Spatio-Temporal Graph Neural Networks for Weather Station Network Forecasting

Your Name

Abstract

Weather forecasting over station networks poses unique challenges due to irregular spatial geometry, non-local dependencies, and temporal non-stationarity. Traditional grid-based convolutional approaches fail to capture relational structure inherent in sensor networks. In this work, we formulate station-based weather forecasting as a spatio-temporal graph learning problem. We construct physically motivated and data-driven graphs over weather stations, and employ spatio-temporal graph neural networks to predict multi-horizon forecasts. We evaluate temporal and spatial generalization, conduct extensive ablation studies on graph construction and model components, and analyze failure modes. Our results demonstrate that explicit graph modeling significantly improves long-horizon forecasts and enables inductive generalization to unseen stations.

1 Introduction

Weather observations are often collected through networks of geographically distributed stations. Unlike gridded reanalysis products, station networks are irregular, sparse, and heterogeneous. Forecasting in such settings requires models that respect both temporal dynamics and relational spatial dependencies.

Conventional time-series models operate independently per station, ignoring spatial coupling. Grid-based convolutional neural networks impose artificial Euclidean structure, which is misaligned with station geometry. Graph-based learning offers a principled framework for modeling signals over irregular domains by explicitly encoding inter-station relationships.

This project investigates spatio-temporal graph neural networks (ST-GNNs) for forecasting weather variables over station networks. We emphasize rigorous graph construction, inductive generalization, and empirical validation.

2 Problem Formulation

Let $V = \{1, \ldots, N\}$ denote a set of weather stations. Each station i produces a multivariate time series

$$x_i(t) \in \mathbb{R}^F, \tag{1}$$

where F is the number of meteorological variables.

Stacking across all stations yields

$$X(t) \in \mathbb{R}^{N\times F}. \tag{2}$$

Given a historical window of length T,

$$\mathcal{X}_t = \{X(t-T+1), \ldots, X(t)\}, \tag{3}$$

the objective is to predict the next H steps,

$$\mathcal{Y}_t = \{X(t+1), \ldots, X(t+H)\}. \tag{4}$$

We seek to learn a function

$$\hat{\mathcal{Y}}_t = f_\theta(\mathcal{X}_t, G), \tag{5}$$

where G is a graph encoding spatial dependencies.

3 Graph Construction

The choice of graph is a critical inductive bias. We explore multiple constructions.

3.1 Geographic kNN Graph

Stations are connected to their k nearest neighbors based on geodesic distance $d(i,j)$. Edge weights are defined using a radial basis function:

$$W_{ij} = \exp\left(-\frac{d(i,j)^2}{\sigma^2}\right). \tag{6}$$

This graph encodes local spatial smoothness consistent with atmospheric continuity.

3.2 Correlation Graph

To capture non-local dependencies, we compute pairwise correlations between station time series using training data only:

$$\rho_{ij} = \mathrm{corr}(x_i, x_j). \tag{7}$$

Edges connect the top-k most correlated stations per node, with weights $|\rho_{ij}|$. This captures teleconnections and shared synoptic drivers.

3.3 Hybrid Graph

We combine physical and data-driven priors:

$$W = \lambda W_{\mathrm{kNN}} + (1-\lambda) W_{\mathrm{corr}}, \tag{8}$$

where $\lambda \in [0,1]$ controls the trade-off.

3.4 Learned Adaptive Adjacency (Optional)

An adaptive adjacency can be learned from node embeddings:

$$\tilde{W}_{ij} = \mathrm{softmax}_j\left(\frac{\phi(h_i)^\top \psi(h_j)}{\sqrt{d}}\right), \tag{9}$$

allowing the model to infer latent spatial structure.

4 Spatial Modeling with Graph Neural Networks

4.1 Graph Convolutional Networks

Given adjacency A and degree matrix D, a GCN layer is defined as:

$$H^{(l+1)} = \sigma\left(D^{-1/2} A D^{-1/2} H^{(l)} W^{(l)}\right). \tag{10}$$

This operation performs learnable diffusion over the graph, smoothing noisy sensor readings.

4.2 Inductive Aggregation via GraphSAGE

To enable generalization to unseen stations, we adopt neighborhood aggregation:

$$h_i^{(l+1)} = \sigma\left(W^{(l)}\left[h_i^{(l)} \| \frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)} h_j^{(l)}\right]\right). \tag{11}$$

This inductive formulation decouples learned parameters from node identity.

5 Temporal Modeling

5.1 Temporal Convolutional Networks

For each station, a temporal encoder maps historical observations to latent features:

$$z_i = \mathrm{TCN}(x_i(t-T+1:t)). \tag{12}$$

TCNs use causal dilated convolutions to capture long-range dependencies with stable gradients.

5.2 Alternative: GRU

As an alternative, gated recurrent units can be employed:

$$r_t = \sigma(W_r x_t + U_r h_{t-1}), \tag{13}$$
$$u_t = \sigma(W_u x_t + U_u h_{t-1}), \tag{14}$$
$$h_t = u_t \odot h_{t-1} + (1-u_t)\odot \tilde{h}_t. \tag{15}$$

6 Model Architecture

The proposed spatio-temporal GNN consists of:

1. A temporal encoder applied independently to each station,
2. A spatial GNN layer to propagate information across stations,
3. A decoder producing multi-horizon forecasts.

7 Training Objective

We minimize mean absolute error:

$$\mathcal{L}_{\mathrm{MAE}} = \frac{1}{NHF}\sum_{i,h,f}\left|Y_{i,h,f} - \hat{Y}_{i,h,f}\right|. \tag{16}$$

MAE is robust to outliers common in meteorological sensors.

8 Evaluation Protocol

8.1 Temporal Generalization

Models are trained on early time periods and evaluated on future unseen periods.

8.2 Spatial Generalization

A subset of stations is held out entirely during training. Performance on these unseen nodes evaluates inductive capability.

8.3 Metrics

We report MAE and RMSE per forecast horizon:

$$\mathrm{MAE}(h) = \frac{1}{NF} \sum_{i,f} |Y_{i,h,f} - \hat{Y}_{i,h,f}|. \tag{17}$$

We additionally compute skill scores relative to persistence baselines.

9 Ablation Studies

We perform ablations over:

- Graph construction (kNN, correlation, hybrid, learned),
- Neighborhood size k,
- Spatial operator (GCN vs GraphSAGE),
- Temporal encoder (TCN vs GRU),
- Forecast horizon.

These isolate the impact of modeling choices.

10 Pitfalls and Failure Modes

Common issues include data leakage in correlation graphs, over-smoothing in deep GNNs, missing-data bias, and seasonal non-stationarity. We mitigate these via training-only statistics, shallow GNNs, explicit masking, and seasonal features.

11 Related Work

This work builds upon diffusion-based spatio-temporal forecasting, inductive graph representation learning, and climate network analysis. Prior studies demonstrate the importance of explicit spatial modeling and motivate our comparative evaluation of physical and learned adjacency structures.

12 Expected Results

We expect graph-based models to outperform temporal-only baselines, with larger gains at longer horizons. Correlation-based and hybrid graphs are anticipated to improve mid-to-long-range forecasts, while inductive aggregation enables robust performance on unseen stations.

13 Conclusion

This project presents a rigorous framework for weather station forecasting using spatio-temporal graph neural networks. By combining principled graph construction, inductive learning, and careful evaluation, the approach demonstrates the necessity of graph-based modeling for irregular Earth observation systems.