Domain-Informed Graph Neural Networks for Climate Factor Forecasting to Support Sustainable Crop Management

Ziyue Sun[1] , **Zixin Jiang**[2] , **Chenkai Xu**[3] and **Xinggao Liu**[1*]
[1]College of Control Science and Engineering, Zhejiang University, Hangzhou, China
[2]College of Environmental & Resource Sciences, Zhejiang University, Hangzhou, China
[3]Ningbo Natural Resources and Planning Big Data Center, Ningbo, China
sunziyue@zju.edu.cn, lxg@zju.edu.cn

Abstract

Forecasting climate factors is critical for anticipating agro-climatic risks and enabling sustainable crop management. However, accurate prediction remains challenging due to complex spatiotemporal variability, heterogeneous seasonal patterns, and intricate interdependencies among climate variables. Inspired by agronomic knowledge, We propose DoIGNN, a Domain-Informed Graph Neural Network that injects a domain-structured graph constraint built from Agro-Climatic Homogeneous Zones (ACHZs). Specifically, we partition stations into agro-climatic zones using long-term climatic statistics and location attributes, and construct a hierarchical ACHZ-guided adjacency. To better capture shared climate dynamics, we introduce a spatiotemporal decomposition module with temporal regularization that factorizes the climate tensor into low-rank global temporal bases and station loadings, yielding a compact station-level global component as auxiliary information for target forecasting. Finally, DoIGNN performs forecasting on both the ACHZ-guided and static-dynamic graphs to learn cross-region dependencies. Experiments on real-world climate datasets demonstrate that DoIGNN consistently improves forecasting accuracy over strong baselines while yielding more interpretable spatial dependency patterns that support climate-informed crop management decisions. Cooperating with Ningbo Natural Resources and Planning Big Data Center, the proposed model has been trained and deployed for local data analysis.

1 Introduction

Climate factors, such as temperature, precipitation and humidity, are climatic drivers that directly influence crop phenology, soil moisture dynamics, and pest & disease pressures. Therefore, predicting changes in these factors is irreplaceable for efficient crop management and food security planning [Nouri and Veysi, 2025]. However, achieving high-resolution and robust predictions remains challenging because local climatic dynamics are inherently influenced by surrounding regions. Without explicitly considering the spatial dependencies among neighboring areas, predictions often fail to capture propagating weather patterns and regional interactions, especially at large spatial scales. Moreover, climate regimes vary substantially across regions, exhibiting non-uniform feature distributions and heterogeneous temporal patterns, which makes it difficult for a single model to generalize effectively across diverse agro-climatic zones.

Traditional climate forecasting has long relied on numerical weather prediction (NWP) and climate models grounded in atmospheric dynamics [Kimura, 2002; Leutbecher and Palmer, 2008; Zhao *et al.*, 2025; Giorgi *et al.*, 2023]. These methods have achieved strong performance in large-scale weather and climate simulation by explicitly modeling physical processes. However, their effectiveness often depends on complex physical parameterizations, high-quality boundary conditions, and substantial computational resources, which may limit their flexibility for regional, station-level, and application-oriented agro-climatic forecasting.

To complement physics-based approaches, data-driven models have been increasingly explored for meteorological and climate prediction. Classical machine learning methods such as SVMs and random forests have been used to model nonlinear relationships among meteorological variables [Das and Ghosh, 2018; He *et al.*, 2021b]. Deep learning models, including CNNs, RNNs, ConvLSTMs, and recent spatiotemporal architectures, further improve the ability to capture temporal evolution and spatial patterns from historical observations [Haggag *et al.*, 2021; Chen *et al.*, 2022; Gong *et al.*, 2024; Utku and Akyol, 2025]. Despite these advances, many existing methods still rely mainly on statistical correlations and may insufficiently encode domain-specific agro-climatic structure.

Recently, graph neural networks (GNNs) have shown strong potential in climate-related forecasting by representing stations, grids, or regions as nodes and modeling spatial dependencies through graph propagation [Humphries *et al.*, 2026; Yousaf *et al.*, 2025; Zhang *et al.*, 2026; Li *et al.*, 2025]. Such models are particularly suitable for multi-station forecasting because they can capture non-Euclidean spatial interactions beyond regular grid assumptions. Nevertheless, most existing graph-based methods construct spatial graphs from geographical distance, statistical similarity, or learnable adjacency matrices, while the integration of agro-climatic domain knowledge into graph construction remains underexplored.

Compared with NWP and traditional machine learning methods, deep learning methods—especially graph-based models—exhibit improved performance in capturing nonlinear spatiotemporal correlations of univariate data. Nevertheless, the prediction of climate data via deep learning still encounters the following challenges:

1. Existing GNN-based methods construct graphs using geographic proximity, statistical similarity, or fully learnable adjacency. Such topologies are often misaligned with agro-climatic regimes.
2. Existing models simply encode all components in a single representation, which may cause representation redundancy.
3. Many methods capture only one type of dependency (static or dynamic) or yield opaque learned edges without clear agro-climatic meaning.

Based on the above challenges, we propose DoIGNN, a Domain-Informed Graph Neural Network, a framework that explicitly integrates agro-climatic domain knowledge into graph-based spatiotemporal modeling. In this approach, climate observation stations are systematically grouped into Agro-Climatic Homogeneous Zones (ACHZs) using long-term meteorological statistics and geospatial attributes. Specifically, intra-zone connections capture localized climatic coherence, while inter-zone links selectively model broader regional interactions and teleconnections. Furthermore, the spatiotemporal decomposition module is introduced to factorizes the multivariate climate tensor into a low-rank set of global temporal bases and station-specific loadings, and then derives a station-level global component that summarizes common spatiotemporal trends for target forecasting. DoIGNN jointly performs prediction over the ACHZ-based and learned graphs, thereby uncovering both cross-variable and cross-regional dependencies. The environmental-resource science domain experts from Ningbo Natural Resources and Planning Big Data Center co-developed the ACHZ construction by specifying agronomically meaningful indicators and thresholding rules. In addition, they provided domain-relevant criteria for more comprehensive and accurate model evaluation.

The main contributions of this article are as follows.

1. DoIGNN partitions stations into ACHZs and constructs a ACHZ-guided hierarchical graph that encourages within-zone edges while selectively allowing cross-zone connections.
2. By spatiotemporal decomposition, DoIGNN exploits shared climate dynamics while avoiding excessive reliance on raw auxiliary variables. The temporal regularization further preserves temporal structure, leading to more stable learning and improved robustness.
3. DoIGNN performs forecasting on real-word climate datasets validate the effectiveness of our model. Compared with several baseline methods under different prediction horizons, our method demonstrates superior performance.

2 Preliminaries and Problem Statement

2.1 Agro-Climatic Homogeneous Zones (ACHZs)

Agro-Climatic Homogeneous Zones (ACHZs) partition monitoring stations into groups sharing similar long-term agro-climatic regimes (e.g., thermal resources, moisture availability, and rainfall seasonality). Unlike purely data-driven clustering, ACHZs are defined by rule-based agro-climatic indicators computed from multi-decadal daily meteorological records, improving interpretability and aligning the graph structure with agro-ecological principles [Fischer *et al.*, 2012].

Given the annual accumulated Growing Degree Days $\overline{\mathrm{GDD}}_i$, the Aridity Index AI_i, and the Seasonality Index SI_i, each station i can be assigned to an ACHZ by discretizing these indicators into ordered classes with agronomically meaningful thresholds:

$$z_i = \phi_{\mathrm{th}}(\overline{\mathrm{GDD}}_i) \ \oplus \ \phi_{\mathrm{mo}}(\mathrm{AI}_i) \ \oplus \ \phi_{\mathrm{se}}(\mathrm{SI}_i), \tag{1}$$

where $\phi_{\mathrm{th}}, \phi_{\mathrm{mo}}, \phi_{\mathrm{se}}$ are rule-based binning functions (e.g., low/medium/high heat, arid-humid regimes, weak-strong seasonality), and $\oplus$ denotes concatenation into a unique ACHZ identifier. This yields zones that are directly interpretable in terms of agricultural climate constraints, enabling principled spatial regularization in subsequent graph modeling.

2.2 Problem Statement

Consider a multivariate spatiotemporal climate dataset $\mathbf{X} \in \mathbb{R}^{T\times N\times C}$, where T denotes the number of time steps (days), N is the number of stations, and C is the number of climate factors. Given a target channel, the goal is to forecast its future values $\mathbf{Y} \in \mathbb{R}^{f\times N}$ over a future horizon of $f = 12$ days using the historical $h = 12$ days of all variables.

Formally, for each station $n \in \{1, \ldots, N\}$ and each reference time t satisfying $L \leq t \leq T - H$, we define the input history tensor and prediction target as:

$$\begin{aligned} \mathbf{X}_t^{(n)} &= \mathbf{X}_{t-h:t-1,\,n,\,:} \in \mathbb{R}^{h\times C}, \\ \mathbf{Y}_t^{(n)} &= \mathbf{Y}_{t:t+f-1,\,n} \in \mathbb{R}^{f}. \end{aligned} \tag{2}$$

We aim to learn a forecasting function f_θ parameterized by θ such that:

$$\hat{\mathbf{y}}_t^{(n)} = f_\theta\left(\mathbf{X}_t^{(n)}, \mathcal{G}\right) \in \mathbb{R}^f, \tag{3}$$

where $\mathcal{G}$ denotes a set of graphs describing inter-station dependencies.

3 Methods

The DoIGNN architecture mainly consists of ACHZs Construction, graph generation and prediction head module. The schematic diagram of DoIGNN can be seen in Figure 1.

3.1 ACHZs Construction

We construct an ACHZ-guided graph to inject domain knowledge into spatial dependency modeling. Each station is first assigned a zone label using rule-based discretization of long-term agro-climatic indicators, and then a weighted adjacency matrix is derived by combining within-zone consistency with cross-zone similarity.

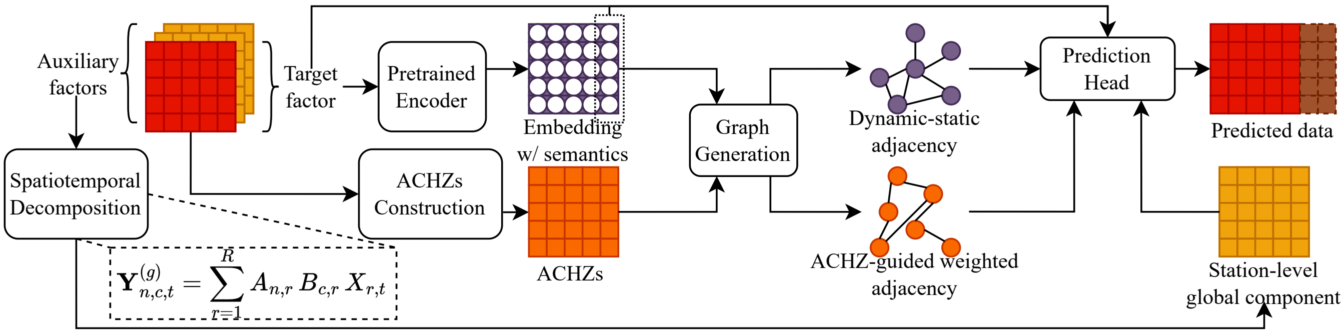

Figure 1: Schematic diagram of DoIGNN

Let $\mathcal{D}_i^{(y)}$ denote the set of days in year y at station i. Using daily mean temperature $T_i(d)$ (°C), daily minimum/maximum temperatures $T_i^{\min}(d), T_i^{\max}(d)$ (°C), daily precipitation $P_i(d)$ (mm), and daily incoming solar radiation $R_{s,i}(d)$ (MJ m^{-2} day^{-1}), we compute the following long-term descriptors.

Station-Wise Agro-Climatic Descriptors

For station i, we compute a climatological descriptor vector

$$\mathbf{c}_i = \left[\overline{T}_i,\ \overline{\mathrm{GDD}}_i(T_b),\ \overline{\mathrm{FFP}}_i,\ \overline{\mathrm{HD}}_i,\ \mathrm{AI}_i,\ \mathrm{SI}_i\right] \in \mathbb{R}^6, \quad (4)$$

where $\overline{T}_i$ is the multi-year mean daily temperature, $\overline{\mathrm{GDD}}_i(T_b)$ is the multi-year mean annual GDD with base temperature T_b, $\overline{\mathrm{FFP}}_i$ is the mean frost-free period, $\overline{\mathrm{HD}}_i$ is the mean number of heat-stress days (e.g., $T^{\max} \geq 35$°C).

Thermal Resources via GDD

For a crop-specific base temperature T_b, the annual accumulated GDD is

$$\mathrm{GDD}_i^{(y)} = \sum_{d \in \mathcal{D}_i^{(y)}} \max\bigl(T_i(d) - T_b,\, 0\bigr). \quad (5)$$

We use the multi-year climatological mean $\overline{\mathrm{GDD}}_i = \frac{1}{Y}\sum_y \mathrm{GDD}_i^{(y)}$ as a proxy of heat availability, which is widely used to quantify crop development and phenological progress.

Moisture Availability via AI

Following a commonly adopted definition in dryland and agro-environmental studies, we define the aridity index as

$$\mathrm{AI}_i = \frac{\overline{P}_i}{\overline{\mathrm{PET}}_i}, \quad (6)$$

where $\overline{P}_i$ is the multi-year mean annual precipitation (mm year^{-1}) and $\overline{\mathrm{PET}}_i$ is the multi-year mean annual potential evapotranspiration (mm year^{-1}). The AI $= P/\mathrm{PET}$ formulation is standard in global aridity characterization [Middleton *et al.*, 1992].

When full meteorological inputs for FAO-56 Penman-Monteith are unavailable, we estimate PET using the Makkink reference evapotranspiration estimator based on solar radiation and pressure-related psychrometric variables, and then aggregate daily ET_0 to annual PET [Allen *et al.*, 1998].

Rainfall Seasonality via SI

To characterize intra-annual rainfall concentration, we compute a seasonality index based on monthly precipitation climatology $\{\overline{P}_{i,m}\}_{m=1}^{12}$:

$$\mathrm{SI}_i = \frac{\sum_{m=1}^{12} \left|\overline{P}_{i,m} - \frac{\overline{P}_i}{12}\right|}{\overline{P}_i}, \quad (7)$$

where $\overline{P}_i$ and $\overline{\mathrm{PET}}_i$ are multi-year mean annual precipitation and potential evapotranspiration, and $\{\overline{P}_{i,m}\}_{m=1}^{12}$ is monthly precipitation climatology. SI measures deviation from uniformly distributed rainfall across months and is commonly used to describe rainfall seasonality patterns [Walsh and Lawler, 1981].

3.2 Spatiotemporal Decomposition

Given a 3-D tensor $\mathbf{Y} \in \mathbb{R}^{T \times N \times C}$, the decomposition module separates $\mathbf{Y}$ into a shared global component and a local residual:

$$\mathbf{Y} = \mathbf{Y}^{(g)} + \mathbf{Y}^{(l)}. \quad (8)$$

We model $\mathbf{Y}^{(g)}$ via a rank-R CANDECOMP/PARAFAC (CP) factorization [Harshman, 1970; Carroll and Chang, 1970; Kolda and Bader, 2009]:

$$\mathbf{Y}^{(g)}_{n,c,t} = \sum_{r=1}^{R} A_{n,r}\, B_{c,r}\, X_{r,t}, \quad (9)$$

where $\mathbf{A} \in \mathbb{R}^{N \times R}$ captures spatial loadings, $\mathbf{B} \in \mathbb{R}^{C \times R}$ captures variable loadings, and $\mathbf{X} \in \mathbb{R}^{R \times T}$ denotes global temporal factors shared across all stations and variables. Equivalently, $\mathbf{Y}^{(g)} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{x}_r$ with $\circ$ being the outer product. The local term is defined as the residual $\mathbf{Y}^{(l)} = \mathbf{Y} - \mathbf{Y}^{(g)}$.

To encourage temporally consistent dynamics in $\mathbf{X}$, we introduce a one-step predictive regularizer using a Temporal Convolutional Network (TCN) $\phi(\cdot)$ [Bai *et al.*, 2018]. For a window length w, we penalize deviations from the one-step prediction:

$$\mathcal{R}_{temp}(\mathbf{X}) = \sum_{t=1}^{T-w} \left\| \mathbf{X}_{:,\,t+1:t+w} - \phi(\mathbf{X}_{:,\,t:t+w-1}) \right\|_F^2 \quad (10)$$

Finally, the factors are estimated by minimizing a regularized reconstruction objective:

$$\min_{\mathbf{A},\mathbf{B},\mathbf{X}} \left\| \mathbf{Y} - \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{x}_r \right\|_F^2 + \lambda \left(\|\mathbf{A}\|_F^2 + \|\mathbf{B}\|_F^2 + \|\mathbf{X}\|_F^2 \right) + \mu \, \mathcal{R}_{\text{temp}}(\mathbf{X}) \tag{11}$$

where $\lambda, \mu > 0$ control weight decay and temporal smoothness, respectively. In practice, we solve ((11)) by alternating updates: fixing $(\mathbf{B}, \mathbf{X})$ to update $\mathbf{A}$, fixing $(\mathbf{A}, \mathbf{X})$ to update $\mathbf{B}$, and fixing $(\mathbf{A}, \mathbf{B})$ to update $\mathbf{X}$, while periodically applying the TCN-based temporal step.

For downstream forecasting, we derive a station-wise global representation as auxiliary information by projecting temporal factors through spatial loadings:

$$\mathbf{X}_{aux} = \mathbf{A}\mathbf{X} \in \mathbb{R}^{N \times T}, \qquad X_{aux,n,t} = \sum_{r=1}^{R} A_{n,r} X_{r,t}. \tag{12}$$

To avoid information leakage, the decomposition is fitted using only the training period.

3.3 Pre-trained Encoder

Our encoder follows the masked autoencoder (MAE) paradigm [He *et al.*, 2021a]. Given the input sequences, we perform patch-wise masking and representation learning for each station. For node i, let $\mathbf{X}_i \in \mathbb{R}^T$ denote the input sequence, we split it into P non-overlapping patches of length L, where the j-th patch is $\mathbf{X}_{i,j} \in \mathbb{R}^L$. Then, a high ratio of patches is randomly masked and fed into the corresponding encoder. The encoder is composed of four Transformer blocks to capture temporal patterns, producing patch embeddings $H_j^i \in \mathbb{R}^d$ for local inputs.

Before decoding, masked positions are filled with trainable mask tokens. The decoder contains one Transformer block followed by an MLP head, which reconstructs the j-th patch as $\hat{\mathbf{X}}_j^i \in \mathbb{R}^L$ from H_j^i. We optimize the reconstruction error only on masked patches using RMSE:

$$\mathcal{L}_{emb}(\mathbf{X}, \hat{\mathbf{X}}) = \sqrt{\frac{1}{M} \sum_{m=1}^{M} \left(\frac{\left\| \left(\mathbf{X} - \hat{\mathbf{X}} \right) \odot M_m \right\|_F^2}{\text{sum}(M_m)} \right)}, \tag{13}$$

where M is the number of masked patches, M_m is the corresponding mask matrix, $\|\cdot\|_F$ denotes the Frobenius norm, and $\text{sum}(\cdot)$ sums all entries of a matrix. After training, the trained encoder can be used to model the overall understanding of the time series, as the autoencoder masks the majority of the input sequence to encourage the learning of high-level semantics.

3.4 Graph Generation

To fully model spatiotemporal dependencies in climate data, DoIGNN constructs (i) a dynamic-static adjacency from node embeddings and raw sequences, and (ii) an ACHZ-guided weighted adjacency. The flowchart is shown in Figure 2.

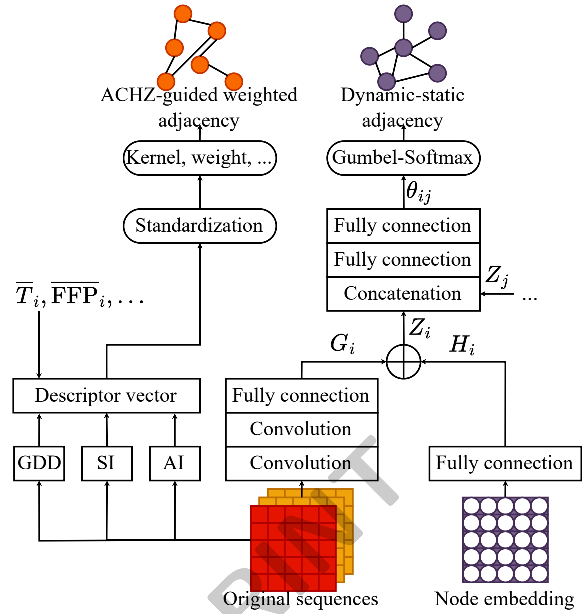

Figure 2: Flowchart of graph generation

Dynamic-Static Adjacency

For node i, we extract a static representation from the input sequence $\mathbf{X}^i$ via a lightweight CNN-MLP encoder:

$$G_i = \text{FC}\left(\text{Conv}(\text{Conv}(\mathbf{X}_i))\right), \tag{14}$$

where each $\text{Conv}(\cdot)$ denotes a convolutional block with convolution, ReLU, and batch normalization, and $\text{FC}(\cdot)$ denotes a fully-connected block with ReLU and batch normalization. We then fuse the static feature G^i with the masked embedding H^i to obtain a dynamic-static representation:

$$Z_i = \text{FC}\left(H_i\right) + G_i. \tag{15}$$

We aim to learn a sparse adjacency by modeling edge existence between node pairs. Given Z^i and Z^j, we compute the Bernoulli logits and probabilities as:

$$\begin{aligned} \Theta_{ij} &= \text{FC}(\text{FC}(Z_i \| Z_j)), \\ \Theta'_{ij} &= \text{softmax}(\Theta_{ij}), \end{aligned} \tag{16}$$

where $\Theta_{ij} \in \mathbb{R}^2$ denotes unnormalized logits and Θ'_{ij} is the corresponding normalized Bernoulli distribution. To enable end-to-end optimization with discrete sampling, we employ the Gumbel-Softmax reparameterization [Jang *et al.*, 2022]:

$$A_{ij}^{ds} = \text{softmax}\left(\left(\Theta_{ij} + g\right)/\tau\right), \tag{17}$$

where $g \in \mathbb{R}^2$ is sampled i.i.d. from $\text{Gumbel}(0, 1)$ and τ is the temperature. To regularize the learned structure and control sparsity, we introduce a kNN prior graph A^k and penalize deviations using a cross-entropy term:

$$\mathcal{R}_{graph} = \sum_{ij} -A_{ij}^k \log \Theta'_{ij} - \left(1 - A_{ij}^k\right) \log\left(1 - \Theta'_{ij}\right), \tag{18}$$

where A^k specifies the kNN connectivity.

ACHZ-Guided Weighted Adjacency

Let $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_N]^\top$ be the descriptor matrix. We first standardize each feature dimension:

$$\tilde{\mathbf{c}}_i = \frac{\mathbf{c}_i - \boldsymbol{\mu}}{\boldsymbol{\sigma}}, \tag{19}$$

and define an agro-climatic similarity kernel

$$s_{ij} = \exp\left(-\frac{\|\tilde{\mathbf{c}}_i - \tilde{\mathbf{c}}_j\|_2^2}{2\sigma^2}\right), \tag{20}$$

where σ controls the kernel bandwidth. The ACHZ-guided weighted adjacency $A^{ACHZ}$is

$$A_{ij}^{ACHZ} = s_{ij} \cdot \begin{cases} 1, & z_i = z_j, \\ \lambda, & z_i \neq z_j, \end{cases} \qquad A_{ii} = 0, \tag{21}$$

where $\lambda \in (0, 1)$ down-weights cross-zone connections to emphasize within-zone homogeneity while still allowing weak long-range interactions.

3.5 Prediction Head

We adopt a downstream GNN, an ACHZ-guided Graph WaveNet, to perform multi-step forecasting. To jointly leverage the learned dynamic-static ACHZ-guided weighted adjacency, we treat A^{ds} and A^{ACHZ} as undirected graph and obtain forward & backward transitions.

ACHZ-guided Graph WaveNet further employs dilated causal convolutions to capture temporal dependencies. In the fusion layer, we combine the latent representation with the last patch embedding H_P to obtain the final prediction:

$$\begin{aligned} H_{gw} &= \mathrm{GW}(\mathbf{X}_{target} || \mathbf{X}_{aux}), \\ \hat{\mathbf{Y}} &= \mathrm{Conv}(\mathrm{Conv}(\mathrm{MLP}(H_P) + H_{gw})), \end{aligned} \tag{22}$$

where $\mathbf{X}_{target} \in \mathbb{R}^{N \times T}$ is the target factor. $||$ denotes concatenation, MLP consists of two fully-connected layers with ReLU, and Conv is a 1×1 convolution with ReLU. The output $\hat{\mathbf{Y}}$ represents the f-step-ahead forecasts for n nodes.

Finally, we optimize the prediction model using MAE together with the graph regularizer in:

$$\begin{aligned} \mathcal{L}_{predict} &= \mathcal{L}(\hat{\mathbf{Y}}, \mathbf{Y}) + \lambda_{graph} \mathcal{R}_{graph} \\ &= \frac{1}{fn} \sum_{j=1}^{f} \sum_{i=1}^{n} \left| \hat{\mathbf{Y}}_{ij} - \mathbf{Y}_{ij} \right| + \lambda_{graph} \mathcal{R}_{graph}, \end{aligned} \tag{23}$$

where $\mathbf{Y}$ is the ground truth, and λ_{graph} controls the strength of structure regularization. An overview of the full training pipeline is provided in Algorithm 1.

Algorithm 1: DoIGNN

```
Input: Past observations: X ∈ R^{T×N×C}.
Output: Future forecasts: Ŷ ∈ R^{f×N}.
/* Spatiotemporal Decomposition */
for i_1 alternating iterations do
    Fix TCN φ(·), A, B update X using Equation (11).
    Fix TCN φ(·), X and update A, B using Equation (11).
    if mod iteration then
        Fix A, B, X and update TCN φ(·) using Equation (10).
    end
end
Obtain station-wise global representation X_aux by Equation (12).
/* Pre-trained Encoder */
Split Xinto P patches and mask them with probability p.
for i_2 epochs do
    Encode X to obtain H, complement masked positions with trainable patches, and decode to yield X̂.
    Update encoder & decoder parameters using Equation (13).
end
Obtain pretrained node embeddings H.
/* Prediction Head */
Compute ACHZ through Equation (4) - (7)
Construct ACHZ-guided weighted adjacency A^{ACHZ} by Equation (19) - (21).
for i_3 epochs do
    Generate dynamic-static adjacency A^{ds} by Equation (14)-(18).
    /* Prediction module */
    Compute transition matrices and perform dilated causal convolutions by Equation (22) to obtain H_gw.
    Fuse H_P and H_gw via (22) to produce Ŷ.
    Update prediction head parameters using (23).
end
```

4 Experiments

4.1 Dataset

We use the China Meteorological Assimilation Driving Datasets for the SWAT model-Long time series (CMADS-L V1.0), obtained from the official CMADS portal.[1] CMADS is a meteorological forcing dataset designed for hydrological and agro-environmental applications, produced by assimilating multi-source observations and numerical background fields [Meng *et al.*, 2018]. CMADS V1.0 provides daily meteorological fields over East Asia (approximately 0°-65°N, 60°-160°E) at $1/3°$ spatial resolution, organized as a 195×300 grid. Climate factors includes: surface atmospheric pressure p (hPa), daily mean temperature $\bar{T}$ (°C), daily maximum/minimum temperatures $T^{\max}, T^{\min}$ (°C), 24-hour accumulated precipitation P (mm), relative humidity RH (%), solar radiation R_s (MJ m^{-2}), specific humidity q ($\mathrm{g\,kg}^{-1}$), and wind speed u ($\mathrm{m\,s}^{-1}$).

In this study, the forecasting target is the daily mean temperature, solar radiation and specific humidity. We subsample $N = 130$ stations on a regular grid by selecting one site every 15 latitude indices and one site every 30 longitude indices, yielding a 13×10 station set.

[1] https://www.cmads.org/

Datasets	Methods	Horizon 3				Horizon 6				Horizon 12			
		MAE	RMSE	NSE	KGE	MAE	RMSE	NSE	KGE	MAE	RMSE	NSE	KGE
Average Temperature	AGCRN	4.5517	6.5774	0.8542	0.8152	4.4600	6.2099	0.8701	0.8285	4.7601	6.5818	0.8540	0.8152
	Crossformer	3.7753	5.0153	0.9152	0.8641	4.1669	5.4702	0.8992	0.8219	4.5749	5.9810	0.8794	0.7998
	DCRNN	3.4406	4.6053	0.9285	0.8777	3.8580	5.1542	0.9105	0.8665	4.2656	5.6267	0.8933	0.8352
	GWNet	3.6785	5.0882	0.7382	0.3256	4.1683	5.6851	0.6733	0.3401	4.8886	6.5952	0.5601	0.3625
	MTGNN	4.3886	5.7360	0.8891	0.7832	5.0663	6.7010	0.8487	0.7458	6.0572	8.0127	0.7836	0.6836
	STAEformer	3.1057	4.2495	0.9391	0.8825	3.8038	5.0490	0.9141	0.8265	4.8632	6.2786	0.8671	0.7442
	StemGNN	3.8485	6.1733	0.8715	0.8301	4.1262	6.4820	0.8584	0.8134	4.4708	6.8545	0.8416	0.8061
	WaveNet	**2.8077**	4.3968	0.2832	-1.2076	**3.0342**	**4.7023**	0.1805	-1.1916	**3.4245**	**5.2627**	-0.0270	-1.1595
	DoIGNN	2.8893	**4.1478**	**0.9420**	**0.9085**	3.4268	4.8750	**0.9199**	**0.8829**	3.9589	5.5587	**0.8958**	**0.8660**
Solar Radiation	AGCRN	3.5236	4.5878	0.6971	0.7941	3.6210	4.7080	0.6812	0.7865	3.8030	4.9276	0.6512	0.7668
	Crossformer	3.4260	4.5561	0.7022	0.8246	3.4260	4.5561	0.7022	**0.8246**	3.7127	4.8695	0.6604	**0.8046**
	DCRNN	**3.1305**	**4.3004**	**0.7338**	**0.8165**	3.3518	4.5150	0.7066	0.7935	3.7899	4.9960	0.6411	0.7579
	GWNet	3.3451	4.4530	0.1437	0.5076	3.3861	4.5145	0.1203	0.5166	3.6021	4.7356	0.0326	0.5372
	MTGNN	3.6169	4.7857	0.6703	0.7309	3.6169	4.7857	0.6703	0.7309	3.9756	5.1814	0.6141	0.6952
	STAEformer	3.8173	5.1648	0.6163	0.7530	4.2601	5.8060	0.5156	0.7079	4.2289	5.5980	0.5500	0.7045
	StemGNN	3.3799	4.4714	0.7123	0.7719	3.4673	4.5763	0.6988	0.7587	3.6776	4.8136	0.6670	0.7381
	WaveNet	3.6020	4.8193	-2.6656	-0.7193	3.7347	4.9346	-2.8409	-0.6890	4.0419	5.2053	-3.2687	-0.6155
	DoIGNN	3.1399	4.3061	0.7329	0.8131	**3.2864**	**4.4670**	**0.7127**	0.7996	**3.4872**	**4.6463**	**0.6895**	0.7704
Specific Humidity	AGCRN	1.7444	2.4009	0.8956	0.8572	1.5971	2.1652	0.9151	0.9036	1.6919	2.2957	0.9046	0.8967
	Crossformer	1.6095	2.2173	0.9110	0.8447	1.6197	2.2797	0.9059	0.8650	1.7733	2.4671	0.8898	0.8393
	DCRNN	1.0926	1.5689	0.9554	0.9469	1.2713	1.8416	0.9386	0.9333	1.5280	2.1906	0.9131	0.9197
	GWNet	1.6941	2.3124	0.7096	0.3025	1.7326	2.3770	0.6931	0.3098	1.8507	2.5165	0.6561	0.3169
	MTGNN	2.1349	2.8370	0.8543	0.7399	2.3742	3.1812	0.8168	0.7109	2.6914	3.6306	0.7614	0.6746
	STAEformer	1.1062	1.5340	0.9574	0.9583	1.2270	1.6899	0.9483	0.9431	1.5017	1.9961	0.9279	0.9074
	StemGNN	1.6801	2.4713	0.8894	0.9276	1.7825	2.5622	0.8812	0.9336	1.8495	2.6814	0.8698	0.9177
	WaveNet	1.1926	1.7617	0.3819	-1.2225	1.2698	1.8571	0.3132	-1.2123	1.4167	2.0427	0.1691	-1.1907
	DoIGNN	**1.0638**	**1.5019**	**0.9592**	**0.9699**	**1.2256**	**1.6847**	**0.9486**	**0.9446**	**1.3746**	**1.8875**	**0.9355**	**0.9358**

Table 1: Experiment results

4.2 Evaluation Criteria and Experiment Settings

Beyond point-wise error metrics, we adopt two metrics that are widely used to assess model performance in environmental science and meteorology: Nash-Sutcliffe Efficiency (NSE) [Nash and Sutcliffe, 1970] and the Kling-Gupta Efficiency (KGE) [Gupta *et al.*, 2009].

Let $y_t(i)$ and $\hat{y}_t(i)$ denote the observed and predicted values at station i and time index t over an evaluation window $\mathcal{T}$, and let $\bar{y}(i) = \frac{1}{|\mathcal{T}|}\sum_{t\in\mathcal{T}} y_t(i)$ be the sample mean. The NSE is defined as

$$\mathrm{NSE}(i) = 1 - \frac{\sum_{t\in\mathcal{T}} \left(y_t(i) - \hat{y}_t(i)\right)^2}{\sum_{t\in\mathcal{T}} \left(y_t(i) - \bar{y}(i)\right)^2} \tag{24}$$

NSE quantifies the relative improvement over a mean-value baseline [Nash and Sutcliffe, 1970].

To provide a more diagnostic assessment that disentangles error sources, we further report KGE, which jointly evaluates linear association, bias, and variability [Gupta *et al.*, 2009]:

$$\mathrm{KGE}(i) = 1 - \sqrt{\left(r(i) - 1\right)^2 + \left(\beta(i) - 1\right)^2 + \left(\gamma(i) - 1\right)^2} \tag{25}$$

where $r(i)$ is the Pearson correlation between $\{y_t(i)\}$ and $\{\hat{y}_t(i)\}$, $\beta(i) = \mu_{\hat{y}}(i)/\mu_y(i)$ measures the mean bias ratio, and $\gamma(i) = \sigma_{\hat{y}}(i)/\sigma_y(i)$ measures the variability ratio, with $\mu(\cdot)$ and $\sigma(\cdot)$ denoting the sample mean and standard deviation over $\mathcal{T}$. Similar to NSE, values closer to 1 imply better overall agreement [Gupta *et al.*, 2009].

In spatiotemporal decomposition, the learning rate is 0.0005. The temporal regularization parameter λ_{GLD} is 0.1. TCN $\phi(\cdot)$, $\mathbf{A}, \mathbf{B}, \mathbf{X}$ are optimized in 50 iterations and alternate training fashion for every 5 iterations. In embedding module, mask probability p is set as 75%. Embedding and prediction module are both trained for $i_2 = i_3 = 100$ times. The learning rate for embedding is 0.0005 without decay, and for prediction is 0.005 with decay. The softmax temperature parameter τ is 0.5. Finally, The graph regularization parameter λ_{graph} is 0.1. Details for ACHZs can be seen in Table 3. Relevant code is available at the project repository.

4.3 Forecasting Results

Figure 3 visualizes the constructed ACHZs, which provide an interpretable partition of stations with similar long-term agroclimatic regimes. Table 1 reports the macro-averaged multi-horizon forecasting results over all stations, where the best scores are highlighted in bold. Overall, DoIGNN achieved the most consistent performance across horizons and variables, especially on skill-oriented metrics (NSE and KGE), indicating improved predictive reliability beyond point-wise errors.

Methods	Horizon 3				Horizon 6				Horizon 12			
	MAE	RMSE	NSE	KGE	MAE	RMSE	NSE	KGE	MAE	RMSE	NSE	KGE
DoIGNN w/o A^{ACHZ}	3.3719	4.5856	0.9291	0.8848	3.8006	5.1667	0.9101	0.8555	4.1501	5.5278	**0.8970**	0.8406
DoIGNN w/o STD	3.8264	5.1587	0.9103	0.8772	4.4102	5.8590	0.8843	0.8321	4.5579	6.0073	0.8783	0.8213
DoIGNN w/o STD, A^{ACHZ}	**2.8077**	4.3968	0.2832	-1.2076	**3.0342**	**4.7023**	0.1805	-1.1916	**3.4245**	**5.2627**	-0.0270	-1.1595
DoIGNN	2.8893	**4.1478**	**0.9420**	**0.9085**	3.4268	4.8750	**0.9199**	**0.8829**	3.9589	5.5587	0.8958	**0.8660**

Table 2: Ablation study

Axis	Indicator (unit)	Discretization rule (ordered bins)
Thermal	$\mathrm{GDD}_i(T_b)$ (°C·day/year), $T_b = 10$°C	$[0, 1000), [1000, 2000), [2000, 3000), [3000, \infty)$
Moisture	$\mathrm{AI}_i = \overline{P}_i / \overline{\mathrm{PET}}_i$ (-)	$[0, 0.2), [0.2, 0.5), [0.5, 0.65), [0.65, 1.0), [1.0, \infty)$
Seasonality	SI_i (-)	$[0, 0.2), [0.2, 0.4), [0.4, 0.6), [0.6, 0.8), [0.8, \infty)$

Table 3: Criteria for Agro-Climatic Homogeneous Zones

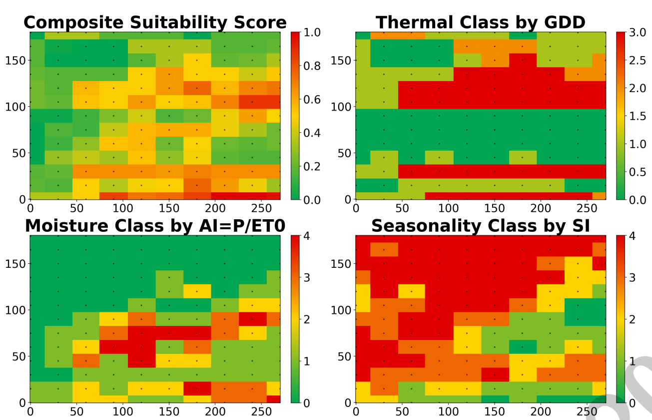

Figure 3: Result of constructing ACHZs

Specifically, in average temperature dataset, DoIGNN attains the best overall skill for temperature forecasting, yielding the highest NSE and KGE at all horizons. Notably, some baselines like WaveNet can obtain relatively small MAE but exhibit substantially degraded NSE and KGE. This discrepancy suggests that those models without graph information may fit the central tendency while failing to reproduce the temporal variability and correlation structure. GWNet, on the other hand, markedly improved prediction errors by integrating GCN rather than TCN for spatial relationship analysis. This underscored the significance of graph relationships from domain knowledge in augmenting predictive performance of climate factor.

Finally, Table 2 demonstrates the ablation study results. Removing the ACHZ-guided adjacency (A^{ACHZ}) causes a clear degradation in NSE/KGE, indicating that domain-consistent spatial constraints are critical for preserving forecast skill and preventing spurious cross-regime message passing. Removing the spatiotemporal decomposition (STD) further deteriorates performance, confirming that explicitly extracting global shared dynamics improves stability and generalization. When both STD and A^{ACHZ} are removed, the model collapses to the WaveNet-style backbone and exhibits the same behavior as the corresponding baseline, demonstrating that the performance gains of DoIGNN primarily come from the proposed domain-informed graph prior and the global–local decomposition rather than architectural scaling alone.

5 Discussion

The proposed forecasting model can be deployed as a decision-support component for crop management. Given multi-station predictions of climate factors over the forthcoming horizon, the system can derive risk-aware indicators at the zone and field scales such as heat & cold stress likelihood, rainfall deficit & excess, and humidity-driven disease favorability. These indicators enable practitioners to (i) schedule irrigation and drainage by anticipating short-term water stress, (ii) adjust sowing, fertilization, and harvest windows to avoid adverse temperature/rainfall regimes, and (iii) trigger early warnings for pest and fungal disease management under high-risk humidity–temperature patterns. By aggregating forecasts within agro-climatic homogeneous zones, the framework also supports coherent regional planning.

6 Conclusion

This paper presented DoIGNN, a domain-informed graph neural network for climate factor forecasting. DoIGNN injects agro-climatic knowledge via an ACHZ-guided adjacency, and further employs a temporally-regularized global–local spatiotemporal decomposition to obtain compact shared dynamics as auxiliary information. Built upon the ACHZ constraint and a learned dynamic-static dependency, the proposed prediction model jointly captures cross-region and cross-variable interactions. Experiments on CMADS demonstrate consistent improvements over strong baselines across horizons. In the future, we will focus on lightweight and scalable modeling for long-horizon forecasting over large-scale climate archives to reduce computational and storage costs without sacrificing predictive skill.

Ethical Statement

There are no ethical issues.

Acknowledgements

This work is supported by National Key R&D Program of China (Grant No. 2021YFC2101100), National Natural Science Foundation of China (62073288, 12075212), and Zhejiang University NGICS Platform, and their supports are thereby acknowledged.

References

[Allen *et al.*, 1998] Richard G. Allen, Luis S. Pereira, Dirk Raes, and Martin Smith. Crop evapotranspiration-guidelines for computing crop water requirements-fao irrigation and drainage paper 56. *Fao, Rome*, 300(9):D05109, 1998.

[Bai *et al.*, 2018] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, April 2018.

[Carroll and Chang, 1970] J. Douglas Carroll and Jih-Jie Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition. *Psychometrika*, 35(3):283–319, September 1970.

[Chen *et al.*, 2022] Chengcheng Chen, Qian Zhang, Mahsa H. Kashani, Changhyun Jun, Sayed M. Bateni, Shahab S. Band, Sonam Sandeep Dash, and Kwok-Wing Chau. Forecast of rainfall distribution based on fixed sliding window long short-term memory. *Engineering Applications of Computational Fluid Mechanics*, December 2022.

[Das and Ghosh, 2018] Monidipa Das and Soumya K. Ghosh. Data-driven approaches for meteorological time series prediction: A comparative study of the state-of-the-art computational intelligence techniques. *Pattern Recognition Letters*, 105:155–164, April 2018.

[Fischer *et al.*, 2012] G. Fischer, F. O. Nachtergaele, S. Prieler, E. Teixeira, G. Toth, H. van Velthuizen, L. Verelst, and D. Wiberg. *Global Agro-Ecological Zones (GAEZ v3.0)- Model Documentation*. IIASA, Laxenburg, Austria and FAO, Rome, Italy., Laxenburg, Austria; Rome, Italy, 2012.

[Giorgi *et al.*, 2023] Filippo Giorgi, Erika Coppola, Graziano Giuliani, James M. Ciarlo', Emanuela Pichelli, Rita Nogherotto, Francesca Raffaele, Piero Malguzzi, Silvio Davolio, Paolo Stocchi, and Oxana Drofa. The fifth generation regional climate modeling system, regcm5: Description and illustrative examples at parameterized convection and convection-permitting resolutions. *Journal of Geophysical Research: Atmospheres*, 128(6):e2022JD038199, 2023.

[Gong *et al.*, 2024] Yongshun Gong, Tiantian He, Meng Chen, Bin Wang, Liqiang Nie, and Yilong Yin. Spatio-temporal enhanced contrastive and contextual learning for weather forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 36(8):4260–4274, August 2024.

[Gupta *et al.*, 2009] Hoshin V. Gupta, Harald Kling, Koray K. Yilmaz, and Guillermo F. Martinez. Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1):80–91, October 2009.

[Haggag *et al.*, 2021] May Haggag, Ahmad S. Siam, Wael El-Dakhakhni, Paulin Coulibaly, and Elkafi Hassini. A deep learning model for predicting climate-induced disasters. *Natural Hazards*, 107(1):1009–1034, May 2021.

[Harshman, 1970] R. A. Harshman. Foundations of the parafac procedure: Models and conditions for an 'explanatory' multi-modal factor analysis. Technical Report PB197824, California Univ., Los Angeles., 1970.

[He *et al.*, 2021a] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners, December 2021.

[He *et al.*, 2021b] Sijie He, Xinyan Li, Timothy DelSole, Pradeep Ravikumar, and Arindam Banerjee. Sub-seasonal climate forecasting via machine learning: Challenges, analysis, and advances. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(1):169–177, May 2021.

[Humphries *et al.*, 2026] Usa Wannasingha Humphries, Muhammad Waqas, and Shakeel Ahmad. Novel deep learning framework for rainfall forecasting integrating generative adversarial and spatiotemporal graph neural networks. *Results in Engineering*, 29:108616, March 2026.

[Jang *et al.*, 2022] Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. In *International Conference on Learning Representations*, July 2022.

[Kimura, 2002] Ryuji Kimura. Numerical weather prediction. *Journal of Wind Engineering and Industrial Aerodynamics*, 90(12):1403–1414, December 2002.

[Kolda and Bader, 2009] Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, August 2009.

[Leutbecher and Palmer, 2008] M. Leutbecher and T. N. Palmer. Ensemble forecasting. *Journal of Computational Physics*, 227(7):3515–3539, March 2008.

[Li *et al.*, 2025] Menglin Li, Ming Yang, and Yixiao Yu. Short-term regional wind power forecast method based on hierarchical graph neural network and extreme value theory. *Energy Conversion and Management*, 341:120002, October 2025.

[Meng *et al.*, 2018] Xianyong Meng, Hao Wang, Chunxiang Shi, Yiping Wu, and Xiaonan Ji. Establishment and evaluation of the china meteorological assimilation driving datasets for the swat model (cmads). *Water*, 10(11):1555, November 2018.

[Middleton *et al.*, 1992] Nick Middleton, David Thomas, and UNEP, editors. *World Atlas of Desertification*. Edward Arnold : UNEP, London ; New York ; Melbourne, Australia ; Auckland, New Zealand, 1992.

[Nash and Sutcliffe, 1970] J. E. Nash and J. V. Sutcliffe. River flow forecasting through conceptual models part i — a discussion of principles. *Journal of Hydrology*, 10(3):282–290, April 1970.

[Nouri and Veysi, 2025] Milad Nouri and Shadman Veysi. Mitigating crop modeling uncertainties through machine learning in drylands. *Scientific Reports*, 15(1):42663, November 2025.

[Utku and Akyol, 2025] Anıl Utku and Sinem Akyol. Hybrid ga-convlstm for data-driven prediction of climate variables: A case study of the most biodiverse cities in india. *Neural Computing and Applications*, 37(24):19549–19578, August 2025.

[Walsh and Lawler, 1981] R. P. D. Walsh and D. M. Lawler. Rainfall seasonality: Description, spatial patterns and change through time. *Weather*, 36(7):201–208, 1981.

[Yousaf *et al.*, 2025] Umair Yousaf, Alessio De Rango, Luca Furnari, Donato D'Ambrosio, Alfonso Senatore, and Giuseppe Mendicino. Spatio-temporal graph neural networks to improve precipitation forecasts from numerical models. *Soft Computing*, 29(9):4481–4494, May 2025.

[Zhang *et al.*, 2026] Gangfeng Zhang, Shuo Wang, Jing Xu, Tim R. McVicar, Yun Cheng, Cesar Azorin-Molina, Lorenzo Minola, Nan Wang, Jiang Zhang, and Peijun Shi. A deep learning approach predicts o3 increase and pm2.5 declines under high emission scenario across the northern china urban agglomeration. *Urban Climate*, 65:102752, February 2026.

[Zhao *et al.*, 2025] Junfan Zhao, Fan Zhao, and Hang Deng. Evaluation of climate prediction models in yunnan, china: Traditional methods and ai approaches. *Scientific Reports*, 15(1):43347, December 2025.