AirDDE: Multifactor Neural Delay Differential Equations for Air Quality Forecasting

Binqing Wu [1,2,*], **Zongjiang Shang** [1,2,*], **Shiyu Liu** [3], **Jianlong Huang** [1,2], **Jiahui Xu** [1,2], **Ling Chen** [1,2,†]

[1] State Key Laboratory of Blockchain and Data Security, Zhejiang University
[2] College of Computer Science and Technology, Zhejiang University
[3] State Key Laboratory of Clean Energy Utilization, Zhejiang University
{binqingwu, zongjiangshang, 22251226, xujiahui19, lingchen}@cs.zju.edu.cn, shiyuliu@zju.edu.cn

Abstract

Accurate air quality forecasting is essential for public health and environmental sustainability, but remains challenging due to the complex pollutant dynamics. Existing deep learning methods often model pollutant dynamics as an instantaneous process, overlooking the intrinsic delays in pollutant propagation. Thus, we propose AirDDE, the first neural delay differential equation framework in this task that integrates delay modeling into a continuous-time pollutant evolution under physical guidance. Specifically, two novel components are introduced: (1) a memory-augmented attention module that retrieves globally and locally historical features, which can adaptively capture delay effects modulated by multifactor data; and (2) a physics-guided delay evolving function, grounded in the diffusion-advection equation, that models diffusion, delayed advection, and source/sink terms, which can capture delay-aware pollutant accumulation patterns with physical plausibility. Extensive experiments on three real-world datasets demonstrate that AirDDE achieves the state-of-the-art forecasting performance with an average MAE reduction of 8.79% over the best baselines.

Code — https://github.com/w2obin/airdde-aaai

Introduction

Rapid industrialization and urbanization over the past decades have exacerbated air pollution, making air quality a critical concern for public health and environmental sustainability (Azimi and Rahman 2024; Geng et al. 2025). This pressing issue underscores the importance of accurate air quality forecasting. Nevertheless, such forecasting is highly challenging owing to the complex pollutant dynamics (Vallero 2025; Bodnar et al. 2025).

Many methods have been proposed to tackle this problem in the past decades. Traditional methods, e.g., physical-chemical simulation models (Xie, Huang, and Wang 2005) and shallow machine learning techniques (Lee et al. 2012), often rely on simplified assumptions and handcrafted features, limiting their ability to capture the latent pollutant dynamics inherent in air quality data. This limitation has driven a paradigm shift toward deep learning methods, which offer stronger representational capacity (Qi et al. 2018; Wu et al. 2024). Early deep learning methods primarily utilize convolutional neural networks (CNNs) (Yan et al. 2021; Chen et al. 2023a) and recurrent neural networks (RNNs) (Xu and Yoneda 2019; Xu et al. 2021) to extract spatial and temporal features of air quality. Building on these foundations, recent works have adopted spatial temporal graph neural networks (STGNNs) (Chen et al. 2023b; Han et al. 2023) and attention mechanisms (Liang et al. 2023; Xia et al. 2025) to more effectively model spatial and temporal dependencies, enabling more comprehensive feature extraction. Despite their advancements, these methods formulate pollutant dynamics as a discrete-time process, where pollutant transitions occur only at fixed temporal intervals. This formulation limits their ability to capture the continuous-time dynamics of pollutants in the real world.

More recently, several methods have explored modeling pollutant dynamics as a continuous-time process by integrating multifactor field data (Wang et al. 2025) and leveraging neural ordinary differential equations (NODEs) (Hettige et al. 2024; Tian et al. 2025). However, these methods often adopt an instantaneous assumption, wherein the system's evolution depends only on the current state. This simplification neglects the transmission time, i.e., delay, during pollutant propagation. In fact, delays are pervasive and essential in real-world air quality systems (Cai, Alam, and Duong 2021; Wu et al. 2025). For example, pollutants emitted in one location may take several hours to be transported by wind before affecting air quality in downstream locations, introducing a non-negligible delay between emission and observable impact.

To model delays in a continuous-time process, a natural approach is to employ neural delay differential equations (NDDEs) (Zhu, Guo, and Lin 2021; Long et al. 2024). NDDEs are an extension of NODEs, which allow the system's evolution to depend not only on the current state but also on historical states. While theoretically appealing, existing NDDEs are constrained to modeling uniform delays, i.e., applying the same delay across all spatial locations, which limits their ability to capture location-specific delays.

Despite the heterogeneity of delays being particularly es-

[*] These authors contributed equally.
[†] Corresponding author.

sential and evident, it is non-trivial to capture them effectively. (1) Delays are modulated by multiple factors. The transport path and arrival time of pollutants are strongly influenced by varying meteorological and geographical factors, e.g., pollutant concentration levels, wind fields, and geographical distances. (2) Delay effects exhibit spatiotemporal accumulation. The pollutant concentration at a given location and time is not only determined by local pollutants but also by pollutants from other locations, each arriving with different delays. These cumulative effects are grounded in atmospheric dynamic processes, which cannot be captured by existing statistics-based or purely data-driven delay modeling techniques (Jiang et al. 2023a; Long et al. 2024).

To this end, we propose AirDDE, a multifactor neural delay differential equation framework for air quality forecasting. To the best of our knowledge, AirDDE is the first physics-guided work that integrates delay modeling into pollutant continuous-time evolution. The main contributions are summarized as follows:

- We introduce a memory-augmented attention (MAA) module. Given the delay-aware transport paths constructed from geographic distances and real-time wind fields, this module adopts a dual attention mechanism to retrieve globally and locally historical features. Such a design enables MAA to capture multifactor-modulated delay effects fully considering spatial heterogeneity.
- We introduce a physics-guided delay evolving (PDE) function. Guided by the diffusion-advection equation, this function models diffusion, delayed advection, and source/sink terms from multifactor features to capture continuous-time pollutant evolution. Such a design enables PDE to capture delay-aware pollutant accumulation patterns in a physically consistent manner.
- We compare AirDDE with 19 competitive baselines on 3 real-world datasets. The results demonstrate that AirDDE achieves the state-of-the-art (SOTA) performance with an average MAE reduction of 8.79% over the best baselines.

Related Work

Deep learning for air quality modeling. Early deep learning methods use CNNs and RNNs to extract spatial and temporal features. For example, DAL (Qi et al. 2018) and AirNet (Yu et al. 2020) combine CNNs with RNNs to forecast and calibrate air quality index (AQI) measurements, respectively. FAIRY (Chen et al. 2023a) utilizes SegNets to learn multiresolution spatial features for air quality estimation. Recent studies adopt STGNNs (Han et al. 2022; Liang et al. 2022) and attention mechanisms (Liang et al. 2023; Geng et al. 2025) for richer dependency modeling. In the STGNN paradigm, for example, PM2.5GNN (Wang et al. 2020), MasterGNN and MasterGNN+ (Han et al. 2021, 2023), and GAGNN (Chen et al. 2023b) enhance spatial and temporal dependency learning through GNN–GRU integration, adversarial training, and hierarchical graph design, respectively. Parallel to the STGNN-based advances, attention-based methods are also a powerful paradigm. For example, AirFormer (Liang et al. 2023) proposes a dartboard-style spatial attention and a causal temporal attention for long-term forecasting. AirRadar (Wang et al. 2025) introduces a masked feature reconstruction framework using spatial attention and temporal causal adjustment to infer air quality. Fuxi-Air (Geng et al. 2025) leverages a Transformer architecture for air pollution forecasting. Nevertheless, these methods model pollutant dynamics as a discrete-time process, overlooking their continuous-time nature.

To address this limitation, some recent methods model pollutant dynamics as a continuous-time process by integrating multifactor field data and leveraging NODEs. For example, STFNN (Feng et al. 2024) unifies field- and graph-based views for fine-grained continuous spatiotemporal inference. AirPhyNet (Hettige et al. 2024) embeds pollutant transport equations into a GNN–NODE framework, while AirDualODE (Tian et al. 2025) uses dual-branch NODEs combining data-driven and physics-informed components. However, these methods often adopt an instantaneous assumption, which largely ignore propagation delays. Although a few methods in general spatiotemporal tasks model delays using shared patterns (Jiang et al. 2023a) or cross-correlations (Long et al. 2024), they assume globally uniform delays and fail to capture delays shaped by heterogeneous conditions.

Thus, we propose AirDDE, a physics-guided framework that integrates delay modeling into continuous-time pollutant evolution. We introduce a memory-augmented attention and a delay-aware evolving function to model delay effects conditioned on location- and time-specific multifactor data. These designs enable AirDDE to make accurate forecasts with strong physical plausibility.

Preliminary

Task Formulation. Given historical air quality observations and auxiliary factors (e.g., meteorological and geographical variables), the goal of air quality forecasting is to predict future air quality for the next time steps, formulated as:

$$\hat{\boldsymbol{X}}^{T+1:T+H} = \mathcal{F}(\boldsymbol{X}^{1:T}, \boldsymbol{M}^{1:T}; \boldsymbol{\Theta}), \tag{1}$$

where $\boldsymbol{X}^{1:T} \in \mathbb{R}^{N \times T}$ and $\boldsymbol{M}^{1:T} \in \mathbb{R}^{N \times T}$ represent historical air quality and auxiliary factors from N locations over past T steps, respectively. $\hat{\boldsymbol{X}}^{T+1:T+H} \in \mathbb{R}^{N \times H}$ represents denotes the predicted air quality for the next H steps. $\mathcal{F}$ is the neural network. $\boldsymbol{\Theta}$ is the learnable parameters of $\mathcal{F}$.

Diffusion-Advection Equation. The diffusion–advection equation describes the transport of a substance (e.g., pollutants) in a fluid (Moreira et al. 1998), combining diffusion, advection, and source/sink effects, formulated as:

$$\frac{\partial u}{\partial t} + \vec{v} \cdot \nabla u = D \nabla^2 u + S, \tag{2}$$

where $\frac{\partial u}{\partial t}$ represents the partial derivative of u with respect to time. $\vec{v}$ is the velocity vector field. ∇u is the gradient of u. $\vec{v} \cdot \nabla u$ represents the advection term of u. D is the diffusion coefficient. $\nabla^2 u$ is the Laplacian of u. $D\nabla^2 u$ represents the diffusion term. S represents the source/sink term.

Recent studies (Hettige et al. 2024; Tian et al. 2025) have modified this equation as a diffusion-advection equation on

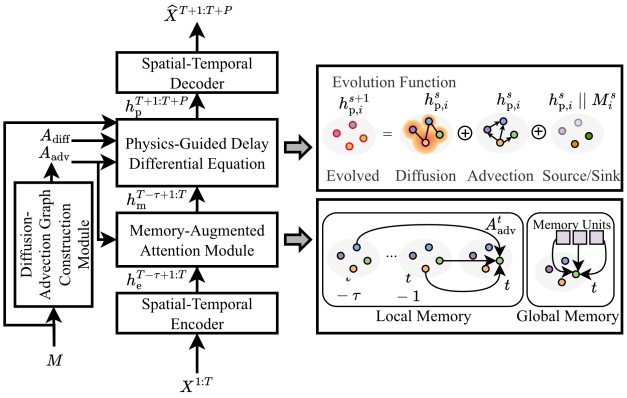

Figure 1: The architecture of AirDDE.

graphs. The Laplacian operators for advection and diffusion are approximated using the Chebyshev GNN (Defferrard, Bresson, and Vandergheynst 2016).

Neural Delay Differential Equations. NDDEs (Zhu, Guo, and Lin 2021; Long et al. 2024) extend NODEs (Chen et al. 2018; Li et al. 2025a) by modeling delayed dynamics, where the state evolution depends not only on the current state but also on historical states. The general form of an NDDE is formulated as:

$$\boldsymbol{F}: \frac{d\boldsymbol{h}^t}{dt} = f(\boldsymbol{h}^t, \boldsymbol{h}^{t-\tau}), \tag{3}$$

where $\boldsymbol{h}^t$ is the state at time step t, τ is a delay, and $f(\cdot)$ is a neural network-based evolution function. Compared to NODEs, NDDEs face more complex initial value and integration problems, but are more effective for delay effects.

NDDEs are trained by solving $\boldsymbol{F}$ forward and optimizing neural network parameters using automatic differentiation, often with the adjoint sensitivity method. Implementation tools, e.g., torchdiffeq (Kidger, Chen, and Lyons 2021), facilitate efficient simulation and backpropagation.

Methodology

Overview

The architecture of AirDDE is illustrated in Fig. 1. AirDDE models continuous-time pollutant propagation with delayed effects via a multifactor-enhanced NDDE framework. Considering the complex diffusion and advection effects in air quality, AirDDE constructs diffusion and advection graphs based on wind fields and distances, which can indicate pollutant transport paths. Given encoded inputs via an STGNN-based encoder and constructed graphs, AirDDE uses the MAA module to obtain initial states that retrieve global and local historical features, which can adaptively capture delay effects modulated by multifactor data. Then, AirDDE formulates the PDE function guided by the diffusion-advection equation, which can model delay-aware pollutant accumulation patterns with physical consistency. Given the historical states and evolution function, AirDDE adopts a DDE solver that maintains a history buffer to account for delayed states and employs numerical integration to get future air quality states. After that, these states are fed into a decoder to generate the final predictions.

Spatiotemporal Encoder

To capture spatiotemporal features, we follow an STGNN-based paradigm (Jiang et al. 2023b; Chen et al. 2024). Specifically, we derive the graph structure from learnable node embeddings and utilize GNNs to replace the MLPs in the GRU's gating mechanisms (Bai et al. 2020; Wu and Chen 2023). This design facilitates adaptive feature extraction by incorporating the underlying graph topology into the temporal updates. The process is formulated as:

$$\begin{aligned} \boldsymbol{A} &= \mathrm{SoftMax}(\mathrm{ReLU}(\boldsymbol{E}_1(\boldsymbol{E}_2)^{\mathrm{T}})), \\ \boldsymbol{h}_{\mathrm{e}}^t &= \text{GNN-GRU}(\boldsymbol{X}^t, \boldsymbol{h}_{\mathrm{e}}^{t-1}, \boldsymbol{A}), \end{aligned} \tag{4}$$

where $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ represents the underlying adjacency matrix, $\boldsymbol{E}_1, \boldsymbol{E}_2 \in \mathbb{R}^{N\times d}$ are two parameterized node embeddings with d dimensions. GNN-GRU$(\cdot)$ denotes a GRU variant where the original MLPs in the update and reset gates are replaced with GNNs. $\boldsymbol{X}^t$ and $\boldsymbol{h}_{\mathrm{e}}^{t-1}$ are the current inputs and historical hidden features, respectively.

Diffusion-Advection Graph Construction

Diffusion Graph. Diffusion is significant for pollutant transport, especially when wind is absent or negligible. Since the impact of diffusion is highly related to geographical proximity, we construct a diffusion graph $\boldsymbol{A}_{\text{diff}} \in \mathbb{R}^{N\times N}$ based on Haversine distances between locations, computed from their longitudes and latitudes. The resulting graph is then normalized using a Gaussian kernel (Li et al. 2018).

Advection Graph. Advection governs pollutant transport under windy conditions, carrying pollutants across space with inherent delays. Unlike previous works (Jiang et al. 2023a,b), which rely on globally averaged delays, we construct an advection graph at each time step to capture delays derived from wind fields and geographical distances. Specifically, a directed edge from location j at time step t_2 to location i at time step t_1 exists if the wind speed at j at t_2 enables an air parcel to reach i at t_1. The time lag $\tau = t_1 - t_2$ represents the travel time of the pollutant from j to i, serving as a looking-back window to identify dependencies. The construction is formulated as:

$$\boldsymbol{A}_{\text{adv},ij}^t = \begin{cases} 1, & \text{if } v_j^\tau \cos(\theta_j^\tau) \cdot \tau \geq d_{ij} \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where v_j^τ and θ_j^τ represent the wind speed and wind direction at location j at the previous τ time step, respectively. d_{ij} represents the distance between location i and j. $\boldsymbol{A}_{\text{adv}}^t \in \mathbb{R}^{N\times N}$ represents the delay-aware dependencies at time step t. Notably, $\boldsymbol{A}_{\text{adv}}^t$ offers greater adaptiveness than unified or implicit delay assumptions, as it explicitly models delays conditioned on location- and time-specific dynamics.

Memory-Augmented Attention Module

The MAA module is introduced to capture delay-aware initial states modulated by multifactor data. Pollution propagation shows dual-scale historical patterns: global background

trends (e.g., persistent high-PM2.5 regions) and local transient events (e.g., sudden AQI spikes from dust storms). To model these dual-scale delay effects under dynamic multifactor conditions, MAA employs a dual attention mechanism that retrieves both global and local historical features.

Global Memory Modeling. To capture global historical patterns, we introduce a set of learnable global memory units denoted as $\boldsymbol{M}_{\mathrm{g}} \in \mathbb{R}^{m \times d_{\mathrm{e}}}$, where m is the number of memory units and d_{e} is the hidden dimension. These memory units aim to memorize global historical patterns, which are randomly initialized and updated during training. We integrate the patterns to the current hidden features $\boldsymbol{h}_{\mathrm{e}}^t$ via attention (Vaswani et al. 2017), which is formulated as:

$$\boldsymbol{h}_{\mathrm{g}}^t = \mathrm{Attention}(\boldsymbol{h}_{\mathrm{e}}^t, \boldsymbol{M}_{\mathrm{g}}, \boldsymbol{M}_{\mathrm{g}}), \tag{6}$$

where $\boldsymbol{h}_{\mathrm{e}}^t$ and $\boldsymbol{M}_{\mathrm{g}}$ are linearly projected to obtain the query, keys, and values, respectively. $\mathrm{Attention}(\cdot)$ first computes similarity scores between query and keys to produce attention weights, which are then used to aggregate values. This allows $\boldsymbol{h}_{\mathrm{e}}^t$ to adaptively incorporate relevant global memories from $\boldsymbol{M}_{\mathrm{g}}$.

Local Memory Modeling. To capture local historical patterns, we leverage advection graphs to define dynamic neighborhoods based on wind-driven pollutant transport. Specifically, for location i at time step t, we define its neighbors $\mathcal{N}(i)^t$ based on the advection graph $\boldsymbol{A}_{\mathrm{adv}}^t$. We then attend over its own and its neighbors' historical features within a time lag τ, which is formulated as:

$$\boldsymbol{h}_{\mathrm{l},i}^t = \mathrm{MLP}(\mathrm{Attention}(\boldsymbol{h}_{\mathrm{e},i}^t, \boldsymbol{h}_{\mathrm{e},j\in\mathcal{N}(i)^t}^{t-\tau+1:t}, \boldsymbol{h}_{\mathrm{e},j\in\mathcal{N}(i)^t}^{t-\tau+1:t})), \tag{7}$$

where $\boldsymbol{h}_{\mathrm{e},i}^t$ is used as the query, while $\boldsymbol{h}_{\mathrm{e},j\in\mathcal{N}(i)^t}^{t-\tau+1:t} \in \mathbb{R}^{|\mathcal{N}(i)^t|\times\tau\times d}$ serves as keys and values. $\boldsymbol{h}_{\mathrm{l},i}^t \in \mathbb{R}^d$ is the output after an MLP layer.

We then concatenate the hidden features with those integrating global and local historical patterns and embed them for the initial states. The process is formulated as $\boldsymbol{h}_{\mathrm{m}}^t = \mathrm{MLP}(\mathrm{concat}(\boldsymbol{h}_{\mathrm{e}}^t, \boldsymbol{h}_{\mathrm{g}}^t, \boldsymbol{h}_{\mathrm{l}}^t))$, where $\boldsymbol{h}_{\mathrm{m}}^t \in \mathbb{R}^{N\times d}$ is the output features of the MAA module, which comprehensively capture dual-scale delay effects considering multifactor data.

Physics-Guided Delay Evolving Function

The PDE function is introduced to formulate the delay-aware evolution with physical consistency. Unlike prior works that assume conservative pollutant transport, we capture real-world non-conservative behavior, e.g., windborne inflows for sources and precipitation-driven removal for sinks. We model such hidden source/sink dynamics from multifactor features, as different factors encode complementary physical signals that together reveal the latent drivers of pollutant variation. Formally, the evolution function for the pollutant is formulated as:

$$\begin{aligned}\boldsymbol{F}: \frac{d\boldsymbol{h}^t}{dt} &= D \cdot \mathrm{GNN}_{\mathrm{diff}}(\boldsymbol{A}_{\mathrm{diff}}, \boldsymbol{h}^t) \\ &\quad + \mathrm{GNN}_{\mathrm{adv}}(\boldsymbol{A}_{\mathrm{adv}}^t, \boldsymbol{h}^{t-\tau}) + f(\boldsymbol{h}^t || \boldsymbol{M}),\end{aligned} \tag{8}$$

where $\mathrm{GNN}(\cdot)$ denotes the K-hop message passing mechanism that approximates K-order Chebyshev GNN, which can achieve a better efficiency and adaptability for dynamics (Hamilton 2020). D is the diffusion coefficient, which is empirically set to 0.1 (Hettige et al. 2024). f is a lightweight MLP-based network. $\boldsymbol{M}$ is the multifactor features.

Given the delay-aware initial states $\boldsymbol{h}_{\mathrm{m}}^T$ and historical states $\boldsymbol{h}_{\mathrm{m}}^{T-\tau:T-1}$, the hidden states of pollutant concentrations from $T+1$ to $T+P$ can be obtained by solving $\boldsymbol{F}$. The process is formulated as:

$$\boldsymbol{h}_{\mathrm{p}}^{T+1:T+P} = \mathrm{DDESolver}(\boldsymbol{F}, \boldsymbol{h}_{\mathrm{m}}^T, \boldsymbol{h}_{\mathrm{m}}^{T-\tau:T}, \boldsymbol{M}), \tag{9}$$

where the solver, following the existing works (Chen et al. 2018; Long et al. 2024), is implemented by torchdiffeq (Kidger, Chen, and Lyons 2021). The solver maintains an explicit memory set of past states during Fourth-order Runge-Kutta integration. By incorporating current and past states grounded in the diffusion-advection equation, the future states offer a more physically consistent representation of delay-aware air quality evolution.

Spatiotemporal Decoder

To consider spatial and temporal dependencies, we adopt the structure $\mathrm{GNN\text{-}GRU}(\cdot)$ (similar to the encoder) as the decoder. Given the adjacency matrix $\boldsymbol{A}$ learned from node embeddings (Eq.4) and the states $\boldsymbol{h}_{\mathrm{p}}^{T+1:T+P}$ derived by the solver, the decoder is formulated as:

$$\hat{\boldsymbol{X}}^t = \mathrm{MLP}(\mathrm{GNN\text{-}GRU}(\boldsymbol{h}_{\mathrm{p}}^t, \boldsymbol{h}_{\mathrm{p}}^{t-1}, \boldsymbol{A})) \tag{10}$$

where $t \in [T+1, T+P]$. $\hat{\boldsymbol{X}}^{T+1:T+P} \in \mathbb{R}^{N\times P}$ are the final predictions.

For the training loss, we adopt the Huber loss, which is commonly used in ODE-based methods (Fang et al. 2021; Chen et al. 2024; Long et al. 2024). Due to its robustness to outliers and smooth optimization behavior, the Huber loss is well-suited for modeling noisy systems, especially air quality dynamics, where measurement noise and extreme values are often significant. The loss is formulated as:

$$\mathcal{L} = \begin{cases} \frac{1}{2}(\boldsymbol{X} - \hat{\boldsymbol{X}})^2 & , |\boldsymbol{X} - \hat{\boldsymbol{X}}| \le \delta \\ \delta|\boldsymbol{X} - \hat{\boldsymbol{X}}| - \frac{1}{2}\delta^2 & , \text{ otherwise} \end{cases} \tag{11}$$

where δ is the threshold to change between delta-scaled L1 and L2 loss, which controls the sensitivity to outliers.

Dataset	# Factors	# Locations	Time Range	Granularity
KnowAir	18	184	1.1.2015-12.31.2018	3h
China-AQI	8	209	1.1.2017-4.30.2019	1h
US-PM	8	175	1.1.2020-12.31.2021	1h

Table 1: Dataset statistics.

Experiments

Experimental Setup

Datasets. We evaluate AirDDE on three real-world air quality datasets. KnowAir is provided by PM2.5GNN (Wang et al. 2020), including PM2.5 data and 17 meteorological

	Dataset	KnowAir			China-AQI			US-PM		
	Metric	MAE	RMSE	SMAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
STGNNs	DCRNN (2018)	24.02*	37.87*	0.53*	24.78	38.05	35.76	7.49	9.72	15.85
	STGCN (2018)	23.64*	32.48*	0.52*	23.87	37.29	35.03	6.99	8.81	14.72
	ASTGCN (2019)	19.92*	31.39*	0.44*	21.91	36.02	34.28	4.86	7.53	12.93
	MTGNN (2020)	18.92	30.34	0.41	21.56	35.80	34.65	4.63	7.03	12.85
	PM25GNN (2020)	19.32*	30.12*	0.43*	22.01	36.21	34.12	5.24	7.86	13.43
	GAGNN (2023)	20.71	32.88	0.42	19.54	33.37	32.95	4.32	6.53	12.43
	MegaCRN (2023)	18.77	29.45	0.42	18.93	32.41	32.23	3.85	5.41	11.72
	HimNet (2024)	20.98	33.00	0.44	20.72	33.78	34.08	5.28	7.79	13.28
Attentions	Corrformer (2023)	21.11	32.97	0.44	21.22	34.93	34.97	5.95	7.71	14.37
	AirFormer (2023)	19.17*	30.19*	0.43*	19.60	33.14	32.86	3.90	5.37	11.55
	PDFormer (2023)	19.06	30.66	0.41	19.07	32.76	32.45	3.81	5.36	11.67
	iTransformer (2024)	21.03	33.14	0.46	20.60	33.54	33.74	4.67	7.33	12.88
	STMFormer (2025)	19.57	31.12	0.44	20.04	33.21	32.52	4.22	6.05	12.69
NODEs	STGODE (2021)	21.40	33.47	0.45	20.53	33.46	33.83	4.56	6.71	13.01
	STG-NCDE (2022)	21.21	33.80	0.45	21.33	35.64	34.53	4.37	6.62	13.12
	STDDE (2024)	22.85	34.26	0.46	21.04	34.36	34.15	4.44	6.57	12.96
	SGODE (2024)	20.02	32.03	0.42	20.17	33.23	32.84	4.22	5.59	12.76
	AirPhyNet (2024)	21.31*	31.77*	0.47*	21.78	35.43	34.80	4.79	6.80	13.10
	AirDualODE (2025)	18.64*	29.37*	0.42*	18.89	32.26	32.06	3.98	5.41	11.86
Ours	AirDDE	**16.92**	**27.78**	**0.38**	**17.03**	**29.91**	**30.82**	**3.53**	**4.87**	**10.94**

Table 2: Results of AirDDE and baselines. The best results are bolded, and the second best results are underlined. The results with $*$ are cited from AirDualODE (Tian et al. 2025), while others are rerun using their official codes under multifactor settings.

factors from 184 cities across China. China-AQI and US-PM are provided by GAGNN (Chen et al. 2023b). China-AQI includes AQI data and 7 meteorological factors from 203 cities across China. US-PM includes PM2.5 data and 7 meteorological factors from 175 counties across the US. The detailed statistics of these datasets are summarized in Table 1. We follow the established preprocessing protocols from the original dataset studies.

Baselines. We compare AirDDE with 19 competitive baselines over 3 groups, including (1) **STGNNs**: DCRNN (Li et al. 2018), STGCN (Yu, Yin, and Zhu 2018), ASTGCN (Guo et al. 2019), MTGNN (Wu et al. 2020), PM25GNN (Wang et al. 2020), GAGNN (Chen et al. 2023b), MegaCRN (Jiang et al. 2023b), and HimNet (Dong et al. 2024); (2) **Attentions**: Crossformer (Zhang and Yan 2023), AirFormer (Liang et al. 2023), PDFormer (Jiang et al. 2023a), iTransformer (Liu et al. 2024), and STMFormer (Li et al. 2025b); and (3) **NODEs**: STGODE (Fang et al. 2021), STGNCDE (Choi et al. 2022), SGODE (Chen et al. 2024), STDDE (Long et al. 2024), AirPhyNet (Hettige et al. 2024), and AirDualODE (Tian et al. 2025).

Settings. We split the datasets following the original dataset studies (Wang et al. 2020; Chen et al. 2023b). KnowAir is divided chronologically for training, validation, and testing in a 2:1:1 ratio due to its ample four-year data span. China-AQI and US-PM are divided chronologically in a 7:1:2 ratio. All experiments are conducted on a single A100 GPU, employing the Adam optimizer with an initial learning rate of 0.005. We set the maximum number of epochs to 100 and employ an early stopping strategy with a tolerance of 10 epochs. For KnowAir, we set the batch size to 64, the input length to 24 (3-day), and the output length to 24 (3-day). For China-AQI and US-PM, we set the batch size to 32, the input length to 96 (4-day), and the output length to 24 (1-day). The number of global memory units is chosen from $\{8, 16, 32, 64\}$. The time lag is chosen from $\{1, 2, 3, 4\}$. We use the AutoML toolkit NNI (Microsoft 2021) and its built-in Bayesian optimizer to efficiently tune the hyperparameters.

Variant	MAA			PDE			AirDDE
	w/o MAA	w/o GM	w/o LM	w/o PDE	w/o SST	w/o ATT	
1st	15.45	15.13	15.37	15.97	15.70	15.00	14.53
2nd	20.18	18.47	17.80	19.93	18.86	19.82	17.16
3rd	21.65	19.57	18.54	21.47	19.79	21.33	18.01
AVG	19.16	17.80	17.44	19.39	18.18	18.78	16.92

Table 3: MAE results of the ablation study on KnowAir.

Overall Comparison

Table 2 summarizes the results of all methods. We can observe that: (1) AirDDE achieves the best performance in all cases, outperforming the second-best with MAE reduction of 9.23%, 9.85%, and 7.3%, on KnowAir, China-AQI, and US-PM, respectively. This improvement highlights the effectiveness of modeling continuous-time pollutant dynamics with delay effects. (2) AirDualODE and PDFormer exhibit competitive results, as they address continuous-time dynamics and unified delay effects, respectively. AirDDE outperforms them by considering delays during continuous-time evolution, yielding more accurate forecasts with improved physical fidelity. (3) AirDDE demonstrates the most significant improvement on China-AQI. Compared to KnowAir and US-PM, China-AQI features finer temporal granularity, higher pollution levels, and more locations, leading to the most complex pollutant dynamics. AirDDE effectively addresses this complexity, as it explicitly integrates transport paths based on distances and winds, achieving better adaptation to environments.

Ablation Study

MAA Module. We design three variants: (1) Removing the entire module (-w/o MAA) and using an MLP to encode in-

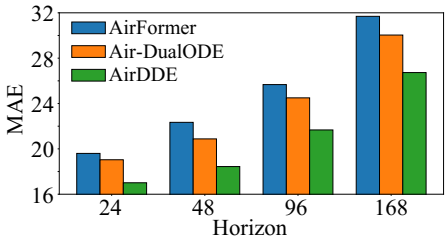

Figure 2: Results of the long-term study on China-AQI.

Dataset	Time lag				Number of global memory unit			
	0	1	2	3	8	16	32	64
China-AQI	18.54	17.03	19.87	21.05	19.66	18.49	17.03	18.78
US-PM	4.32	3.98	3.53	4.06	3.93	3.53	4.07	4.25

Table 4: Results of the hyperparameter study on China-AQI.

puts to get the initial state (Long et al. 2024). (2) Removing the global memory modeling (-w/o GM). (3) Removing the local memory modeling (-w/o LM). As shown in Table 3, AirDDE outperforms its variants -w/o MAA, -w/o GM, and -w/o LM, showing the contributions of memory augmentation and both global and local historical patterns to delay modeling. In addition, -w/o MAA causes a significant drop in 3rd-day forecasts, where historical pollution patterns are more influential, highlighting MAA's crucial role in long-term forecasting.

PDE Function. We design three variants: (1) Removing the entire module (-w/o PDE) and directly feeding encoder states to the decoder. (2) Removing the source/sink term (-w/o SST). (3) Removing the physics priors and replacing with attention-based evolving function (Tian et al. 2025) (-ATT). As shown in Table 3, AirDDE performs better than -w/o PDE, -w/o SST, and -ATT, showing the effectiveness of the delay-aware evolution, multifactor enhancement, and physics priors, respectively, for continuous-time modeling. In addition, the physics-guided variants, i.e., -SST and AirDDE, outperform purely data-driven variants, i.e., -w/o PDE and -ATT. This is because the physical priors offer structural guidance to capture delay effects and maintain consistency with real-world pollutant dynamics.

Long-Term Study

To evaluate the long-term forecasting ability of AirDDE, we fix the input length to $T = 96$ and extend the output horizon to $H = \{24, 48, 96, 168\}$. As shown in Fig. 2, AirDDE consistently achieves the best performance, with its advantage becoming more pronounced at longer forecasting horizons. These results demonstrate its superior long-term forecasting capability, highlighting the effectiveness of delay-aware continuous modeling with physical priors over extended horizons.

Hyperparameter Study

We evaluate the effect of key hyperparameters in AirDDE. To ensure fairness, all other hyperparameters are held fixed when varying a specific one.

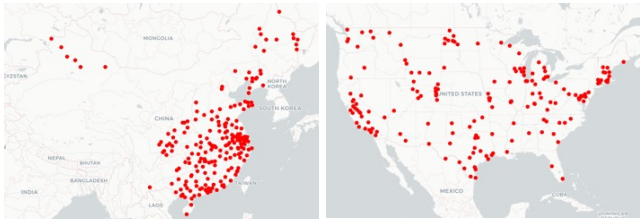

Figure 3: Station distribution of China-AQI (left) and US-PM (right).

Method	Conf.	GPU Memory (GB)	Training Time (Min/Epoch)	MAE
STGODE	KDD 2021	14.88	6.37	20.53
STG-NCDE	AAAI 2022	5.89	39.14	21.33
AirFormer	AAAI 2023	8.78	2.95	19.60
PDFormer	AAAI 2023	15.71	5.04	19.07
STDDE	WWW 2024	21.43	24.64	21.04
SGODE	AAAI 2024	12.22	11.06	20.17
AirPhyNet	ICLR 2024	14.31	4.78	21.78
AirDualODE	ICLR 2025	11.14	10.09	18.89
AirDDE	Ours	10.46	9.24	17.03

Table 5: Results of the efficiency study on China-AQI.

Time Lag τ. τ governs the construction of transport paths and directly influences the advection process. We vary τ from 0 to 3 with a step size of 1. As summarized in Table 4, the optimal values are $\tau = 1$ on China-AQI and $\tau = 2$ on US-PM. This is because a too-small τ fails to capture sufficient transport delays, whereas an overly large τ introduces outdated or less relevant historical states. In addition, the discrepancy of τ between the two datasets reflects dataset-specific pollutant dynamics. As illustrated in Fig. 3, the denser station distribution in China-AQI results in rapid and localized pollutant propagation, favoring shorter time lags. In contrast, the sparser station distribution in US-PM requires longer τ to model long-range transport delays.

Number of Global Memory Units m. m controls the model's capacity to store global historical patterns. We evaluate m in $\{8, 16, 32, 64\}$. As summarized in Table 4, the optimal values are $m = 32$ for China-AQI and $m = 16$ for US-PM, while both larger and smaller values degrade performance. This is because a too-small m fails to capture diverse patterns, whereas an overly large m introduces redundancy or noise, increasing the risk of overfitting. In addition, the discrepancy of m reflects the different characteristics of the two countries. Due to higher industrial activity and denser population centers, air quality patterns in China exhibit more complex pollution dynamics than those in the US, requiring greater memory capacity.

Efficiency Study

We compare the GPU memory usage, training time, and MAE of AirDDE against competitive baselines and NODEs on the China-AQI dataset. As shown in Table 5, we can observe that: (1) AirDDE achieves the best prediction performance with relatively reasonable computational cost. Although the second-best method, AirDualODE, also shows strong accuracy, its dual ODE solvers introduce heavy overhead and reduce efficiency. (2) Compared with condi-

Method	Original	Missing Rate			SNR		
		10%	30%	50%	80db	60db	40db
AirFormer	19.17	23.09	27.83	38.44	24.68	31.85	45.84
PDFormer	19.06	22.43	29.86	41.23	25.86	36.93	48.42
STDDE	22.85	26.11	32.54	40.78	28.75	35.92	49.34
AirPhyNet	21.31	25.20	31.30	37.79	27.82	33.60	46.42
AirDualODE	18.64	21.10	25.83	34.82	23.46	31.14	38.42
AirDDE	16.92	18.39	22.44	29.45	21.02	26.68	32.51
Impro.	9.23%	12.84%	13.12%	15.42%	14.06%	14.32%	15.38%

Table 6: MAE Results of the robustness study on KnowAir.

tioned NODEs, i.e., STG-NCDE and STDDE, AirDDE improves in efficiency and accuracy. This is mainly because AirDDE incorporates transport paths directly into the advection term, avoiding the expensive continuous-path encoding while still modeling delay-aware dependencies. (3) Compared with competitive baselines, i.e., AirFormer and PDFormer, AirDDE incurs longer training time due to historical state maintenance for delay modeling. However, it achieves significant MAE reductions, i.e., 13.11% over AirFormer and 10.70% over PDFormer. Moreover, AirDDE remains more memory-efficient than PDFormer, demonstrating a favorable efficiency–performance trade-off.

Robustness Study

In real-world scenarios, air quality data are often irregular due to sensor failures and noise. To evaluate the robustness of AirDDE, we compare it with competitive baselines on KnowAir. For missing data, following STNCDE (Choi et al. 2022), we randomly drop 10% to 50% of values for each sensor independently. For noisy data, following CrossGNN (Huang et al. 2023), we inject Gaussian white noise with varying intensities, progressively decreasing the signal-to-noise ratio (SNR) from 80 dB to 40 dB. As shown in Table 6, AirDDE consistently outperforms all baselines under both settings, with MAE improvements increasing as data quality deteriorates, i.e., from 9.23% to 15.42% with higher missing rates, and from 14.06% to 15.38% under stronger noise. This trend highlights AirDDE's superior robustness, particularly under more challenging conditions. This advance stems from its global memory and physics-guided evolution, which enables effective recovery and denoising via global historical patterns while maintaining physical consistency with real-world pollutant dynamics.

Case Study

City-Wise Advection with Delay Effects. Fig. 4 (a) illustrates the AQI of Taicang, Shanghai, and Ningbo from Nov.14 10:00, 2018 to Nov.19 10:00, 2018 in the test dataset of China-AQI. As shown in red cycles, since the average wind direction in Shanghai and Taicang during Nov.15 and Nov. 18 is north and northeast, the pollutant is driven to the downstream city, i.e, Ningbo, resulting in AQI peaks with lags in Ningbo. As shown in the predicted results in Fig. 4 (b), compared to SOTA methods, AirDDE can effectively capture these peaks with lags, as AirDDE explicitly considers this wind-driven delays in the advection process.

Region-Wise Advection with Delay Effects. Figure 5 illustrates the PM2.5 concentration of a region in Shanxi

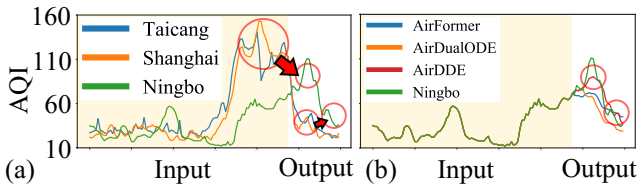

Figure 4: Case of city-wise advection with delay effects.

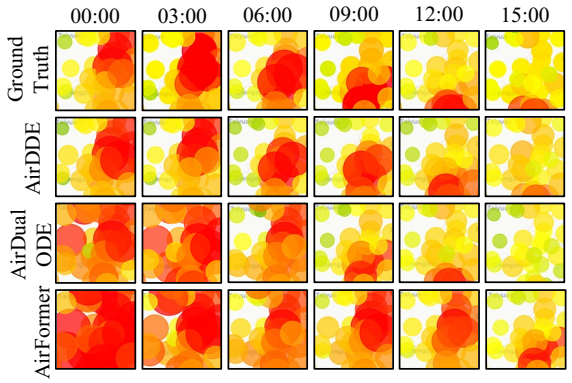

Figure 5: Case of region-wise advection with delay effects.

Province from 00:00 to 15:00 on December 3, 2018, using data from the KnowAir test dataset. Each circle represents the PM2.5 concentration at a location, with a larger radius and deeper red color indicating higher concentration levels. During this period, wind-driven advection transports pollutants from northeast to southwest. Downwind areas receive pollutants from upwind regions after a certain time lag, reflecting the inherent delay effect. Compared to predicted results from SOTA methods, AirDDE effectively captures the transport path of pollutants, showcasing its ability to represent delay-aware pollutant dynamics.

Conclusions and Future Work

In this work, we introduce AirDDE, the first neural delay differential equation framework that captures delay effects during continuous-time pollutant evolution. We introduce two novel blocks guided by physics priors: the MAA module to capture delay effects modulated by historical multifactor data and the PDE function to capture delay-aware pollutant accumulation patterns. Extensive experimental results demonstrate that AirDDE outperforms 19 competitive baselines, reducing average MAE by 8.79% over the best baselines, while demonstrating its practical strength in long-term forecasting and robustness.

Despite such advantages, several open directions remain for more comprehensive delay modeling. We highlight three key areas: (1) improving the efficiency of delay state maintenance; (2) incorporating uncertainty into delay estimation, given the stochastic nature of wind fields; and (3) modeling compound delays to capture complex transport paths involving intermediate regions.

References

Azimi, M. N.; and Rahman, M. M. 2024. Unveiling the health consequences of air pollution in the world's most polluted nations. *Scientific Reports*, 14(1): 9856.

Bai, L.; Yao, L.; Li, C.; Wang, X.; and Wang, C. 2020. Adaptive graph convolutional recurrent network for traffic forecasting. *Advances in neural information processing systems*, 33: 17804–17815.

Bodnar, C.; Bruinsma, W. P.; Lucic, A.; Stanley, M.; Allen, A.; Brandstetter, J.; Garvan, P.; Riechert, M.; Weyn, J. A.; Dong, H.; et al. 2025. A foundation model for the Earth system. *Nature*, 641(8065): 1180–1187.

Cai, Q.; Alam, S.; and Duong, V. N. 2021. A spatial–temporal network perspective for the propagation dynamics of air traffic delays. *Engineering*, 7(4): 452–464.

Chen, L.; Long, H.; Xu, J.; Wu, B.; Zhou, H.; Tang, X.; and Peng, L. 2023a. Deep citywide multisource data fusion-based air quality estimation. *IEEE Transactions on Cybernetics*, 54(1): 111–122.

Chen, L.; Wu, K.; Lou, J.; and Liu, J. 2024. Signed graph neural ordinary differential equation for modeling continuous-time dynamics. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 8292–8301.

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

Chen, R. T.; Rubanova, Y.; Bettencourt, J.; and Duvenaud, D. K. 2018. Neural ordinary differential equations. *Advances in Neural Information Processing Systems*, 31: 6572–6583.

Choi, J.; Choi, H.; Hwang, J.; and Park, N. 2022. Graph neural controlled differential equations for traffic forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 6367–6374.

Defferrard, M.; Bresson, X.; and Vandergheynst, P. 2016. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in Neural Information Processing Systems*, 29: 3844–3852.

Dong, Z.; Jiang, R.; Gao, H.; Liu, H.; Deng, J.; Wen, Q.; and Song, X. 2024. Heterogeneity-informed meta-parameter learning for spatiotemporal time series forecasting. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 631–641.

Fang, Z.; Long, Q.; Song, G.; and Xie, K. 2021. Spatial-temporal graph ode networks for traffic flow forecasting. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 364–373.

Feng, Y.; Wang, Q.; Xia, Y.; Huang, J.; Zhong, S.; and Liang, Y. 2024. Spatio-temporal field neural networks for air quality inference. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 7260–7268.

Geng, Z.; Fan, X.; Lu, X.; Zhang, Y.; Yu, G.; Huang, C.; Wang, Q.; Li, Y.; Ma, W.; Yu, Q.; et al. 2025. FuXi-Air: Urban air quality forecasting based on emission-meteorology-pollutant multimodal machine learning. *arXiv preprint arXiv:2506.07616*.

Guo, S.; Lin, Y.; Feng, N.; Song, C.; and Wan, H. 2019. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 922–929.

Hamilton, W. L. 2020. *Graph representation learning*. Morgan & Claypool Publishers.

Han, J.; Liu, H.; Xiong, H.; and Yang, J. 2022. Semi-supervised air quality forecasting via self-supervised hierarchical graph neural network. *IEEE Transactions on Knowledge and Data Engineering*, 35(5): 5230–5243.

Han, J.; Liu, H.; Zhu, H.; and Xiong, H. 2023. Kill two birds with one stone: A multi-view multi-adversarial learning approach for joint air quality and weather prediction. *IEEE Transactions on Knowledge and Data Engineering*, 35(11): 11515–11528.

Han, J.; Liu, H.; Zhu, H.; Xiong, H.; and Dou, D. 2021. Joint air quality and weather prediction based on multi-adversarial spatiotemporal networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 4081–4089.

Hettige, K. H.; Ji, J.; Xiang, S.; Long, C.; Cong, G.; and Wang, J. 2024. Airphynet: Harnessing physics-guided neural networks for air quality prediction. *International Conference on Learning Representations*.

Huang, Q.; Shen, L.; Zhang, R.; Ding, S.; Wang, B.; Zhou, Z.; and Wang, Y. 2023. Crossgnn: Confronting noisy multivariate time series via cross interaction refinement. *Advances in Neural Information Processing Systems*, 36: 46885–46902.

Jiang, J.; Han, C.; Zhao, W. X.; and Wang, J. 2023a. Pdformer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 4365–4373.

Jiang, R.; Wang, Z.; Yong, J.; Jeph, P.; Chen, Q.; Kobayashi, Y.; Song, X.; Fukushima, S.; and Suzumura, T. 2023b. Spatio-temporal meta-graph learning for traffic forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 8078–8086.

Kidger, P.; Chen, R. T. Q.; and Lyons, T. J. 2021. Hey, that's not an ODE: Faster ODE Adjoints via Seminorms. *International Conference on Machine Learning*.

Lee, M. H.; Abd Rahman, N. H.; Latif, M. T.; Nor, M. E.; Kamisan, N. A. B.; et al. 2012. Seasonal ARIMA for forecasting air pollution index: A case study. *American Journal of Applied Sciences*, 9(4): 570–578.

Li, X.; Zhao, C.; Zhang, X.; and Duan, X. 2025a. Symbolic neural ordinary differential equations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 18511–18519.

Li, Y.; Yu, R.; Shahabi, C.; and Liu, Y. 2018. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *International Conference on Learning Representations*.

Li, Z.; Hu, Z.; Han, P.; Gu, Y.; and Cai, S. 2025b. SSL-STMFormer self-supervised learning spatio-temporal entanglement Transformer for traffic flow prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 12130–12138.

Liang, Y.; Ouyang, K.; Wang, Y.; Pan, Z.; Yin, Y.; Chen, H.; Zhang, J.; Zheng, Y.; Rosenblum, D. S.; and Zimmermann, R. 2022. Mixed-order relation-aware recurrent neural networks for spatio-temporal forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 35(9): 9254–9268.

Liang, Y.; Xia, Y.; Ke, S.; Wang, Y.; Wen, Q.; Zhang, J.; Zheng, Y.; and Zimmermann, R. 2023. Airformer: Predicting nationwide air quality in china with transformers. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 14329–14337.

Liu, Y.; Hu, T.; Zhang, H.; Wu, H.; Wang, S.; Ma, L.; and Long, M. 2024. iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. In *International Conference on Learning Representations*.

Long, Q.; Fang, Z.; Fang, C.; Chen, C.; Wang, P.; and Zhou, Y. 2024. Unveiling delay effects in traffic forecasting: a perspective from spatial-temporal delay differential equations. In *Proceedings of the ACM Web Conference*, 1035–1044.

Microsoft. 2021. Neural Network Intelligence.

Moreira, D.; Rizza, U.; Degrazia, G. A.; Mangia, C.; Tirabassi, T.; et al. 1998. An analytical air pollution model: development and evaluation. *Contributions to Atmospheric Physics*, 71(3): 315–320.

Qi, Z.; Wang, T.; Song, G.; Hu, W.; Li, X.; and Zhang, Z. 2018. Deep air learning: Interpolation, prediction, and feature analysis of fine-grained air quality. *IEEE Transactions on Knowledge and Data Engineering*, 30(12): 2285–2297.

Tian, J.; Liang, Y.; Xu, R.; Chen, P.; Guo, C.; Zhou, A.; Pan, L.; Rao, Z.; and Yang, B. 2025. Air quality prediction with physics-informed dual neural odes in open systems. *International Conference on Learning Representations*.

Vallero, D. A. 2025. *Fundamentals of air pollution*. Academic press.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in Neural Information Processing Systems*, 30: 6000–6010.

Wang, Q.; Xia, Y.; Zhong, S.; Li, W.; Wu, Y.; Cheng, S.; Zhang, J.; Zheng, Y.; and Liang, Y. 2025. AirRadar: Inferring nationwide air quality in China with deep neural networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(27): 28467–28475.

Wang, S.; Li, Y.; Zhang, J.; Meng, Q.; Meng, L.; and Gao, F. 2020. Pm2. 5-gnn: A domain knowledge enhanced graph neural network for pm2. 5 forecasting. In *Proceedings of the International Conference on Advances in Geographic Information Systems*, 163–166.

Wu, B.; and Chen, L. 2023. DSTCGCN: Learning dynamic spatial-temporal cross dependencies for traffic forecasting. *arXiv preprint arXiv:2307.00518*.

Wu, B.; Chen, W.; Wang, W.; Peng, B.; Sun, L.; and Chen, L. 2024. WeatherGNN: Exploiting meteo-and spatial-dependencies for local numerical weather prediction bias-correction. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 2433–2441.

Wu, B.; Shang, Z.; Huang, J.; and Chen, L. 2025. MillGNN: Learning Multi-Scale Lead-Lag Dependencies for Multi-Variate Time Series Forecasting. In *Proceedings of the ACM International Conference on Information and Knowledge Management*, 3344–3354.

Wu, Z.; Pan, S.; Long, G.; Jiang, J.; Chang, X.; and Zhang, C. 2020. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 753–763.

Xia, H.; Chen, X.; Chen, B.; and Hu, Y. 2025. Dynamic synchronous graph transformer network for region-level air-quality forecasting. *Neurocomputing*, 616: 128924.

Xie, X.; Huang, Z.; and Wang, J.-s. 2005. Impact of building configuration on air quality in street canyon. *Atmospheric Environment*, 39(25): 4519–4530.

Xu, J.; Chen, L.; Lv, M.; Zhan, C.; Chen, S.; and Chang, J. 2021. HighAir: A hierarchical graph neural network-based air quality forecasting method. *arXiv preprint arXiv:2101.04264*.

Xu, X.; and Yoneda, M. 2019. Multitask air-quality prediction based on LSTM-autoencoder model. *IEEE Transactions on Cybernetics*, 51(5): 2577–2586.

Yan, R.; Liao, J.; Yang, J.; Sun, W.; Nong, M.; and Li, F. 2021. Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering. *Expert Systems with Applications*, 169: 114513.

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

Yu, H.; Li, Q.; Geng, Y.-a.; Zhang, Y.; and Wei, Z. 2020. Airnet: A calibration model for low-cost air monitoring sensors using dual sequence encoder networks. In *Proceedings of the AAAI conference on artificial intelligence*, 1129–1136.

Zhang, Y.; and Yan, J. 2023. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *International Conference on Learning Representations*.

Zhu, Q.; Guo, Y.; and Lin, W. 2021. Neural Delay Differential Equations. In *International Conference on Learning Representations*.