

Physics-Guided Inductive Spatiotemporal Kriging for PM_{2.5} with Satellite Gradient Constraints

Shuo Wang^{a,c}, Mengfan Teng^{b,*}, Yun Cheng^c, Lothar Thiele^c, Olga Saukh^d,
Shuangshuang He^e, Yuanting Zhang^e, Jiang Zhang^{a,*}, Gangfeng Zhang^f, Xingyuan
Yuan^e, Jingfang Fan^a

^a*School of Systems Science, Beijing Normal University, Beijing, China*

^b*College of Information Engineering, Jiangxi University of Science and Technology, Ganzhou, China*

^c*D-ITET, ETH Zurich, Zurich, Switzerland*

^d*Graz University of Technology, Graz, Austria*

^e*ColorfulClouds Technology Co.,Ltd., Beijing, China*

^f*State Key Laboratory of Earth Surface Processes and Disaster Risk Reduction, Beijing Normal University, Beijing, China*

Abstract

High-resolution mapping of fine particulate matter (PM_{2.5}) is a cornerstone of sustainable urbanism but remains critically hindered by the spatial sparsity of ground monitoring networks. While traditional data-driven methods attempt to bridge this gap using satellite Aerosol Optical Depth (AOD), they often suffer from severe, non-random data missingness (e.g., due to cloud cover or nighttime) and inversion biases. To overcome these limitations, this study proposes the **S**patiotemporal **P**hysics-Guided **I**nductive **N**etwork (SPIN), a novel framework designed for inductive spatiotemporal kriging. Unlike conventional approaches, SPIN synergistically integrates domain knowledge into deep learning by explicitly modeling physical advection and diffusion processes via parallel graph kernels. Crucially, we introduce a paradigm-shifting training strategy: rather than using error-prone AOD as a direct input, we repurpose it as a *spatial gradient constraint* within the loss function. This allows the model to learn structural pollution patterns from satellite data while remaining robust to data voids. Validated in the highly polluted Beijing-Tianjin-Hebei and Surrounding Areas (BTHSA), SPIN achieves a new state-of-the-art with a Mean Absolute Error (MAE) of 9.52 $\mu\text{g}/\text{m}^3$, effectively generating continuous, physically

*Corresponding authors.

Email addresses: shuowang.ai@gmail.com (Shuo Wang), tengmengfan@whu.edu.cn (Mengfan Teng), zhangjiang@bnu.edu.cn (Jiang Zhang)

plausible pollution fields even in unmonitored areas. This work provides a robust, low-cost, and all-weather solution for fine-grained environmental management.

Keywords: PM_{2.5} Inference, Physics-Informed Deep Learning, Spatiotemporal Kriging, Graph Neural Networks, Aerosol Optical Depth, AI for Earth Science

1. Introduction

1.1. *From Forecasting to Inference: The Spatial Gap in Air Quality Intelligence*

Fine particulate matter (PM_{2.5}) remains a formidable threat to global public health and ecological stability (Keller et al., 2015; Burnett et al., 2018). Long-term exposure to elevated PM_{2.5} levels is intrinsically linked to cardiovascular (Chen et al., 2017; Xi et al., 2022) and respiratory diseases (Chai et al., 2019), necessitating precise monitoring for effective policymaking. Recently, breakthrough efforts in gapless global and regional mapping (Geng et al., 2021; Southerland et al., 2022; Wei et al., 2023) have revealed a critical discrepancy: sparsely distributed ground stations frequently fail to capture fine-grained pollution hotspots, thereby fundamentally limiting the accuracy of community-level exposure assessments.

While the past decade has witnessed remarkable progress in *spatiotemporal forecasting at fixed monitoring sites*, driven by advanced deep learning models (Qi et al., 2019; Wang et al., 2020; Xiao et al., 2022; Ma et al., 2025; Wang et al., 2025; Teng et al., 2023, 2024), a pervasive challenge persists: the *spatial data gap*. Although these approaches have achieved high precision in predicting future states, they are predominantly **transductive** in nature—fundamentally relying on historical local data streams and fixed graph topologies of existing stations. This dependency renders them ineffective for the vast unmonitored "blind spots" where no historical ground truth exists. Consequently, to achieve holistic environmental intelligence, the research frontier must expand from **fixed-point forecasting** to **inductive spatiotemporal kriging**—the capability to reconstruct continuous, high-resolution pollution fields for arbitrary unobserved locations based on sparse point measurements.

1.2. *Challenges in Current Modeling Paradigms*

Current approaches to high-resolution mapping predominantly fall into two categories, each with distinct limitations that hinder real-time, all-weather application.

Chemical Transport Models (CTMs): Physics-based models, such as CMAQ and WRF-Chem, represent the traditional gold standard, offering interpretability by numerically solving atmospheric dynamic equations. However, their operational utility is often constrained by prohibitive computational costs for real-time, high-resolution simulations and a heavy reliance on emission inventories that often lag

behind real-time changes (Woody et al., 2016; Wang et al., 2025). As noted in Wang et al. (2025), while surrogate models are emerging to address computational bottlenecks, explicitly integrating physical constraints into data-driven inference remains a grand challenge in Earth system science (Reichstein et al., 2019).

Data-Driven Deep Learning & AOD Utilization: In the data-driven domain, Graph Neural Networks (GNNs) have emerged as the dominant paradigm for capturing complex spatiotemporal dependencies (Li et al., 2018; Wang et al., 2020; Xiao et al., 2022; Ma et al., 2025). However, a critical limitation of most existing architectures is their *transductive* nature—they typically rely on fixed graph structures defined during training and lack the inductive capability to generalize to unobserved locations outside the training graph.

To overcome station sparsity, many studies incorporate satellite Aerosol Optical Depth (AOD) as an auxiliary spatial input (Van Donkelaar et al., 2016; Wei et al., 2021b; Teng et al., 2023). Pioneering works like Wei et al. (2021a) and Xiao et al. (2017) have successfully utilized AOD for high-quality historical reconstruction. However, relying on AOD as a *direct input feature* presents a fundamental operational dilemma: satellite retrievals suffer from severe, non-random missingness due to cloud cover and high surface reflectance (e.g., snow/ice) (Levy et al., 2013; Zhang et al., 2019). Consequently, models treating AOD as a mandatory input often fail or produce significant artifacts in these data-void regions, severely limiting their robustness for real-time, all-weather operational systems.

1.3. A Physics-Guided Inductive Framework

To bridge these gaps and answer the call for Physics-Informed Machine Learning (PIML) (Karniadakis et al., 2021), this study proposes the **S**patiotemporal **P**hysics-Guided **I**nterference **N**etwork (SPIN). SPIN distinguishes itself by embedding physical laws into the deep learning architecture and treating satellite data as a constraint rather than an input. The primary contributions are fourfold:

1. **Physics-Embedded Architecture:** We propose a novel graph network architecture that explicitly embeds fundamental atmospheric processes. By designing parallel graph kernels for advection (wind-driven transport) and diffusion, the model learns spatiotemporal dependencies consistent with physical laws, surpassing the "black-box" nature of traditional deep learning.
2. **Inductive Inference Mechanism:** The framework is explicitly designed for inductive kriging. Through a dynamic training strategy that randomly partitions observed and unobserved nodes, SPIN learns a generalizable interpolation function. This enables it to accurately predict $\text{PM}_{2.5}$ concentrations at any arbitrary unobserved location (*Station Inference*) and generate complete maps

(*Grid Inference*), effectively extending prediction capabilities from discrete sites to the global space.

3. **Paradigm-Shifting AOD Fusion:** We introduce a novel *Masked AOD Spatial Gradient Loss*. Instead of using AOD as a direct, error-prone input feature (which fails under cloud cover), we use the *spatial gradients* of available AOD data to constrain the model’s output structure during training. This approach robustly leverages the structural patterns in satellite imagery while elegantly circumventing the pervasive issue of missing data.
4. **SOTA Performance:** Through extensive experiments in the highly polluted Beijing-Tianjin-Hebei and Surrounding Areas (BTHSA), demonstrate that SPIN significantly outperforms comprehensive baselines—spanning feature-based, temporal, and state-of-the-art spatiotemporal kriging methods—achieving a new benchmark with a Mean Absolute Error (MAE) of $9.52 \mu\text{g}/\text{m}^3$.

2. Materials and Methods

2.1. Study Domain and Multi-Source Datasets

Our study targets the **BTHSA**, specifically anchoring on the strategic "**2+26**" **cities** air pollution transmission channel (Lu et al., 2021) (Figure 1). This region is characterized by highly heterogeneous topography, flanked by the **Taihang Mountains** to the west and the **Bohai Sea** to the east, enclosing the vast North China Plain. Such complex terrain induces distinct meteorological patterns (e.g., mountain blocking effects and land-sea circulation), resulting in **sharp pollution gradients** and frequent accumulation episodes. These challenging environmental dynamics render the BTHSA an ideal **testbed** for developing and validating the robustness of our physics-guided inference framework.

To comprehensively capture these spatiotemporal dynamics, we constructed a multi-source dataset integrating ground observations, meteorological reanalysis, and emission inventories. As summarized in Table 1, all datasets were processed to a unified hourly temporal resolution and spatially aligned to a uniform $0.25^\circ \times 0.25^\circ$ grid to support the grid inference task.

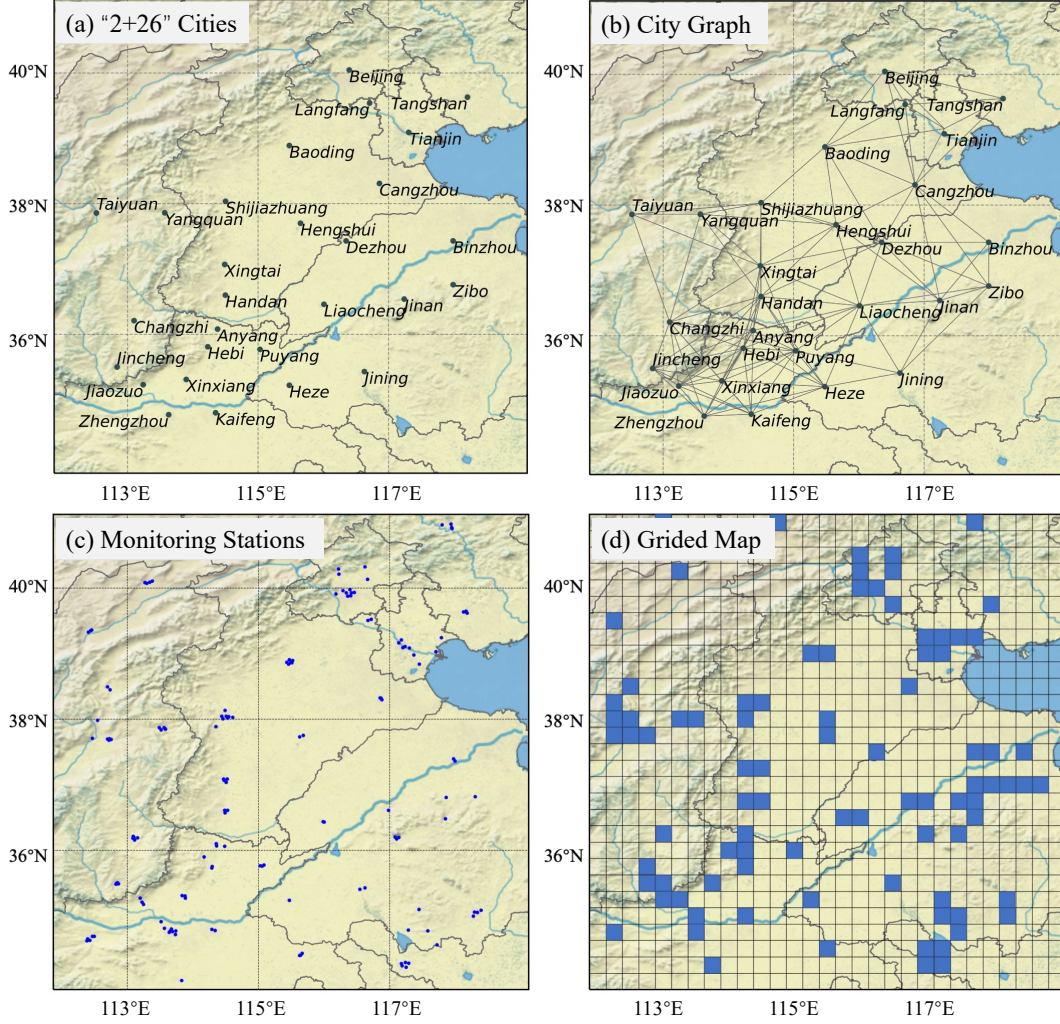


Figure 1: **Study area and task setup.** (a) The "2+26" cities in the BTHSA region. (b) A spatial adjacency graph constructed by connecting monitoring stations within a **200km geodesic threshold** (Qi et al., 2019; Wang et al., 2021). (c) Sparse distribution of the 152 ground monitoring stations. (d) The 0.25° grid used for dense map inference, where most cells lack ground truth.

- **Air Pollutants (X):** Hourly $PM_{2.5}$ concentrations were sourced from the China National Environmental Monitoring Center (CNEMC). Following rigorous quality control, data from 152 national monitoring stations (Figure 1(c)) were retained to serve as the ground truth.

Table 1: Summary of multi-source datasets used in BTHSA. Notations **X**, **P**, and **Q** denote Air Pollutants, Meteorology, and Emissions, respectively.

Category (Notation)	Variable	Description	Unit	Source
Target (X)	PM _{2.5}	Fine particulate matter	$\mu\text{g}/\text{m}^3$	CNEMC
Meteorology (P)	$t2m, d2m$	2m Temperature & Dewpoint	K	ERA5
	$rh2m, sp$	Relative Humidity & Surface Pressure	$\%, hPa$	
	tp, blh	Total Precipitation & Boundary Layer Height	m	
	u_{100}, v_{100}	100m U/V Wind Components	m/s	
	rad	Surface Short-wave Radiation	W/m^2	
Emissions (Q)	Gases	NO _x , SO ₂ , NH ₃ , VOC	ton	MEIC
	Particulates	primary PM _{2.5} , PM ₁₀	ton	
Auxiliary	AOD	Aerosol Optical Depth	-	JMA (Himawari-8)

- **Meteorology (P):** Key meteorological variables governing pollutant transport and diffusion—including wind components (u_{100}, v_{100}), temperature ($t2m$), precipitation (tp), and boundary layer height (blh)—were sourced from the ERA5 reanalysis dataset. These variables represent the atmospheric conditions driving PM_{2.5} formation and accumulation (Chen et al., 2017).
- **Emissions (Q):** Regional anthropogenic emissions of NO_x, VOC, SO₂, NH₃, PM₁₀, and PM_{2.5} were obtained from the **Multi-resolution Emission Inventory for China (MEIC)** (Li et al., 2017; Geng et al., 2024). While these monthly inventories are statistical estimates, they effectively capture general emission trends. To align with hourly air quality and meteorological data, we downscaled them to an hourly resolution using the temporal profile methodology described in the technical guideline (CSES, 2024). Emission variables include primary PM_{2.5} and PM₁₀, which are distinct from the ambient concentration target.
- **Satellite AOD (X^{AOD}):** Hourly AOD products from the Himawari-8 satellite were utilized as an auxiliary spatial constraint. While AOD is highly correlated with surface PM_{2.5}, it suffers from frequent data missingness due to cloud cover, a challenge we specifically address via our proposed loss function.

2.2. Problem Formulation: Inductive Spatiotemporal Kriging

We formulate the inference task as an *inductive spatiotemporal kriging* problem. The core objective is to learn a generalized mapping function capable of inferring PM_{2.5} concentrations at arbitrary unobserved locations by leveraging sparse monitoring data and auxiliary physical fields. This problem is instantiated through two distinct downstream tasks:

- **Station Inference:** Inferring time-series data for a held-out subset of monitoring stations based on the remaining observed stations. This task assesses the model’s ability to interpolate and generalize within the sparse monitoring network.
- **Grid Inference:** Generating a continuous, high-resolution (0.25°) $\text{PM}_{2.5}$ spatial field by treating all grid cells without stations as unobserved target nodes. This task aims to reconstruct the full spatial distribution of pollutants to eliminate monitoring blind spots.

Formally, let $G = (\mathcal{V}, \mathcal{E})$ denote a graph where $\mathcal{V}$ represents the set of all spatial locations (stations or grid cells). The node set is dynamically partitioned into observed nodes $\mathcal{V}_{obs}$ and unobserved target nodes $\mathcal{V}_{target}$. Given the historical pollutant observations at $\mathcal{V}_{obs}$, along with meteorological conditions $\mathbf{P}$ and emissions $\mathbf{Q}$ for all nodes, we aim to learn a mapping function $\mathcal{F}$ to infer $\text{PM}_{2.5}$ concentrations at $\mathcal{V}_{target}$:

$$\hat{\mathbf{X}}_{target} = \mathcal{F}(\mathbf{X}_{obs}, \mathbf{P}, \mathbf{Q}, G) \quad (1)$$

A critical challenge in **Grid Inference** is ensuring the physical plausibility of the generated field over vast unmonitored regions. While satellite-derived Aerosol Optical Depth (AOD) offers valuable spatial guidance, it suffers from high spatiotemporal sparsity due to cloud cover and nighttime limitations. To address this, our framework introduces AOD data ($\mathbf{X}^{AOD}$) not as a direct input feature, but as a *spatial structural constraint* during training. This strategy allows the model to leverage satellite spatial patterns while effectively overcoming the pervasive data missingness challenge.

2.3. The SPIN Architecture

The proposed SPIN (Figure 2) embodies a "Physics-Embedded" design philosophy, seamlessly integrating deep representation learning with explicit atmospheric dynamic laws. The architecture consists of three hierarchical stages: temporal feature encoding, physics-guided propagation, and constrained optimization.

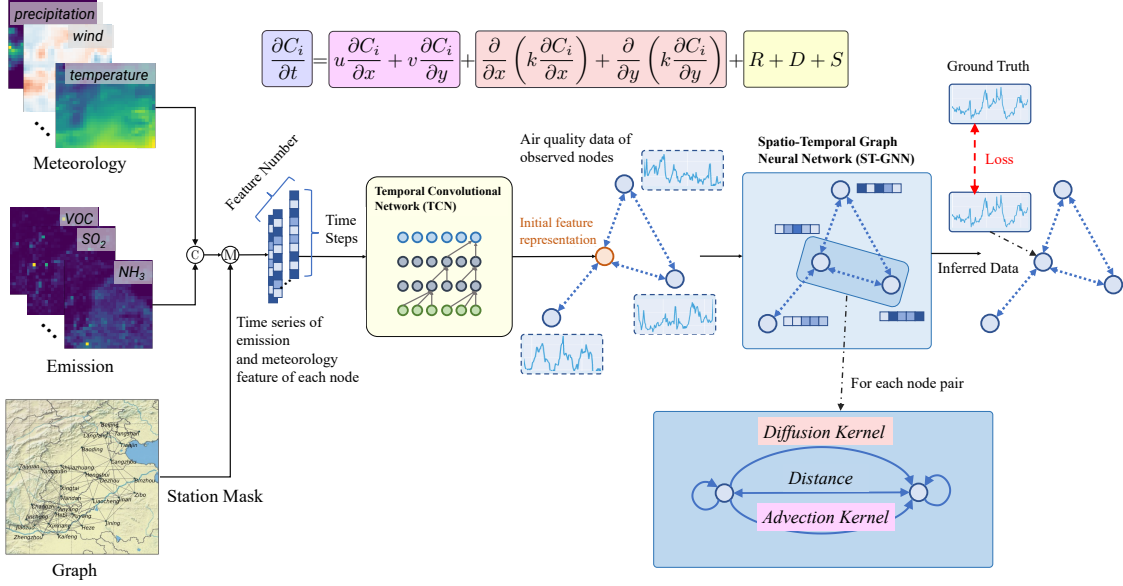


Figure 2: **Architecture of SPIN.** The model first encodes local temporal dynamics via TCN. It then propagates spatial information via a Physics-Guided Graph Network, driven by two explicit kernels: a *Diffusion Kernel* (isotropic spreading) and an *Advection Kernel* (wind-driven transport). Finally, the training is constrained by a novel Masked AOD Gradient Loss to ensure structural consistency with satellite observations.

2.3.1. Temporal Feature Encoding (TCN)

In the first stage, we employ a Temporal Convolutional Network (TCN) to encode the local temporal dynamics for each node. Let $\mathcal{X}_i^{in} = [\mathbf{P}_i, \mathbf{Q}_i]$ denote the input time-series of meteorology and emissions for node i . The TCN maps this sequence to a high-dimensional latent representation $\mathbf{H}_i^{(0)}$:

$$\mathbf{H}_i^{(0)} = \text{TCN}(\mathcal{X}_i^{in}) \quad (2)$$

Compared to RNNs, TCNs employ dilated causal convolutions to capture long-range temporal dependencies (e.g., cumulative emission effects) more efficiently with parallel computing (Bai et al., 2018). This step provides a physics-informed initialization for all nodes, offering a robust starting point even for unmonitored grid cells where no historical pollution data is available.

2.3.2. Physics-Guided Spatiotemporal Propagation

A single graph structure is insufficient to capture the multifaceted interactions in atmospheric systems. Therefore, we explicitly model the spatiotemporal dependen-

cies using a set of three distinct graphs, each encoding a specific physical or statistical relationship:

1. Geospatial Graph (G^S): This graph captures the fundamental principle of spatial autocorrelation (Tobler’s First Law of Geography). The adjacency matrix A^S is constructed based on the pairwise geographic distances between nodes v_i and v_j using a Gaussian kernel, thresholded by a distance ξ :

$$A_{ij}^S = \begin{cases} \exp\left(-\frac{\text{dist}(v_i, v_j)^2}{\sigma^2}\right), & \text{if } \text{dist}(v_i, v_j) < \xi \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where σ^2 is the variance of distances. This graph serves as the structural basis for modeling isotropic interactions.

2. Diffusion Graph (G^D): Based on G^S , this graph models the physical process of diffusion, where pollutants spread from high to low concentration areas. Mathematically, this corresponds to the Laplacian operator on the graph. We define the diffusion operator $\tilde{A}^D$ by symmetrically normalizing A^S :

$$\tilde{A}^D = \mathbf{D}_S^{-\frac{1}{2}} A^S \mathbf{D}_S^{-\frac{1}{2}} \quad (4)$$

where $\mathbf{D}_S$ is the degree matrix of A^S . This operator allows the GNN to simulate the smoothing effect of atmospheric diffusion.

3. Advection Graph (G^A): This kernel explicitly models anisotropic transport driven by wind fields. A directed edge exists from node j to i only if j is upwind. As visually illustrated in Figure 3(a), the edge weight is determined by the projection of the wind vector $\vec{v}$ onto the node-pair vector $\vec{e}_{ji}$:

$$A_{ij}^A = \begin{cases} \text{ReLU}\left(\frac{|\vec{v}|}{d_{ij}} \cdot \cos(\alpha)\right), & \text{if } d_{ij} < \xi \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where α is the angle between the wind direction and the vector connecting node j to i , and d_{ij} is the distance. This kernel encodes the physical prior that pollutants transfer significantly faster and stronger along the wind direction.

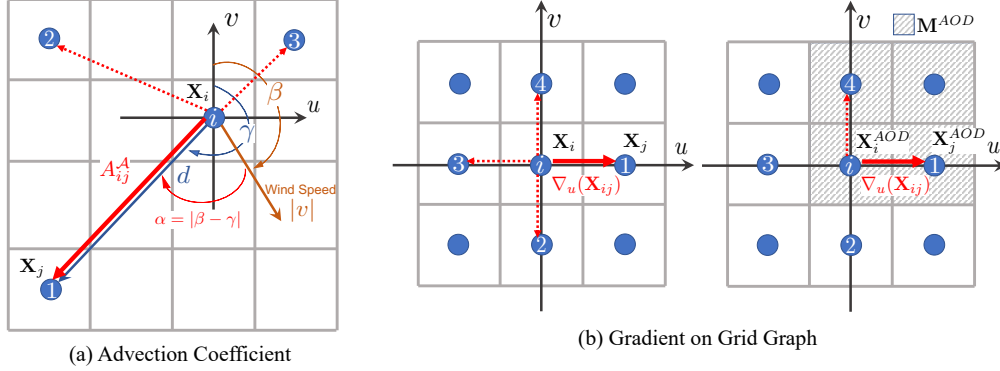


Figure 3: **Physical Kernels and AOD Constraint.** (a) The **Advection Kernel** weights edges based on wind velocity projections, modeling directional transport. (b) The **AOD Gradient Loss** aligns the spatial difference of predictions with AOD, active only on valid pixels (shaded area), ensuring robustness to missing data.

Propagation Layer: The node representation $\mathbf{H}^{(l)}$ at layer l is updated by aggregating messages from both the Diffusion and Advection operators:

$$\mathbf{H}^{(l)} = \sigma \left(\left(\tilde{A}^D \mathbf{H}^{(l-1)} + A^A \mathbf{H}^{(l-1)} \right) \mathbf{W}^{(l)} \right) \quad (6)$$

where $\mathbf{W}^{(l)}$ is the learnable weight matrix and σ is the activation function. This parallel propagation disentangles the isotropic diffusion (local spreading) and anisotropic advection (directional transport), enabling the model to learn physically interpretable patterns.

2.3.3. Readout Layer and Final Prediction

The final node representations $\mathbf{H}^{(L)}$ synergize *local temporal dynamics* (encoded by TCN) with *spatial context* aggregated from neighbors via physics-guided propagation. Since target nodes lack historical ground truth, we employ a node-wise shared Multi-Layer Perceptron (MLP) to directly map these integrated features to the pollutant concentration:

$$\hat{\mathbf{X}}_i^t = \text{MLP}(\mathbf{H}_i^{(L),t}) \quad (7)$$

This mapping enables inference based on the full spatiotemporal context—local interactions combined with regional transport—effectively filling data gaps in unobserved regions.

2.4. Training Strategy

2.4.1. Inductive Node Masking

To enforce inductive generalization, we employ a dynamic masking strategy (Wu et al., 2021). In each training iteration, we randomly mask a subset (e.g., 50%) of the observed stations as "unknown" target nodes. The model is forced to reconstruct the values of these masked nodes by combining their **local meteorological and emission features** with **pollution context propagated from the remaining observed nodes**. This mechanism prevents the model from overfitting to specific station IDs and accurately simulates the real-world scenario of inferring unmonitored locations where historical air quality data is absent.

2.4.2. Composite Loss with AOD Gradient Constraint

A critical innovation of SPIN is the integration of multi-stage supervision and satellite constraints. We define a composite loss function $\mathcal{L}$:

$$\mathcal{L} = \mathcal{L}_{infer} + \lambda_1 \mathcal{L}_{init} + \lambda_2 \mathcal{L}_{AOD} \quad (8)$$

where λ_1 and λ_2 are hyperparameters balancing the contribution of initialization and spatial constraints.

1. Inference Loss ($\mathcal{L}_{infer}$): This is the primary objective. After the physics-guided GNN aggregates real pollution information from observed neighbors to refine the initial features, the final prediction $\hat{X}$ is generated. We minimize the L1 error on the **masked target nodes** ($\mathcal{V}_{target}$) to simulate the inference scenario:

$$\mathcal{L}_{infer} = \sum_{t=1}^T \sum_{i \in \mathcal{V}_{target}} \left| \hat{X}_i^t - X_i^t \right| \quad (9)$$

This ensures the model learns to accurately infer unknown locations by combining propagated neighbor information with local features.

2. Initialization Loss ($\mathcal{L}_{init}$): To enable robust inference, we require the TCN to generate a reasonable "initial proposal" ($\hat{X}^{init}$) based solely on the node's local meteorology and emissions, *before* any neighbor information is aggregated. Crucially, this loss is also calculated on the **target nodes**:

$$\mathcal{L}_{init} = \sum_{t=1}^T \sum_{i \in \mathcal{V}_{target}} \left| \text{MLP}(\mathbf{H}_i^{(0),t}) - X_i^t \right| \quad (10)$$

This constraint forces the TCN to act as a solid "prior" generator. Even without neighbor information, the model can produce a physics-based baseline estimate,

which is then refined by the GNN layers. This follows the inductive training paradigm described in Wu et al. (2021).

3. Masked AOD Spatial Gradient Loss ($\mathcal{L}_{AOD}$): This term introduces spatial structure constraints, particularly for Grid Inference. We align the *spatial gradient* of the final prediction with that of the satellite AOD field. A binary mask $\mathbf{M}$ ensures robustness against missing satellite data:

$$\mathcal{L}_{AOD} = \sum_{t=1}^T \sum_{(i,j) \in \mathcal{E}} \left| \nabla_{ij}(\hat{\mathbf{X}}) - \nabla_{ij}(\mathbf{X}^{AOD}) \right| \cdot \mathbf{M}_{ij}^t \quad (11)$$

As visualized in Figure 3(b), $\nabla_{ij}(\mathbf{X}) = X_j - X_i$ approximates the spatial gradient between adjacent nodes. This loss injects satellite-derived structural knowledge (e.g., pollution plume shapes) into the model without requiring AOD as a direct input, ensuring continuous and physically consistent mapping.

3. Results

To comprehensively validate the **inductive generalization** and **physical consistency** of SPIN, the experimental evaluation is structured into three hierarchical stages. First, we establish the model’s quantitative superiority in *Station Inference* against state-of-the-art baselines. Second, we conduct a "stress test" under extreme data scarcity (50% missing stations) to evaluate robustness. Finally, we provide a forensic mechanism analysis of *Grid Inference*, demonstrating how the physics-informed kernels and AOD constraints cooperate to generate continuous, high-fidelity pollution maps.

3.1. Quantitative Superiority in Station Inference

We first evaluate the core inductive capability by partitioning the 152 monitoring stations into a training set (70%, 106 stations) and a **completely unobserved** test set (30%, 46 stations). To rigorously prevent temporal data leakage, we employed a strict chronological split strategy:

- **General Evaluation:** To strictly prevent temporal data leakage, data from **2016 to 2018** were used for training, **2019** for validation, and **2020** for testing.
- **Seasonal Evaluation:** To assess adaptability to extreme seasonal variations, we constructed specific sub-datasets:

- **Winter (Heating Season):** Defined as November to March. The training set includes winters from Jan 2016 to Mar 2018, while the validation and test sets cover the winters of **2018/2019** (Nov 2018–Mar 2019) and **2019/2020** (Nov 2019–Mar 2020), respectively.
- **Summer (Warm Season):** Defined as May to September. The training set spans the summers of **2016 to 2018**, with validation and testing performed on the summers of **2019** and **2020**, respectively.

Table 2 summarizes the performance benchmarks. While spatiotemporal models generally outperform traditional feature-based and purely temporal methods, SPIN establishes a new state-of-the-art.

- **Overall Performance:** SPIN achieves a yearly Mean Absolute Error (MAE) of **9.52** $\mu\text{g}/\text{m}^3$, representing a substantial **25.2% improvement** over the strongest baseline, IGNNK (Wu et al., 2021). This confirms that explicitly modeling physical transport yields superior generalization compared to purely statistical interpolation (Kriging).
- **Seasonal Robustness (The Winter Challenge):** The performance gap is most pronounced in winter (MAE 15.09 $\mu\text{g}/\text{m}^3$), a season characterized by severe pollution episodes and complex meteorological conditions. Traditional models like STGCN struggle here (MAE 23.43 $\mu\text{g}/\text{m}^3$). The fundamental limitation is that STGCN relies on **static graphs based solely on geographic distance**, implicitly assuming isotropic influence between locations. It fails to capture the **anisotropic** nature of wind-driven transport (i.e., upwind sources affect downwind targets, not vice-versa), which is the dominant driver of winter pollution. In contrast, SPIN utilizes the dynamic *Advection Kernel* to explicitly model this directional transport, ensuring robust inference.
- **Baseline Analysis:** Feature-based models (e.g., XGBoost) and temporal models (e.g., LSTM) fail to capture spatial correlations, resulting in high errors ($> 20 \mu\text{g}/\text{m}^3$). Even advanced inductive baselines like IGNNK, while capable of dynamic graph modeling, still construct connections based on **spatial proximity**. Both STGCN and IGNNK lack an explicit mechanism to encode **physical dynamics** (e.g., diffusion vs. advection). SPIN’s success validates that embedding domain knowledge—specifically separating wind direction from distance—provides a stronger inductive bias than learning from proximity-based data alone.

Table 2: **Comparison of PM_{2.5} inference performance on unobserved stations (30% hold-out).** SPIN achieves the lowest error (MAE) across all seasons, demonstrating superior inductive generalization compared to feature-based, temporal, and data-driven spatiotemporal baselines.

Model Category	Methods	All Year MAE↓	Winter MAE↓	Summer MAE↓
Feature-based	XGBoost	24.21	35.97	15.49
	MLP	25.71	38.67	15.30
Temporal	LSTM	24.65	34.96	14.49
	GRU	22.79	34.17	15.00
Spatiotemporal	STGCN	17.84	23.43	11.23
	IGNNK	12.73	16.29	8.53
Physics-informed	SPIN (Ours)	9.52	15.09	7.65

3.2. Robustness Verification under Sparse Observation

To probe the limits of SPIN in data-scarce scenarios, we increased the difficulty by randomly masking **50%** of the stations. This "stress test" evaluates whether the physics-informed kernels can maintain structural integrity when local information is severely depleted. The evaluation is conducted across three dimensions: regional consistency, error bounds analysis, and edge-case stability.

1. Regional Generalization (Central Urban Agglomeration): Figure 4 illustrates inference results for three representative unobserved stations in **Beijing, Tianjin, and Shijiazhuang**—the core cities of the BTHSA pollution transmission channel. Even with half the network missing, the model accurately reconstructs synoptic-scale variability ($R^2 > 0.85$) across these geographically distinct centers. It captures both the rapid accumulation and dissipation phases of regional episodes (e.g., late January 2020), demonstrating that the learned physical kernels effectively generalize across the major transport corridor.

2. Error Analysis (Topology Dependence): Figure 5 provides a forensic contrast between best and worst-performing cases, revealing the influence of local station topology:

- **Best Case (Station 2409A):** Located in a dense cluster, this station achieves near-perfect alignment ($R^2 = 0.96$). The dense connectivity allows the *Diffusion Kernel* to effectively aggregate gradients, resulting in high-fidelity reconstruction.
- **Worst Case (Station 2396A):** This station exhibits larger residuals ($R^2 = 0.644$) during rapid shifts, likely due to micro-scale emission bursts smoothed

out by neighbor averaging. However, the model still correctly identifies the *onset timing* and *trend direction*, proving that the *Advection Kernel* successfully captures macro-scale transport even when local magnitude is underestimated.

3. Extreme Sparsity (Physics as a Safety Mechanism): A critical test is shown in Figure 6, focusing on the challenging Station 1057A (Zhangjiakou), an edge node with almost no upstream sensors. While the quantitative metric drops ($R^2 = 0.132$) due to the loss of high-frequency variance, a significant physical property emerges: the model successfully recovers the *low-frequency trend* and correct background concentration levels (20-50 $\mu\text{g}/\text{m}^3$). It does not "hallucinate" zero values or random noise. This validates that our *Diffusion Kernel* acts as a conservative low-pass filter in data vacuums, ensuring inference remains physically plausible even when information is extremely scarce.

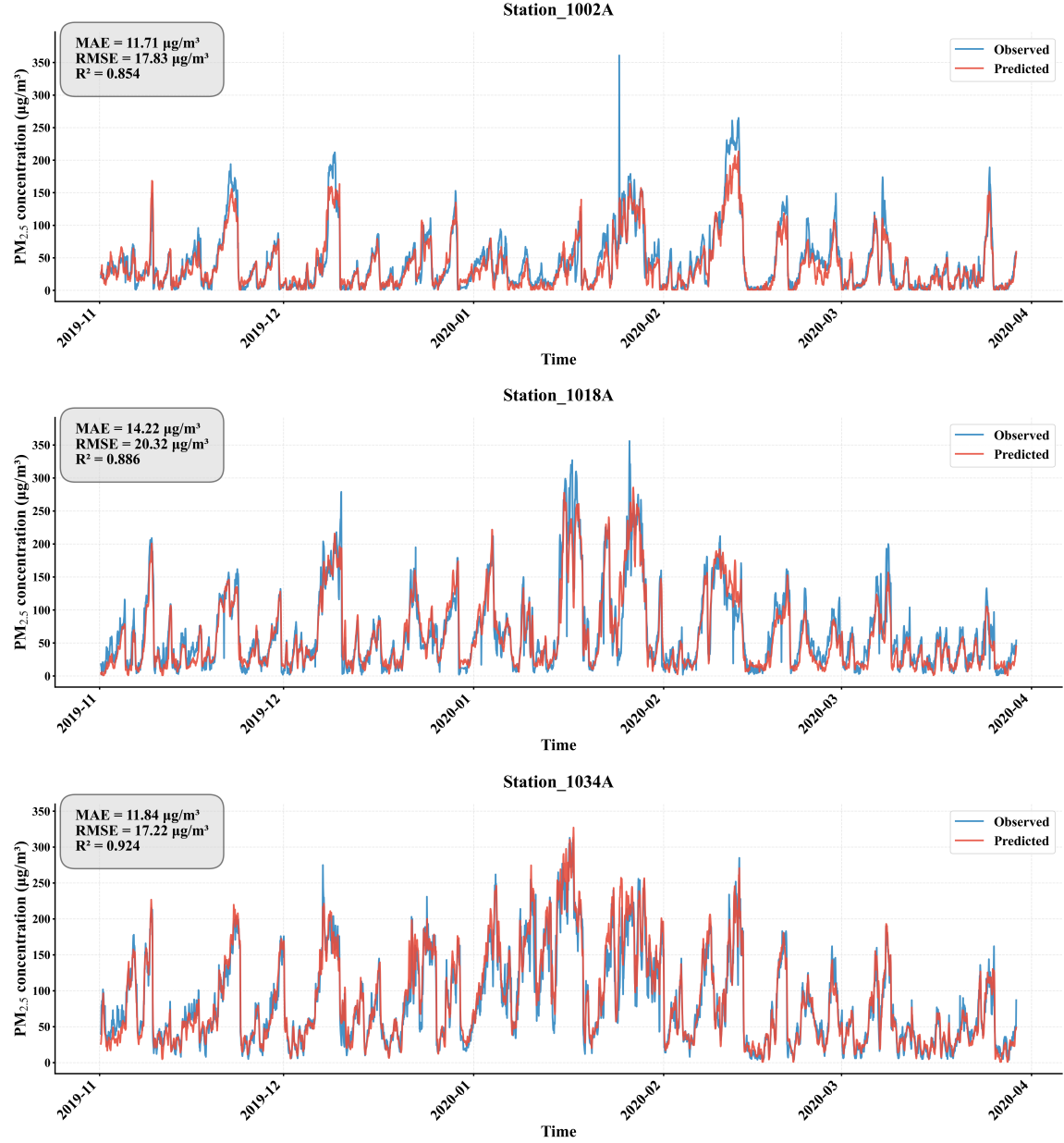


Figure 4: **Inference performance across core cities in the BTHSA region.** From top to bottom: **Beijing** (Station 1002A), **Tianjin** (Station 1018A), and **Shijiazhuang** (Station 1034A). Despite being completely unobserved during training, the model accurately tracks the ground truth (blue) with high correlation ($R^2 > 0.85$) across these geographically distinct urban centers, successfully capturing the complex winter pollution cycles.

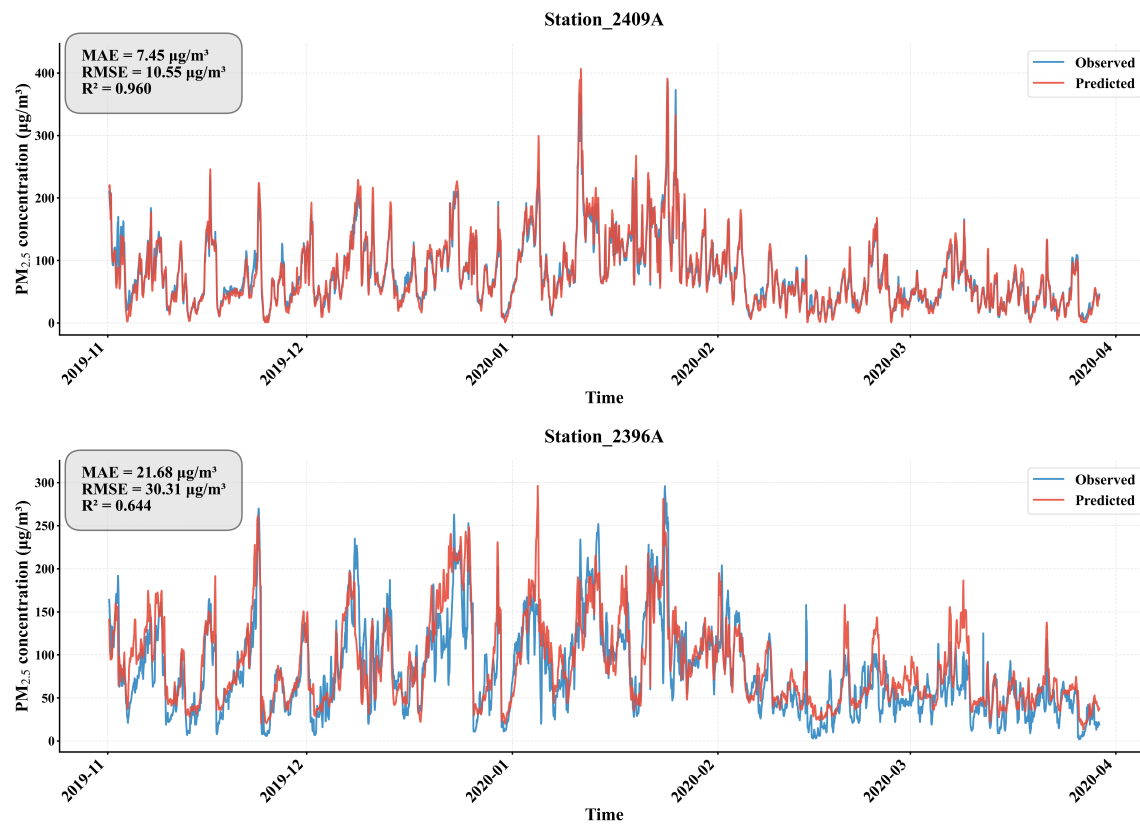


Figure 5: **Performance bounds analysis.** Top: The best-performing case (Station 2409A) shows near-perfect reconstruction due to dense neighbor support. Bottom: The worst-case scenario (Station 2396A) still maintains correct trend directionality, demonstrating stability under challenging topology.

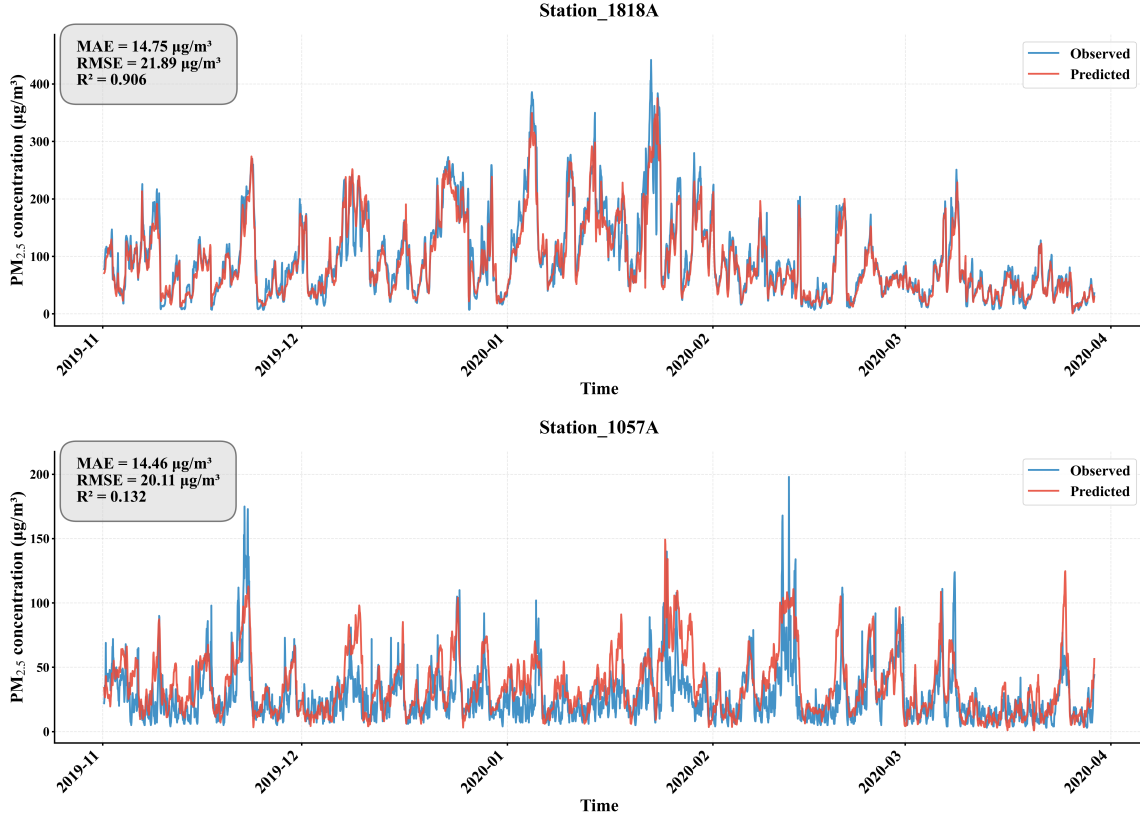


Figure 6: **Inference in spatially isolated regions.** Station 1057A (bottom) represents an extreme challenge with minimal upstream sensors. While high-frequency details are lost (lower R^2), the physics-based kernels ensure the inferred baseline trend remains physically plausible, acting as a conservative filter against hallucination.

3.3. AOD-Guided Grid Inference: Mechanism Analysis

The ultimate objective of SPIN is to generate continuous, high-resolution PM_{2.5} maps (Grid Inference) that are both physically plausible and consistent with sparse ground observations. Figure 7 presents a forensic analysis of the model’s inference capabilities across four distinct scenarios.

It is crucial to clarify that in our framework, **AOD is not used as a direct input feature** during inference; rather, it serves as a spatial gradient constraint during training. Therefore, the AOD panels in Figure 7 act as an external reference to validate whether the model has successfully internalized the spatial structural patterns of pollution.

- **Scenario (a): Ideal Consistency (Structure Transfer).** In this baseline

scenario, the satellite AOD field shows a clear pollution plume that is structurally consistent with the sparse ground measurements. SPIN successfully reproduces this complex spatial pattern, generating smooth transitions between stations. This indicates that the *Masked AOD Spatial Gradient Loss* has effectively taught the model to bridge sparse points using realistic, satellite-derived spatial gradients when data sources align.

- **Scenario (b): Total Data Absence (Physics Fallback).** This case represents a "nighttime" or "heavy cloud" scenario where AOD data is completely unavailable (shown as a blank field). A key advantage of our framework is that it does not "crash" or output artifacts in the absence of satellite data. Instead, the model robustly falls back to its internal physics-informed kernels—specifically the *Advection* and *Diffusion* operators. Driven by meteorological fields and ground stations alone, SPIN generates a physically smooth and continuous pollution field, demonstrating its capability to function independently of satellite inputs.
- **Scenario (c): Structural Conflict (Robustness).** Here, the AOD data exhibits a spatial trend that conflicts with the high-fidelity ground observations (likely due to surface reflectance errors or aerosol type mismatch). The results show that SPIN is not misled by these potential satellite artifacts. By prioritizing the hard constraints from the ground truth (via $\mathcal{L}_{infer}$) and only using AOD for gradient guidance where appropriate, the model avoids "hallucinating" false hotspots, ensuring the inference remains anchored to reliable ground-level reality.
- **Scenario (d): Magnitude Bias (Calibration).** In this scenario, the AOD field correctly captures the spatial extent of the pollution but significantly overestimates the intensity (indicated by the purple-yellow scale mismatch). SPIN demonstrates a sophisticated decoupling capability: it extracts the valid *spatial shape* (gradient structure) from the AOD signal while simultaneously calibrating the *absolute magnitude* using the ground stations. This proves that our gradient-based loss formulation effectively filters out systematic biases in satellite retrieval, retaining only the useful structural information.

In summary, these cases confirm that SPIN provides a "best-of-both-worlds" solution. It leverages the fine-grained spatial structure of satellite data when available and consistent, yet seamlessly reverts to robust physical interpolation when satellite data is missing or unreliable.

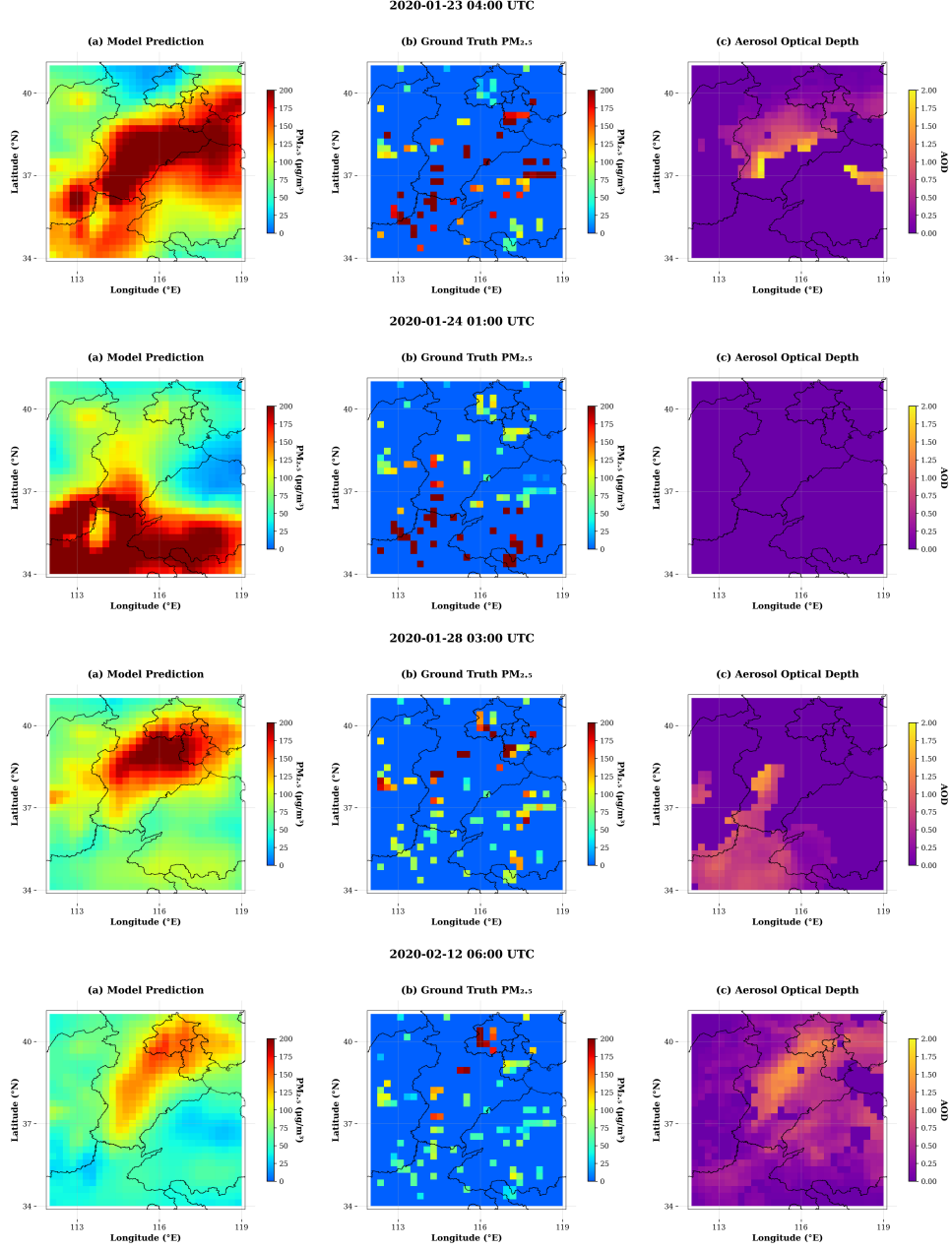


Figure 7: **Robustness analysis of Grid Inference across varying satellite data conditions.** Columns from left to right: Model Prediction, Ground Truth Stations, and Reference Satellite AOD. Note that AOD is used only for training constraints, not as a direct input. (a) **Ideal Case:** When AOD and ground truth are consistent, the model accurately reconstructs the plume structure. (b) **Missing AOD:** In the complete absence of satellite observations (e.g., nighttime or heavy cloud cover), the model robustly generates a smooth field driven by physical diffusion and advection kernels. (c) **Conflicting Data:** The model resists artifacts from misleading AOD patterns by anchoring to ground truth. (d) **Biased AOD:** The model adopts the correct spatial shape from AOD but corrects its magnitude bias.

4. Discussion

The development of SPIN represents a fundamental paradigm shift in urban air quality analytics—transitioning from *point-based temporal forecasting* to *field-based spatiotemporal inference*. By synergizing deep representation learning with atmospheric physical laws, this study **effectively mitigates** the "blind spot" dilemma that has long plagued environmental monitoring networks.

4.1. Interpretation of Physical-Data Synergy

The superior performance of SPIN, particularly in the challenging winter season (MAE reduced by 25.2% vs. SOTA), provides empirical validation for the "Physics-Embedded" design philosophy.

- **Decoupling Transport Mechanisms:** Unlike traditional GNNs that rely on static geometric distances, SPIN explicitly disentangles pollutant transport into *isotropic diffusion* and *anisotropic advection*. This allows the model to capture the dominant wind-driven transmission corridors in the BTHSA region—a critical feature for modeling winter haze episodes driven by monsoon winds.
- **Robustness via Structural Constraints:** A critical insight from this work is the advantage of using satellite data as a *loss constraint* rather than an *input feature*. Our analysis (Figure 7) demonstrates that this strategy effectively decouples the model’s inference capability from the availability of real-time satellite data. By learning the "spatial gradient patterns" during training, the model internalizes the structural logic of pollution distribution. This enables a **"physics fallback" mechanism**: when satellite "eyes" are blinded by clouds or darkness, the model seamlessly reverts to physics-driven extrapolation based on meteorology and ground observations, ensuring continuous operation.

4.2. Implications for Urban Environmental Management

The inductive inference capability of SPIN **enables more effective and innovative applications** for smart cities:

- **Virtual Sensing Network:** The model effectively functions as a cost-efficient "virtual sensor," providing reliable air quality estimates for unmonitored suburbs and rural areas. This capability is crucial for identifying hidden pollution hotspots and assessing environmental justice issues where low-income communities often lack monitoring infrastructure.

- **Real-Time Decision Support:** Compared to chemical transport models (CTMs) like CMAQ, which require heavy computational resources and hours of runtime, SPIN can generate regional inference maps in seconds. This efficiency allows policymakers to perform rapid assessment and emergency response during pollution events.

4.3. Limitations and Future Prospects

Despite these advancements, challenges remain. First, as seen in the extreme case of Station 1057A (Figure 6), the model’s ability to capture high-frequency fluctuations degrades in regions with extremely sparse upstream neighbors. While the physical kernels ensure a correct baseline trend, capturing rapid local variations still requires minimal data support. Second, the model’s reliance on reanalysis meteorology (ERA5) implies that errors in wind field data could propagate into the inference results, highlighting the need for high-precision local meteorological inputs. Future research could explore: (1) integrating real-time emission proxies (e.g., traffic flow) to enhance local responsiveness; and (2) extending the physics-guided framework to reactive pollutants like Ozone (O_3) by incorporating chemical reaction modules.

5. Conclusion

To address the practical challenge of **spatiotemporal inference** in unmonitored urban areas, this study proposes SPIN, a physics-guided inductive graph neural network. We successfully reconstruct high-resolution $PM_{2.5}$ fields from sparse ground observations by innovating in three dimensions:

1. **Architecture:** We embed atmospheric fluid dynamics into the deep learning structure via parallel Advection and Diffusion kernels, ensuring the model respects physical transport laws rather than merely fitting statistical correlations.
2. **Strategy:** We adopt a dynamic node-masking training strategy that empowers the model with inductive reasoning, enabling generalization to arbitrary unobserved locations.
3. **Data Fusion:** We pioneer a *Masked AOD Spatial Gradient Loss*, transforming satellite data from an error-prone input into a robust structural constraint. This novel approach effectively solves the pervasive "missing data" problem caused by cloud cover, achieving all-weather inference capability.

Extensive validation in Beijing-Tianjin-Hebei and Surrounding Areas demonstrates that SPIN establishes a new state-of-the-art, reducing inference error by over 25% compared to leading baselines and generating physically plausible pollution

maps even under total satellite data loss. This work provides a scalable, low-cost, and robust solution for next-generation environmental sensing, paving the way for data-driven, precision-guided sustainable urban governance.

Acknowledgments

We extend our gratitude to the Ministry of Ecology and Environment of the People’s Republic of China, the Japan Meteorological Agency (JMA), and the European Centre for Medium-Range Weather Forecasts (ECMWF) for providing the essential data supporting this study.

This work was supported by the National Natural Science Foundation of China (Grant Nos. 42450183, 12275020, 12135003, 12205025, and 42461144209) and the Ministry of Science and Technology of China (Grant No. 2023YFE0109000). Jingfang Fan is supported by the Fundamental Research Funds for the Central Universities. Shuo Wang acknowledges financial support from the China Scholarship Council (CSC) under Grant No. 202106040117. Mengfan Teng acknowledges support from the Jiangxi University of Science and Technology High-level Talent Project (Grant No. jxust-66) and the Jiangxi Province Youth Science Foundation (Grant No. 20252BAC200051).

References

- Bai, S., Kolter, J.Z., Koltun, V., 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271 .
- Burnett, R., Chen, H., Szyszkowicz, M., Fann, N., Hubbell, B., Pope III, C.A., Apte, J.S., Brauer, M., Cohen, A., Weichenthal, S., et al., 2018. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proceedings of the National Academy of Sciences* 115, 9592–9597.
- Chai, G., He, H., Sha, Y., Zhai, G., Zong, S., 2019. Effect of pm_{2.5} on daily outpatient visits for respiratory diseases in lanzhou, china. *Science of the total Environment* 649, 1563–1572.
- Chen, Z., Cai, J., Gao, B., Xu, B., Dai, S., He, B., Xie, X., 2017. Detecting the causality influence of individual meteorological factors on local pm_{2.5} concentration in the jing-jin-ji region. *Scientific Reports* 7, 40735.

- CSES, 2024. Technical guideline on city-level air pollutant emission inventory development. Standard T/CSES 144—2024 <https://www.ttbz.org.cn/StandardManage/Detail/108461/> by the Chinese Society for Environmental Sciences.
- Geng, G., Liu, Y., Liu, Y., Liu, S., Cheng, J., Yan, L., Wu, N., Hu, H., Tong, D., Zheng, B., et al., 2024. Efficacy of china’s clean air actions to tackle pm2.5 pollution between 2013 and 2020. *Nature Geoscience* , 1–8.
- Geng, G., Xiao, Q., Liu, S., Liu, X., Cheng, J., Zheng, Y., Xue, T., Tong, D., Zheng, B., Peng, Y., et al., 2021. Tracking air pollution in china: near real-time pm2. 5 retrievals from multisource data fusion. *Environmental Science & Technology* 55, 12106–12115.
- Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S., Yang, L., 2021. Physics-informed machine learning. *Nature Reviews Physics* 3, 422–440.
- Keller, J.P., Olives, C., Kim, S.Y., Sheppard, L., Sampson, P.D., Szpiro, A.A., Oron, A.P., Lindström, J., Vedal, S., Kaufman, J.D., 2015. A unified spatiotemporal modeling approach for predicting concentrations of multiple air pollutants in the multi-ethnic study of atherosclerosis and air pollution. *Environmental health perspectives* 123, 301–309.
- Levy, R.C., Mattoo, S., Munchak, L., Remer, L., Sayer, A., Patadia, F., Hsu, N., 2013. The collection 6 modis aerosol products over land and ocean. *Atmospheric Measurement Techniques* 6, 2989–3034.
- Li, M., Liu, H., Geng, G., Hong, C., Liu, F., Song, Y., Tong, D., Zheng, B., Cui, H., Man, H., et al., 2017. Anthropogenic emission inventories in china: a review. *National Science Review* 4, 834–866.
- Li, Y., Yu, R., Shahabi, C., Liu, Y., 2018. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting, in: *International Conference on Learning Representations*.
- Lu, J., Zhang, Y., Chen, M., Wang, L., Zhao, S., Pu, X., Chen, X., 2021. Estimation of monthly 1 km resolution pm2. 5 concentrations using a random forest model over “2+26” cities, china. *Urban Climate* 35, 100734.
- Ma, J., Cui, Z., Wang, B., Wang, P., Zhou, Z., Zhao, Z., Wang, Y., 2025. Causal learning meet covariates: Empowering lightweight and effective nationwide air quality forecasting, in: *International Joint Conference on Artificial Intelligence*.

- Qi, Y., Li, Q., Karimian, H., Liu, D., 2019. A hybrid model for spatiotemporal forecasting of pm2.5 based on graph convolutional neural network and long short-term memory. *Science of the Total Environment* 664, 1–10.
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, F., 2019. Deep learning and process understanding for data-driven earth system science. *Nature* 566, 195–204.
- Southerland, V.A., Brauer, M., Mohegh, A., Hammer, M.S., van Donkelaar, A., Martin, R.V., Apte, J.S., Anenberg, S.C., 2022. Global urban temporal trends in fine particulate matter (pm2.5) and attributable health burdens: estimates from global datasets. *The Lancet Planetary Health* 6, e139–e146.
- Teng, M., Li, S., Xing, J., Fan, C., Yang, J., Wang, S., Song, G., Ding, Y., Dong, J., Wang, S., 2023. 72-hour real-time forecasting of ambient pm2. 5 by hybrid graph deep neural network with aggregated neighborhood spatiotemporal information. *Environment International* 176, 107971.
- Teng, M., Li, S., Yang, J., Chen, J., Fan, C., Ding, Y., 2024. A new hybrid deep neural network for multiple sites pm2. 5 forecasting. *Journal of Cleaner Production* 473, 143542.
- Van Donkelaar, A., Martin, R.V., Brauer, M., Hsu, N.C., Kahn, R.A., Levy, R.C., Lyapustin, A., Sayer, A.M., Winker, D.M., 2016. Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. *Environmental science & technology* 50, 3762–3772.
- Wang, C., Zhu, Y., Zang, T., Liu, H., Yu, J., 2021. Modeling inter-station relationships with attentive temporal graph convolutional network for air quality prediction, in: *ACM WSDM*, pp. 616–634.
- Wang, S., Cheng, Y., Meng, Q., Saukh, O., Zhang, J., Fan, J., Zhang, Y., Yuan, X., Thiele, L., 2025. Pcdcnet: A surrogate model for air quality forecasting with physical-chemical dynamics and constraints. *arXiv preprint arXiv:2505.19842* .
- Wang, S., Li, Y., Zhang, J., Meng, Q., Meng, L., Gao, F., 2020. Pm2. 5-gnn: A domain knowledge enhanced graph neural network for pm2. 5 forecasting, in: *Proceedings of the 28th international conference on advances in geographic information systems*, pp. 163–166.

- Wei, J., Li, Z., Lyapustin, A., Sun, L., Peng, Y., Xue, W., Su, T., Cribb, M., 2021a. Reconstructing 1-km-resolution high-quality pm_{2.5} data records from 2000 to 2018 in china: spatiotemporal variations and policy implications. *Remote Sensing of Environment* 252, 112136.
- Wei, J., Li, Z., Lyapustin, A., Sun, L., Peng, Y., Xue, W., Su, T., Cribb, M., 2023. First close insight into global daily gapless 1 km pm_{2.5} pollution, variability, and health impact. *Nature Communications* 14, 8349.
- Wei, J., Li, Z., Pinker, R.T., Wang, J., Sun, L., Xue, W., Li, R., Cribb, M., 2021b. Himawari-8-derived diurnal variations in ground-level pm_{2.5} pollution across china using the fast space-time light gradient boosting machine (lightgbm). *Atmospheric Chemistry and Physics* 21, 7863–7880.
- Woody, M., Wong, H.W., West, J., Arunachalam, S., 2016. Multiscale predictions of aviation-attributable pm_{2.5} for us airports modeled using cmaq with plume-in-grid and an aircraft-specific 1-d emission model. *Atmospheric Environment* 147, 384–394.
- Wu, Y., Zhuang, D., Labbe, A., Sun, L., 2021. Inductive graph neural networks for spatiotemporal kriging. *Proceedings of the AAAI Conference on Artificial Intelligence* 35, 4478–4485.
- Xi, Y., Richardson, D.B., Kshirsagar, A.V., Wade, T.J., Flythe, J.E., Whitsel, E.A., Peterson, G.C., Wyatt, L.H., Rappold, A.G., 2022. Effects of short-term ambient pm_{2.5} exposure on cardiovascular disease incidence and mortality among us hemodialysis patients: a retrospective cohort study. *Environmental Health* 21, 33.
- Xiao, Q., Wang, Y., Chang, H.H., Meng, X., Geng, G., Lyapustin, A., Liu, Y., 2017. Full-coverage high-resolution daily pm_{2.5} estimation using maiac aod in the yangtze river delta of china. *Remote Sensing of Environment* 199, 437–446.
- Xiao, X., Jin, Z., Wang, S., Xu, J., Peng, Z., Wang, R., Shao, W., Hui, Y., 2022. A dual-path dynamic directed graph convolutional network for air quality prediction. *Science of The Total Environment* 827, 154298.
- Zhang, Z., Wu, W., Fan, M., Tao, M., Wei, J., Jin, J., Tan, Y., Wang, Q., 2019. Validation of himawari-8 aerosol optical depth retrievals over china. *Atmospheric environment* 199, 32–44.