervals around the predictions. For instance, it is critical to account for measurement noise in the readings obtained from various sensors like temperature and soil moisture sensors installed in agricultural fields. If these uncertainties are overlooked and the features are assumed to be deterministic, it can lead to unreliable predictions.

It is well known that the widespread installation of sensors can be excessively costly and can place a substantial burden on communication networks. This work, for the first time, presents an uncertainty guided AL framework for predicting measurements across the entire field by using

sparsely available data. The uncertainty associated with the model prediction of individual nodes is used as the heuristic for iterative selection of nodes in the context of AL. This allows to learn a robust decision model by utilizing significantly lesser number of labeled nodes for training, as compared to conventional supervised learning approaches. Such a framework is highly beneficial for developing predictive engines using information from the agricultural fields, as the data is frequently impeded by absence of information owing to multiple factors like sensor dependability, data collection, communication disruptions, maintenance problems, and field operations.

Proposed Framework

This work proposes a novel uncertainty-guided framework to implement AL on spatio-temporal graphs. The learning scheme utilized in the proposed methodology involves a combination of DC-CNN and GNN. While DC-CNN addresses the temporal relations associated with individual nodes, GNN captures the spatial dependencies between different nodes of the given spatio-temporal graph G. The proposed framework is illustrated in Figure 1. A small fraction of nodes (~ 10%) from G is chosen for initial training of the decision model. During this process, we systematically propagate the uncertainties associated with node features (assumed as sensor or measurement noise) through different layers of DC-CNN and GNN in order to finally obtain the uncertainty in model predictions. This is achieved by making use of ADF, followed by moment matching. We consider propagation and matching the means and variances of the output probability density functions (PDFs) of all DC-CNN and GNN layers. The uncertainty in model predictions is then leveraged to iteratively select the most informative nodes in the context of AL. This process is repeated until the query budget is exhausted.

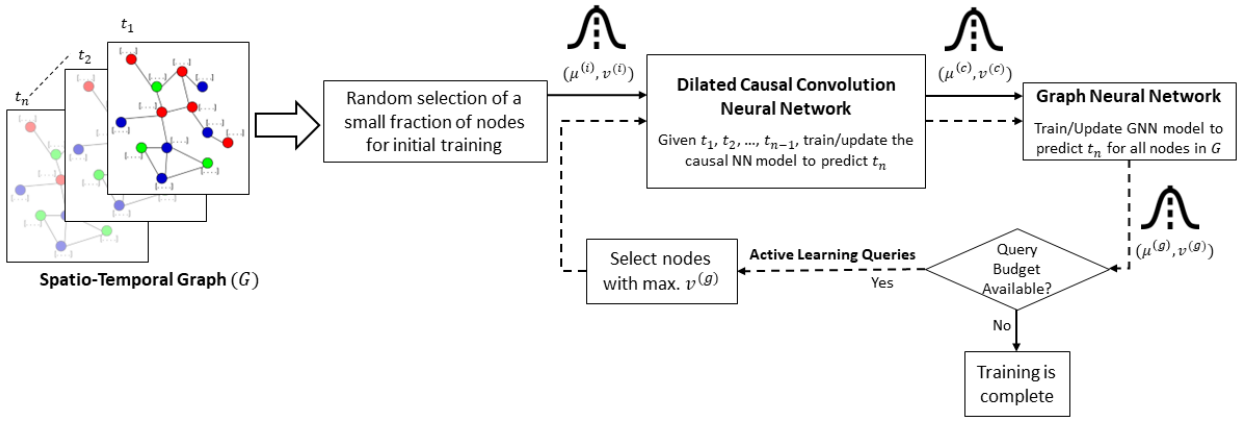

Fig 1. Proposed framework for uncertainty-guided AL on spatio-temporal graphs.

The following result formalizes the propagation of uncertainty across DC-CNN and GNN layers.

Theorem 1: The expected value $(\mu_{u,t})$ and variance $(v_{u,t})$ of node embedding for node u at time t, accounting for aleatoric uncertainty with mean aggregation and linear activation functions are:

$$\mu_{u,t}^{(i)} = \theta_{C,t}^{(i)} \mu_{u,t}^{(i-1)} + \theta_{A,t}^{(i)} \frac{n}{k|N(u)|D(u)} \sum_{\substack{s=0 \\ v \in N(u)}}^{k-1} p_{uv,t} \mu_{v,t}^{(i-1)}(t - \eta s) \tag{1}$$

$$v_{u,t}^{(i)} = \theta_{C,t}^{(i)^2} v_{u,t}^{(i-1)} + \theta_{A,t}^{(i)^2} \frac{n}{k|N(u)|D(u)} \sum_{\substack{s=0 \\ v \in N(u)}}^{k-1} p_{uv,t}^2 v_{v,t}^{(i-1)}(t - \eta s) \tag{2}$$

where, (i) corresponds to quantities of i^{th} layer, θ_C denotes the combination parameters, θ_A denotes the aggregation parameters, $N(u)$ is the neighborhood of the node u, $D(u)$ is the degree of node u and η is the dilation factor. $p_{uv,t}$ represents the probability of the link between nodes u

and v at time t.

Proof:

For a node u, let the noisy feature vector be: $h_u \sim \mathcal{N}(h_u^*, \Sigma)$. The aggregation and the combination operations at the i^{th} node embedding layer is written as:

$$h_u^{(i)} = f\left(h_u^{(i-1)}, h_{N(u)}^{(i-1)}\right)$$
$$= g\left(\theta_C^{(i)}\, h_u^{(i-1)}, \theta_A^{(i)} \tilde{A}\left(h_{N(u)}^{(i-1)}\right)\right) \quad (3)$$

where, $g[\cdot]$ is the activation function and $\tilde{A}$ is the aggregation operator. At time t, for node u, the temporal representation learned by the causal convolution can be written as:

$$h_{u,t}^{(i)} = h_u^{(i-1)} * k^{(i)}(t)$$
$$= \sum_{s=0}^{k-1} k^{(i)}(s) \odot h_u^{(i-1)}(t - \eta s) \quad (4)$$

where k is the kernel and $\odot$ is the Hadamard product.

Using ADF, the joint density of all embeddings for a node u, i.e., $p(h_{u,(t-ns):t}^{(o:l)})$ can be approximated (Gast & Roth, 2018), (Munikoti et al., 2023). This can be done by minimizing KL divergence and moment matching (first two moments) (Munikoti et al., 2023):

$$\mu_{u,t}^{(i)} = \mathbb{E}_q\left(h_{u,t}^{(i-1)}\right)\left[f^{(i)}\left(h_{u,t}^{(i-1)}, h_{N(u)}^{(i-1)}(t - ns)\right)\right] \quad (5)$$

$$\nu_{u,t}^{(i)} = var_q\left(h_{u,t}^{(i-1)}\right)\left[f^{(i)}\left(h_{u,t}^{(i-1)}, h_{N(u)}^{(i-1)}(t - ns)\right)\right] \quad (6)$$

where, $\mathbb{E}$ is the expectation operator and var is the variance operator.

The following assumptions were made: $\tilde{A}$ is a mean operator and $g[\cdot]$ is linear activation. As a result of these assumption, (i) the variance of mean is normalized form of mean of variances; and (ii) the mean of expectations is same as expectation of means.

By substituting equation (3) and (4) in equation (5) and (6) the mean and variance can be obtained as:

$$\mu_{u,t}^{(i)} = \theta_{C,t}^{(i)} \mu_{u,t}^{(i-1)} + \theta_{A,t}^{(i)} \frac{n}{k|N(u)|D(u)} \sum_{\substack{s=0 \\ v \in N(u)}}^{k-1} p_{uv,t} \mu_{v,t}^{(i-1)}(t - \eta s)$$

$$v_{u,t}^{(i)} = \theta_{C,t}^{(i)^2} v_{u,t}^{(i-1)} + \theta_{A,t}^{(i)^2} \frac{n}{k|N(u)|D(u)} \sum_{\substack{s=0 \\ v \in N(u)}}^{k-1} p_{uv,t}^2 v_{v,t}^{(i-1)}(t - \eta s)$$

This proves Theorem 1.

Experimental Evaluation

We evaluate the proposed methodology over a field-scale sensor network data set for monitoring and modeling the spatial and temporal variation of soil health in a dryland agricultural field (Gasch et al., 2017). We consider hourly measurements of volumetric water content and soil temperature collected at 42 monitoring locations and 5 depths (30, 60, 90, 120, and 150 cm) across Cook

Agronomy Farm (Gasch et al., 2017). A window of 1000 measurements is chosen for each measurement location at every depth. The next step is to create a connected graph G with 210 nodes (i.e., 42 locations at 5 depth). The Euclidean distances d_{ij} between all node pairs i and j were calculated, and an edge was considered between the ones where d_{ij} was found to be less than a pre-defined threshold distance d_T. The selection criteria for d_T was to ensure that the resulting graph is connected. Therefore, every node in G corresponds to a location with two sensors, soil temperature and volumetric water content, each with a hourly time-series measurement for 1000 time instances.

The node features comprise of the truncated Singular Value Decomposition (SVD) of the weighted adjacency matrix (Lingam et al., 2021). This helps to improve the learning scheme by incorporating higher order neighbor embeddings, while keeping self-embedding separate from the neighbor embeddings in aggregation function of GNN (Lingam et al., 2021). The detailed architecture used in our experiments is as follows: (i) DC-CNN: number of convolution layers: 1; number of LSTM layers: 1; number of linear layers: 3; hidden dimensions: 48, 24; kernel size: 3; window size: 100 (ii) GNN: depth (i.e., number of node embedding modules): 2; number of neurons in 2 layers: 48, 36; number of multi-layer perceptron (MLP) layers: 2; activation function: linear (except last layer with softmax); aggregation function: mean. The training of GNN is carried out in a mini-batch manner.

An initial labeled dataset was curated by randomly selecting around 10% of nodes in G. It is assumed that both soil temperature and volumetric water content data are available for all these nodes for the first 100-time instances. Subsequently, this data is used to train an initial model as per the aforementioned architecture. During this process, the uncertainties arising from sensor or measurement noise are also propagated systematically through different layers of DC-CNN and GNN, as outlined in Theorem 1. After training an initial model, the querying procedure is commenced, and the decision model is updated after each query step. We use the uncertainty in model predictions obtained during the previous step as a heuristic to iteratively select the most informative nodes at each successive query step. Around 30% of the nodes in G are used for AL queries and 40% are held out separately for testing.

We have designed four tasks in order to demonstrate the utility of our proposed framework. The first two tasks are intended to predict soil temperature (TR) and volumetric water content (VWR) for the test nodes during all time instances by using truncated SVD values as node features. The third task is designed to predict TR by using both truncated SVD values and VWR measurements as node features. The last task is to predict VWR by using a combination of truncated SVD values and TR measurements as node features. We use Mean Squared Error (MSE) and Predictive Performance (PP) as metrics to evaluate the performance of the proposed methodology over different tasks. PP is calculated using eq. (7). Here, A_i and P_i are actual and predicted values, respectively, and n is the total number of observations.

$$Predictive\ performance\ (PP) = 100 - \frac{100}{n}\sum_{i=1}^{n}\left|\frac{A_i - P_i}{A_i}\right| \tag{7}$$

Firstly, we establish the benefits of utilizing AL for strategic selection of nodes for model training during each query step. The random selection of nodes from the dataset is treated as the baseline. Table 1 illustrates the percentage of node labels required to achieve 90% predictive performance. It can be observed that the proposed framework achieves a predictive performance of 90% using 30% or lesser node labels for all the tasks. On the other hand, 65% - 70% of node labels are required to achieve the same level of performance using baseline, i.e., if the node labels are selected at random for model training. This is because the proposed framework strategically selects the nodes with highest uncertainty at each query step, thereby improving the label efficiency. Such an approach is highly beneficial for developing predictive engines using information from the agricultural fields, where data availability is sparse. It can also be seen that the percentage of node labels required to achieve 90% PP is lesser for tasks 3 and 4, as compared to those for tasks 1 and 2. This is intuitive because a larger amount of information is utilized as

node features for model training in these tasks, namely, VWR for task 3 and TR for task 4, respectively, which is missing in tasks 1 and 2.

Table 1. Percentage of node labels required to achieve 90% predictive performance.

Sr. No.	Task	Baseline	Proposed Framework
1	Prediction of TR	71.42%	30.95%
2	Prediction of VWR	70.47%	29.52%
3	Prediction of TR based on VWR	64.28%	28.57%
4	Prediction of VWR based on TR	65.23%	27.61%

Secondly, we demonstrate the efficacy of AL approach by evaluating performance for all the tasks after specific query steps. The values of MSE and PP after regular intervals of query steps are provided in Table 2. It is clearly observed that the values of MSE reduce and PP improves as more queries are made. This indicates the incremental learning of the decision model, as it is updated after each query step. Once again, it can be observed that the performance of decision model is better for tasks 3 and 4 as compared to tasks 1 and 2. This is because additional information is incorporated as node features during model training for these tasks.

We also capture the uncertainty associated with model predictions for all the tasks after specific query steps, as outlined in Table 3. These uncertainty values were calculated by considering that the sensor measurements are corrupted with 10% Gaussian noise, i.e., an input uncertainty of 0.10. It was observed that the values of uncertainties reduce as more queries are made. This is intuitive because learning of the decision model improves with each query step, as the information about strategically selected uncertain nodes is incorporated in the learning scheme.

Table 2. Performance Evaluation at different stages of Active Learning.

Sr. No.	Task	Performance Metrics after number of queries					
		20		40		60	
		MSE	PP	MSE	PP	MSE	PP
1	Prediction of TR	0.1123	52.23%	0.0911	77.90%	0.0537	92.07%
2	Prediction of VWR	0.1345	51.54%	0.1032	75.15%	0.0625	92.85%
3	Prediction of TR based on VWR	0.1056	55.87%	0.0744	81.33%	0.0498	94.67%
4	Prediction of VWR based on TR	0.0988	56.12%	0.0779	80.47%	0.0386	96.58%

Table 3. Uncertainty Evaluation at different stages of Active Learning (considering 0.10 at input).

Sr. No.	Task	Uncertainty after number of queries		
		20	40	60
1	Prediction of TR	0.0823	0.0621	0.0408
2	Prediction of VWR	0.0965	0.0612	0.0536
3	Prediction of TR based on VWR	0.0809	0.0577	0.0245
4	Prediction of VWR based on TR	0.0767	0.0448	0.0287

Finally, the effectiveness of the proposed framework is demonstrated by carrying out longitudinal studies corresponding to specific nodes in G for task of predicting TR based on VWR. A few nodes were selected at random and the variations in PP and uncertainty values are examined. Table 4 presents a summary of the results for four randomly selected nodes from G. It can be seen that the predictive performance improves, and uncertainty reduces as more queries are made. These observations illustrate the systematic propagation of uncertainties across all layers of DC-CNN and GNN, as well as incremental learning of the decision model after subsequent query steps in

the context of AL.

Table 4. Longitudinal Studies for specific nodes: Prediction of TR based on VWR (uncertainty measured by considering 0.10 at input).

Sr. No.	Node Instance	Performance Metrics after number of queries					
		20		40		60	
		PP	Uncertainty	PP	Uncertainty	PP	Uncertainty
1	13	55.41%	0.0798	82.08%	0.0542	95.31%	0.0265
2	46	56.22%	0.0877	81.53%	0.0603	94.87%	0.0247
3	88	54.85%	0.0843	83.16%	0.0528	94.63%	0.0316
4	162	55.91%	0.0812	81.74%	0.0537	93.96%	0.0255

Conclusion

In this paper, we propose a novel framework to perform AL on spatio-temporal graphs. Specifically, we have developed an uncertainty-guided learning scheme involving a combination of DC-CNN and GNN. While DC-CNN addresses the temporal relations associated with individual nodes, GNN captures the spatial dependencies between different nodes of the spatio-temporal graph. The proposed framework supports systematic propagation of uncertainty in node features through the DC-CNN and GNN layers using ADF, followed by moment matching in order to finally obtain the uncertainty in model predictions. This allows for more reliable predictions by quantifying uncertainties in data through confidence intervals around the predictions. The uncertainty in model predictions is then leveraged to iteratively select the most informative nodes in the context of AL. It is worthwhile to note that the proposed methodology is generic and hence, agnostic to learning architecture, AL setup and query strategy.

The proposed methodology is evaluated by conducting experiments over a field-scale sensor network data set for monitoring and modeling the spatial and temporal variation of soil health in a dryland agricultural field. Experimental results demonstrate the applicability of the proposed methodology to train a decision model for soil moisture and temperature monitoring using a limited number of sensor measurements which are strategically selected to minimize uncertainty in model predictions. Such a framework is highly beneficial for developing predictive engines driving precision agriculture applications using spatially sparse information from the agricultural fields, The future extension of this work will involve extending the framework for (1) monitoring other soil parameters; (2) uncertainty propagation aspect across non-linear activation functions, and (3) evaluating the proposed methodology across more dynamic setups related to spatio-temporal graphs.

References

Agarwal, D., & Natarajan, B. (2022). Active Learning for Node Classification using a Convex Optimization approach. *2022 IEEE Eighth International Conference on Big Data Computing Service and Applications (BigDataService)*, 96–102. https://doi.org/10.1109/BigDataService55688.2022.00022

Agarwal, D., Srivastava, P., Martin-del-Campo, S., Natarajan, B., & Srinivasan, B. (2021). Addressing Uncertainties within Active Learning for Industrial IoT. *2021 IEEE 7th World Forum on Internet of Things (WF-IoT)*, 557–562. https://doi.org/10.1109/WF-IoT51360.2021.9595397

Basha, S. M., Rajput, D. S., J, J., Somula, R. S., & Ram, S. (2020). Principles and Practices of Making Agriculture Sustainable: Crop Yield prediction using Random Forest. *Scalable Computing: Practice and Experience*, *21*(4), Article 4. https://doi.org/10.12694/scpe.v21i4.1714

Chang, S., Zhang, Y., Han, W., Yu, M., Guo, X., Tan, W., Cui, X., Witbrock, M., Hasegawa-Johnson, M. A., & Huang, T. S. (2017). Dilated Recurrent Neural Networks. *Advances in Neural Information Processing Systems*, *30*.

https://proceedings.neurips.cc/paper_files/paper/2017/hash/32bb90e8976aab5298d5da10fe66f21d-Abstract.html

Chen, H., & Wang, J. (2023). Active Learning for Efficient Soil Monitoring in Large Terrain with Heterogeneous Sensor Network. *Sensors (Basel, Switzerland)*, *23*(5), 2365. https://doi.org/10.3390/s23052365

Del Mondo, G., Peng, P., Gensel, J., Claramunt, C., & Lu, F. (2021). Leveraging Spatio-Temporal Graphs and Knowledge Graphs: Perspectives in the Field of Maritime Transportation. *ISPRS International Journal of Geo-Information*, *10*(8), Article 8. https://doi.org/10.3390/ijgi10080541

Duan, S., Yang, W., Wang, X., Mao, S., & Zhang, Y. (2021). Temperature Forecasting for Stored Grain: A Deep Spatiotemporal Attention Approach. *IEEE Internet of Things Journal*, *8*(23), 17147–17160. https://doi.org/10.1109/JIOT.2021.3078332

Fan, J., Bai, J., Li, Z., Ortiz-Bobea, A., & Gomes, C. P. (2022). A GNN-RNN Approach for Harnessing Geospatial and Temporal Information: Application to Crop Yield Prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, *36*(11), 11873–11881. https://doi.org/10.1609/aaai.v36i11.21444

Gałęzewski, L., Jaskulska, I., Jaskulski, D., Lewandowski, A., Szypłowska, A., Wilczek, A., & Szczepańczyk, M. (2021). Analysis of the need for soil moisture, salinity and temperature sensing in agriculture: A case study in Poland. *Scientific Reports*, *11*(1), 16660. https://doi.org/10.1038/s41598-021-96182-1

Gasch, C. K., Brown, D. J., Campbell, C. S., Cobos, D. R., Brooks, E. S., Chahal, M., & Poggio, M. (2017). A Field-Scale Sensor Network Data Set for Monitoring and Modeling the Spatial and Temporal Variation of Soil Water Content in a Dryland Agricultural Field. *Water Resources Research*, *53*(12), 10878–10887. https://doi.org/10.1002/2017WR021307

Gast, J., & Roth, S. (2018). Lightweight Probabilistic Deep Networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3369–3378. https://doi.org/10.1109/CVPR.2018.00355

Gupta, A., Matta, P., & Pant, B. (2021). Graph neural network: Current state of Art, challenges and applications. *Materials Today: Proceedings*, *46*, 10927–10932. https://doi.org/10.1016/j.matpr.2021.01.950

Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive Representation Learning on Large Graphs. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper_files/paper/2017/hash/5dd9db5e033da9c6fb5ba83c7a7ebea9-Abstract.html

Jain, A., Zamir, A. R., Savarese, S., & Saxena, A. (2016). Structural-RNN: Deep Learning on Spatio-Temporal Graphs. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5308–5317. https://doi.org/10.1109/CVPR.2016.573

Joshi, C., & Mohanty, B. P. (2010). Physical controls of near-surface soil moisture across varying spatial scales in an agricultural landscape during SMEX02. *Water Resources Research*, *46*(12). https://doi.org/10.1029/2010WR009152

Lablack, M., & Shen, Y. (2023). Spatio-temporal graph mixformer for traffic forecasting. *Expert Systems with Applications*, *228*, 120281. https://doi.org/10.1016/j.eswa.2023.120281

Laurini, M. P. (2019). A spatio-temporal approach to estimate patterns of climate change. *Environmetrics*, *30*(1), e2542. https://doi.org/10.1002/env.2542

Lingam, V., Ragesh, R., Iyer, A., & Sellamanickam, S. (2021). *Simple Truncated SVD based Model for Node Classification on Heterophilic Graphs*.

Lloret, J., Sendra, S., Garcia, L., & Jimenez, J. M. (2021). A Wireless Sensor Network Deployment for Soil Moisture Monitoring in Precision Agriculture. *Sensors*, *21*(21), Article 21. https://doi.org/10.3390/s21217243

Ma, M., Xie, P., Teng, F., Wang, B., Ji, S., Zhang, J., & Li, T. (2023). HiSTGNN: Hierarchical spatio-temporal graph neural network for weather forecasting. *Information Sciences*, *648*, 119580. https://doi.org/10.1016/j.ins.2023.119580

Montgomery, D. R., Rabinowitz, P., Sipos, Y., & Wheat, E. E. (2024). Soil health: A common focus for one health and planetary health interventions. *One Health*, *18*, 100673. https://doi.org/10.1016/j.onehlt.2023.100673

Munikoti, S., Agarwal, D., Das, L., & Natarajan, B. (2023). A general framework for quantifying aleatoric and epistemic uncertainty in graph neural networks. *Neurocomputing*, *521*, 1–10. https://doi.org/10.1016/j.neucom.2022.11.049

Mythili, K., & Al, E. (2021). A Swarm based Bi-directional LSTM-Enhanced Elman Recurrent Neural Network Algorithm for Better Crop Yield in Precision Agriculture. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(10), Article 10. https://doi.org/10.17762/turcomat.v12i10.5656

Oord, A. van den, Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). WaveNet: A Generative Model for Raw Audio. *ArXiv*. https://www.semanticscholar.org/paper/WaveNet%3A-A-Generative-Model-for-Raw-Audio-Oord-Dieleman/df0402517a7338ae28bc54acaac400de6b456a46

Pereira, G. W., Valente, D. S. M., de Queiroz, D. M., Santos, N. T., & Fernandes-Filho, E. I. (2022). Soil mapping for precision agriculture using support vector machines combined with inverse distance weighting. *Precision Agriculture*, *23*(4), 1189–1204. https://doi.org/10.1007/s11119-022-09880-9

Qu, Z., Zhang, Y., Hong, C., Zhang, C., Dai, Z., Zhao, Y., Wu, X., gao, Y., Jiang, X., Qian, J., & Gu, Z. (2023). Temperature forecasting of grain in storage: A multi-output and spatiotemporal approach based on deep learning. *Computers and Electronics in Agriculture*, *208*, 107785. https://doi.org/10.1016/j.compag.2023.107785

Saleem, M. H., Potgieter, J., & Arif, K. M. (2021). Automation in Agriculture by Machine and Deep Learning Techniques: A Review of Recent Developments. *Precision Agriculture*, *22*(6), 2053–2091. https://doi.org/10.1007/s11119-021-09806-x

San Emeterio de la Parte, M., Martínez-Ortega, J.-F., Hernández Díaz, V., & Martínez, N. L. (2023). Big Data and precision agriculture: A novel spatio-temporal semantic IoT data management framework for improved interoperability. *Journal of Big Data*, *10*(1), 52. https://doi.org/10.1186/s40537-023-00729-0

Santana, S. R., Ortega, L. A., Hernández-Lobato, D., & Zaldívar, B. (2022). *Correcting Model Bias with Sparse Implicit Processes*. https://doi.org/10.48550/ARXIV.2207.10673

Scarselli, F., Gori, M., Ah Chung Tsoi, Hagenbuchner, M., & Monfardini, G. (2009). The Graph Neural Network Model. *IEEE Transactions on Neural Networks*, *20*(1), 61–80. https://doi.org/10.1109/TNN.2008.2005605

Sishodia, R. P., Ray, R. L., & Singh, S. K. (2020). Applications of Remote Sensing in Precision Agriculture: A Review. *Remote Sensing*, *12*(19), Article 19. https://doi.org/10.3390/rs12193136

Tharzeen, A., Munikoti, S., Prakash, P., Kim, J., & Natarajan, B. (2023). A General Spatiotemporal Imputation Framework for Missing Sensor Data. *2023 IEEE Conference on Artificial Intelligence (CAI)*, 55–58. https://doi.org/10.1109/CAI54212.2023.00032

Tully, J., Haight, R., Hutchinson, B., Huang, S., Lee, J.-Y., & Katipamula, S. (2023). Dilated causal convolutional neural networks for forecasting zone airflow to estimate short-term energy consumption. *Energy and Buildings*, *286*, 112890. https://doi.org/10.1016/j.enbuild.2023.112890

Van Ness, M., Bosschieter, T. M., Halpin-Gregorio, R., & Udell, M. (2023). The Missing Indicator Method: From Low to High Dimensions. *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 5004–5015. https://doi.org/10.1145/3580305.3599911

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). *Graph Attention Networks* (arXiv:1710.10903). arXiv. http://arxiv.org/abs/1710.10903

Wang, Y., Jing, C., Xu, S., & Guo, T. (2022). Attention based spatiotemporal graph attention networks for traffic flow forecasting. *Information Sciences*, *607*, 869–883. https://doi.org/10.1016/j.ins.2022.05.127

Wu, Y., Xu, Y., Singh, A., Yang, Y., & Dubrawski, A. (2019). Active Learning for Graph Neural

Networks via Node Feature Propagation. In *arXiv e-prints*. https://doi.org/10.48550/arXiv.1910.07567

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2021). A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(1), 4–24. https://doi.org/10.1109/TNNLS.2020.2978386

Xu, Y., Cai, X., Wang, E., Liu, W., Yang, Y., & Yang, F. (2023). Dynamic traffic correlations based spatio-temporal graph convolutional network for urban traffic prediction. *Information Sciences*, *621*, 580–595. https://doi.org/10.1016/j.ins.2022.11.086

Yu, L., Du, B., Hu, X., Sun, L., Han, L., & Lv, W. (2021). Deep spatio-temporal graph convolutional network for traffic accident prediction. *Neurocomputing*, *423*, 135–147. https://doi.org/10.1016/j.neucom.2020.09.043

Zhang, Q., Li, B., Zhang, Y., & Wang, S. (2022). Suitability Evaluation of Crop Variety via Graph Neural Network. *Computational Intelligence and Neuroscience*, *2022*, 5614974. https://doi.org/10.1155/2022/5614974

Zhang, S., Tong, H., Xu, J., & Maciejewski, R. (2019). Graph convolutional networks: A comprehensive review. *Computational Social Networks*, *6*(1), 11. https://doi.org/10.1186/s40649-019-0069-y

Zhang, Y., & Thorburn, P. J. (2021). A dual-head attention model for time series data imputation. *Computers and Electronics in Agriculture*, *189*, 106377. https://doi.org/10.1016/j.compag.2021.106377

Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M., & Li, H. (2020). T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction. *IEEE Transactions on Intelligent Transportation Systems*, *21*(9), 3848–3858. https://doi.org/10.1109/TITS.2019.2935152

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI Open*, *1*, 57–81. https://doi.org/10.1016/j.aiopen.2021.01.001