MIRAGE: MULTIMODAL RADAR-CONDITIONED GRAPH LEARNING FOR RAINFALL INTERPOLATION AND FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Precipitation interpolation and forecasting over sparse weather station networks are challenging because rainfall exhibits highly localized and non-stationary dynamics, while ground observations are spatially limited. We propose MIRAGE, a radar-conditioned dynamic graph framework that integrates dense radar fields with sparse station observations for precipitation estimation at locations without local precipitation histories. MIRAGE constructs station-specific multimodal representations through cross-attention and uses them to adapt graph computation in two complementary ways: Dynamic Graph Construction determines *where* information should propagate, while a Meteorological Hypernetwork determines *how* propagated information should be transformed. Experiments on KMA and MeteoNet datasets show that MIRAGE achieves strong performance in both spatial interpolation and multi-step forecasting, with particularly consistent improvements at unobserved stations. MIRAGE also maintains competitive performance as observation sparsity increases. Ablation studies further demonstrate the complementary contributions of radar conditioning, cross-modal attention, and adaptive graph computation.

1 INTRODUCTION

Short-range precipitation estimation and forecasting are challenging because precipitation systems exhibit highly non-stationary dynamics and pronounced spatial heterogeneity. Reliable rainfall information is essential for flood warning, disaster mitigation, and hydrological risk assessment. Weather radar and Automated Weather Station (AWS) networks provide complementary observations: radar offers dense spatial coverage of evolving precipitation structures, whereas AWS networks provide accurate ground-level measurements at sparse and irregularly distributed locations.

Integrating these two modalities is non-trivial because they are defined on fundamentally different spatial domains. Radar observations are represented on dense, regular grids, while AWS measurements are collected at sparse point locations. In addition, conventional radar-based rainfall estimation often relies on an empirical Z–R relationship, $Z = aR^b$, which imposes a predefined mapping from radar reflectivity to rainfall rate. In practice, however, radar-derived precipitation and ground-level station measurements can differ in localized regions, as illustrated by the MeteoNet example in Figure 1. This suggests that radar should not necessarily be regarded as a direct substitute for station observations. Rather, radar can provide complementary information about the spatial organization and evolution of precipitation systems.

Recent grid-based weather forecasting and vision-based precipitation models have demonstrated strong performance in learning large-scale spatiotemporal dynamics from dense meteorological fields Pathak et al. (2022); Bi et al. (2023); Lam et al. (2022); Gao et al. (2022; 2023); Gong et al. (2024); Lin et al. (2025). However, these methods are primarily designed for regular spatial domains and do not naturally accommodate irregular ground-observation networks. Graph Neural Networks (GNNs), in contrast, provide a natural representation for sparse station networks Ejurothu et al. (2023); Chen et al. (2023b); Hettige et al. (2024); Wu et al. (2023). Graph-based interpolation methods further demonstrate that spatial information can be propagated toward locations with limited or missing observations Appleby et al. (2020); Wu et al. (2021); Li et al. (2023); Zheng et al. (2025).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

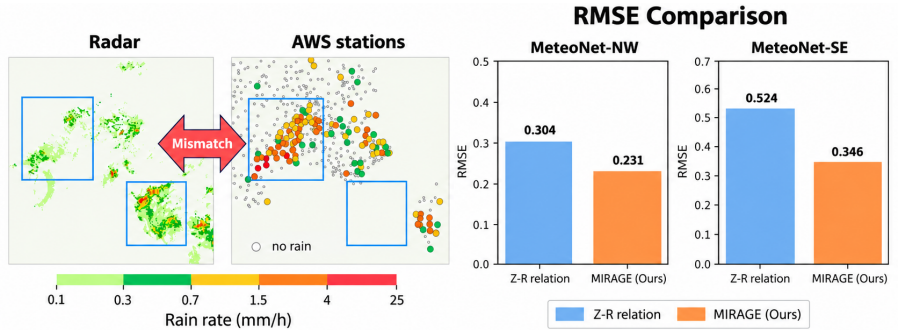


Figure 1: Comparison between radar-derived precipitation and ground-level precipitation observations. (a) Radar-derived precipitation obtained from an empirical Z–R relationship and accumulated AWS precipitation observations in the MeteoNet dataset, illustrating localized regions where the two measurements differ. (b) Average RMSE comparison between the empirical Z–R baseline and MIRAGE on the KMA and MeteoNet datasets.

Nevertheless, existing station-based graph models typically define spatial relationships using fixed geographical structures or information available only within the station network. Such representations may be insufficient for precipitation modeling. The effective relationship between two stations can depend on the location, scale, and evolution of a precipitation system and therefore need not coincide with geographical proximity. Meanwhile, the surrounding radar field contains dense spatial information that may not be recoverable from sparse station measurements alone. This motivates a different use of radar: rather than merely augmenting station features, radar can provide external meteorological context that adapts graph computation over the station network.

This perspective leads to two fundamental questions. First, *where* should information be aggregated under the current precipitation state? Second, *how* should the aggregated information be transformed? The first question concerns adaptive spatial aggregation over geographical neighborhoods, whereas the second concerns adaptive message transformation. Hypernetworks provide a principled mechanism for generating context-dependent parameters Ha et al. (2016), while cross-attention enables selective interaction between heterogeneous spatial representations.

We propose **MIRAGE**, a radar-conditioned graph learning framework for rainfall interpolation and forecasting at locations without local precipitation histories. MIRAGE separately encodes radar observations, station measurements, and spatial coordinates, and integrates them through cross-attention to construct a shared *Multimodal Context*. Rather than using this context merely as an additional node feature, MIRAGE uses it to jointly modulate two complementary aspects of graph computation. First, *Dynamic Graph Construction* adapts the aggregation weights over geographical neighborhoods, determining *where* information should propagate. Second, a *Hypernetwork* generates context-dependent parameters for graph message transformation, determining *how* the propagated information should be transformed. Consequently, both spatial aggregation and message transformation adapt to the current meteorological state.

We evaluate MIRAGE for both spatial interpolation and multi-step rainfall forecasting. To explicitly assess spatial generalization, a subset of target stations is fixed before training, and their local precipitation histories are completely excluded from the model inputs. MIRAGE must therefore estimate rainfall at these locations using radar information and observations from the remaining station network. Beyond this station-level holdout setting, we systematically vary the observation sparsity to examine robustness as station measurements become increasingly limited. We further conduct component ablations to investigate the roles of multimodal conditioning, adaptive spatial aggregation, and adaptive message transformation. Experiments are conducted across three geographical regions from the KMA and MeteoNet datasets.

Our contributions are summarized as follows:

- We introduce **MIRAGE**, a radar-conditioned graph learning framework for rainfall interpolation and forecasting at locations without local precipitation histories. MIRAGE combines

dense radar fields with sparse ground observations while preserving their distinct spatial representations.

- We develop a shared *Multimodal Context* that jointly modulates two complementary components of graph computation: the aggregation weights over geographical neighborhoods and the parameters used to transform graph messages. This enables both *where* information is aggregated and *how* it is transformed to adapt to the current meteorological state.
- We evaluate MIRAGE through station-level holdouts, varying observation sparsity, and component ablations across three geographical regions from South Korea (KMA) and France (MeteoNet), covering both spatial interpolation and multi-step rainfall forecasting.

2 METHODOLOGY

2.1 PROBLEM FORMULATION

Let $V = V_o \cup V_u$ denote a weather station network of N stations, where V_o and V_u are the observed and unobserved (target) stations, respectively. At prediction time t, MIRAGE receives a sequence of radar observations $\mathbf{X}_{\mathrm{radar}}$, station precipitation histories $\mathbf{X}_{\mathrm{aws}}$, station coordinates $\mathbf{X}_{\mathrm{coord}}$, and temporal information $\mathbf{X}_{\mathrm{time}}$.

Precipitation histories at V_u are unavailable to the model, while their coordinates remain known. Let $\mathcal{S}_t \subseteq V_o$ denote the set of stations whose precipitation histories are available as input. Our goal is to estimate precipitation $\hat{\mathbf{Y}} \in \mathbb{R}^{N \times T_{\mathrm{out}}}$ over the complete station network, including locations without local ground observations. We consider both spatial interpolation and multi-step forecasting.

A geographical k-nearest-neighbor graph provides an initial spatial structure among stations. Rather than treating geographical proximity as a fixed representation of spatial dependence, MIRAGE adapts graph computation to the current radar and station observations.

2.2 OVERVIEW OF MIRAGE

MIRAGE is a radar-conditioned dynamic graph framework that integrates dense radar observations with sparse ground measurements (Figure 2). Our central idea is to use radar not as a direct substitute for station observations, but as spatial context that controls graph computation. MIRAGE first constructs a station-level *Multimodal Context* from radar, station precipitation, geographical, and temporal information. This context controls graph computation through two complementary mechanisms: *Dynamic Graph Construction* determines *where* information propagates by adapting geographical connections, while a context-conditioned *Hypernetwork* determines *how* source information is transformed during propagation. Only stations in $\mathcal{S}_t$ provide ground-observation messages, while unobserved targets retain their radar-conditioned spatial and temporal context.

2.3 MULTIMODAL CONTEXT ENCODING

Radar provides dense gridded observations, whereas ground measurements are available only at sparse and irregular station locations. To effectively bridge this spatial discrepancy, MIRAGE encodes the two modalities separately before performing cross-modal fusion.

The radar sequence is first transformed into a compact set of spatiotemporal radar tokens. These tokens are generated by a radar encoder that is pretrained independently of MIRAGE and kept frozen during training. Specifically, the encoder is trained as a self-supervised spatiotemporal forecasting model that predicts six future radar frames from six past radar frames using solely radar reflectivity sequences (i.e., without any station or graph supervision). This pretraining task allows the encoder to capture the spatial organization and short-term temporal evolution of precipitation systems prior to interacting with the sparse station network. Once pretrained, only the encoder path (consisting of the spatial encoder and ConvLSTM) is retained and frozen, while the decoder used during pretraining is discarded. In parallel, each station's precipitation history is encoded using a shared station encoder, while station coordinates and temporal information are encoded separately to preserve essential spatial and periodic cues.

To resolve the structural mismatch between dense radar grids and discrete station locations, MIRAGE applies a cross-attention mechanism for station-specific radar retrieval. Each station forms a query by combining its precipitation, coordinate, and temporal representations, and attends to the pretrained radar tokens. The retrieved radar information is then adaptively gated and fused with the station representations, yielding the multimodal context matrix

$$\mathbf{C}_t = [\mathbf{c}_{1,t}, \ldots, \mathbf{c}_{N,t}]^\top \in \mathbb{R}^{N \times d}, \tag{1}$$

where $\mathbf{c}_{i,t}$ denotes the multimodal context vector of station i.

Crucially, because each query inherently retains coordinate and temporal information, stations lacking local precipitation histories (i.e., unobserved nodes) can still retrieve location-dependent radar context. Furthermore, $\mathbf{C}_t$ is not simply concatenated to the graph node features; instead, it serves as a conditioning representation for both Dynamic Graph Construction and the Hypernetwork. Complete specifications of the encoding and cross-attention architectures are provided in the Appendix.

2.4 Dynamic Graph Construction

We construct a geographical k-nearest-neighbor graph $\mathbf{A}_{\text{base}}$ from station coordinates. This graph provides a stable geographical prior by defining plausible candidate connections, but geographical proximity alone cannot capture the evolving spatial organization of precipitation.

MIRAGE therefore constructs a *Dynamic Mask* for each graph head using the station-level Multimodal Context and relative station geometry. The mask captures context-dependent relationships between candidate source and target stations while explicitly accounting for their relative locations and distances. The effective graph is

$$\boxed{\mathbf{A}^{(h)}_{\text{dyn},t} = \text{Norm}\left(\mathbf{A}_{\text{base}} \odot \mathbf{M}^{(h)}_t\right)}. \tag{2}$$

Only stations with available ground observations are allowed to act as message sources during propagation.

The Base Adjacency therefore specifies geographically plausible pathways, while the Dynamic Mask adjusts their strengths according to the current meteorological state. Dynamic Graph Construction determines *where* information propagates. All edge functions are shared across station pairs and do not introduce station-specific graph parameters.

2.5 Context-Conditioned Hypernetwork

Dynamic Graph Construction adapts spatial connectivity, but the appropriate transformation of ground observations may also vary across precipitation systems. MIRAGE therefore uses the Multimodal Context to condition the graph operator itself.

We first aggregate the station-level contexts into a *Global Context* $\mathbf{g}_t$ using permutation-invariant average pooling. A Hypernetwork then generates head-specific message-transformation parameters,

$$\boxed{\left\{\mathbf{W}^{(h)}_t, \mathbf{b}^{(h)}_t\right\}_{h=1}^{H} = \text{HyperNet}(\mathbf{g}_t)}. \tag{3}$$

Because these parameters depend on the current multimodal state, the model does not apply a single fixed transformation across different precipitation systems. The Hypernetwork therefore determines *how* source information is transformed during graph propagation.

2.6 Context-Adaptive Propagation and Target-Aware Decoding

Ground-observation messages originate only from stations in $\mathcal{S}_t$. Their precipitation histories are transformed using the parameters generated by the Hypernetwork and propagated according to the Dynamic Adjacency. The two adaptive mechanisms therefore control complementary aspects of graph computation:

$$\underbrace{\mathbf{A}_{\text{dyn},t}}_{\text{where}} \quad \text{and} \quad \underbrace{(\mathbf{W}_t, \mathbf{b}_t)}_{\text{how}}. \tag{4}$$

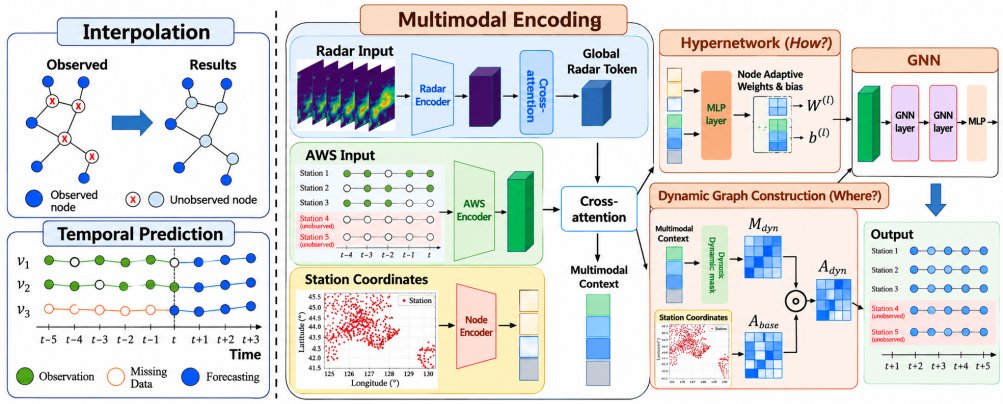


Figure 2: **Overview of MIRAGE.** Left: spatial interpolation at unobserved stations and temporal forecasting over the station network. Middle: radar observations, AWS precipitation histories, and station coordinates are encoded and integrated through cross-attention to construct a station-level *Multimodal Context*. Right: the Multimodal Context controls two complementary mechanisms: *Dynamic Graph Construction* adapts the geographical graph to determine *where* information propagates, while the *Hypernetwork* generates context-adaptive message transformations that determine *how* information is propagated. The resulting representations are used to estimate precipitation at both observed and unobserved stations. Auxiliary temporal features are omitted for visual clarity.

Messages from multiple graph heads are subsequently integrated for each target station. The propagated ground information is then combined with the target's own Multimodal Context through a learned gate. This target-aware fusion is particularly important for unobserved stations, where local precipitation histories are unavailable but radar, geographical, and temporal context remain accessible. A node-wise decoder finally maps the fused representation to the precipitation prediction $\hat{\mathbf{Y}} \in \mathbb{R}^{N \times T_{\text{out}}}$. Detailed definitions of the source transformation, multi-head aggregation, target-aware gating, and prediction decoder are provided in the Appendix.

2.7 Training Objective

MIRAGE is trained end-to-end by minimizing the mean squared error (MSE) between the predicted sequence $\hat{\mathbf{Y}}$ and the ground-truth observations $\mathbf{Y}$. Measurements at the fixed target stations V_u are excluded from the model input and from the training loss. Accordingly, the loss is computed over the observed station set V_o.

To improve generalization to unobserved locations, we additionally mask a random subset of stations in V_o from the model input at each training iteration. These temporarily masked stations remain part of the training targets, and the loss is evaluated over all stations in V_o:

$$\mathcal{L} = \frac{1}{|V_o| T_{\text{out}}} \sum_{i \in V_o} \sum_{\tau=1}^{T_{\text{out}}} \left(\hat{Y}_{i,t+\tau} - Y_{i,t+\tau} \right)^2 . \tag{5}$$

For spatial interpolation, we set $T_{\text{out}} = 1$ and reconstruct the contemporaneous precipitation field at time t. For temporal forecasting, the model predicts precipitation over $t+1, \ldots, t+T_{\text{out}}$.

This masking strategy prevents the model from relying exclusively on each station's local precipitation history. It encourages the use of neighboring observations and radar-derived multimodal context, thereby improving transfer to fixed target stations whose ground-level precipitation measurements are unavailable during training.

3 EXPERIMENTS

3.1 DATASET AND PREPROCESSING

We evaluate MIRAGE on three meteorological datasets: the Korea Meteorological Administration (KMA) dataset and the MeteoNet datasets from the Northwest (NW) and Southeast (SE) regions. Both datasets contain radar reflectivity and ground-level precipitation observations. The KMA dataset is recorded at 10-minute intervals, whereas MeteoNet is recorded at 30-minute intervals. For KMA, we use data from June to September, which covers the primary rainfall season in Korea. For MeteoNet, the full annual records are used. Additional dataset statistics and preprocessing details are provided in the Appendix.

3.2 BASELINES

We compare MIRAGE with representative baselines spanning classical spatial interpolation, neural spatial reconstruction, and spatio-temporal forecasting. This selection covers methods based on spatial covariance, implicit spatial representations, graph-based message passing, and temporal sequence modeling.

- **Classical Spatial Interpolation:** Kriging Wackernagel (1995) is included as a non-learning baseline that estimates precipitation from spatial covariance among observed stations.
- **Neural Spatial Interpolation:** SUICA Zhu et al. (2025) and SINR Cole et al. (2023) represent learning-based spatial reconstruction methods designed to infer continuous fields or values at spatially unobserved locations.
- **Spatio-Temporal Forecasting:** DualCast Su et al. (2025), ReDYNet Gao et al. (2025), MIGN Zheng et al. (2025), and STGFormer Ju et al. (2026) represent neural forecasting approaches that model spatial interactions together with temporal dynamics. In particular, STGFormer combines adaptive graph-based interaction modeling with Transformer-based multi-scale temporal modeling, providing a structured graph forecasting baseline.

Together, these baselines provide comparisons against both spatial reconstruction methods and models that explicitly capture spatio-temporal dependencies. For a fair comparison, all methods use the same station partitions, temporal splits, and evaluation protocol.

3.3 EXPERIMENTAL SETUP

For KMA, the data from 2019-2021, 2022, and 2023 are used for training, validation, and testing, respectively. For MeteoNet, the corresponding partitions are 2016, 2017, and 2018. To evaluate spatial generalization, we randomly select 30% of the stations before training as fixed unobserved target stations. Their ground-level precipitation observations are excluded throughout training, validation, and testing. Their precipitation inputs are set to 0 with a binary missingness indicator set to 1, while their geographic coordinates and corresponding radar observations remain available. The remaining 70% of stations form the observed set. For graph-based methods, each station is connected to its 10 nearest neighbors. The station partition is fixed within each run and shared across all methods. Each experiment is repeated over four independent station splits using seed $\in \{12, 22, 32, 42\}$, and the mean performance is reported. Standard deviations and sensitivity to target-station selection are provided in Appendix H. All experiments are conducted on an Intel Xeon® E5-2620 v3 CPU and a single NVIDIA GeForce RTX 5070 Ti GPU with 16GB VRAM.

3.4 MAIN RESULTS

We evaluate MIRAGE on two complementary tasks: (i) spatial interpolation at stations without local precipitation observations and (ii) multi-step precipitation forecasting. RMSE and MAE are used as the primary evaluation metrics.

Table 1: Spatial interpolation performance at unobserved target stations. Lower RMSE and MAE indicate better performance. Best results are shown in **bold** and second-best results are underlined.

Method	KMA		MeteoNet-NW		MeteoNet-SE	
	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓
Kriging	1.461	0.315	0.248	0.061	0.351	0.081
SINR (2023)	2.094	0.469	0.309	0.069	0.467	0.092
SUICA (2025)	1.801	0.349	0.302	0.059	0.458	0.075
ReDYNet (2025)	1.584	0.322	0.254	0.046	0.394	0.086
DualCast (2025)	1.602	0.347	0.284	0.065	0.409	0.090
MIGN (2025)	1.510	0.267	0.278	0.045	0.397	0.069
STGFormer (2026)	1.713	0.348	0.249	0.057	0.392	0.091
MIRAGE (Ours)	**1.442**	**0.240**	**0.231**	**0.041**	**0.346**	**0.063**

3.4.1 Spatial Interpolation at Unobserved Stations

For spatial interpolation, MIRAGE takes six consecutive historical multimodal observations $\mathcal{M}^{t-5:t}$ and estimates precipitation at the unobserved target stations V_u at time t. Precipitation observations from V_u are unavailable to the model, while geographic information and radar observations covering these locations remain accessible.

As shown in Table 1, MIRAGE obtains the lowest RMSE and MAE on all three datasets. The improvement is observed across KMA, MeteoNet-NW, and MeteoNet-SE, despite their different station configurations and precipitation characteristics. In particular, the results indicate that combining station observations with dense radar information is useful when local precipitation histories are unavailable. These results support the intended role of MIRAGE in spatial interpolation: radar provides complementary information about the precipitation field, while graph-based propagation transfers information from observed stations to unobserved locations.

Table 2: Performance comparison on MeteoNet datasets (NW and SE) for Unobserved Nodes and All Nodes across 30-minute and 60-minute forecasting horizons. Lower RMSE and MAE indicate better performance. Best results are shown in **bold** and second-best results are underlined.

Node Type	Model	NW				SE			
		30 min		60 min		30 min		60 min	
		RMSE (↓)	MAE (↓)	RMSE (↓)	MAE (↓)	RMSE (↓)	MAE (↓)	RMSE (↓)	MAE (↓)
All Nodes	SINR (2023)	0.282	0.053	0.290	0.057	0.425	0.073	0.437	0.077
	SUICA (2025)	0.282	0.052	0.295	0.056	0.418	0.063	0.439	0.071
	ReDYNet (2025)	**0.269**	0.043	0.299	0.048	0.428	0.068	0.429	0.075
	DualCast (2025)	0.311	0.074	0.309	0.072	0.470	0.094	0.471	0.088
	MIGN (2025)	0.274	0.045	**0.289**	0.051	0.401	0.061	0.421	0.073
	STGFormer (2026)	0.271	0.048	0.287	0.053	0.428	0.069	0.432	0.077
	MIRAGE (Ours)	0.273	**0.040**	0.284	**0.046**	**0.391**	**0.055**	**0.418**	**0.064**
Unobserved Nodes	SINR (2023)	0.309	0.069	0.309	0.069	0.467	0.092	0.467	0.092
	SUICA (2025)	0.303	0.062	0.303	0.060	0.457	0.078	0.457	0.079
	ReDYNet (2025)	0.331	0.055	0.313	0.056	0.465	0.074	0.464	0.075
	DualCast (2025)	0.316	0.081	0.317	0.081	0.466	0.125	0.469	0.104
	MIGN (2025)	0.297	0.049	0.299	0.051	0.421	0.078	0.438	0.077
	STGFormer (2026)	0.298	0.058	0.298	0.061	0.454	0.076	0.453	0.077
	MIRAGE (Ours)	**0.295**	**0.044**	**0.296**	**0.048**	**0.415**	**0.060**	**0.426**	**0.067**

Table 3: Performance comparison on All Nodes and Unobserved Nodes across 10–60 minute forecasting horizons. Lower RMSE and MAE indicate better performance. Best results are shown in **bold** and second-best results are underlined.

Node Type	Model	10 min		20 min		30 min		40 min		50 min		60 min	
		RMSE (↓)	MAE (↓)	RMSE (↓)	MAE (↓)	RMSE (↓)	MAE (↓)	RMSE (↓)	MAE (↓)	RMSE (↓)	MAE (↓)	RMSE (↓)	MAE (↓)
All Nodes	SINR (2023)	1.789	0.351	1.799	0.353	1.815	0.357	1.833	0.361	1.850	0.366	1.867	0.371
	SUICA (2025)	1.143	0.173	1.220	0.202	1.321	0.232	1.435	0.260	1.560	0.289	1.678	0.317
	ReDYNet (2025)	1.642	0.658	1.648	0.682	1.698	0.711	1.691	0.689	1.642	0.678	1.752	0.722
	DualCast (2025)	1.663	0.419	1.663	0.399	1.671	0.401	1.672	0.404	1.673	0.397	1.773	0.464
	MIGN (2025)	1.231	0.247	1.286	0.255	1.328	0.265	1.438	0.277	1.596	0.292	1.641	0.306
	STGFormer (2026)	1.258	0.199	1.298	0.212	1.395	0.245	1.468	0.278	1.608	0.310	1.735	0.342
	MIRAGE (Ours)	**0.955**	**0.139**	**1.065**	**0.171**	**1.194**	**0.202**	**1.341**	**0.234**	**1.497**	**0.269**	**1.634**	**0.300**
Unobserved Nodes	SINR (2023)	2.093	0.469	2.092	0.469	2.090	0.468	2.089	0.468	2.088	0.467	2.087	0.467
	SUICA (2025)	1.797	0.347	1.797	0.346	1.799	0.347	1.804	0.348	1.810	0.350	1.819	0.353
	ReDYNet (2025)	2.093	0.528	2.094	0.529	2.097	0.522	2.099	0.512	2.102	0.562	2.051	0.570
	DualCast (2025)	1.992	0.424	1.991	0.404	1.998	0.406	1.998	0.409	1.999	0.401	1.998	0.468
	MIGN (2025)	1.647	0.273	1.663	0.278	1.689	0.287	1.730	0.296	1.774	0.308	1.809	0.321
	STGFormer (2026)	1.976	0.409	1.974	0.410	1.974	0.410	1.975	0.411	1.975	0.410	1.977	0.411
	MIRAGE (Ours)	**1.573**	**0.254**	**1.589**	**0.260**	**1.620**	**0.270**	**1.662**	**0.284**	**1.718**	**0.302**	**1.763**	**0.318**

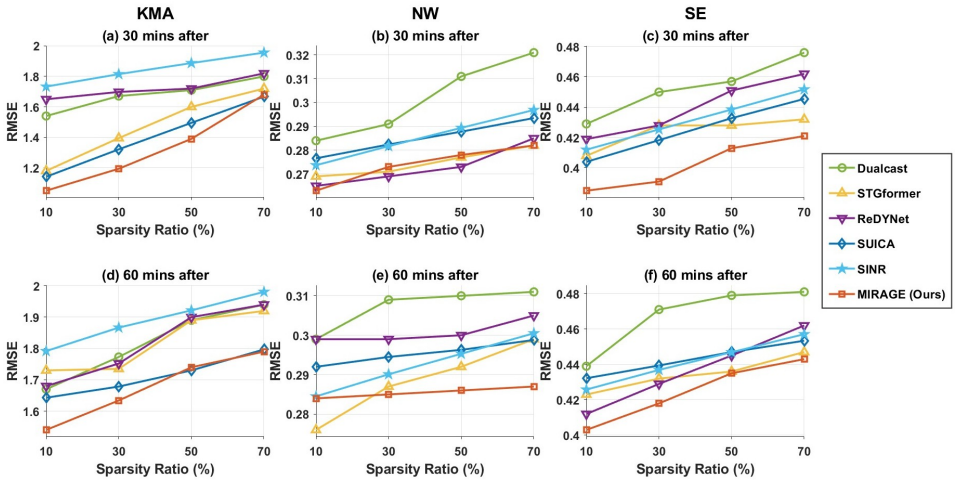


Figure 3: **Robustness to increasing spatial sparsity.** RMSE under 10%–70% unobserved stations at the 30-minute (top) and 60-minute (bottom) forecasting horizons on KMA, MeteoNet-NW, and MeteoNet-SE. Lower RMSE indicates better performance.

3.4.2 Multi-Step Precipitation Forecasting

For forecasting, MIRAGE uses historical multimodal observations to predict precipitation at future time steps. We report results on both the complete station network V_{all} and the fixed unobserved target set V_u. The former evaluates overall forecasting accuracy, whereas the latter measures forecasting performance at locations without local precipitation histories. Tables 2 and 3 show that MIRAGE performs particularly well at unobserved stations. On KMA, MIRAGE obtains the lowest RMSE and MAE at all forecasting horizons from 10 to 60 minutes. The error increases with lead time, as expected, but the advantage over the compared methods is maintained throughout the evaluated horizon.

A similar pattern is observed on MeteoNet. At unobserved nodes, MIRAGE achieves the lowest RMSE and MAE for both 30-minute and 60-minute forecasts in NW and SE. On the complete network, the results are more mixed: several baselines achieve comparable or slightly lower RMSE in NW, whereas MIRAGE obtains the lowest MAE and performs strongly on SE.

The difference between All Nodes and Unobserved Nodes is relevant to the intended setting of MIRAGE. At observed locations, models can use local precipitation histories directly. At unobserved locations, prediction depends more strongly on spatial information transferred from neighboring stations and the radar field. The results show that MIRAGE is most consistently effective in this latter setting.

3.4.3 Robustness to Spatial Sparsity

We further examine robustness by increasing the ratio of unobserved stations from 10% to 70%. Figure 3 reports RMSE at the 30-minute and 60-minute forecasting horizons.

As the sparsity ratio increases, forecasting error generally increases across the evaluated methods. MIRAGE achieves low RMSE over the tested sparsity levels, with particularly clear performance on KMA and MeteoNet-SE. On MeteoNet-NW, the differences among several methods are smaller, and MIRAGE remains competitive rather than uniformly achieving the lowest error. Overall, MIRAGE maintains stable forecasting accuracy as fewer ground-level observations become available.

3.5 Ablation Study

We evaluate the contribution of the key components of MIRAGE, including radar conditioning, the Meteorological Hypernetwork (HyperNet), Dynamic Mask (DynMask), and cross-attention. As

Table 4: Ablation study of the three key components of MIRAGE: radar conditioning, the Meteorological Hypernetwork (HyperNet), and the dynamic topology mask (DynMask). Average RMSE (↓) is reported for All Nodes and Unobserved Nodes across different forecasting horizons on the KMA and MeteoNet datasets.

Dataset	Node Type	Model Variant	Radar	HyperNet	DynMask	10 min	20 min	30 min	40 min	50 min	60 min
KMA	All Nodes	Precip-only	X	O	O	1.201	1.271	1.379	1.491	1.600	1.721
		w/o HyperNet	O	X	O	1.854	1.877	1.900	1.919	1.936	1.949
		w/o DynMask	O	O	X	1.304	1.383	1.419	1.577	1.692	1.829
		Concat+MLP (w/o Cross-Attn)	O	O	O	1.214	1.231	1.342	1.426	1.577	1.712
		MIRAGE	O	O	O	**0.955**	**1.065**	**1.194**	**1.341**	**1.497**	**1.634**
	Unobserved Nodes	Precip-only	X	O	O	1.712	1.751	1.786	1.795	1.805	1.819
		w/o HyperNet	O	X	O	2.034	2.040	2.047	2.054	2.059	2.064
		w/o DynMask	O	O	X	1.651	1.675	1.742	1.810	1.953	2.029
		Concat+MLP (w/o Cross-Attn)	O	O	O	1.612	1.679	1.756	1.791	1.801	1.815
		MIRAGE	O	O	O	**1.573**	**1.589**	**1.620**	**1.662**	**1.718**	**1.763**

						MeteoNet (NW Region)		MeteoNet (SE Region)	
Dataset	Node Type	Model Variant	Radar	HyperNet	DynMask	30 min	60 min	30 min	60 min
MeteoNet	All Nodes	Precip-only	X	O	O	0.275	0.299	0.421	0.441
		w/o HyperNet	O	X	O	0.349	0.362	0.452	0.461
		w/o DynMask	O	O	X	0.322	0.332	0.420	0.456
		Concat+MLP (w/o Cross-Attn)	O	O	O	0.279	0.295	0.412	0.424
		MIRAGE	O	O	O	**0.273**	**0.284**	**0.391**	**0.418**
	Unobserved Nodes	Precip-only	X	O	O	0.299	0.305	0.469	0.451
		w/o HyperNet	O	X	O	0.375	0.391	0.511	0.520
		w/o DynMask	O	O	X	0.361	0.387	0.474	0.450
		Concat+MLP (w/o Cross-Attn)	O	O	O	0.299	0.304	0.419	**0.425**
		MIRAGE	O	O	O	**0.295**	**0.296**	**0.415**	0.426

shown in Table 4, the full model generally provides the most consistent performance across datasets, node types, and forecasting horizons.

Removing radar information generally increases the prediction error, with a more noticeable degradation at unobserved nodes. This indicates that dense radar observations provide complementary spatial information when local precipitation histories are unavailable. Removing either HyperNet or DynMask also leads to clear performance degradation across the evaluated settings. These results support their complementary roles: DynMask adapts *where* information is propagated, while HyperNet controls *how* the propagated information is transformed according to the current meteorological context.

Replacing cross-attention with direct feature concatenation generally results in higher errors than the full model. This suggests that the benefit of radar is not solely due to adding an additional modality, but also depends on how radar and station representations are integrated. Overall, the ablation results support the joint use of radar conditioning, cross-modal attention, and adaptive graph computation in MIRAGE.

4 Conclusion

We introduced **MIRAGE**, a radar-conditioned dynamic graph framework for precipitation interpolation and forecasting over sparse weather station networks. MIRAGE combines dense radar observations with sparse station measurements through cross-attention and uses the resulting multimodal representation to adapt graph computation. Dynamic Graph Construction determines *where* information is propagated, while the Meteorological Hypernetwork determines *how* it is transformed.

Experiments on KMA and MeteoNet show that MIRAGE performs consistently well at unobserved stations, where local precipitation histories are unavailable. It achieves the lowest interpolation RMSE and MAE across the three evaluated regions and provides strong multi-step forecasting performance, particularly at unobserved locations. Under increasing spatial sparsity, MIRAGE maintains competitive performance across the evaluated datasets, with particularly low forecasting errors on KMA and MeteoNet-SE. Ablation results further show that radar conditioning, dynamic connectivity, context-dependent graph transformation, and cross-attention make complementary contributions to the overall performance.

AI USE STATEMENT

In this work, generative AI tools were used solely for language editing, grammar checking, and proofreading to improve the readability and clarity of the manuscript. We have not used generative AI tools for research conception, model architecture design, mathematical derivations, or the analysis of experimental results. All text and contents were thoroughly reviewed and verified by the authors. We take full responsibility for the final content of this work.

ETHICS STATEMENT

This study utilizes publicly available meteorological datasets strictly for scientific research. It does not involve human subjects, private personal information, or commercial tracking data. The primary objective of our framework is to enhance localized precipitation forecasting and extreme weather preparedness, presenting no foreseeable negative societal or environmental risks.

REPRODUCIBILITY STATEMENT

To ensure the reproducibility of our results, detailed descriptions of the model architecture, mathematical formulations, and training hyperparameters are provided in Appendix. The data preprocessing pipelines, spatio-temporal alignment protocols, and evaluation procedures are elaborated in Appendix. The complete source code, training scripts, and pre-trained checkpoints will be made publicly available on GitHub.

REFERENCES

G. Appleby, L. Liu, and L.P. Liu. Kriging convolutional networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020.

Lei Bai, Lina Yao, Can Li, Xianzhi Wang, and Can Wang. Adaptive graph convolutional recurrent network for traffic forecasting. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pp. 17804–17815, 2020.

Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xiao Chen, Xiaotao Gu, and Qi Tian. Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, 619(7970):533–538, 2023.

Filippo Maria Bianchi, Simone Scardapane, Sigurd Løkse, and Robert Jenssen. Recurrent neural networks for multivariate time series with missing values: An application to anomaly detection. *IEEE Transactions on Neural Networks and Learning Systems*, 2019.

Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. Recurrent neural networks for multivariate time series with missing values. *Scientific Reports*, 8(1):6085, 2018.

Chen Chen et al. Multi-scale spatiotemporal graph neural network for multivariate time series forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 2023a.

L. Chen, J. Xu, B. Wu, and J. Huang. Group-aware graph neural network for nationwide city air quality forecasting. *ACM Transactions on Knowledge Discovery from Data*, 18(3):1–20, 2023b.

A. Cini, I. Marisca, and C. Alippi. Filling the g_ap_s: Multivariate time series imputation by graph neural networks. In *International Conference on Learning Representations*, January 2022.

A. Cini, I. Marisca, F. M. Bianchi, and C. Alippi. Scalable spatiotemporal graph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 7218–7226, 2023.

Elijah Cole, Grant Van Horn, Christian Lange, Alexander Shepard, Patrick Leary, Pietro Perona, Scott Loarie, and Oisin Mac Aodha. Spatial Implicit Neural Representations for Global-Scale Species Mapping. In *ICML*, 2023.

Zhicheng Cui, Wenlin Chen, and Yixin Chen. Multi-scale convolutional neural networks for time series classification. *arXiv preprint arXiv:1603.06995*, 2016.

P. S. S. Ejurothu, S. Mandal, and M. Thakur. Forecasting pm2. 5 concentration in india using a cluster based hybrid graph neural network approach. *Asia-Pacific Journal of Atmospheric Sciences*, 59(5):545–561, 2023.

Qiang Gao, Zekai Wang, Lisi Huang, Goce Trajcevski, Guanfeng Liu, and Xuan Chen. Responsive dynamic graph disentanglement for metro flow forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 11690–11698, 2025. doi: 10.1609/aaai.v39i11.33272. URL https://doi.org/10.1609/aaai.v39i11.33272.

Zhihan Gao, Xingjian Shi, Hao Wang, Yi Zhu, Yuyang Wang, Mu Li, and Dit-Yan Yeung. Earthformer: Exploring space-time transformers for earth system forecasting. In *NeurIPS*, 2022.

Zhihan Gao, Xingjian Shi, Boran Han, Hao Wang, Xiaoyong Jin, Danielle C Maddix, Yi Zhu, Mu Li, and Bernie Wang. Prediff: Precipitation nowcasting with latent diffusion models. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

Junchao Gong, Lei Bai, Peng Ye, Wanghan Xu, Na Liu, Jianhua Dai, Xiaokang Yang, and Wanli Ouyang. Cascast: Skillful high-resolution precipitation nowcasting via cascaded modelling. *arXiv preprint arXiv:2402.04290*, 2024.

David Ha, Andrew Dai, and Quoc V Le. Hypernetworks. *arXiv preprint arXiv:1609.09106*, 2016.

K. H. Hettige, J. Ji, S. Xiang, C. Long, G. Cong, and J. Wang. Airphynet: Harnessing physics-guided neural networks for air quality prediction. *arXiv preprint arXiv:2402.03784*, 2024.

Cheng Ju, Yuansha Xie, Zhongrong Wang, Yu Zhao, Wenyao Yan, Rongjun Chai, Juan Duan, Yali Cao, and Yuxin Chang. STGFormer: A pyramidal spatio-temporal graph transformer with cross-disciplinary feature fusion for semantic-rich trajectory prediction in heterogeneous autonomy traffic. *Expert Systems with Applications*, 310:131304, 2026. ISSN 0957-4174.

Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Ferran Alet, Sumanth Ravuri, Timo Ewalds, Zachary Eaton-Rosen, Weihua Hu, et al. Graphcast: Learning skillful medium-range global weather forecasting. *arXiv preprint arXiv:2212.12794*, 2022.

J. Li, Y. Shen, L. Chen, and C.W.W. Ng. Rainfall spatial interpolation with graph neural networks. In X. Wang et al. (eds.), *Database Systems for Advanced Applications (DASFAA 2023)*, volume 13946 of *Lecture Notes in Computer Science*, Cham, 2023. Springer. doi: 10.1007/978-3-031-30678-5_14.

Kenghong Lin, Baoquan Zhang, Demin Yu, Wenzhi Feng, Shidong Chen, Feifan Golds Gao, Xutao Li, and Yunming Ye. Alphapre: Amplitude-phase disentangle model for precipitation nowcasting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17841–17850, 2025.

Ivan Marisca, Andrea Cini, and Cesare Alippi. Learning to reconstruct missing data from spatiotemporal graphs with sparse observations. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pp. 32069–32082, 2022.

Jaideep Pathak, Shashank Subramanian, Peter Harrington, Surya Raja, Ashesh Chattopadhyay, Morteza Mardani, Thorsten Kurth, David Hall, Zongyi Li, Kamyar Azizzadenesheli, et al. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*, 2022.

Syama Sundar Rangapuram, Lucien D Werner, Konstantinos Benidis, Pedro Mercado, Jan Gasthaus, and Tim Januschowski. End-to-end learning of coherent probabilistic forecasts for hierarchical time series. In *International Conference on Machine Learning (ICML)*, pp. 8832–8843. PMLR, 2021.

Syama Sundar Rangapuram, Shubham Kapoor, Rashika S Nirwan, Pedro Mercado, Tim Januschowski, Yining Wang, and Michael Bohlke-Schneider. Coherent probabilistic forecasting of temporal hierarchies. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 9362–9376. PMLR, 2023.

Youngjoo Seo, Michaël Defferrard, Pierre Vandergheynst, and Xavier Bresson. Structured sequence modeling with graph convolutional recurrent networks. In *Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13–16, 2018, Proceedings, Part I 25*, pp. 362–373. Springer, 2018.

Abdul-Saboor Sheikh, Kashif Rasul, Andreas Merentitis, and Urs Bergmann. Stochastic maximum likelihood optimization via hypernetworks. *arXiv preprint arXiv:1712.01141*, 2017.

S. N. Shukla and B. Marlin. Multi-time attention networks for irregularly sampled time series. In *International Conference on Learning Representations*, 2020.

D. I. Shuman, M. J. Faraji, and P. Vandergheynst. A multi-scale pyramid transform for graph signals. *IEEE Transactions on Signal Processing*, 64(8):2119–2134, 2015.

Gilbert Strang and Truong Nguyen. *Wavelets and filter banks*. Wellesley-Cambridge Press, 1996.

X. Su, F. Liu, Y. Chang, E. Tanin, M. Sarvi, and J. Qi. Dualcast: A model to disentangle aperiodic events from traffic series. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2025.

Xianfeng Tang, Huaxiu Yao, Yiwei Sun, Yiqi Wang, Jiliang Tang, Charu Aggarwal, Prasenjit Mitra, and Suhang Wang. Joint modeling of missingness and measurement process for time series forecasting. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2133–2141, 2020.

N. Tremblay and P. Borgnat. Subgraph-based filterbanks for graph signals. *IEEE Transactions on Signal Processing*, 64(15):3827–3840, 2016.

Hans Wackernagel. Ordinary kriging. In *Multivariate Geostatistics*. Springer, Berlin, Heidelberg, 1995. doi: 10.1007/978-3-662-03098-1_11.

A. Wang, Y. Ye, X. Song, S. Zhang, and J. James. Traffic prediction with missing data: A multi-task learning approach. *IEEE Transactions on Intelligent Transportation Systems*, 24(4):4189–4202, 2023.

Y. Wang and C. Jing. Spatiotemporal graph convolutional network for multi-scale traffic forecasting. *ISPRS International Journal of Geo-Information*, 11(2):102, 2022.

H. Wu, H. Zhou, M. Long, and J. Wang. Interpretable weather forecasting for worldwide stations with a unified deep model. *Nature Machine Intelligence*, 5(6):602–611, 2023.

Yuankai Wu, Dingyi Zhuang, Aurelie Labbe, and Lijun Sun. Inductive graph neural networks for spatiotemporal kriging. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 4478–4485, 2021.

Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. Graph wavenet for deep spatial-temporal graph modeling. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1907–1913, 2019.

B. Yu, H. Yin, and Z. Zhu. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pp. 3634–3640, 2018.

Jiani Zhang, Xingjian Shi, Junyuan Xie, Hao Ma, Irwin King, and Dit-Yan Yeung. Gaan: Gated attention networks for learning on large and spatiotemporal graphs. In *Proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence*, pp. 339–349, 2018.

Z. Zheng, Y. Liu, and J. Li. Mesh interpolation graph network for dynamic and spatially irregular global weather forecasting. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

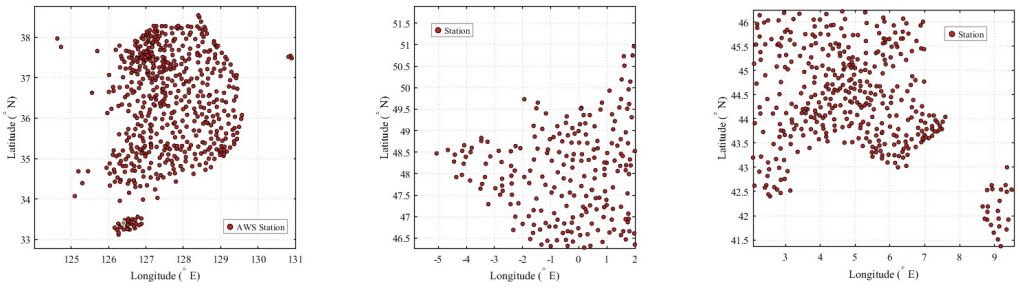
Huilong Zhong, Ran Mu, Gang Xu, Yong Zhang, and Dong Yuan. Traffic speed forecasting with missing values of multiple detectors using spatial-temporal graph neural networks. *IEEE Transactions on Intelligent Transportation Systems*, 23(8):13091–13102, 2021.

Qingtian Zhu, Yumin Zheng, Yuling Sang, Yifan Zhan, Ziyan Zhu, Jun Ding, and Yinqiang Zheng. SUICA: Learning super-high dimensional sparse implicit neural representations for spatial transcriptomics. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 80448–80462. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/zhu25af.html.

A Dataset

A.1 KMA

The Korea Meteorological Administration (KMA) dataset provides nationwide radar coverage for South Korea. It consists of various types of radar reflectivity, such as Plan-Position Indicator (PPI), Constant Altitude PPI (CAPPI), and Hybrid Surface Rainfall (HSR). We used HSR observations obtained every 10 minutes over the central region of South Korea, ranging from longitude 120.5°E to 137.5°E and latitude 29.0°N to 42.0°N. We normalize each image by clipping negative values to 0 and dividing by 8,000, which results in data values in the range $[0, 1)$ (0.01 dBZ). As shown in the figure, we cropped the data into a 1440×1240 image to cover the entire station range, and then downsampled it by a factor of 5 to yield a 288×248 input image. The KMA radar dataset



(a) KMA (b) NW region (c) SE region

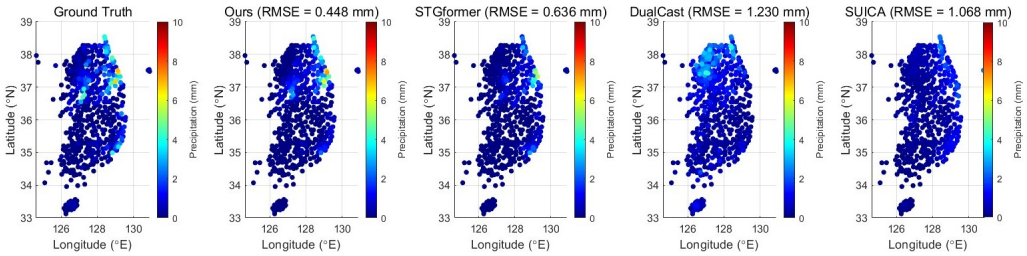
Figure 4: Spatial distributions of weather stations across three different domains: (a) KMA, (b) NW region, and (c) SE region. The red markers indicate the observation stations, while the coordinate axes represent the respective radar coverage domain limits.

Table 5: Summary of meteorological datasets

Dataset	#Nodes	#Time steps	Interval	Duration	Output Horizon
KMA	583	87,840	10 min	2019–2023 (Jun-Sep)	6 steps (60 min)
MeteoNet (NW)	213	52,608	30 min	2016–2018 (Jan-Dec)	2 steps (60 min)
MeteoNet (SE)	404	52,608	30 min	2016–2018 (Jan-Dec)	2 steps (60 min)

spans from 2014 to 2023, with data collected every 10 minutes. For the Automatic Weather Station (AWS) data, observations were collected at the same 10-minute intervals over the identical period. Out of approximately 700 stations, we selected 583 stations with less than 5% missing values for our experiments. Missing (NaN) values were replaced with 0, and the precipitation range was scaled between 0 and 1 using log-normal normalization.

Due to the climatological characteristics of South Korea, where the vast majority of precipitation is concentrated during the summer season, we restricted the KMA dataset to the summer months(6-9). For our experiments, we used a time-split approach: data from 2019 to 2021 was used for training, 2022 for validation, and 2023 for testing. The model takes a sequence of 6 frames (one hour of past observations) as input. It is configured to predict the future 1-hour short-term horizon (6 frames), with a temporal resolution of 10 minutes per frame.



A.2 METEONET

We also use MeteoNet, an open meteorological dataset released by Météo-France, the French national meteorological service. The dataset provides a clean, comprehensive collection of meteorological data; here, we integrate both the radar observation and ground measurement components spanning over three years from January 2016 to December 2018. Our study encompasses two distinct geographical areas: the north-western and south-eastern quarters of France.

Originally, the radar and ground station data possess different temporal resolutions of 5 minutes and 6 minutes, respectively. To achieve temporal alignment between the two modalities, we standardize the temporal frame to a 30-minute interval. Specifically, the radar images are sub-sampled every 30 minutes, while the ground precipitation data are aggregated as the accumulated rainfall sum over each 30-minute window.

For spatial preprocessing, the radar frames, which originally have a spatial resolution of $0.01°$, are resized to a uniform dimension of 256×256 pixels for both regions. Regarding the ground stations, we filter out unreliable stations to ensure data quality, selecting only those with a missing value rate of less than 10%. Consequently, 213 out of 287 stations in the north-western region and 404 out of 575 stations in the south-eastern region are selected and utilized for model training.

Missing values in both the radar and station datasets are imputed with 0. To improve training stability, both modalities are scaled into the range of $[0, 1]$; we apply Min-Max scaling to the radar images, whereas a log transformation is applied to the ground station measurements. The forecasting setup is configured to take an input sequence of 6 frames (corresponding to a 3-hour lookback window) to predict an output sequence of the subsequent 2 frames (corresponding to a 1-hour lead time).

B RELATED WORK

B.1 SPATIOTEMPORAL GRAPH NEURAL NETWORKS AND MISSING DATA

Spatiotemporal Graph Neural Networks (STGNNs) have emerged as a dominant paradigm for modeling multivariate time series collected from distributed sensor networks. By representing synchronous temporal signals over inter-connected topology, early STGNNs integrated spatial message passing (SMP) into recurrent architectures Seo et al. (2018); Zhang et al. (2018); Bai et al. (2020) or applied space-time convolutions and attention mechanisms Yu et al. (2018); Wu et al. (2019). Despite their success, these conventional STGNNs implicitly assume complete observation sequences and lack explicit mechanisms to handle missing or irregular data inputs.

To alleviate missingness in relational time series, recent architectures such as GRIN Cini et al. (2022) and related graph frameworks Wu et al. (2021); Marisca et al. (2022) have incorporated spatial dependencies; however, their primary objective remains restricted to missing value imputation rather than downstream forecasting. While a few domain-specific studies in traffic analytics have tackled predictive tasks under missingness Zhong et al. (2021); Wang et al. (2023), and general sequence models like GRU-D Che et al. (2018) handle incomplete observations across tasks Shukla & Marlin (2020); Bianchi et al. (2019); Tang et al. (2020), they critically omit spatial correlation modeling among sensors.

Separately, hierarchical and pyramidal representations have long been leveraged in signal processing Shuman et al. (2015); Tremblay & Borgnat (2016); Strang & Nguyen (1996) and deep learning Cui et al. (2016); Wang & Jing (2022); Chen et al. (2023a); Rangapuram et al. (2021; 2023) to

process multi-scale structures under missing or noisy signals. In particular, learnable graph pooling techniques Cini et al. (2023) cluster time series into coarse-to-fine abstractions to capture multi-resolution temporal dynamics. Nevertheless, these existing hierarchical STGNNs construct representations based solely on static relational structures or data-driven feature similarities, failing to adapt to rapidly evolving atmospheric states. In contrast, our proposed **MIRAGE** conditions graph propagation directly on multimodal atmospheric contexts, enabling dynamic topology masking and adaptive message passing to effectively handle non-stationary meteorological phenomena.

B.2 HYPERNETWORKS

Hypernetworks, as pioneered by Ha et al. (2016), function as auxiliary models designed to generate the mathematical weights for a separate, primary target network. This framework enables a significantly smaller parameter footprint while allowing for the synthesis of a diverse ensemble of target architectures to approximate highly complex, non-linear functions Sheikh et al. (2017).

C STATISTICAL OVERVIEW OF DATASETS

To quantify the temporal progression and distributional traits of rainfall events, we conducted a comprehensive statistical analysis for each dataset. Following the operational standards set by the Korea Meteorological Administration (KMA) and the World Meteorological Organization (WMO):

- For the **KMA dataset**, *normal rain* is defined within the range of 1-10 mm/h, while *heavy rain* corresponds to intensities of ≥ 10 mm/h.
- For the **MeteoNet dataset**, *normal rain* is bounded between 0.5-2 mm/h, and *heavy rain* is defined as intensities of ≥ 2 mm/h.

This categorical evaluation ensures that the model's capability in capturing highly volatile, high-impact extreme weather events is rigorously and realistically verified.

Based on these operational thresholds, we formally defined the duration of a continuous rainfall event. Specifically, an event is initiated when the maximum precipitation intensity across all stations first reaches or exceeds the threshold for *normal rain* (1.0 mm/h for KMA, 0.5 mm/h for MeteoNet), and concludes when it falls back below this value.

As illustrated in Figure 5, the mode (most frequent duration) for the KMA dataset is 60 minutes, whereas both MeteoNet (NW) and MeteoNet (SE) exhibit a shorter mode of 30 minutes. Furthermore, the KMA dataset displays a high frequency of long-duration, persistent rainfall events. In contrast, MeteoNet (NW) is characterized by notably lower precipitation amounts and a sparser frequency of rainfall events compared to the other datasets. While MeteoNet (SE) exhibits higher precipitation intensities and cumulative amounts than MeteoNet (NW), its events remain relatively transient compared to the prolonged rainfall patterns observed in the KMA dataset.

D BASELINES IMPLEMENTATION

- ReDyNet (2025) is implemented based on the official code `https://github.com/wangzz-yyzz/ReDyNet`.
- DualCast (2025) is implemented based on the official code `https://github.com/suzy0223/DualCast`.
- SUICA(2025) is implemented based on the official code `https://github.com/Szym29/SUICA`.
- SINR(2023) is implemented based on the official code `https://github.com/elijahcole/sinr`.
- STGFormer (2026) is adapted from the architecture described in Ju et al. (2026). Since STGFormer was originally developed for multi-agent trajectory prediction, we adapt it to precipitation forecasting by treating weather stations as graph nodes and replacing the trajectory prediction output with precipitation prediction. We retain its main graph-based spatial interaction and Transformer-based temporal modeling components.

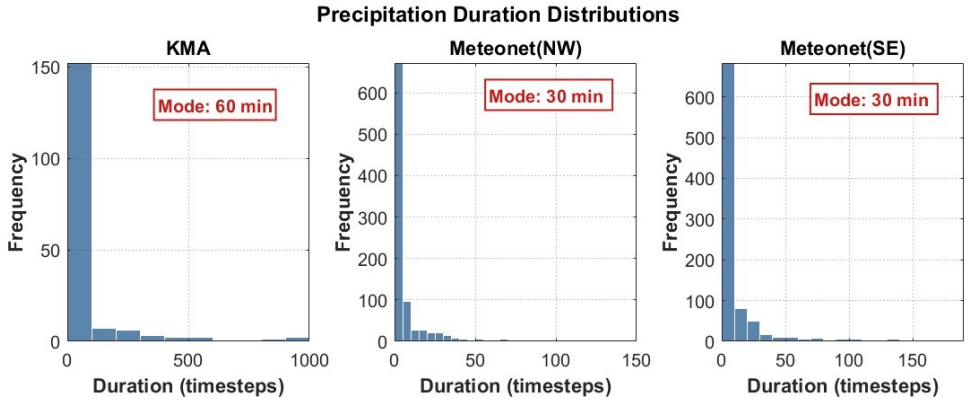


Figure 5: Comparison of continuous rainfall event durations across the KMA, MeteoNet (NW), and MeteoNet (SE) datasets

- MIGN (2025) is implemented based on the official code `https://github.com/compasszzn/MIGN`.

For each learning-based baseline, we tune the main training hyperparameters using Bayesian hyperparameter optimization with Weights & Biases Sweeps. The selected hyperparameters are summarized in Table 6.

Table 6: Selected training hyperparameters for the baseline models and MIRAGE.

Model	ReDyNet	DualCast	SUICA	MIGN	STGFormer	SINR	MIRAGE
Learning rate	0.0014	0.0015	1×10^{-4}	1×10^{-4}	0.001	1×10^{-3}	1×10^{-5}
Batch size	4	4	64	32	16	16	6
Hidden size	256	128	256	256	128	256	256

E Model Configuration and Architectural Details

This section provides the architectural and implementation details of MIRAGE. While the main paper focuses on the overall roles of the two adaptive mechanisms—*where* information propagates and *how* it is transformed—we describe here the radar and station encoders, cross-modal fusion, Dynamic Graph Construction, context-conditioned Hypernetwork, and target-aware reconstruction.

E.1 Radar Encoder

MIRAGE uses a pretrained radar encoder to extract a compact spatiotemporal representation from six consecutive radar frames. The radar encoder is frozen during MIRAGE training, so the subsequent graph model operates on a fixed radar representation.

Each radar frame is first processed independently by two strided convolutional layers. The first layer maps the input to 32 channels using a 5×5 kernel with stride 2, and the second maps the representation to 64 channels using a 3×3 kernel with stride 2. The resulting sequence of spatial feature maps is processed by a ConvLSTM with hidden dimension 64 and a 3×3 kernel. The final hidden state therefore summarizes both the spatial organization and short-term temporal evolution of the radar sequence.

Rather than globally pooling this representation, MIRAGE preserves its spatial structure. Let the final ConvLSTM state be $\mathbf{R}_t^{\text{map}} \in \mathbb{R}^{64 \times H_r \times W_r}$. Its spatial dimensions are flattened to obtain

$$\mathbf{R}_t = [\mathbf{r}_{1,t}, \ldots, \mathbf{r}_{N_r,t}]^\top \in \mathbb{R}^{N_r \times 64}, \qquad N_r = H_r W_r. \tag{6}$$

A learned two-dimensional positional encoding is added to each token to preserve its location within the radar field.

Table 7: Model configuration used in MIRAGE.

Component	Setting
Radar input	6 consecutive frames
Radar spatial encoder	Conv 5×5/2, 32 ch → Conv 3×3/2, 64 ch
Radar temporal encoder	ConvLSTM, hidden dim 64, kernel 3×3
Radar latent dimension	64
Global radar tokens	128 learned queries, 64-d
Radar attention heads	4
Station history	6 precipitation values
AWS embedding	64-d
Coordinate embedding	64-d
Temporal embedding	64-d
Multimodal Context	256-d
Graph heads	4
Graph neighbors	$k = 10$
Edge embedding	64-d
Graph hidden dimension	128

Directly attending from every station to all radar tokens can be computationally expensive. MIRAGE therefore introduces 128 learned global radar tokens. These tokens act as queries over the complete radar latent map:

$$\mathbf{R}_t^g = \mathrm{CrossAttn}\left(\mathbf{Q}^g, \mathbf{R}_t, \mathbf{R}_t\right), \tag{7}$$

where $\mathbf{Q}^g \in \mathbb{R}^{128\times 64}$. All spatial radar tokens participate as keys and values, while the resulting compact token set $\mathbf{R}_t^g \in \mathbb{R}^{128\times 64}$ is used for subsequent station–radar interaction.

E.2 Station, Coordinate, and Temporal Encoding

For each station i, the six-step precipitation history

$$\mathbf{x}_{i,t} = [x_{i,t-5}, \ldots, x_{i,t}] \tag{8}$$

is processed by a shared recurrent encoder, producing a 64-dimensional precipitation representation $\mathbf{h}_{i,t}^{\mathrm{aws}}$. The same encoder parameters are shared across all stations and therefore do not depend on station identity or the number of stations.

Station coordinates are normalized and mapped independently to a 64-dimensional representation $\mathbf{h}_i^{\mathrm{coord}}$ through a shared multilayer perceptron.

Temporal information is represented using four periodic features corresponding to sine and cosine encodings of the daily and annual cycles. The temporal features over the six input steps are aggregated and mapped to a 64-dimensional representation $\mathbf{h}_t^{\mathrm{time}}$, which is shared across stations within the same sample.

For an unobserved or temporarily masked station, its precipitation history is masked in the model input, while its coordinate and temporal representations remain available. Consequently, different target locations remain distinguishable even when local precipitation observations are unavailable.

E.3 Radar–Station Cross-Attention

MIRAGE retrieves radar information separately for each station. A station query is constructed from its precipitation, coordinate, and temporal representations,

$$\mathbf{q}_{i,t} = f_q\left([\mathbf{h}_{i,t}^{\mathrm{aws}}, \mathbf{h}_i^{\mathrm{coord}}, \mathbf{h}_t^{\mathrm{time}}]\right) \in \mathbb{R}^{64}. \tag{9}$$

The station query attends to the compact global radar tokens:

$$\mathbf{h}_{i,t}^{\mathrm{radar}} = \mathrm{CrossAttn}\left(\mathbf{q}_{i,t}, \mathbf{R}_t^g, \mathbf{R}_t^g\right). \tag{10}$$

Because coordinate and temporal information are included in the query, stations with unavailable precipitation histories can still retrieve different radar representations according to their spatial and temporal context.

The retrieved radar representation is further regulated by a learned gate conditioned on the station precipitation representation. The gated radar feature is then combined with precipitation, coordinate, and temporal representations and projected to a 256-dimensional station-level Multimodal Context,

$$\mathbf{c}_{i,t} \in \mathbb{R}^{256}. \tag{11}$$

Collecting all stations gives $\mathbf{C}_t \in \mathbb{R}^{N\times 256}$.

The Multimodal Context is subsequently used for two distinct purposes: it controls the effective spatial connectivity through Dynamic Graph Construction and conditions the message transformation through the Hypernetwork.

E.4 Dynamic Graph Construction

Online geographical graph. MIRAGE constructs the Base Adjacency directly from the coordinates of the stations in the current network. Pairwise geographical distances are computed online, and each station is connected to its $k = 10$ nearest neighbors. Candidate edges are weighted using inverse squared distance, and self-connections are included.

For a network containing N stations, the effective number of neighbors is bounded by the available number of nodes. Consequently, the graph construction does not assume a fixed station count.

Context-dependent edge modulation. The geographical graph defines candidate propagation pathways, while their effective strengths are determined dynamically from the current Multimodal Context.

For each graph head h, shared query and key projections map $\mathbf{c}_{i,t}$ and $\mathbf{c}_{j,t}$ into a 64-dimensional edge space. In addition to this context-dependent relation, MIRAGE explicitly encodes the relative geometry of each station pair:

$$\mathbf{e}_{ij} = [\mathbf{s}_i - \mathbf{s}_j,\ \|\mathbf{s}_i - \mathbf{s}_j\|_2]. \tag{12}$$

A shared nonlinear geometry encoder produces a head-specific geometric bias. The Dynamic Mask is then computed as

$$m_{ij,t}^{(h)} = \sigma\left(\frac{(\mathbf{W}_q^{(h)}\mathbf{c}_{i,t})^\top(\mathbf{W}_k^{(h)}\mathbf{c}_{j,t})}{\sqrt{64}} + f_{\text{geo}}^{(h)}(\mathbf{e}_{ij})\right). \tag{13}$$

The mask therefore combines the current meteorological relationship between two stations with their relative geometry.

The Dynamic Adjacency is obtained by modulating the geographical candidate graph:

$$\mathbf{A}_{\text{dyn},t}^{(h)} = \text{Norm}\left(\mathbf{A}_{\text{base}} \odot \mathbf{M}_t^{(h)} \odot \mathbf{S}_t\right), \tag{14}$$

where $\mathbf{S}_t$ denotes source availability. The normalization is performed row-wise after masking.

Source availability. Only stations with available precipitation histories are allowed to send ground-observation messages. Fixed unobserved targets and temporarily masked stations are therefore removed as message sources. Operationally, their corresponding source columns in the adjacency are set to zero before row normalization.

Importantly, this restriction applies only to their role as ground-observation sources. These stations remain target nodes and retain their coordinate, temporal, and radar-conditioned Multimodal Context.

E.5 Context-Conditioned Hypernetwork

While Dynamic Graph Construction determines *where* information is propagated, MIRAGE also adapts *how* source information is transformed according to the current meteorological state.

The station-level Multimodal Context is summarized through global average pooling:

$$\mathbf{g}_t = \frac{1}{N}\sum_{i=1}^{N} \mathbf{c}_{i,t} \in \mathbb{R}^{256}. \tag{15}$$

The resulting Global Context is permutation invariant and has a fixed dimension independent of the number of stations. Two parameter generators conditioned on $\mathbf{g}_t$ produce the message transformation matrix and bias for each of the four graph heads:

$$\left\{\mathbf{W}_t^{(h)}, \mathbf{b}_t^{(h)}\right\}_{h=1}^{4} = \mathrm{HyperNet}(\mathbf{g}_t). \tag{16}$$

The generated parameters are shared across station identities but vary across samples according to the current Multimodal Context. Thus, Dynamic Graph Construction adapts spatial connectivity, whereas the Hypernetwork adapts the transformation applied to the information propagated along those connections.

E.6 Conditional Source-to-Target Propagation

The information propagated through the graph originates exclusively from station precipitation histories. In particular, radar, coordinate, and temporal features are not directly used as source messages.

For each station, the six precipitation values are mapped by a shared source encoder to

$$\mathbf{h}_{j,t}^{\mathrm{src}} = f_{\mathrm{src}}(\mathbf{x}_{j,t}) \in \mathbb{R}^{128}. \tag{17}$$

For each graph head, the source representation is transformed using the parameters generated by the Hypernetwork:

$$\widetilde{\mathbf{h}}_{j,t}^{(h)} = \phi\left(\mathbf{h}_{j,t}^{\mathrm{src}}\mathbf{W}_t^{(h)} + \mathbf{b}_t^{(h)}\right). \tag{18}$$

Unavailable sources are masked again after this transformation so that the generated bias cannot introduce nonzero messages at stations whose ground observations are unavailable.

Each receiving station then aggregates the transformed source messages according to its head-specific Dynamic Adjacency:

$$\mathbf{z}_{i,t}^{(h)} = \sum_j A_{\mathrm{dyn},t,ij}^{(h)} \widetilde{\mathbf{h}}_{j,t}^{(h)}. \tag{19}$$

The messages from the four graph heads are concatenated and projected to obtain the received graph representation

$$\mathbf{m}_{i,t} = f_{\mathrm{merge}}\left([\mathbf{z}_{i,t}^{(1)}, \ldots, \mathbf{z}_{i,t}^{(4)}]\right) \in \mathbb{R}^{128}. \tag{20}$$

This construction separates the two adaptive mechanisms explicitly: the Dynamic Adjacency determines *where* observed information is propagated, while the Hypernetwork-generated transformation determines *how* that information is represented before aggregation.

E.7 Target-Aware Reconstruction

Graph propagation provides each target with information from available ground observations. However, each target also retains its own Multimodal Context, which contains radar-conditioned spatial and temporal information.

The target context is first projected to a 128-dimensional representation,

$$\mathbf{z}_{i,t}^{\mathrm{tgt}} = f_{\mathrm{target}}(\mathbf{c}_{i,t}). \tag{21}$$

A learned gate then balances the received ground-observation message $\mathbf{m}_{i,t}$ and the target-specific context:

$$\boldsymbol{\alpha}_{i,t} = \sigma\left(f_{\text{gate}}([\mathbf{z}_{i,t}^{\text{tgt}}, \mathbf{m}_{i,t}])\right), \tag{22}$$

$$\mathbf{h}_{i,t} = \boldsymbol{\alpha}_{i,t} \odot \mathbf{m}_{i,t} + (1 - \boldsymbol{\alpha}_{i,t}) \odot \mathbf{z}_{i,t}^{\text{tgt}}. \tag{23}$$

The fused representation is subsequently refined by a target update with residual target-context information and passed to a node-wise decoder. The decoder independently produces the required output horizon for every station.

No additional graph propagation is performed after the target-aware reconstruction. Thus, the final prediction is based on the combination of propagated ground evidence and the target's own Multimodal Context, rather than on a second graph layer.

E.8 Training with Random Station Masking

The fixed target set V_u is excluded from both the precipitation input and the training objective. During training, MIRAGE additionally masks a randomly sampled subset $V_m \subset V_o$ of otherwise observed stations.

For stations in V_m, precipitation histories are removed from the model input and the stations are prevented from acting as graph message sources. Their coordinates, temporal information, and radar-conditioned Multimodal Context remain available.

Unlike the fixed target set, however, the ground-truth precipitation values of the randomly masked stations remain in the training objective. The training loss is therefore evaluated over the observed station set V_o, including the temporarily masked stations.

The random subset is resampled across training samples, exposing the model to different observation patterns and encouraging it to infer precipitation from neighboring ground measurements and multimodal context rather than relying exclusively on each station's local precipitation history.

E.9 Variable-Size Station Networks

MIRAGE does not associate learnable parameters with individual station identities or with a predetermined number of nodes. The station, coordinate, and source encoders are shared across nodes; station–radar cross-attention accepts a variable number of station queries; Dynamic Graph Construction evaluates shared edge functions; and the Hypernetwork operates on a fixed-dimensional pooled Global Context.

The Base Adjacency is constructed online from the coordinates of the current station set. Consequently, changing the station configuration changes the sizes of $\mathbf{C}_t$, $\mathbf{M}_t^{(h)}$, and $\mathbf{A}_{\text{dyn},t}^{(h)}$ but does not change the dimensions of the learnable parameters.

This makes the architecture structurally compatible with station networks containing different numbers and spatial arrangements of nodes. This architectural property does not by itself imply generalization to an unseen geographical region, which remains an empirical question.

F Evaluation Metrics

We evaluate precipitation estimates using RMSE and MAE as the primary continuous-valued metrics. In addition, we report the Critical Success Index (CSI) to evaluate categorical precipitation detection at different rainfall intensities.

For ground-truth precipitation y_i and prediction $\hat{y}_i$, RMSE and MAE are defined as

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}, \qquad \text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|. \tag{24}$$

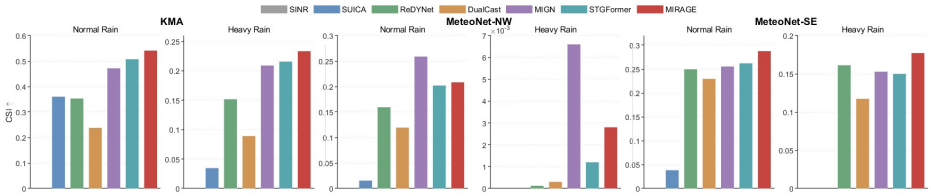


Figure 6: CSI comparison on KMA, MeteoNet-NW, and MeteoNet-SE for normal- and heavy-rain categories. Normal rain is defined as 1–10 mm/h for KMA and 0.5–2 mm/h for MeteoNet, while heavy rain is defined as $\geq$ 10 mm/h and $\geq$ 2 mm/h, respectively. Higher CSI indicates better categorical detection.

RMSE places greater weight on large errors, whereas MAE measures the average absolute prediction error.

F.1 Categorical Rainfall Evaluation

CSI complements RMSE and MAE by evaluating whether precipitation events in a specified intensity range are correctly detected. For a rainfall category, we count true positives (TP), false positives (FP), and false negatives (FN), and compute

$$\mathrm{CSI} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}. \tag{25}$$

Correct negatives are not included in CSI. This is useful for precipitation data because non-rain observations can be substantially more frequent than rainfall events.

We evaluate two precipitation categories for each dataset. For KMA, the normal-rain category is defined as $1 \leq y < 10$ mm/h and the heavy-rain category as $y \geq 10$ mm/h. For MeteoNet, the corresponding categories are $0.5 \leq y < 2$ mm/h and $y \geq 2$ mm/h.

For each category, the same interval or threshold is independently applied to the prediction and ground truth. For example, for MeteoNet normal rain,

$$\begin{aligned} P_i &= \mathbb{I}(0.5 \leq \hat{y}_i < 2), \\ G_i &= \mathbb{I}(0.5 \leq y_i < 2), \end{aligned} \tag{26}$$

whereas heavy rain is evaluated using $P_i = \mathbb{I}(\hat{y}_i \geq 2)$ and $G_i = \mathbb{I}(y_i \geq 2)$. The KMA categories are evaluated analogously using 1 and 10 mm/h.

This categorical analysis is reported in addition to RMSE and MAE and is intended to examine whether the models identify rainfall occurrences within the corresponding intensity regimes.

Figure 6 reports CSI separately for normal and heavy rainfall. These results provide information complementary to the continuous RMSE and MAE metrics: a model can achieve a relatively small average numerical error while still missing or falsely detecting rainfall events around an intensity threshold. CSI instead summarizes the balance between correctly detected events, missed events, and false alarms.

We therefore use the CSI results as an additional evaluation of rainfall-event detection rather than as a replacement for the continuous-valued error metrics. The normal- and heavy-rain results also allow performance to be inspected separately across different precipitation intensity regimes.

G Block Masking

To evaluate the spatial robustness of MIRAGE under catastrophic communication failures, we design a block masking experiment where all ground-truth AWS observations within an entire geographical region are completely lost during inference over a full one-year evaluation period. To thoroughly measure temporal prediction performance in the complete absence of local station observations, we

select four distinct target regions exhibiting diverse geographical and meteorological characteristics across South Korea:

- **Region 1:** A physically isolated volcanic island dominated by extreme mountain-induced orographic precipitation, where geographical isolation fundamentally prevents spatial message passing with mainland stations;
- **Region 2:** A mountainous northern border area subject to spatial information truncation due to the complete lack of observations from the north and fast-moving frontal systems;
- **Region 3:** An expansive southeastern inland basin enclosed by complex mountain ranges, ideal for evaluating information severance during large-scale regional blackouts; and
- **Region 4:** A mid-western coastal area serving as the primary landfall passage for convective storms originating over the Yellow Sea.

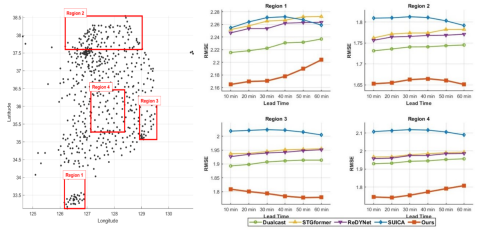


Figure 7: (a) Spatial distribution of weather stations and designated target regions (Regions 1–4) (b) Comprehensive evaluation framework consisting of geographical boundaries of experimental station networks and corresponding regional predictive performance under varying lead times.

The feature matrices of the stations within these designated boundaries are completely zeroed out using a binary masking matrix $\mathbf{M} \in \{0,1\}^{N\times F}$. During testing, the model must predict future precipitation in the masked region for up to a year without any localized station observations, relying solely on unmasked surrounding contexts and global radar images. This setup rigorously validates the capability of our Meteorological Hypernetwork to reconstruct the adjacency matrix and edge weights on-the-fly without relying on fixed spatial priors.

As illustrated in Figure 7, all evaluated models exhibit the highest RMSE in Region 1 compared to other regions. This performance degradation stems from the prevailing southwesterly wind patterns characteristic of South Korea's climate. Because Region 1 is located in the extreme southwesterly boundary, there are virtually no upstream (upwind) stations available to pass predictive spatial messages before weather systems reach this region. Consequently, the lack of spatial context from preceding nodes naturally creates a fundamental bottleneck for spatial message passing. Despite this intrinsic limitation, our proposed MIRAGE consistently outperforms all baseline models across all regions, demonstrating its superior capability to infer regional dynamics even under severe spatial information truncation.

While MIRAGE performs robustly under severe masking, its spatial propagation faces physical bottlenecks in boundary zones lacking upstream stations under prevailing winds (e.g., Region 1). Moreover, while the dynamic mask captures rain movement via radar echoes, it lacks explicit physical wind field constraints. Future work will integrate numerical wind vectors and physical advection priors into the graph formulation to better align message passing with atmospheric transport.

H Statistical Stability and Sensitivity Analysis

To examine the sensitivity of the results to the selection of unobserved stations, we repeat the experiments over multiple random station splits. Specifically, four independent runs are performed using seed $\in \{12, 22, 32, 42\}$. For each run, the target nodes corresponding to unobserved stations are randomly sampled at a fixed proportion from the full station network, while the same training and evaluation protocol is maintained across methods.

Table 8: Performance comparison on All Nodes and Unobserved Nodes across 10–60 minute forecasting horizons. Results are reported as mean ± standard deviation. Lower RMSE and MAE indicate better performance.

	Model	10 min		20 min		30 min		40 min		50 min		60 min	
		RMSE ↓	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓
All Nodes	SUICA	1.143 ± 0.0139	0.173 ± 0.0022	1.220 ± 0.0124	0.202 ± 0.0027	1.321 ± 0.0117	0.232 ± 0.0029	1.435 ± 0.0119	0.260 ± 0.0027	1.560 ± 0.0118	0.289 ± 0.0035	1.678 ± 0.0104	0.317 ± 0.0053
	SINR	1.789 ± 0.0081	0.351 ± 0.0022	1.799 ± 0.0077	0.353 ± 0.0023	1.815 ± 0.0074	0.357 ± 0.0023	1.833 ± 0.0073	0.361 ± 0.0024	1.850 ± 0.0071	0.366 ± 0.0024	1.867 ± 0.0068	0.371 ± 0.0022
	DualCast	1.663 ± 0.0244	0.419 ± 0.0047	1.663 ± 0.0286	0.399 ± 0.0005	1.671 ± 0.0469	0.401 ± 0.0053	1.672 ± 0.0748	0.404 ± 0.0083	1.673 ± 0.0962	0.397 ± 0.0040	1.773 ± 0.2638	0.464 ± 0.1450
	ReDYNet	1.642 ± 0.0814	0.658 ± 0.0062	1.648 ± 0.0917	0.682 ± 0.0008	1.698 ± 0.0569	0.711 ± 0.0031	1.691 ± 0.0689	0.689 ± 0.0015	1.642 ± 0.0107	0.678 ± 0.0008	1.752 ± 0.1818	0.722 ± 0.5499
	MIGN	1.231 ± 0.0205	0.247 ± 0.0018	1.286 ± 0.0209	0.255 ± 0.0011	1.328 ± 0.0206	0.265 ± 0.0015	1.438 ± 0.0200	0.277 ± 0.0013	1.596 ± 0.0202	0.292 ± 0.0018	1.641 ± 0.0203	0.306 ± 0.0015
	STGFormer	1.258 ± 0.0929	0.199 ± 0.0035	1.298 ± 0.0757	0.212 ± 0.0053	1.395 ± 0.0012	0.245 ± 0.0017	1.468 ± 0.0451	0.278 ± 0.0054	1.608 ± 0.0005	0.310 ± 0.0026	1.735 ± 0.1455	0.342 ± 0.8530
	MIRAGE	**0.955 ± 0.0258**	**0.139 ± 0.00403**	**1.065 ± 0.0141**	**0.171 ± 0.00235**	**1.194 ± 0.0146**	**0.202 ± 0.00285**	**1.341 ± 0.0140**	**0.234 ± 0.00240**	**1.497 ± 0.0081**	**0.269 ± 0.00145**	**1.634 ± 0.0050**	**0.300 ± 0.0034**
Unobserved Nodes	SUICA	1.797 ± 0.0257	0.347 ± 0.0069	1.797 ± 0.0274	0.346 ± 0.0077	1.799 ± 0.0283	0.347 ± 0.0074	1.804 ± 0.0284	0.348 ± 0.0060	1.810 ± 0.0282	0.350 ± 0.0077	1.819 ± 0.0277	0.353 ± 0.0123
	SINR	2.093 ± 0.0332	0.469 ± 0.0078	2.092 ± 0.0332	0.469 ± 0.0078	2.090 ± 0.0332	0.468 ± 0.0078	2.089 ± 0.0332	0.468 ± 0.0078	2.088 ± 0.0331	0.467 ± 0.0078	2.087 ± 0.0332	0.467 ± 0.0078
	DualCast	1.992 ± 0.0197	0.424 ± 0.0059	1.991 ± 0.0380	0.404 ± 0.0093	1.998 ± 0.0162	0.406 ± 0.0026	1.998 ± 0.0229	0.409 ± 0.0008	1.999 ± 0.0817	0.401 ± 0.0043	1.998 ± 0.8693	0.468 ± 0.3510
	ReDYNet	2.093 ± 0.0350	0.528 ± 0.0083	2.094 ± 0.0754	0.529 ± 0.0078	2.097 ± 0.0337	0.522 ± 0.0060	2.099 ± 0.0084	0.512 ± 0.0100	2.102 ± 0.0775	0.562 ± 0.0080	2.051 ± 0.1361	0.570 ± 0.6221
	MIGN	1.647 ± 0.0202	0.273 ± 0.0016	1.663 ± 0.0203	0.278 ± 0.0017	1.689 ± 0.0207	0.287 ± 0.0017	1.730 ± 0.0205	0.296 ± 0.0011	1.774 ± 0.0202	0.308 ± 0.0019	1.809 ± 0.0202	0.321 ± 0.0018
	STGFormer	1.976 ± 0.0251	0.409 ± 0.0055	1.974 ± 0.0568	0.410 ± 0.0013	1.974 ± 0.0794	0.410 ± 0.0065	1.975 ± 0.0913	0.411 ± 0.0044	1.975 ± 0.0869	0.410 ± 0.0091	1.977 ± 0.5797	0.411 ± 0.5132
	MIRAGE	**1.573 ± 0.0325**	**0.254 ± 0.00523**	**1.589 ± 0.0412**	**0.260 ± 0.00625**	**1.620 ± 0.0343**	**0.270 ± 0.00461**	**1.662 ± 0.0376**	**0.284 ± 0.00442**	**1.718 ± 0.0259**	**0.302 ± 0.00374**	**1.763 ± 0.0371**	**0.318 ± 0.00462**

Table 9: Performance comparison on MeteoNet datasets (NW and SE) for Unobserved Nodes (unseen stations) and All Nodes (entire spatial domain) across 30-minute and 60-minute forecasting horizons. Lower RMSE and MAE indicate better performance. Best results are shown in **bold** and second-best results are underlined.

Region	Node Type	Model	30 min		60 min	
			RMSE (↓)	MAE (↓)	RMSE (↓)	MAE (↓)
NW	All Nodes	SINR (2023)	0.282 ± 0.0004	0.053 ± 0.0002	0.290 ± 0.0002	0.057 ± 0.0002
		SUICA (2025)	0.282 ± 0.0001	0.052 ± 0.0003	0.295 ± 0.0003	0.056 ± 0.0006
		ReDYNet (2025)	**0.269 ± 0.0012**	0.043 ± 0.0023	0.299 ± 0.0018	0.048 ± 0.0012
		DualCast (2025)	0.311 ± 0.0002	0.074 ± 0.0004	0.309 ± 0.0003	0.072 ± 0.0002
		MIGN (2025)	0.274 ± 0.00066	0.045 ± 0.00068	0.289 ± 0.00069	0.051 ± 0.00064
		STGFormer (2026)	0.271 ± 0.00849	0.048 ± 0.00369	0.287 ± 0.00441	0.053 ± 0.00370
		MIRAGE (Ours)	0.273 ± 0.0012	**0.040 ± 0.0013**	**0.284 ± 0.0009**	**0.046 ± 0.0008**
	Unobserved Nodes	SINR (2023)	0.309 ± 0.0060	0.069 ± 0.0006	0.309 ± 0.0060	0.069 ± 0.0006
		SUICA (2025)	0.303 ± 0.0005	0.062 ± 0.0002	0.303 ± 0.00044	0.060 ± 0.0001
		ReDYNet (2025)	0.331 ± 0.0077	0.055 ± 0.0052	0.313 ± 0.0044	0.056 ± 0.0038
		DualCast (2025)	0.316 ± 0.00042	0.081 ± 0.0034	0.317 ± 0.00023	0.081 ± 0.0039
		MIGN (2025)	0.297 ± 0.00017	0.049 ± 0.00015	0.299 ± 0.00012	0.051 ± 0.00018
		STGFormer (2026)	0.298 ± 0.00015	0.058 ± 0.00089	0.298 ± 0.00011	0.061 ± 0.00012
		MIRAGE (Ours)	**0.295 ± 0.0060**	**0.044 ± 0.0015**	**0.296 ± 0.0059**	**0.048 ± 0.0011**
SE	All Nodes	SINR (2023)	0.425 ± 0.0009	0.073 ± 0.0003	0.437 ± 0.0007	0.077 ± 0.0003
		SUICA (2025)	0.418 ± 0.0080	0.063 ± 0.0016	0.439 ± 0.0070	0.071 ± 0.0024
		ReDYNet (2025)	0.428 ± 0.0038	0.068 ± 0.0068	0.429 ± 0.0026	0.075 ± 0.0025
		DualCast (2025)	0.470 ± 0.0044	0.094 ± 0.0028	0.471 ± 0.0075	0.088 ± 0.0084
		MIGN (2025)	0.401 ± 0.00071	0.061 ± 0.00080	0.421 ± 0.00075	0.073 ± 0.00077
		STGFormer (2026)	0.428 ± 0.0077	0.069 ± 0.0066	0.432 ± 0.0051	0.077 ± 0.0081
		MIRAGE (Ours)	**0.391 ± 0.0019**	**0.055 ± 0.0012**	**0.418 ± 0.0089**	**0.064 ± 0.0093**
	Unobserved Nodes	SINR (2023)	0.467 ± 0.0058	0.092 ± 0.0010	0.467 ± 0.0058	0.092 ± 0.0010
		SUICA (2025)	0.457 ± 0.0071	0.078 ± 0.0059	0.457 ± 0.0015	0.079 ± 0.0062
		ReDYNet (2025)	0.465 ± 0.0045	0.074 ± 0.0096	0.464 ± 0.0055	0.075 ± 0.0020
		DualCast (2025)	0.466 ± 0.0049	0.125 ± 0.0050	0.469 ± 0.0096	0.104 ± 0.0035
		MIGN (2025)	0.421 ± 0.00043	0.078 ± 0.00042	0.438 ± 0.00043	0.077 ± 0.00050
		STGFormer (2026)	0.454 ± 0.0065	0.076 ± 0.0034	0.453 ± 0.0014	0.077 ± 0.0025
		MIRAGE (Ours)	**0.415 ± 0.0075**	**0.060 ± 0.0022**	**0.426 ± 0.0026**	**0.067 ± 0.0047**

Tables 8 and 9 report the mean ± standard deviation of RMSE and MAE across the four runs for KMA and MeteoNet, respectively. The standard deviations measure the variation induced by different selections of unobserved target stations.

On KMA, MIRAGE achieves the lowest mean RMSE and MAE at all forecasting horizons from 10 to 60 minutes for both All Nodes and Unobserved Nodes. For All Nodes, the RMSE increases from 0.955 ± 0.0258 at 10 minutes to 1.634 ± 0.0050 at 60 minutes, while for Unobserved Nodes it increases from 1.573 ± 0.0325 to 1.763 ± 0.0371. The relatively small standard deviations across most horizons indicate that the reported performance is not strongly affected by the particular random selection of unobserved stations.

The MeteoNet results show a similar pattern, although MIRAGE is not the best method for every All-Node metric. On MeteoNet-NW, ReDYNet obtains the lowest All-Node RMSE at 30 minutes, whereas MIRAGE obtains the lowest All-Node MAE at both horizons and the lowest RMSE and MAE on Unobserved Nodes at both 30 and 60 minutes. On MeteoNet-SE, MIRAGE obtains the lowest mean RMSE and MAE for both All Nodes and Unobserved Nodes at both forecasting horizons.

Overall, the repeated-split results show that the main performance trends reported in the paper are preserved across the four random selections of unobserved stations. In particular, the advantage of MIRAGE is more consistent at Unobserved Nodes, which are the primary setting considered in this work.

Table 10: Comparison of inference latency per sample across models on KMA, NW, and SE datasets. Lower latency indicates faster inference. Best results are shown in **bold** and second-best results are underlined.

Model	Latency per Sample (ms) ↓		
	KMA	NW	SE
SINR (2023)	4.75	1.79	1.92
SUICA (2025)	24.00	1.60	1.70
ReDYNet (2025)	1.91	1.12	1.41
DualCast (2025)	**1.31**	**0.95**	**1.11**
MIGN (2025)	1.91	1.04	1.51
STGFormer (2026)	3.15	1.59	1.79
MIRAGE (Ours)	1.71	1.01	1.17

H.0.1 Mesh neighbors analysis

In this study, we quantitatively measured the performance variations across three datasets (KMA, NW, and SE) and different prediction horizons (t) by varying the number of neighbors (k), a key spatial graph parameter, from 5 to 50.

Figure 8 shows distinctly contrasting trends in how changes in the number of neighbors (k) affect the RMSE performance depending on the unique characteristics of each dataset. For the KMA dataset, which consists of summer monsoon data, short-term predictions ($t = 10$ min) achieved the optimal RMSE when the number of neighbors was set to 10 ($k = 10$), with errors increasing slightly beyond this threshold. In contrast, for long-term predictions ($t = 60$ min), the RMSE consistently followed a downward curve as the number of neighbors expanded toward $k = 50$, yielding the best performance.

Conversely, completely opposite results were observed for the MeteoNet (both NW and SE) datasets, which encompass full-year observations across all seasons. Regardless of the short-term ($t = 30$ min) or long-term ($t = 60$ min) prediction horizons, both MeteoNet NW and SE recorded the lowest RMSE when the neighborhood size was configured to be small ($k = 5$ or $k = 10$).

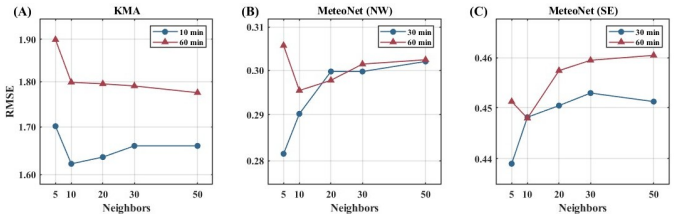


Figure 8: Comparison of model performance across different prediction horizons and neighborhood sizes

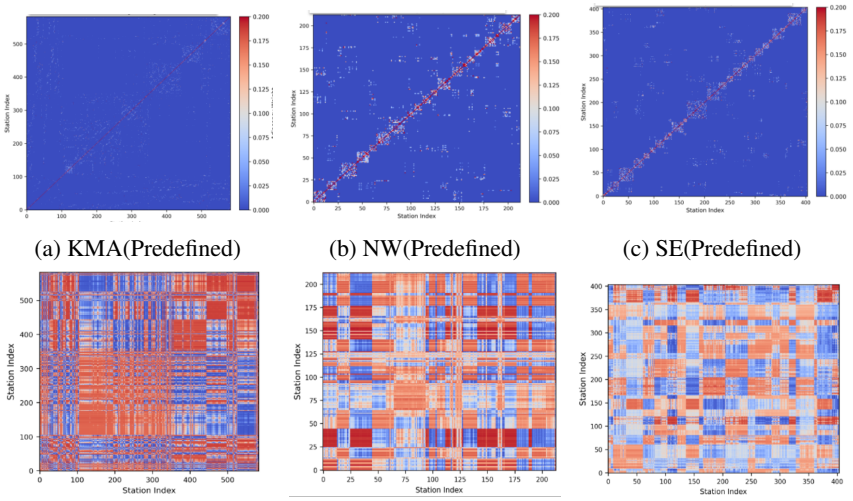
H.1 Computational Efficiency

We evaluated pure forward inference latency across all datasets under identical GPU settings. As shown in Table 10, our model achieves competitive low-latency inference while delivering superior predictive accuracy compared to baselines. This optimal balance between accuracy and computational efficiency demonstrates its strong applicability for real-time forecasting.

I Adjacency Interpretability

Prior efforts usually rely on a predefined adjacency matrix to model spatial connectivity, while our method aims to construct a dynamically responsive structure based on real-time weather conditions. Fig. 9 shows that the learned mask dynamically adapts across different regions according to local

radar reflectivity and precipitation levels. Likewise, the model's weights and biases adaptively shift in response to regional weather changes, illustrating that our approach can capture more complex, data-driven meteorological relations based on evolving rain dynamics.



(a) KMA(Predefined) (b) NW(Predefined) (c) SE(Predefined)

(d) KMA(Dynamic) (e) NW(Dynamic) (f) SE(Dynamic)

Figure 9: Comparison of spatial structure representations across KMA, NW, and SE datasets. Top row (a)–(c): Predefined adjacency matrices. Bottom row (d)–(f): Corresponding dynamic masks learned by the model. For all subfigures, both axes represent the station indices.