International Journal of Geo-Information

MDPI

Article

A Novel Wind-Aware Dynamic Graph Neural Network for Urban Ground-Level Ozone Concentration Prediction

Wenjie Wu †, Xinyue Mo *,† and Huan Li

School of Cyberspace Security (School of Cryptology), Hainan University, Haikou 570228, China; 23210839000006@hainanu.edu.cn (W.W.); lihuan@hainanu.edu.cn (H.L.)

* Correspondence: moxinyue@hainanu.edu.cn

† These authors contributed equally to this work.

Abstract

Ground-level ozone pollution poses significant risks to public health and ecosystems and remains a major environmental challenge worldwide. Accurate forecasting is difficult due to the nonlinear formation mechanisms of ozone and its strong dependence on meteorological conditions. This study proposes a Wind Speed and Direction-Based Dynamic Spatiotemporal Graph Attention Network (WSDST-GAT) for multi-step hourly ground-level ozone prediction. The model integrates a wind-aware dynamic graph to represent anisotropic pollutant transport and a Transformer-based temporal encoder to capture long-range dependencies. Meteorological variables are incorporated to enhance physical interpretability and predictive robustness. A co-kriging module is further employed to reconstruct continuous spatial ozone fields with quantified uncertainty. Using hourly observations from 35 monitoring stations in Beijing, WSDST-GAT achieves a Coefficient of Determination of 0.957, with a Mean Absolute Error of 5.25 $\mu g/m^3$, and a Root Mean Square Error of 9.58 $\mu g/m^3$. The prediction intervals demonstrate strong reliability with a Prediction Interval Coverage Probability of 94.01% and a Prediction Interval Normalized Average Width of 0.174. These results indicate that the proposed framework provides an accurate and physically informed solution for ozone forecasting and air quality management.

Keywords: ground-level ozone concentration prediction; spatiotemporal graph attention network; uncertainty quantification; co-kriging

Check for updates

Academic Editors: Wolfgang Kainz, Suzana Dragicevic, Shivanand Balram and Eric Vaz

Received: 15 January 2026
Revised: 26 February 2026
Accepted: 27 February 2026
Published: 28 February 2026

Copyright: © 2026 by the authors. Published by MDPI on behalf of the International Society for Photogrammetry and Remote Sensing. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

On a global scale, environmental pollution has been markedly intensified by urbanization and industrialization [1,2]. In urban systems, atmospheric pollution is inherently a spatial–temporal phenomenon, shaped by heterogeneous emission sources, land-use patterns, population exposure, and rapidly changing meteorological conditions. Traditional geospatial technologies—including ground monitoring networks, remote sensing, geographic information systems (GISs), and meteorological reanalysis products—provide essential infrastructures for observing, mapping, and interpreting the spatial and temporal dimensions of urban air quality [3–5]. Among various pollutants, ground-level ozone (O_3) is a secondary pollutant formed through photochemical reactions in the troposphere and has emerged as a critical environmental challenge across many metropolises [6]. Unlike stratospheric ozone, which protects against ultraviolet radiation, ground-level ozone is harmful to human health and ecosystems. Prolonged exposure to elevated ground-level ozone has been associated with adverse health outcomes, including respiratory diseases

such as chronic obstructive pulmonary disease (COPD), and is linked to premature mortality [7]. These health burdens translate into substantial socioeconomic costs and highlight the need for decision-relevant information that is both spatially explicit and temporally actionable. China has made sustained efforts in air pollution control, and substantial progress has been achieved for several pollutants [8–10]. However, ozone remains persistent and challenging due to its nonlinear formation mechanisms and strong sensitivity to meteorological variability, which often leads to pronounced spatial contrasts within and between cities. Therefore, developing accurate forecasting frameworks that leverage geospatial observations to capture urban-scale spatiotemporal dynamics of ground-level ozone is essential for air-quality management, public health protection, and sustainable urban environmental governance [11].

The models used to predict ozone concentrations mainly included three classifications: numerical forecast, statistical forecast, and machine learning models [12–14]. In recent years, numerical forecast models have been universally used in the prediction of air pollutants, including the Community Multi-Scale Air Quality Model (CMAQ) and the Weather Research and Forecasting Model with Chemistry (WRF-Chem) [15]. Uncertainties in initial and boundary meteorological conditions, along with imperfect parameter configurations and dataset limitations, frequently introduce inaccuracies into model predictions [16–19].

Statistical models were initially used for their flexible and practically applicable advantages [20], such as the Autoregressive Moving Average (ARMA) model [21], Autoregressive Integrated Moving Average (ARIMA) model [22], and Multiple Linear Regression (MLR) model [23]. These models are based on linear assumptions and only handle one-dimensional data; they can capture linear relationships but cannot handle complex nonlinear problems [24,25].

Recently, machine learning has received attention as a tool for predicting ozone concentration with strong data-mining capabilities. Compared with traditional statistical models, machine learning models can process complex multidimensional data and nonlinear problems. Machine learning models include artificial neural networks (ANNs) [26], Support Vector Machines (SVMs) [27], and Random Forests (RFs) [28]. These models are easy to use, can easily access data, and have flexible input and output features. However, as the scale and complexity of the data have grown substantially, the computational challenges posed by enormous high-dimensional datasets hinder the effective performance of traditional machine learning models. As an emerging form of machine learning models, deep learning models have quickly become an efficient way to process complex multidimensional data [29]. For example, Ebrahim Eslami developed a Convolutional Neural Network (CNN) model for 24 h ozone predictions in Seoul, achieving high accuracy with a mean Index of Agreement (IOA) of 0.84–0.89 across 25 stations [30]. Building on this line of work, subsequent studies have increasingly sought not only to model temporal patterns directly from observations but also to harness and correct information from chemistry transport models, thereby moving from purely data-driven forecasting to hybrid, model-guided approaches. Ahmed Khan Salman develops and evaluates a Temporal Convolutional Neural Network (TCNN) to bias-correct hourly ozone forecasts from the CMAQ model over South Korea for a 3-day horizon [31]. A complementary thread leveraged recurrent architectures whose gating can preserve and reuse context across many hours. Matteo Sangiorgio employs Long Short-Term Memory (LSTM) networks to predict hourly concentrations at 20 stations around the Alpine region [32]. However, purely recurrent setups, even with gated memory, can under-represent cross-site interactions and uneven spatial influences [33]. To explicitly encode these inter-station relations and adaptively weight salient temporal cues, Zhang and Hou propose a spatiotemporal neural network for 24 h ozone prediction. The model fuses graph convolution, bidirectional LSTM, and an attention mechanism, achieving the

best performance [34]. Extending this progression from spatiotemporal fusion toward greater interpretability, recent work has coupled explicit time series decomposition with sequence models to separate structure from noise before prediction. Mu and Bi develop an interpretable hybrid model that combines Seasonal-Trend decomposition using Loess (STL) with a Transformer. The model first decomposes the ozone time series into trend, seasonal, and residual components via STL, achieving promising results [35].

Recent GNN-based ozone prediction studies generally define inter-station connectivity using fixed geographic proximity, effectively freezing the graph structure along the temporal dimension. Meanwhile, recent advances have explored learning adaptive or dynamic spatial dependencies from data, such as adaptive hierarchical graph convolution networks [36], dynamic graph neural networks with learnable edge attributes [37], and adaptive adjacency matrix-based graph convolutional recurrent networks [38]. Although these approaches improve the flexibility of spatial dependency modeling, most existing studies still do not explicitly incorporate wind-driven transport physics when constructing time-varying inter-station connectivity for ozone forecasting.

However, ozone transport exhibits significant anisotropy: air masses transport pollutants downwind, creating essential differences in mixing effects between upwind and downwind directions. As wind speed and direction dynamically modulate horizontal advection and diffusion, the influence of each station on its neighbors fluctuates in real time with the evolving flow field. This urgently necessitates replacing static topologies with a dynamic wind-sensitive graph, where edge connections and weights update in response to instantaneous meteorological conditions. One concrete pathway is a GNN–Transformer architecture in which graph operations ingest the time-varying, directionally weighted topology to represent instantaneous inter-station coupling, while a Transformer captures longer-range temporal dependencies and, via attention, adaptively emphasizes periods and sources that matter most. Although such a design is well aligned with the physics of pollutant transport and is likely to enhance predictive skill, few ozone prediction studies explicitly construct dynamic wind-driven graphs based on real-time wind fields.

The main aim of this study is to develop a novel spatiotemporal hybrid deep learning model called the Wind Speed and Direction-Based Dynamic Spatial Graph Attention Network (WSDST-GAT). Concretely, WSDST-GAT constructs a time-varying directed monitoring graph whose connectivity is determined by the instantaneous wind field, orienting edges along the downwind pathway and scaling their strengths as a function of wind speed and direction. On this evolving topology, a Graph Attention Network computes adaptive inter-station coefficients to aggregate information from dynamically relevant neighbors, enabling rich extraction of spatial signals under changing flow conditions. For the temporal backbone, we employ a Transformer encoder–decoder that leverages multi-head attention to capture long-range dependencies in the hourly series, assigning greater weight to pivotal periods while maintaining sequence coherence. The resulting architecture supports both point forecasting of hourly ozone concentrations and the range estimation of concentration intervals. To translate station-level forecasts into operational, city-wide ozone fields with quantified uncertainty, we integrate a co-kriging module. Rather than serving as a mere visualization step, this spatial component models a physically informed continuous surface conditioned on meteorology, thereby bridging point forecasts and geospatial decision needs.

2. Materials and Methods

2.1. Study Area

Beijing, the economic, social, and cultural center of China, is selected as the study area in this study. Beijing is located in the northern part of the North China Plain, with

the Yanshan Mountains in the north and the Taihang Mountains in the west. The terrain of Beijing is higher in the northwest and flatter in the east. Beijing's climate is a typical northern temperate semi-humid continental monsoon climate. It is hot and rainy in summer, cold and dry in winter; spring and fall are relatively short. Ozone is easily generated through chemical reactions of nitrogen oxides and volatile organic compounds under favorable meteorological conditions, including intense solar radiation, low humidity, high temperature, and low winds. Due to the particular location and local geographical conditions, ozone pollution once became the main component of air pollutants in Beijing. The study area encompasses Beijing's 16 administrative districts: Yanqing (YQ), Miyun (MY), Huairou (HR), Changping (CP), Pinggu (PG), Shunyi (SY), Haidian (HD), Chaoyang (CY), Tongzhou (TZ), Daxing (DX), Fangshan (FS), Mentougou (MTG), Shijingshan (SJS), Fengtai (FT), Dongcheng (DC), and Xicheng (XC).

2.2. Data Sources

Air pollutant concentration from 1 January 2022to 31 December 2022 was obtained from the Real-time Air Quality Release System of the Beijing Ecological Environment Monitoring Center (https://www.bjmemc.com.cn/ (accessed on 11 October 2025)). The system provides hourly ground-level air pollutants, including O_3, CO, NO_2, SO_2, $PM_{2.5}$, and PM_{10}. Meteorological data were derived from the ERA5-Land dataset, provided by the European Center for Medium-Range Weather Forecasting (ECMWF) (https://cds.climate.copernicus.eu/ (accessed on 11 October 2025)) [39]. ERA5-Land is a global reanalysis product that ingests vast worldwide observations from sources such as satellites, radars, and weather stations and combines them with model data through the laws of physics to generate a globally complete and consistent dataset. This dataset was used to capture key atmospheric variables: 2 m dewpoint temperature (d2m), 2 m temperature (t2m), surface pressure (sp), 10 m u-component of wind (u10), and 10 m v-component of wind (v10). The monitoring stations in the study area are shown in Figure 1. All spatial data were processed in the WGS 84 geographic coordinate reference system. The great-circle distance and azimuth angle required for wind-aligned graph construction were derived from these coordinates.

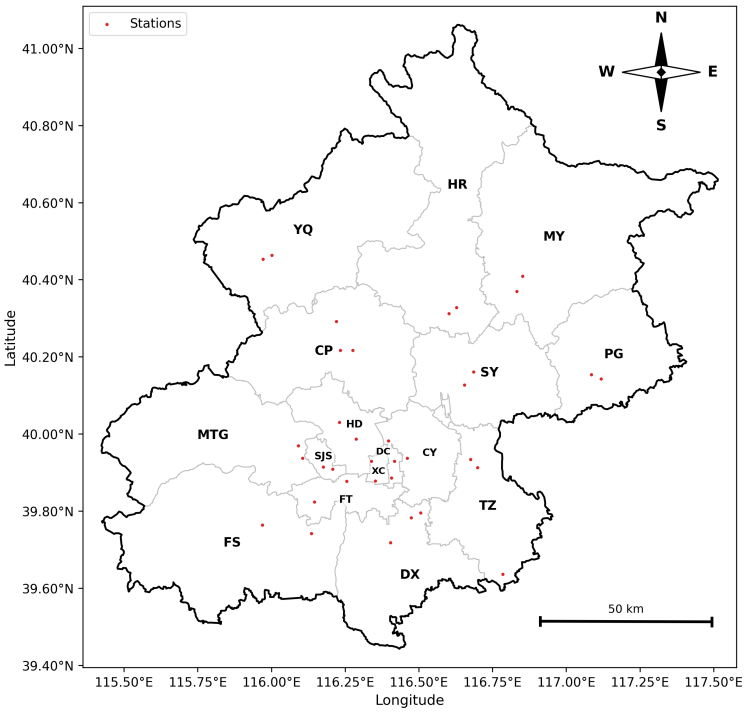


Figure 1. Spatial distribution of environmental monitoring stations in the study area.

2.3. Data Preprocessing

Hourly air quality observations were obtained from 35 monitoring stations in Beijing. All stations measure six pollutants: CO, NO_2, PM_{10}, $PM_{2.5}$, SO_2, and O_3. The latitude and longitude coordinates of each station are listed in Table A1.

In addition to pollutant concentrations, key meteorological variables were collected, including temperature, humidity, surface pressure, and wind components. These variables describe different atmospheric states and reflect the underlying climatic conditions that influence ozone formation and dispersion. From a statistical perspective, their distributions exhibit substantial variability across seasons and time scales, providing essential explanatory signals for modeling atmospheric dynamics.

All records were converted to a unified hourly timeline in UTC. Obvious invalid values, such as sentinel codes and negative concentrations, were treated as missing. This preprocessing step was applied independently to each variable at each station to ensure data consistency. Table 1 summarizes the statistical characteristics of both pollution and meteorological variables, including minimum values (Min), maximum values (Max), median values (Median), and standard deviation (St.Dev).

From a climatic perspective, these statistical measures quantitatively describe the variability in atmospheric states during the study year. The wide ranges observed in temperature and wind components reflect seasonal transitions and synoptic-scale variability, while dispersion statistics capture the dynamic fluctuations that influence ozone formation and transport.

These descriptive statistics provide a mathematical characterization of the climatic background under which the proposed model is evaluated.

Table 1. Experimental pollution and climatic conditions in this study.

Variables	Units	Min	Max	Mean	Median	St.Dev
$PM_{2.5}$	$\mu g/m^3$	1.00	376.00	30.39	20.00	32.12
PM_{10}	$\mu g/m^3$	1.00	1839.00	56.11	43.00	61.60
SO_2	$\mu g/m^3$	1.00	210.00	2.72	3.00	1.91
NO_2	$\mu g/m^3$	1.00	173.00	23.36	17.00	19.16
O_3	$\mu g/m^3$	1.00	517.00	64.85	57.00	49.15
CO	mg/m^3	0.10	9.00	0.50	0.40	0.32
d2m	K	238.16	300.02	274.58	275.03	14.22
t2m	K	261.58	313.20	285.77	286.66	12.34
sp	Pa	98,138.94	103,332.53	100,749.93	100,802.96	1085.60
u10	m/s	−4.77	5.38	0.22	0.14	1.14
v10	m/s	−8.03	5.81	−0.13	−0.17	1.90

2.3.1. Outlier Processing

To mitigate spurious spikes, we applied a Hampel filter within a centered 24 h window to each station–variable series.

$$\mathrm{MAD}_t = \underset{|k|\le w/2}{\mathrm{median}} \left| x_{t+k} - \tilde{x}_t \right| \tag{1}$$

$$x_t - \tilde{x}_t > \kappa \cdot 1.4826 \cdot \mathrm{MAD}_t, \qquad \kappa = 3 \tag{2}$$

where x_t is the hourly value at time t; $\tilde{x}_t$ is the rolling median within a centered window of width w; MAD_t is the rolling median absolute deviation computed within the same window; and κ is the outlier threshold.

2.3.2. Missing-Value Imputation

After outlier handling, data gaps for each station and variable are filled using a two-stage sequential strategy. First, forward/backward filling is applied to bridge short-term gaps. If gaps persist, a 24 h rolling median is then used for interpolation. This approach balances local temporal coherence with robustness in handling longer data gaps.

2.3.3. Normalization

Due to significant differences in the significance and numerical ranges of different feature variables, normalization is necessary before entering the model. Data normalization can effectively accelerate the convergence speed of the model and improve its predictive performance. All variables were normalized using min–max scaling prior to model training.

2.4. Supporting Statistical and Geospatial Methods

This subsection introduces the supporting methodological components used for exploratory dependency characterization and input construction, volatility-stratified evaluation, as well as spatial field reconstruction.

2.4.1. Dependency Characterization and Input Construction

Spatiotemporal Dependency Characterization

To characterize temporal persistence and spatial coherence in ozone observations prior to model development, two standard exploratory statistics were computed. First, for each station, the autocorrelation function (ACF) was calculated to assess temporal dependence within a 100 h lag window. Second, to quantify inter-station association, pairwise Pearson correlation coefficients were computed across all station pairs using contemporaneously aligned hourly ozone series.

Input Variable Screening via Mutual Information

To construct an informative yet minimally redundant feature set, candidate predictors were screened based on their nonlinear dependence with ozone concentration using mutual information (MI) [40]. For continuous variables X and Y with joint density $p(x, y)$ and marginals $p(x)$ and $p(y)$, MI is defined as

$$I(X;Y) = \iint p(x,y)\, \log \frac{p(x,y)}{p(x)\, p(y)}\, dx\, dy. \tag{3}$$

MI was estimated pairwise between ozone and each candidate predictor using the training split only, thereby avoiding temporal leakage. Estimation was performed using samples with at least 24 overlapping hourly observations, while variables with zero variance were excluded. Predictors with negligible MI values were removed, and the retained variables were used as inputs to the proposed model. Detailed exploratory results and visualizations are reported in Section 3.

2.4.2. Meteorological Volatility Index

To quantify the atmospheric instability and group the test hours as discussed in Section 3, we introduce the meteorological volatility index ($\mathcal{V}_t$). This index measures the intensity of temporal fluctuations in wind speed over a sliding window. It is defined mathematically as the rolling standard deviation:

$$\mathcal{V}_t = \sqrt{\frac{1}{W} \sum_{i=0}^{W-1} (x_{t-i} - \bar{x}_t)^2} \tag{4}$$

where: x_t denotes the observed meteorological value at time step t; W is the size of the sliding window; $\bar{x}_t$ represents the moving average over the window W.

Based on the distribution of $\mathcal{V}_t$, we categorize the test samples into different stability groups. Specifically, periods where $\mathcal{V}_t$ exceeds the 80th percentile of the historical distribution are classified as high volatility, whereas periods falling below the 20th percentile represent steady conditions.

2.4.3. Co-Kriging Method

We improve ozone interpolation by co-kriging with physically related meteorology: 2 m dewpoint temperature, 2 m temperature, surface pressure, 10 m u-component of wind, and 10 m v-component of wind. All variables are time-aligned and standardized.

Let the primary variable be ozone $Z^{(0)}(\mathbf{s})$ and the secondary set $\{Z^{(q)}(\mathbf{s})\}_{q=1}^{Q}$. Under second-order stationarity of residuals, direct and cross semivariograms follow a Linear Model of Coregionalization (LMC) [41]:

$$\gamma^{pq}(h) \;=\; \tfrac{1}{2}\,\mathrm{Var}\big(Z^{(p)}(\mathbf{s}) - Z^{(q)}(\mathbf{s}+\mathbf{h})\big) \;=\; \sum_{m=1}^{M} b_{pq}^{(m)}\,\gamma_m(h), \qquad p,q \in \{0,\dots,Q\}, \tag{5}$$

where $\gamma_m(h)$ are authorized basic structures, and each coregionalization matrix $B^{(m)} = [b_{pq}^{(m)}]$ is positive semidefinite.

The ordinary co-kriging predictor at $\mathbf{s}_0$ is

$$\hat{Z}^{(0)}(\mathbf{s}_0) = \sum_{p=0}^{Q} \sum_{i=1}^{n_p} \lambda_i^{(p)}\, Z^{(p)}(\mathbf{s}_i^{(p)}), \tag{6}$$

with weights solving

$$\begin{bmatrix} \Gamma^{00} & \Gamma^{01} & \cdots & \Gamma^{0Q} & \mathbf{1}^{(0)} \\ \Gamma^{10} & \Gamma^{11} & \cdots & \Gamma^{1Q} & \mathbf{1}^{(1)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Gamma^{Q0} & \Gamma^{Q1} & \cdots & \Gamma^{QQ} & \mathbf{1}^{(Q)} \\ (\mathbf{1}^{(0)})^\top & (\mathbf{1}^{(1)})^\top & \cdots & (\mathbf{1}^{(Q)})^\top & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda}^{(0)} \\ \boldsymbol{\lambda}^{(1)} \\ \vdots \\ \boldsymbol{\lambda}^{(Q)} \\ \mu \end{bmatrix} = \begin{bmatrix} \boldsymbol{\gamma}_0^{00} \\ \boldsymbol{\gamma}_0^{10} \\ \vdots \\ \boldsymbol{\gamma}_0^{Q0} \\ 1 \end{bmatrix}, \tag{7}$$

where $(\Gamma^{pq})_{ij} = \gamma^{pq}(\|\mathbf{s}_i^{(p)} - \mathbf{s}_j^{(q)}\|)$, $\boldsymbol{\gamma}_0^{p0} = \big[\gamma^{p0}(\|\mathbf{s}_i^{(p)} - \mathbf{s}_0\|)\big]$, $\mathbf{1}^{(p)}$ is a vector of ones, and μ enforces $\sum_{p,i} \lambda_i^{(p)} = 1$. The co-kriging variance is

$$\sigma^2_{\mathrm{cok}}(\mathbf{s}_0) = \sum_{p=0}^{Q} \sum_{i=1}^{n_p} \lambda_i^{(p)}\, \gamma^{p0}(\|\mathbf{s}_i^{(p)} - \mathbf{s}_0\|) + \mu. \tag{8}$$

Anisotropy aligned with prevailing winds can be used in $\gamma_m(h)$ to capture directional transport.

2.5. Proposed WSDST-GAT Model

Figure 2 shows the detailed architecture of WSDST-GAT.

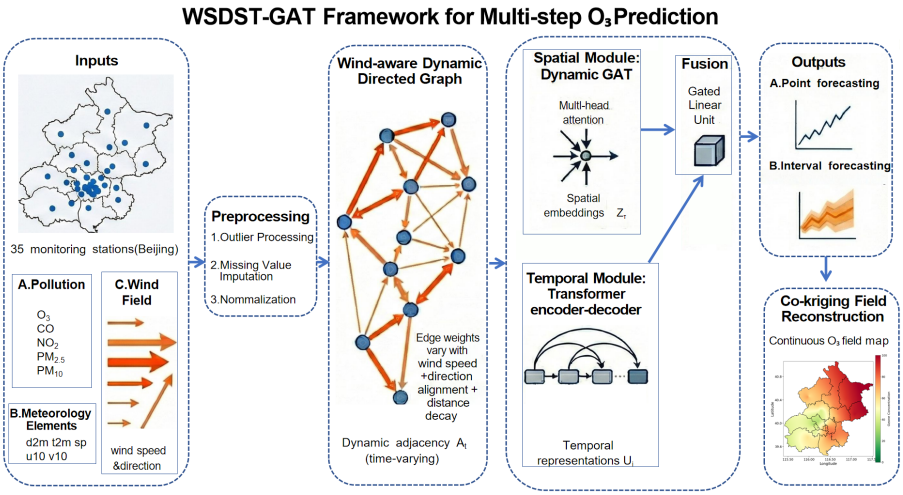


Figure 2. The proposed model WSDST-GAT architecture diagram.

2.5.1. Problem Definition

The current time is t. We aim to predict hourly ozone concentrations over a future horizon of H steps,

$$\boldsymbol{Y}_{t+1:t+H} = [y_{t+1}, y_{t+2}, \ldots, y_{t+H}] \in \mathbb{R}^{H \times N} \tag{9}$$

where $y_{t+h} \in \mathbb{R}^N$ denotes the ozone concentrations at all N stations at time $t+h$. The prediction is conditioned on a historical window of length L (from $t-L+1$ to t) that includes air-quality features, meteorological drivers, and dynamic graphs driven by wind.

Specifically, we partition the inputs as

$$\boldsymbol{X}^a_{t-L+1:t} \in \mathbb{R}^{L \times N \times F_a}, \tag{10}$$

$$\boldsymbol{X}^m_{t-L+1:t} \in \mathbb{R}^{L \times N \times F_m}, \tag{11}$$

$$\mathcal{G}_{t-L+1:t} = \{\boldsymbol{A}_\tau \in \mathbb{R}^{N \times N}\}^t_{\tau=t-L+1} \tag{12}$$

Here, $\boldsymbol{X}^a$ collects air-quality predictors at the station level (CO, NO_2, $PM_{2.5}$, PM_{10}) with $F_a = 4$; $\boldsymbol{X}^m$ contains meteorology, including 2 m dewpoint temperature, 2 m temperature, surface pressure, 10 m u-component of wind, and 10 m v-component of wind ($F_m = 5$). $\mathcal{G}_{t-L+1:t}$ denotes the sequence of time-varying directed adjacency matrices constructed from instantaneous wind speed/direction and inter-station distances.

The proposed model is a mapping

$$\boldsymbol{Y}_{t+1:t+H} = F_\Theta(\boldsymbol{X}^a_{t-L+1:t}, \boldsymbol{X}^m_{t-L+1:t}, \mathcal{G}_{t-L+1:t}) \tag{13}$$

where F_Θ is parameterized by the wind-aware dynamic GAT and Transformer, and $\boldsymbol{Y}^\tau$ denotes the quantile forecasts τ when prediction intervals are required. N denotes the number of stations; L denotes the input length; H denotes the forecast horizon; F_a and F_m denote the characteristic dimensions of the air-quality and meteorological input, respectively; $\boldsymbol{A}_\tau$ denotes the directed adjacency at hour τ; and Θ denotes the training parameters of WSDST-GAT.

2.5.2. Spatial Feature Extraction Module

At each hour $\tau \in [t-L+1, t]$, we operate on a directed graph of wind awareness and time-varying $\mathcal{G}_\tau = (\mathcal{V}, \mathcal{E}_\tau, \boldsymbol{A}_\tau)$, whose adjacency $\boldsymbol{A}_\tau$ encodes downwind-oriented connectivity, distance decay, and wind-speed gain. Given node features $\boldsymbol{X}_\tau$, a dynamic Graph Attention Network computes layer-wise embeddings $\boldsymbol{H}^{(\ell)}_\tau$ via edge-biased attention [42]:

$$e_{ij,\tau}^{(\ell)} = \phi\left(\boldsymbol{W}^{(\ell)}\boldsymbol{h}_{i,\tau}^{(\ell)} \parallel \boldsymbol{W}^{(\ell)}\boldsymbol{h}_{j,\tau}^{(\ell)}\right) + \beta \cdot g\left([\boldsymbol{A}_\tau]_{i\to j}\right) \tag{14}$$

$$\alpha_{ij,\tau}^{(\ell)} = \frac{\exp\{e_{ij,\tau}^{(\ell)}\}}{\sum_{k\in\mathcal{N}_\tau(i)}\exp\{e_{ik,\tau}^{(\ell)}\}}, \tag{15}$$

$$\boldsymbol{h}_{i,\tau}^{(\ell+1)} = \sigma\left(\sum_{j\in\mathcal{N}_\tau(i)} \alpha_{ij,\tau}^{(\ell)}\boldsymbol{W}^{(\ell)}\boldsymbol{h}_{j,\tau}^{(\ell)}\right) \tag{16}$$

with parallel multi-head attention and residual connections plus layer normalization for stability. The outputs $\boldsymbol{Z}_\tau = \boldsymbol{H}_\tau^{(L_s)}$ serve as spatial embeddings for the temporal module. Here, $\boldsymbol{h}_{i,\tau}^{(\ell)}$ is the embedding of node i at layer ℓ and hour τ; $\boldsymbol{W}^{(\ell)}$ is a linear projection; $\phi(\cdot)$ is the attention scorer; $g(\cdot)$ maps the edge weight; β is a scalar bias-injecting wind topology; $\sigma(\cdot)$ is a nonlinearity; $\mathcal{N}_\tau(i)$ is the neighbor set of i at hour τ; and L_s is the number of spatial layers.

2.5.3. Temporal Feature Extraction Module

For each station i, we form a length-L sequence $\boldsymbol{Z}_{i,t-L+1:t} = [\boldsymbol{z}_{i,t-L+1}, \ldots, \boldsymbol{z}_{i,t}]$ from the spatial module and encode it with a Transformer to capture multi-scale temporal dynamics and long-range dependencies [43]:

$$\boldsymbol{U}_i = \text{TransformerEncoder}(\text{PE}(\boldsymbol{Z}_i)), \quad \boldsymbol{y}_{i,t+1:t+H} = \text{Decoder}(\boldsymbol{U}_i, \text{PE}(\boldsymbol{Q})) \tag{17}$$

with sinusoidal positional encodings on inputs and queries and a pre-norm Transformer. For multi-step forecasting, a non-autoregressive decoder with H queries maps the encoded history to the horizon. Here, $\boldsymbol{Z}_i$ is the station-wise embedding sequence; $\text{PE}(\cdot)$ is positional encoding; $\boldsymbol{U}_i$ is the encoded representation; $\boldsymbol{Q}$ is the H queries; and $\boldsymbol{y}_{i,t+1:t+H}$ are the H-step predictions at station i.

2.5.4. Fusion Module

We fuse the spatial context $\boldsymbol{s}_i$ and temporal encoding $\boldsymbol{u}_i$ with a gated linear unit (GLU):

$$\boldsymbol{h}_i = \text{GLU}(\boldsymbol{s}_i \parallel \boldsymbol{u}_i) = (\boldsymbol{s}_i \parallel \boldsymbol{u}_i)\boldsymbol{W}_h \odot \sigma\big((\boldsymbol{s}_i \parallel \boldsymbol{u}_i)\boldsymbol{W}_g\big). \tag{18}$$

A point head and optional quantile heads then produce H-step forecasts:

$$\hat{\boldsymbol{y}}_{i,t+1:t+H} = \text{Head}_{\text{point}}(\boldsymbol{h}_i), \tag{19}$$

$$\hat{\boldsymbol{y}}_{i,t+1:t+H}^{\tau} = \text{Head}_{\text{quant}}^{\tau}(\boldsymbol{h}_i), \quad \tau \in \{0.1, 0.5, 0.9\}. \tag{20}$$

Here, $\parallel$ is concatenation, $\odot$ is the Hadamard product, $\sigma(\cdot)$ is the logistic sigmoid, and $\boldsymbol{W}_h, \boldsymbol{W}_g$ are learnable matrices. $\text{Head}_{\text{point}}$ and $\text{Head}_{\text{quant}}^{\tau}$ denote linear mappings that produce, respectively, mean forecasts and τ-quantiles.

2.6. Process Details

The overall experimental workflow of the proposed WSDST-GAT model comprises four primary stages: data preprocessing, spatial–temporal representation learning, model training, and evaluation and visualization. The detailed procedure is described below.

2.6.1. Construction of Dynamic Spatial Graphs

At each hourly step t, a dynamic graph driven by the wind $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t, \mathbf{A}_t)$ was constructed, where nodes correspond to air-quality stations, and the edges represent the dynamic directional influence determined by the instantaneous wind field.

Specifically, the wind direction φ_t determines the orientation of pollutant transport. For any pair of stations (i, j), the azimuth angle $\theta_{ij,t}$ describes the direction from station i to station j. When the angular difference $\theta_{ij,t} - \varphi_t$ is small, station j lies approximately in the downwind direction of station i, indicating potential pollutant transport from i to j.

Meanwhile, wind speed controls the overall strength of atmospheric advection, thereby modulating the intensity of inter-station influence. Stronger winds enhance horizontal transport and amplify the effective connectivity between directionally aligned stations.

Accordingly, the edge weights were defined as

$$[\mathbf{A}_t]_{i,j} = \exp\left(-\frac{d_{ij}}{\lambda}\right) \cdot \max\left(0, \cos(\theta_{ij,t} - \varphi_t)\right) \tag{21}$$

where d_{ij} denotes the great-circle distance between stations i and j; $\theta_{ij,t}$ represents the azimuth angle between the two stations; φ_t denotes the instantaneous wind direction; and λ controls the spatial decay scale.

This formulation ensures that wind direction determines edge orientation through downwind alignment, while wind speed regulates the magnitude of spatial interaction under varying meteorological conditions. The resulting time-varying adjacency matrix $\mathbf{A}_t$ dynamically adapts to the evolving atmospheric transport pathways.

2.6.2. Model Training

The learning process integrates spatial and temporal dependencies through an end-to-end framework. The dynamic Graph Attention Network (GAT) extracts spatial correlations among stations under varying flow fields, while the Transformer captures long-range temporal dependencies from sequential embeddings. These two feature streams are fused via a GLU and passed into fully connected layers for final prediction. The model is optimized using a composite loss function:

$$\mathcal{L} = \lambda_{\text{MSE}} \cdot \text{MSE}(\hat{\boldsymbol{Y}}, \boldsymbol{Y}) + \lambda_{\text{Q}} \cdot \sum_{\tau \in \{0.1, 0.5, 0.9\}} \text{Pinball}_\tau(\hat{\boldsymbol{Y}}^\tau, \boldsymbol{Y}) \tag{22}$$

where MSE represents the mean squared error for point forecasts and Pinball_τ represents the quantile regression loss for probabilistic range estimation. Adam optimizer was used with an initial learning rate of 1.00×10^{-3} and a batch size of 32. Early stopping was applied based on Root Mean Square Error (RMSE) validation to prevent overfitting.

2.7. Experimental Settings

All experiments were implemented using Python 3.9 and PyTorch 1.12 and executed on a workstation equipped with an NVIDIA RTX 4060ti GPU (16 GB memory), Intel Core i7-10700K CPU, and 32 GB RAM running Windows 11. The computational environment was configured to ensure reproducibility and consistency across all runs.

2.7.1. Dataset Division

The complete dataset, consisting of hourly air quality and meteorological records from January 2022 to December 2022, was divided into three subsets in chronological order:

- **Training set (70%)**—used for model parameter learning;
- **Validation set (15%)**—used for hyperparameter tuning and early stopping;
- **Test set (15%)**—used exclusively for performance evaluation.

The chronological split avoids temporal leakage and ensures that the model generalizes to unseen future periods.

2.7.2. Baseline Models for Comparison

A comprehensive evaluation was conducted to validate the proposed WSDST-GAT, which included a comparison with several widely adopted baseline models.

- **LSTM** [44]: captures temporal dependencies via a recurrent structure;
- **GRU** [45]: simplified recurrent model for short-term dynamics;
- **GCN-LSTM** [46]: incorporates static spatial correlations with temporal recurrence;
- **ST-GAT** [47]: Graph Attention Network with static adjacency for spatiotemporal learning;
- **ST-Transformer** [48]: Transformer-based spatiotemporal sequence model without dynamic graph adaptation.
- **STD-GCN** [49]: a dynamic directed graph convolutional network with wind-field-based adjacency.

All baseline models were trained under identical data splits, optimization settings, and evaluation metrics for fair comparison.

2.7.3. Evaluation Metrics

In order to fully assess point and interval forecast performance, multiple evaluation indicators were employed in this study.

Point Forecast Metrics

For deterministic prediction of ozone concentration, we utilized three widely used statistical indicators: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2).

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{23}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{24}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{25}$$

where y_i and $\hat{y}_i$ denote the observed and predicted ozone concentrations, respectively; $\bar{y}$ is the mean of the observed values; and N is the number of samples.

Interval Forecast Metrics

For probabilistic or range-based forecasting, we evaluated the model using Prediction Interval Coverage Probability (PICP) and Prediction Interval Normalized Average Width (PINAW).

Prediction Interval Coverage Probability (PICP):

$$\mathrm{PICP} = \frac{1}{N}\sum_{i=1}^{N} c_i, \quad c_i = \begin{cases} 1, & \text{if } y_i \in [y_i^L, y_i^U] \\ 0, & \text{otherwise} \end{cases} \tag{26}$$

where $[y_i^L, y_i^U]$ denotes the predicted lower and upper bounds. PICP measures the percentage of actual observations covered by the predicted interval.

Normalized Average Width (PINAW) Prediction Interval:

$$\mathrm{PINAW} = \frac{1}{NR_y}\sum_{i=1}^{N}(y_i^U - y_i^L) \tag{27}$$

where $R_y = \max(y_i) - \min(y_i)$ represents the range of observed ozone concentrations. PINAW reflects the average width of the interval, normalized to the data range—smaller PINAW values indicate tighter intervals.

3. Results and Discussion

3.1. Exploratory Spatiotemporal Characteristics and Feature Screening Results

To better understand the intrinsic statistical properties of ozone concentration and justify the modeling design, we first examine its temporal persistence, spatial coherence, and nonlinear dependency with candidate predictors. The results of these exploratory analyses are presented below.

3.1.1. Temporal Dependency Analysis

Figure 3 illustrates the ACF of hourly ozone concentration over a 100 h lag window.

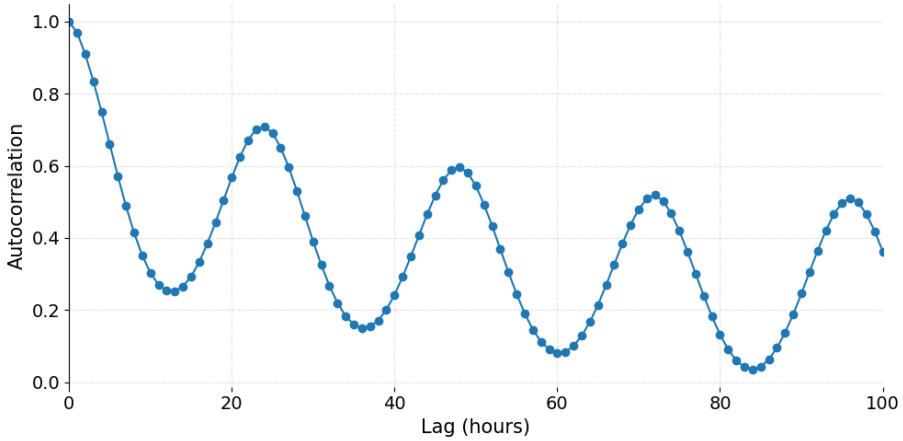


Figure 3. Autocorrelation of O_3 concentration for 100 h.

Distinct peaks are observed at lag intervals of approximately 24 and 48 h, indicating pronounced daily periodicity. In several stations, the autocorrelation coefficients exceed 0.5 at these lags, reflecting strong short-term temporal persistence.

This cyclical behavior is consistent with the photochemical formation mechanism of ozone, which is modulated by solar radiation and temperature variations. The presence of substantial temporal dependency supports the necessity of incorporating long-range sequence modeling mechanisms within the predictive framework.

3.1.2. Spatial Correlation Analysis

The spatial association between monitoring stations is visualized in Figure 4, which presents the Pearson correlation matrix of hourly ozone concentrations across all 35 stations.

The majority of station pairs exhibit moderate-to-strong positive correlations, indicating substantial spatial coherence in ozone variability across the study area. This pattern suggests that ozone concentrations are influenced not only by local emissions but also by regional transport and atmospheric mixing processes.

Such spatial dependency justifies the adoption of graph-based modeling structures capable of explicitly encoding inter-station interactions.

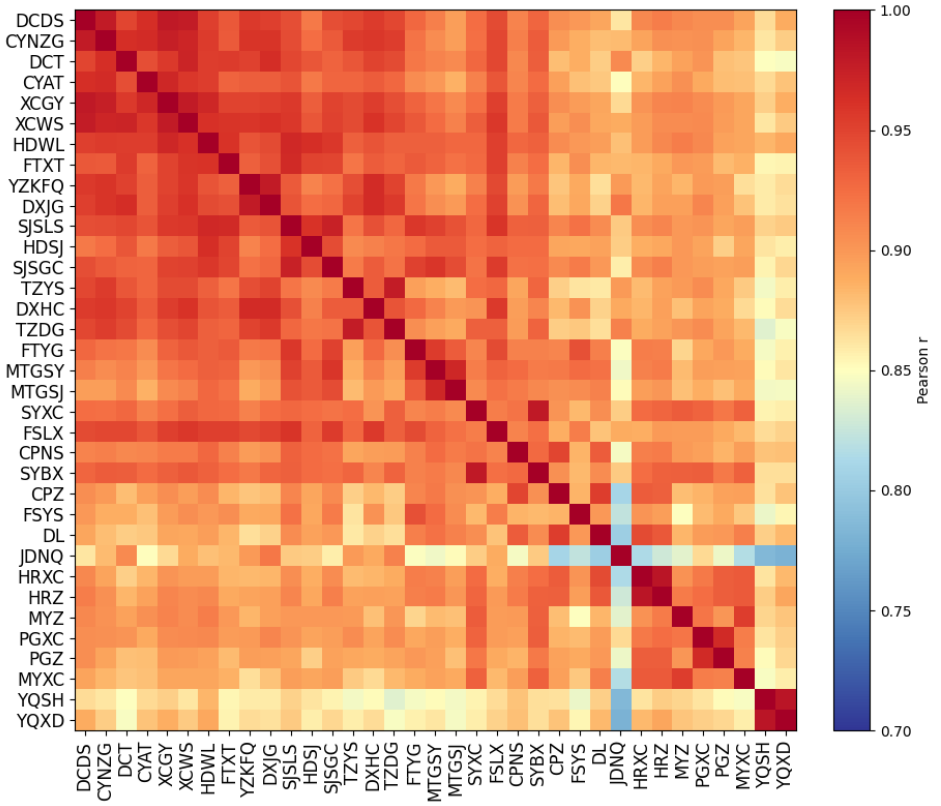


Figure 4. Results of ozone-based Pearson correlation analysis between 35 monitoring stations.

3.1.3. Mutual Information-Based Feature Screening

Figure 5 reports the estimated mutual information (MI) values between ozone concentration and candidate predictors.

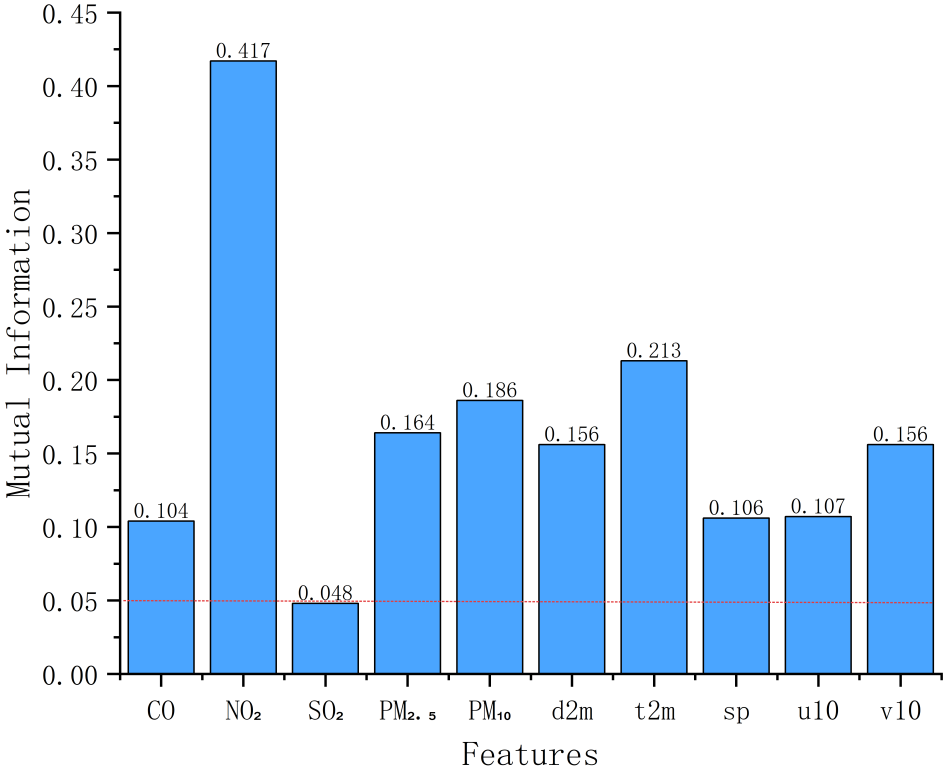


Figure 5. Mutual information of the feature variables for feature selection.

Most meteorological and pollutant variables exhibit non-negligible nonlinear dependence with ozone, indicating their potential explanatory relevance. In contrast, SO_2 demonstrates near-zero MI values, suggesting minimal contribution to ozone variability during the study period.

Based on these results, SO_2 was excluded from the input feature set, and nine predictors were retained per station. This selection balances information richness and redundancy control, improving generalization capacity while maintaining computational efficiency.

3.2. Overall Performance Evaluation

The comprehensive evaluation of the proposed WSDST-GAT's predictive capability relied on both point and interval forecasting techniques. Evaluation metrics include MAE, RMSE, and R^2 for point prediction, as well as PICP and PINAW for interval prediction. These metrics jointly assess the accuracy, reliability, and uncertainty calibration of the model.

3.2.1. Point Forecasting: Comparative Benchmarking Across Models

Table 2 summarizes the quantitative results of predicting ozone concentration in all models. Meanwhile, Figure 6 provides a visual comparison where MAE, RMSE, and R^2 are shown as bars. The proposed WSDST-GAT demonstrates overall superior performance to all baseline methods, namely LSTM, GRU, GCN-LSTM, ST-GAT, ST-Transformer, and STD-GCN, with an MAE of 5.25 $\mu g/m^3$ and RMSE of 9.58 $\mu g/m^3$. These results demonstrate that integrating wind-speed and direction-driven dynamic graphs effectively captures evolving spatial dependencies, while the Transformer-based temporal encoder improves the extraction of long-term patterns.

Table 2. Comparison of point forecasting results between the proposed model and the baseline models.

Model	MAE	RMSE	R^2
LSTM	14.83	21.02	0.864
GRU	14.26	20.38	0.873
GCN-LSTM	12.94	18.72	0.891
ST-GAT	11.75	17.33	0.917
ST-Transformer	10.63	16.48	0.924
STD-GCN	6.87	9.91	0.933
WSDST-GAT	5.25	9.58	0.957

WSDST-GAT demonstrates a marked improvement over static graph models such as ST-GAT, underscoring the benefits of employing a wind-aware dynamic topology in which edges between monitoring stations adjust adaptively based on hourly wind field data.

Conventional static distance-based graphs implicitly assume isotropic and time-invariant spatial interactions, where inter-station influence depends solely on geographic proximity. However, atmospheric pollutant transport is inherently anisotropic and strongly modulated by evolving meteorological conditions, particularly wind speed and direction. Static topologies therefore fail to reflect the directional and time-varying nature of advection-driven dispersion processes.

In contrast, the proposed dynamic graph explicitly incorporates wind-aware connectivity, allowing spatial interactions to vary according to real-time flow conditions. This dynamic structure captures both the directional alignment of downwind transport and the intensity of wind-driven dispersion, leading to spatial representations that are more physically consistent and better aligned with atmospheric transport mechanisms. Con-

sequently, the model achieves improved predictive performance and enhanced physical interpretability.

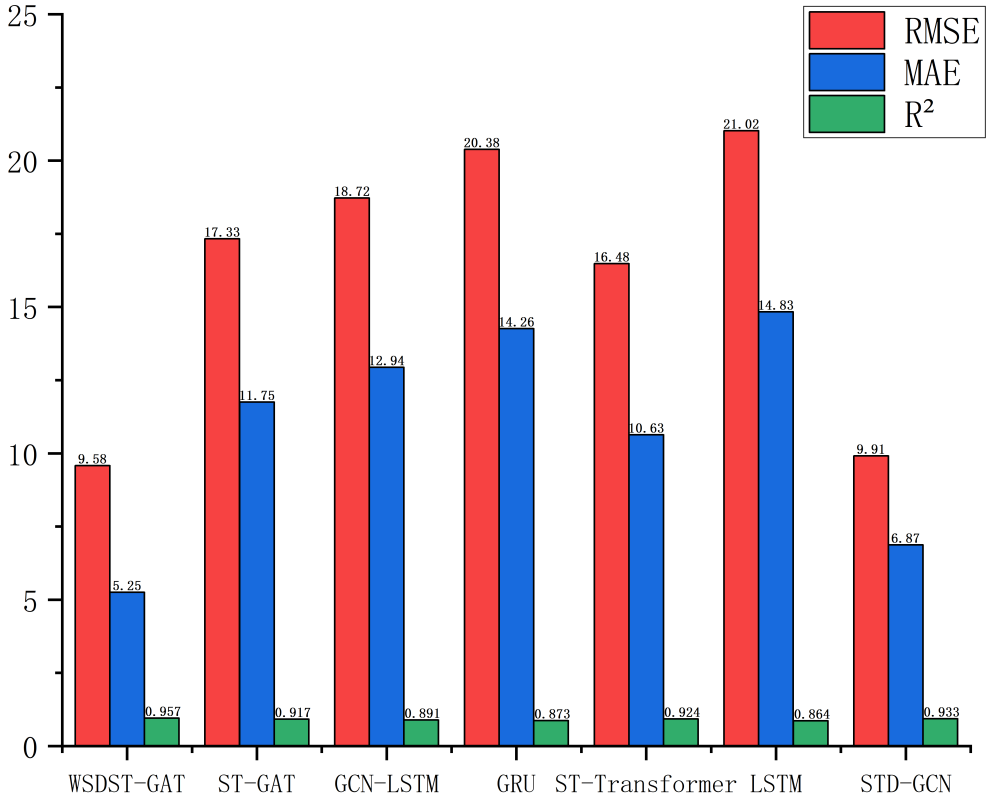


Figure 6. Performance comparison of different models for point forecasting.

3.2.2. Prediction Intervals: Comparative Benchmarking Across Models

WSDST-GAT quantifies the prediction uncertainty by generating prediction intervals for each predicted time step. A well-calibrated prediction interval achieves an optimal balance between coverage reliability and sharpness, maintaining narrow widths while consistently encompassing the true observations. To evaluate the performance of the proposed model in uncertainty quantification, Table 3 compares the interval forecasting results of the WSDST-GAT model with the baseline models.

Table 3. Comparison of interval forecasting results between the proposed model and the baseline models.

Model	PICP	PINAW
LSTM	88.12	0.214
GRU	89.47	0.207
GCN-LSTM	93.40	0.176
ST-GAT	91.32	0.191
ST-Transformer	92.48	0.186
STD-GCN	93.58	0.181
WSDST-GAT	94.01	0.174

As demonstrated in the results, Figure 7 provides a visual comparison of interval forecast performance. WSDST-GAT produces the most reliable and compact prediction intervals, achieving the highest PICP of 94.01% and the lowest PINAW of 0.174, reflecting the superior calibration of uncertainty. A high PICP indicates that the constructed prediction intervals successfully cover the majority of true observations, demonstrating strong reliability. Meanwhile, a relatively low PINAW suggests that the intervals remain sufficiently

narrow, avoiding an overly conservative uncertainty estimation. The combination of high coverage and compact width implies that the model achieves a favorable balance between reliability and precision in probabilistic forecasting. Notably, by incorporating wind field dynamics, the model adaptively widens its uncertainty bands under unstable meteorological conditions, such as rapid changes in wind speed or direction, thereby enhancing robustness in complex atmospheric scenarios.

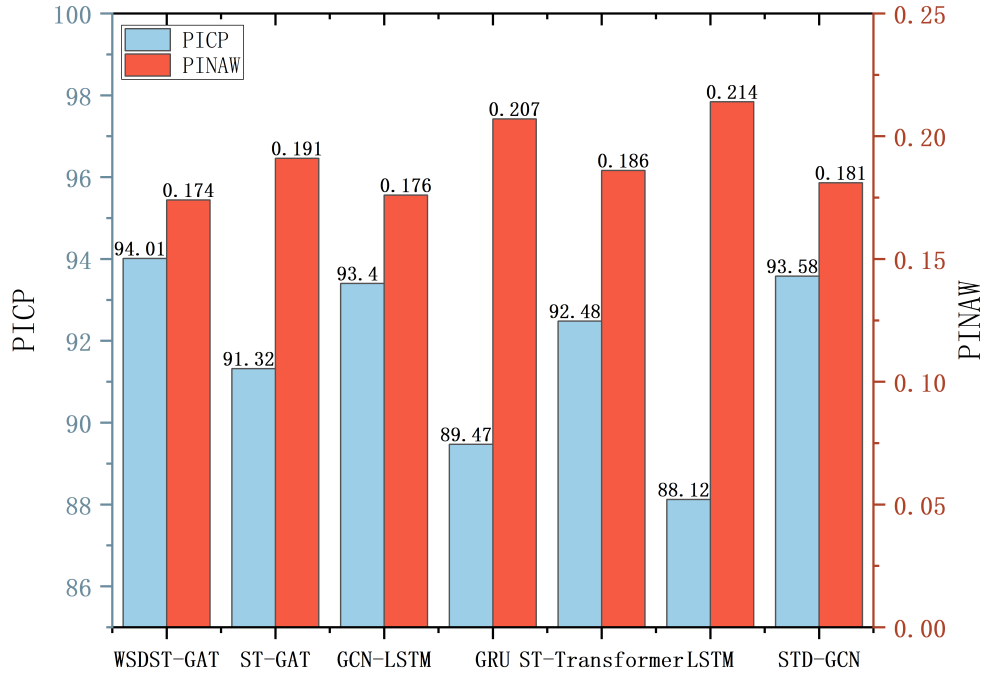


Figure 7. Performance comparison of different models for interval forecasting.

In general, these findings validate that WSDST-GAT achieves high deterministic accuracy and strong probabilistic reliability, offering a physically interpretable and data-driven framework suitable for real-time ozone prediction and early-warning systems.

3.3. Temporal Forecasting Analysis

To provide the government with more time to formulate policy responses to potential severe ozone-related issues, long-term prediction of ozone concentrations is crucial. Consequently, the forecast horizon is extended from the next hour to the entire following day. The present study further investigates the temporal modeling capacity of the proposed WSDST-GAT model by evaluating its performance across multiple forecasting horizons and examining its efficacy in capturing short-term fluctuations and diurnal patterns of ozone concentrations. Taking the XCGY station as an example, the 24 h prediction results are shown in Figures 8 and 9, which present the forecast accuracy of the proposed model and baselines across different horizons. As expected, the prediction errors gradually increase with longer forecast horizons due to the accumulation of temporal uncertainty. However, the proposed WSDST-GAT consistently outperforms all comparison models on each horizon, showing superior stability and robustness.

The results indicate that WSDST-GAT maintains stable predictive performance even at longer forecast intervals. The combination of a Transformer-based temporal encoder and dynamic graph-driven spatial representation allows the model to effectively capture both long-term temporal dependency and short-term variations in the ozone concentration series. In particular, the multi-head self-attention mechanism enables the model to assign greater weights to the most informative historical periods, thereby improving temporal prediction coherence.

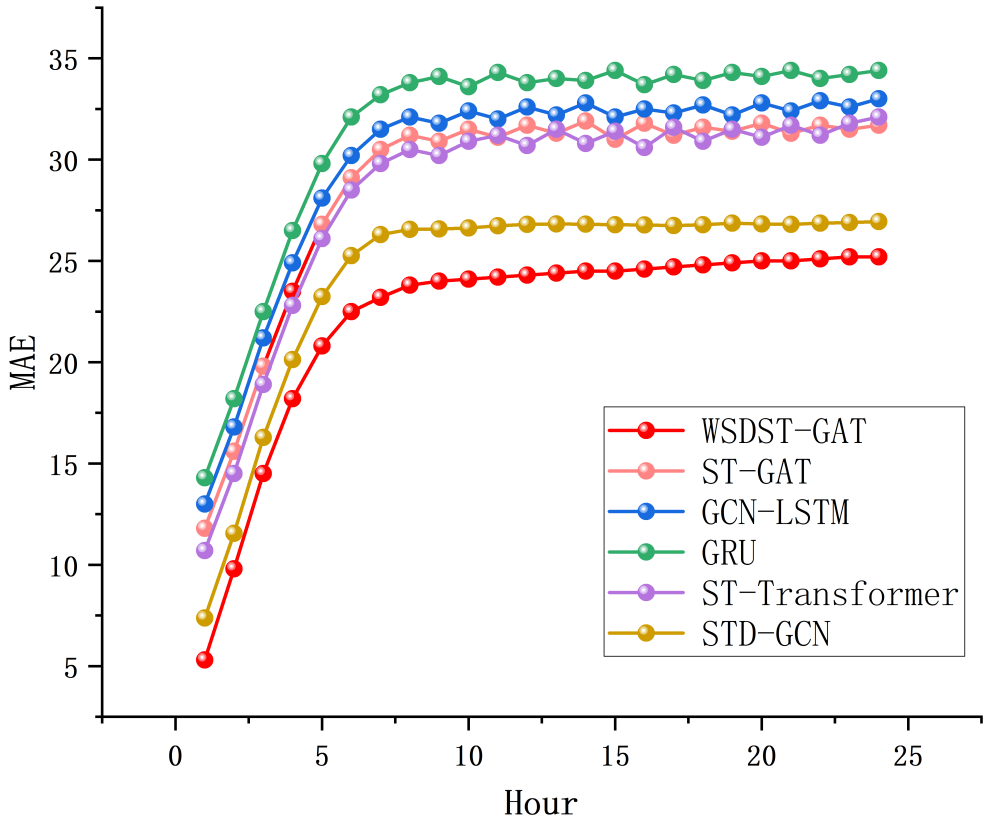


Figure 8. The MAEs of different models in 24 h prediction.

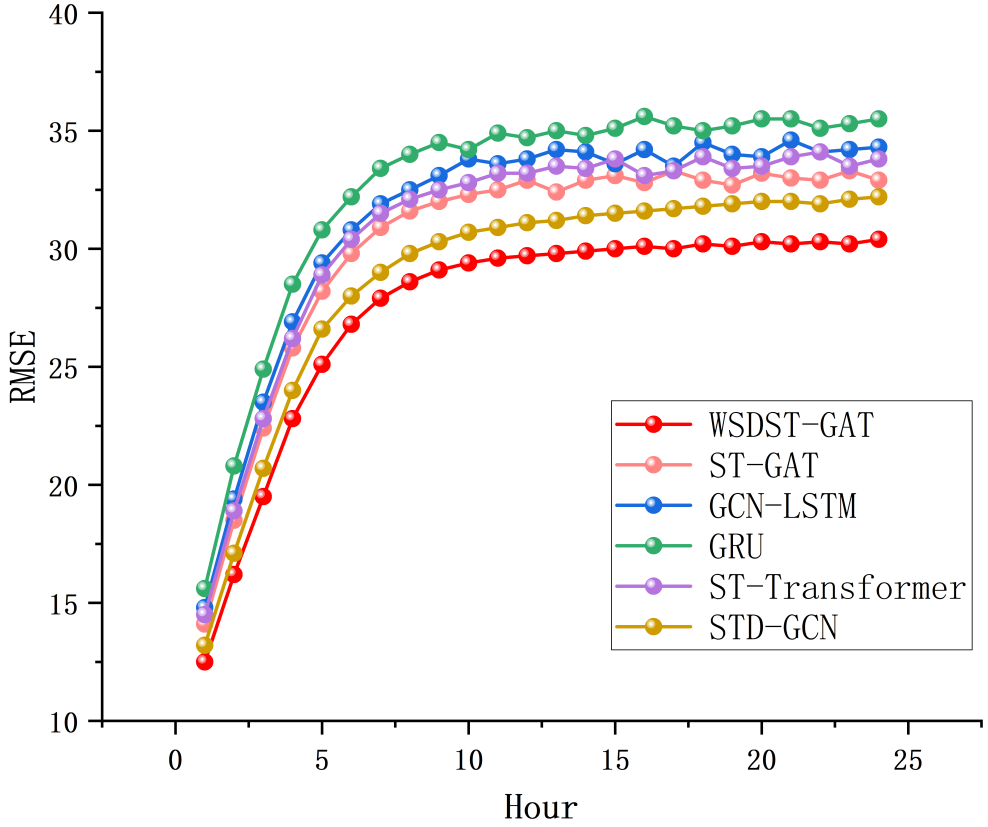


Figure 9. The RMSE of different models in 24 h prediction.

3.4. Spatial Prediction Performance

To comprehensively evaluate the spatial generalization ability of the proposed WSDST-GAT model, we assess its predictive performance at 35 air-quality monitoring stations spread throughout Beijing. For each model, the station-level MAEs are calculated. Figure 10 illustrates the MAE values and R^2 of the baseline approaches compared to the proposed WSDST-GAT.

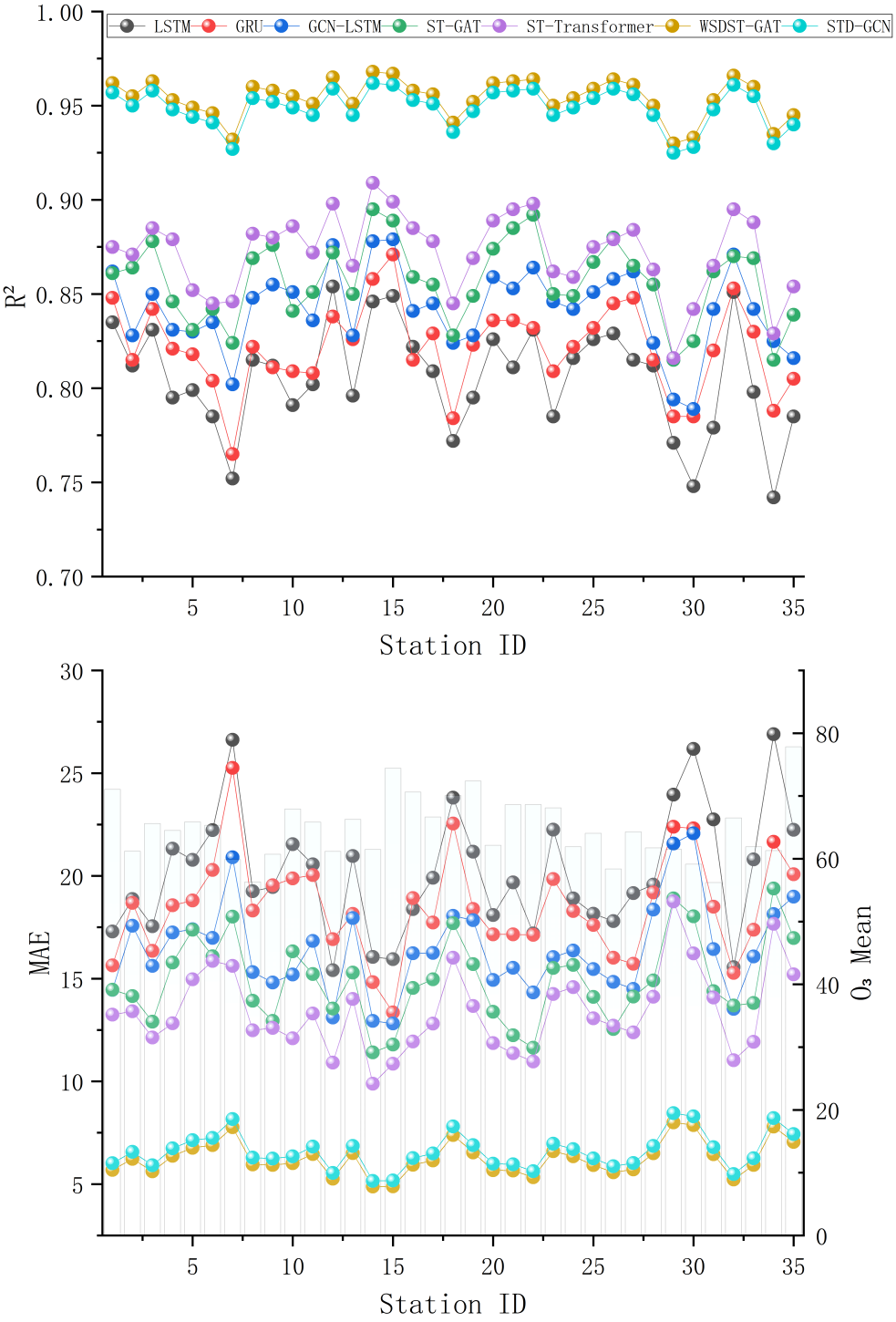


Figure 10. Station-wise spatial prediction performance of 35 monitoring stations in Beijing.

The results demonstrate that WSDST-GAT consistently achieves the lowest MAE among all competing models, confirming its superior robustness when applied to geographically heterogeneous monitoring sites. Conventional recurrent models such as LSTM

exhibit the largest errors, primarily due to their inability to encode explicit inter-station spatial dependencies. GCN-LSTM and ST-GAT partially alleviate this limitation by incorporating graph structures; however, their spatial graphs remain static and therefore cannot accommodate the inherently time-varying nature of atmospheric transport. ST-Transformer achieves further improvements by integrating spatiotemporal attention mechanisms, yet it still relies on a fixed or quasi-static representation of spatial connectivity.

In contrast, WSDST-GAT employs a dynamically evolving, wind-aware directed graph, where edge directions are aligned with downwind flows, and edge weights are adaptively modulated by wind speed and directional alignment. This enables the model to better capture the anisotropic and temporally varying transport processes that govern ozone formation and dispersion. Consequently, WSDST-GAT yields more spatially coherent predictions not only within densely populated urban areas but also at suburban and peripheral stations that are strongly influenced by regional atmospheric transport. These findings underscore the importance of dynamically modeling wind-driven topological structures to improve spatial prediction accuracy in ozone forecasting tasks.

3.5. Prediction Intervals: Case Study and Volatility-Stratified Analysis

While point forecasts provide estimates of the expected ozone concentration, they do not convey the inherent uncertainty arising from complex meteorological dynamics and stochastic emission processes. To overcome this limitation, the proposed WSDST-GAT model is extended to generate prediction intervals, offering probabilistic representations of the forecast uncertainty at each time step.

Figure 11 presents an example of the predicted ozone concentration intervals at Xicheng Guanyuan station in July 2022. The blue shaded area represents the 90% prediction interval, while the solid line indicates the predicted mean value. The actual observations mostly lie within the shaded region, confirming that the intervals are well calibrated.

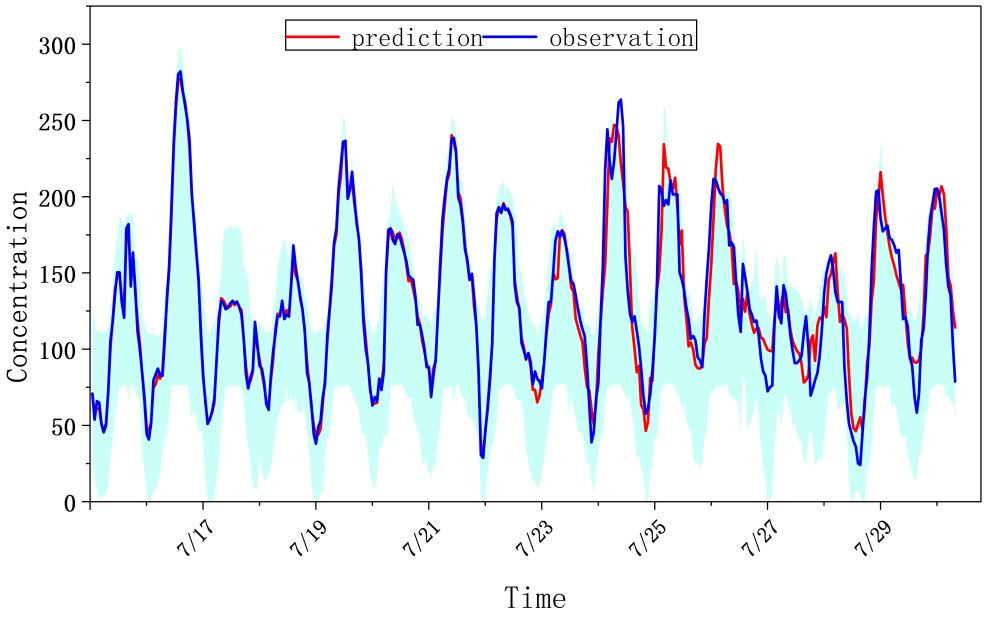


Figure 11. Prediction intervals of ozone concentration at XCGY station.

The predictive bands widen adaptively during periods associated with strong convective activity or abrupt changes in wind. Table 4 shows that grouping of test hours by the meteorological volatility index shows that the mean interval width increases from 99.44 to 112.61, indicating that a higher atmospheric instability translates into a greater uncertainty of the forecast. By contrast, under relatively steady conditions, the intervals are narrower,

showing that the model dynamically modulates its confidence according to the underlying atmospheric stability.

Table 4. Interval performance grouped by meteorological volatility.

Volatility Bin	Mean Width	Median Width
Steady conditions	99.44	96.61
High volatility	112.61	108.51

3.6. Ablation Study

To evaluate the contribution of each major component within the proposed WSDST-GAT, a comprehensive ablation study was conducted. Three key modules were selectively removed or replaced to analyze their individual impact on model performance under identical training and evaluation settings. The following variants were designed:

1. **Without Dynamic Graph (Static-GAT)**: The dynamic wind-driven adjacency matrix was replaced with a static spatial distance matrix.
2. **Without Transformer (GAT-only)**: The Transformer-based temporal encoder was replaced with a multi-layer LSTM.
3. **Without Meteorological Fusion**: Meteorological variables were removed, leaving only pollutant concentrations as input.

The quantitative results of these variants are reported in Table 5.

Table 5. Ablation results for major components of the proposed WSDST-GAT.

Model	MAE	RMSE	R^2
Full WSDST-GAT	6.02	10.12	0.91
(1) Static-GAT (No Dynamic Graph)	6.78	11.35	0.88
(2) GAT-only (No Transformer)	7.03	11.92	0.86
(3) No Meteorological Fusion	6.49	10.88	0.90

Ablation results consistently validate the necessity of each core component within the WSDST-GAT architecture. Removing any of the three modules leads to a noticeable decrease in predictive accuracy, demonstrating that each component contributes uniquely to the overall modeling capability.

From a climatic and statistical perspective, the removal of meteorological variables leads to a consistent increase in RMSE and MAE and a reduction in R^2. This performance degradation indicates that ozone concentration cannot be sufficiently explained by pollutant autocorrelation alone.

Meteorological variables introduce additional explanatory variance related to atmospheric thermodynamics and transport processes. Their exclusion reduces the model's capacity to capture climate-driven variability, thereby confirming the quantitative contribution of climatic information to predictive performance.

Specifically, replacing the dynamic graph with a static matrix weakens the model's ability to represent wind-driven anisotropic transport, while substituting the Transformer with an LSTM limits long-range temporal dependency modeling. The removal of meteorological inputs reduces the model's capacity to capture statistically meaningful climate–pollution interactions, confirming that atmospheric variability plays a critical role in ozone prediction.

Overall, these findings demonstrate that the integration of dynamic spatial modeling, temporal attention mechanisms, and meteorological feature fusion forms a coherent framework that jointly captures spatial transport dynamics, temporal persistence, and

climate-driven variability. This synergy enables WSDST-GAT to achieve superior prediction accuracy and robust generalization in ozone forecasting.

3.7. Spatial Enhancement Module: Co-Kriging Field Reconstruction Based on Meteorological Covariates

Spatial enhancement is implemented as a post-processing step. Specifically, the co-kriging module does not participate in model parameter optimization; instead, it operates on the predicted station-level ozone concentrations to reconstruct spatially continuous city-scale fields with quantified uncertainty. In this study, we employ the advanced co-kriging interpolation method. This technique not only utilizes geographic coordinate information such as longitude and latitude to model spatial correlations but also fully leverages multi-source data by integrating key meteorological covariates, including d2m, t2m, sp, u10, and v10.

Specifically, the co-kriging method thoroughly exploits the spatial cross-correlation between the target variable, ozone, and auxiliary variables, namely the meteorological fields. Based on point forecasts from monitoring stations, it considers not only the spatial distance and orientation between observation points through longitude and latitude but also incorporates closely related, continuously distributed meteorological field information. This process reconstructs a physically meaningful and information-rich continuous concentration surface. The resulting fine-scale regional field reflects the underlying dynamics of atmospheric dispersion and transport, while extending reliable estimates to areas with sparse or no monitoring stations, thereby improving spatial completeness and supporting spatially informed environmental decision-making.

Figure 12 shows the three-hour continuous Kriging interpolation forecast results. These spatially enhanced outputs provide high-value support for environmental policymaking and air-quality early warning. The continuous concentration surface enables precise identification of pollution hotspots and facilitates spatial exceedance detection and analysis against regulatory thresholds. This offers a direct, scientific basis for issuing health alerts, targeted deployment of regulatory inspections, and optimized allocation of emission reduction resources. By converting discrete point forecasts into complete, continuous regional concentration fields, this framework successfully bridges station-based observations with macro-level geospatial decision-making needs, delivering high-resolution maps that are ready for actionable urban air-quality management and governmental environmental decision support.

First step

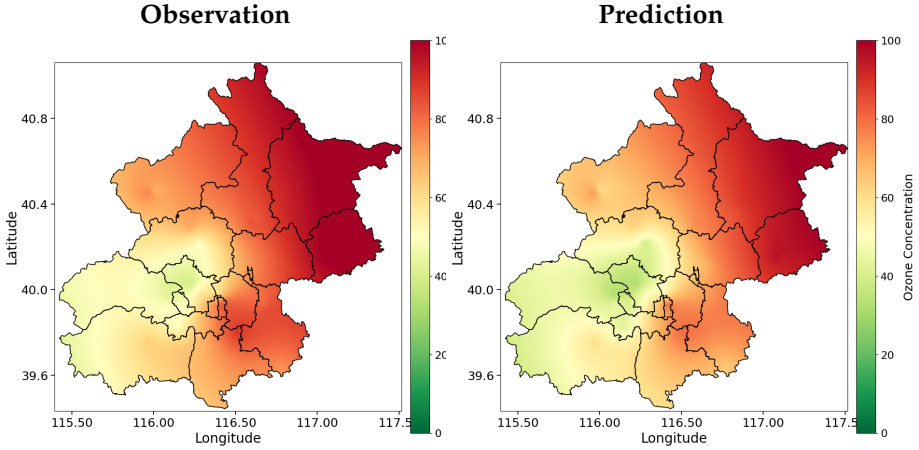


Figure 12. *Cont.*

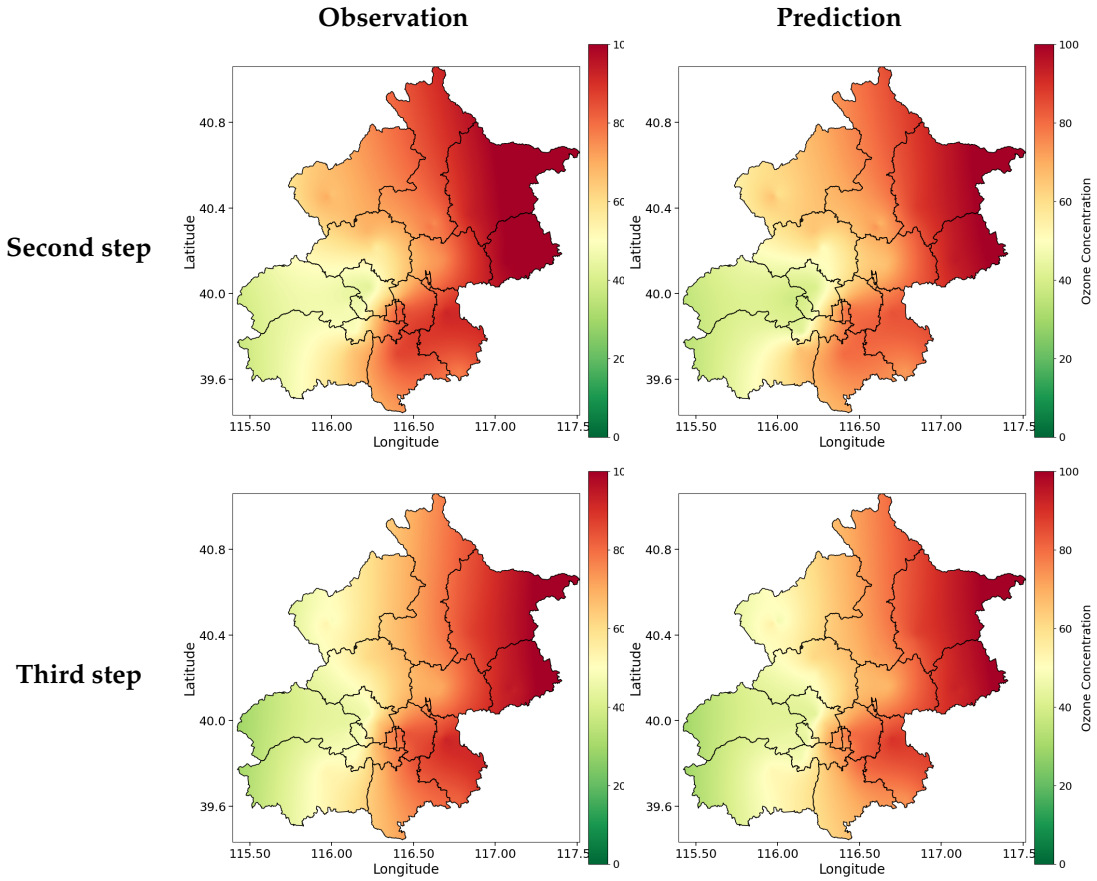


Figure 12. The visualization of the co-kriging interpolation results.

4. Conclusions

This study proposed a novel hybrid deep learning framework, the Wind Speed and Direction-Based Dynamic Spatiotemporal Graph Attention Network (WSDST-GAT), for the hourly prediction and interval estimation of ozone concentration in Beijing. By integrating physical meteorological knowledge with advanced neural network architectures, the model effectively captures both the spatial anisotropy of pollutant transport and the temporal dependencies of atmospheric processes.

The main innovations and contributions of this work are summarized as follows: (1) a wind-driven, time-varying directed graph is constructed to encode downwind transport, where edge directions follow the instantaneous wind field and edge weights reflect wind speed; meanwhile, a Graph Attention Network adaptively aggregates information from relevant neighbors; (2) a Transformer encoder–decoder is employed to capture multi-scale temporal patterns and long-range dependencies in hourly ozone series via multi-head attention; (3) a spatiotemporal model is developed to jointly support deterministic point forecasting and probabilistic interval estimation for risk-aware early warning; (4) a co-kriging module is integrated to convert station-level forecasts into continuous city-wide ozone fields with quantified uncertainty by incorporating key meteorological covariates.

Comprehensive experiments based on hourly data from 35 air-quality monitoring stations in Beijing demonstrated that the proposed model achieved superior performance, with an R^2 of 0.957, an MAE of 5.25 $\mu g/m^3$, and an RMSE of 9.58 $\mu g/m^3$ for point prediction. Furthermore, the results of the probabilistic interval prediction indicated high calibration and sharpness, with a PICP of 94.01% and PINAW of 0.174, confirming the model's reliability for uncertainty estimation. Comparative analysis against classical models and

ablation studies verified that each structural component—dynamic graph, Transformer, and meteorological fusion—contributed significantly to the model's predictive capability.

This study develops and evaluates the proposed model based on the existed dataset. Due to the accessibility of data, the performance assessment may be limited to some extent. Future work will incorporate larger-scale, multi-year data to enable a more comprehensive evaluation of model performance and to further examine its generalization capability and stability.

Although this study focuses on Beijing, the proposed framework is not region-specific. The wind-aware dynamic graph construction and spatiotemporal modeling architecture can be extended to other cities or regions, provided that fundamental air-quality and meteorological monitoring data are available. When applied to different geographical settings, minor adjustments, such as recalibration of spatial decay parameters or adaptation to local monitoring network density, may be required. Nevertheless, the overall modeling strategy remains applicable across regions with diverse terrain and meteorological characteristics.

The proposed WSDST-GAT provides an effective data-driven framework for fine-grained ozone forecasting under dynamic meteorological conditions. By explicitly encoding wind-driven transport and spatial neighborhood interactions, the model links atmospheric dynamics with the spatial–temporal observations of urban air-quality evolution. In this sense, WSDST-GAT supports understanding the spatial and temporal dimensions of urban systems through geospatial technologies: it can ingest observations from routine monitoring networks and meteorological fields from multi-source data and translate them into location-specific, time-resolved ozone forecasts that are necessary for urban contexts.

Overall, WSDST-GAT demonstrates a strong synergy between meteorological drivers and deep learning, offering a practical tool for urban air pollution early warning, smart environmental protection, and sustainable atmospheric governance.

Author Contributions: Funding acquisition, Xinyue Mo and Huan Li; methodology, Wenjie Wu, Xinyue Mo and Huan Li; software, Wenjie Wu; writing—original draft, Wenjie Wu, Xinyue Mo and Huan Li; writing—review and editing, Wenjie Wu, Xinyue Mo and Huan Li. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by Hainan Provincial Natural Science Foundation of China (Grant number: 623RC455, 623RC457, 425QN244), Scientific Research Fund of Hainan University (Grant number: KYQD (ZR)-22096, KYQD(ZR)-22097), Lanzhou University-Hainan University Technical Service Project (HD-KYH-2024424).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets supporting the findings of this study are available as follows: (1) Air Quality Data: The hourly pollutant concentration data for all studied areas are uniformly obtained from the Beijing Municipal Ecological and Environmental Monitoring Center (https://www.bjmemc.com.cn/ (accessed on 11 October 2025)); (2) Meteorological Data: The meteorological datasets are obtainable from the European Center for Medium-Range Weather Forecasting (https://cds.climate.copernicus.eu/ (accessed on 11 October 2025)). All data access dates were on 11 October 2025.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Distribution of air-quality monitoring stations.

Monitoring Station	Longitude	Latitude
Dongcheng Dongsi (DCDS)	116.417° E	39.929° N
Dongcheng Tiantan (DCT)	116.407° E	39.886° N
Fengtai Yungang (FTYG)	116.146° E	39.824° N
Fengtai Xiaotun (FTXT)	116.25528° E	39.87694° N
Yizhuang Kaifaqu (YZKFQ)	116.506° E	39.795° N
Jingdongnan Quyudian (JDNQ)	116.78437° E	39.63606° N
Daxing Jiugong (DXJG)	116.47456° E	39.78284° N
Daxing Huangcun (DXHC)	116.404° E	39.718° N
Dingling (DL)	116.22° E	40.292° N
Miyun Xincheng (MYXC)	116.85152° E	40.4088° N
Miyun Zhen (MYZ)	116.832° E	40.37° N
Pinggu Xincheng (PGXC)	117.0854° E	40.15353° N
Pinggu Zhen (PGZ)	117.118° E	40.143° N
Yanqing Xiadu (YQXD)	115.972° E	40.453° N
Yanqing Shiheying (YQSH)	116.00138° E	40.46327° N
Huairou Xincheng (HRXC)	116.6018° E	40.3118° N
Huairou Zhen (HRZ)	116.628° E	40.328° N
Fangshan Yanshan (FSYS)	115.96916° E	39.76419° N
Fangshan Liangxiang (FSLX)	116.136° E	39.742° N
Changping Nanshao (CPNS)	116.27603° E	40.21651° N
Changping Zhen (CPZ)	116.234° E	40.217° N
Chaoyang Nongzhanguan (CYNZG)	116.461° E	39.937° N
Chaoyang Aotizhongxin (CYAT)	116.397° E	39.982° N
Haidian Wanliu (HDWL)	116.287° E	39.987° N
Haidian Sijiqing (HDSJ)	116.23052° E	40.03° N
Shijingshan Gucheng (SJSGC)	116.176° E	39.914° N
Shijingshan Laoshan (SJSLS)	116.20764° E	39.90886° N
Xicheng Wanshou Xigong (XCWS)	116.352° E	39.878° N
Xicheng Guanyuan (XCGY)	116.339° E	39.929° N
Tongzhou Dongguan (TZDG)	116.6996° E	39.9131° N
Tongzhou Yongshun (TZYS)	116.67503° E	39.93435° N
Mentougou Sanjiadian (MTGSJ)	116.09122° E	39.96926° N
Mentougou Shuangyu (MTGSY)	116.106° E	39.937° N
Shunyi Beixiaoying (SYBX)	116.6853° E	40.16087° N
Shunyi Xincheng (SYXC)	116.655° E	40.127° N

References

1. Saxena, V. Water quality, air pollution, and climate change: Investigating the environmental impacts of industrialization and urbanization. *Water Air Soil Pollut.* **2025**, *236*, 73. [CrossRef]
2. Appannagari, R.R. Environmental pollution causes and consequences: A study. *North Asian Int. Res. J. Soc. Sci. Humanit.* **2017**, *3*, 151–161.
3. Yafouz, A.; AlDahoul, N.; Birima, A.H.; Ahmed, A.N.; Sherif, M.; Sefelnasr, A.; Allawi, M.F.; Elshafie, A. Comprehensive comparison of various machine learning algorithms for short-term ozone concentration prediction. *Alex. Eng. J.* **2022**, *61*, 4607–4622. [CrossRef]
4. Mo, X.; Li, H.; Zhang, L. Design a regional and multistep air quality forecast model based on deep learning and domain knowledge. *Front. Earth Sci.* **2022**, *10*, 995843. [CrossRef]
5. Brauer, M.; Roth, G.A.; Aravkin, A.Y.; Zheng, P.; Abate, K.H.; Abate, Y.H.; Abbafati, C.; Abbasgholizadeh, R.; Abbasi, M.A.; Abbasian, M.; et al. Global burden and strength of evidence for 88 risk factors in 204 countries and 811 subnational locations, 1990–2021: A systematic analysis for the Global Burden of Disease Study 2021. *Lancet* **2024**, *403*, 2162–2203. [CrossRef] [PubMed]
6. Gao, J.; Woodward, A.; Vardoulakis, S.; Kovats, S.; Wilkinson, P.; Li, L.; Xu, L.; Li, J.; Yang, J.; Cao, L.; et al. Haze, public health and mitigation measures in China: A review of the current evidence for further policy response. *Sci. Total Environ.* **2017**, *578*, 148–157. [CrossRef]

7. Atkinson, R.W.; Butl, B.K.; Dimitroulopoulou, C.; Heal, M.R.; Stedman, J.R.; Carslaw, N.; Jarvis, D.; Heaviside, C.; Vardoulakis, S.; Walton, H.; et al. Long-term exposure to ambient ozone and mortality: A quantitative systematic review and meta-analysis of evidence from cohort studies. *BMJ Open* **2016**, *6*, e009493. [CrossRef] [PubMed]
8. Wei, J.; Li, Z.; Lyapustin, A.; Sun, L.; Peng, Y.; Xue, W.; Su, T.; Cribb, M. Reconstructing 1-km-resolution high-quality PM2.5 data records from 2000 to 2018 in China: Spatiotemporal variations and policy implications. *Remote Sens. Environ.* **2021**, *252*, 112136. [CrossRef]
9. Mo, X.; Li, H.; Zhang, L.; Qu, Z. Environmental impact estimation of PM2.5 in representative regions of China from 2015 to 2019: Policy validity, disaster threat, health risk, and economic loss. *Air Qual. Atmos. Health* **2021**, *14*, 1571–1585. [CrossRef]
10. Mandal, S.; Boppani, S.; Dasari, V.; Thakur, M. A bivariate simultaneous pollutant forecasting approach by Unified Spectro-Spatial Graph Neural Network (USSGNN) and its application in prediction of O_3 and NO_2 for New Delhi, India. *Sustain. Cities Soc.* **2024**, *114*, 105741. [CrossRef]
11. Wang, Q.; Liu, H.; Li, Y.; Li, W.; Sun, D.; Zhao, H.; Tie, C.; Gu, J.; Zhao, Q. Predicting plateau atmospheric ozone concentrations by a machine learning approach: A case study of a typical city on the southwestern plateau of China. *Environ. Pollut.* **2024**, *363*, 125071. [CrossRef]
12. Wang, S.; Sun, Y.; Gu, H.; Cao, X.; Shi, Y.; He, Y. A deep learning model integrating a wind direction-based dynamic graph network for ozone prediction. *Sci. Total Environ.* **2024**, *946*, 174229. [CrossRef] [PubMed]
13. López, J.C.S.; Silva, J.S.Z.; Ruiz-Vanoye, J.A.; Simancas-Acevedo, E.; Salgado-Ramírez, J.C.; Díaz-Parra, O. Prediction of PM_{10}, SO_2, NO_2, O_3, and CO Concentrations in Guadalajara Using ARIMA and Open Data with Python. *Int. J. Comb. Optim. Probl. Inform.* **2025**, *16*, 25.
14. Gao, Z.; Zhou, X. A review of the CAMx, CMAQ, WRF-Chem and NAQPMS models: Application, evaluation and uncertainty factors. *Environ. Pollut.* **2024**, *343*, 123183. [CrossRef] [PubMed]
15. Chuang, M.T.; Zhang, Y.; Kang, D. Application of WRF/Chem-MADRID for real-time air quality forecasting over the Southeastern United States. *Atmos. Environ.* **2011**, *45*, 6241–6250. [CrossRef]
16. Luque, S.; Pineda Rojas, A.L.; Fita, L.; Borge, R. Sensitivity analysis of NO_2 and O_3 concentrations modeled with WRF-CMAQ to boundary and initial conditions and first model layer height in the Metropolitan Area of Buenos Aires, Argentina. *Air Qual. Atmos. Health* **2025**, *18*, 3843–3855. [CrossRef]
17. Fang, C.; Wang, Y.; Han, S.; Wang, X.; Wang, J. Attribution of a Typical Ozone Pollution Episode in Handan: WRF-CMAQ modeling and Process Analysis Based on a Local Emission Inventory. *Atmos. Pollut. Res.* **2025**, *16*, 102862. [CrossRef]
18. da Silva, L.F.M.; Deroubaix, A.; Brasseur, G.P.; Gioda, A. Application of WRF-Chem for predicting air quality resulting from the formation of photochemical compounds in a subtropical urban environment. *Urban Clim.* **2026**, *65*, 102733. [CrossRef]
19. Mousavinezhad, S.; Choi, Y.; Pouyaei, A.; Ghahremanloo, M.; Nelson, D.L. A comprehensive investigation of surface ozone pollution in China, 2015–2019: Separating the contributions from meteorology and precursor emissions. *Atmos. Res.* **2021**, *257*, 105599. [CrossRef]
20. Quispe, F.; Salcedo, E.; Iftikhar, H.; Zafar, A.; Khan, M.; Turpo-Chaparro, J.E.; Rodrigues, P.C.; López-Gonzales, J.L. Multi-step ahead ozone level forecasting using a component-based technique: A case study in Lima, Peru. *AIMS Environ. Sci.* **2024**, *11*, 3. [CrossRef]
21. Yan, Z. Comparative Analysis of ARMA and ARIMA Models for Air Quality Prediction. *Adv. Eng. Technol. Res.* **2025**, *14*, 1349. [CrossRef]
22. Salazar, L.; Nicolis, O.; Ruggeri, F.; Kisel'ák, J.; Stehlík, M. Predicting hourly ozone concentrations using wavelets and ARIMA models. *Neural Comput. Appl.* **2019**, *31*, 4331–4340. [CrossRef]
23. Napi, N.N.L.M.; Abdullah, S.; Mansor, A.A.; Ghazali, N.A.; Ahmed, A.N.; Dom, N.C.; Ismail, M. Different approaches of multiple linear regression (MLR) model in predicting ozone (O_3) concentration in industrial area. *Int. J. Integr. Eng.* **2023**, *15*, 106–117. [CrossRef]
24. Yang, Z.; Li, Z.; Cheng, F.; Lv, Q.; Li, K.; Zhang, T.; Zhou, Y.; Zhao, B.; Xue, W.; Wei, J. Two-decade surface ozone (O_3) pollution in China: Enhanced fine-scale estimations and environmental health implications. *Remote Sens. Environ.* **2025**, *317*, 114459. [CrossRef]
25. Lu, Y.; Mo, X.; Li, H. GNADET: A geospatial neural advection–diffusion equation framework with graph transformer for surface temperature forecasting. *Mach. Learn. Sci. Technol.* **2026**, *7*, 015006. [CrossRef]
26. Kapadia, D.; Jariwala, N. Prediction of tropospheric ozone using artificial neural network (ANN) and feature selection techniques. *Model. Earth Syst. Environ.* **2022**, *8*, 2183–2192. [CrossRef]
27. Su, X.; An, J.; Zhang, Y.; Zhu, P.; Zhu, B. Prediction of ozone hourly concentrations by support vector machine and kernel extreme learning machine using wavelet transformation and partial least squares methods. *Atmos. Pollut. Res.* **2020**, *11*, 51–60. [CrossRef]
28. Hosseinpour, F.; Kumar, N.; Tran, T.; Knipping, E. Using machine learning to improve the estimate of US background ozone. *Atmos. Environ.* **2024**, *316*, 120145. [CrossRef]

29. Ahmed, S.F.; Alam, M.S.B.; Hassan, M.; Rozbu, M.R.; Ishtiak, T.; Rafa, N.; Mofijur, M.; Shawkat Ali, A.B.M.; Gandomi, A.H. Deep learning modelling techniques: Current progress, applications, advantages, and challenges. *Artif. Intell. Rev.* **2023**, *56*, 13517–13521. [CrossRef]
30. Eslami, E.; Choi, Y.; Lops, Y.; Sayeed, A. A real-time hourly ozone prediction system using deep convolutional neural network. *Neural Comput. Appl.* **2020**, *32*, 8783–8797. [CrossRef]
31. Salman, A.K.; Choi, Y.; Singh, D.; Kayastha, S.G.; Dimri, R.; Park, J. Temporal CNN-based 72-h ozone forecasting in South Korea: Explainability and uncertainty quantification. *Atmos. Environ.* **2025**, *343*, 120987. [CrossRef]
32. Sangiorgio, M.; Guariso, G. Transfer learning in environmental data-driven models: A study of ozone forecast in the Alpine region. *Environ. Model. Softw.* **2024**, *177*, 106048. [CrossRef]
33. Huang, G.; Li, H.; Mo, X. A novel point-interval prediction model for wind speed based on hybrid deep learning and RIME optimization algorithms. *Energy Rep.* **2025**, *14*, 3977–3992. [CrossRef]
34. Zhang, D.; Hou, Y.; Zhao, B.; Wang, X. A prediction method for ozone concentration based on GCN-BiLSTM-Attention. In Proceedings of the 2024 10th International Conference on Computer and Communications (ICCC), Chengdu, China, 13–16 December 2024; pp. 105–110.
35. Mu, L.; Bi, S.; Ding, X.; Xu, Y. Transformer-based ozone multivariate prediction considering interpretable and priori knowledge: A case study of Beijing, China. *J. Environ. Manag.* **2024**, *366*, 121883. [CrossRef] [PubMed]
36. Hu, W.; Zhang, Z.; Zhang, S.; Chen, C.; Yuan, J.; Yao, J.; Zhao, S.; Guo, L. Learning spatiotemporal dependencies using adaptive hierarchical graph convolutional neural network for air quality prediction. *J. Clean. Prod.* **2024**, *459*, 142541. [CrossRef]
37. Xu, J.; Wang, S.; Ying, N.; Xiao, X.; Zhang, J.; Jin, Z.; Cheng, Y.; Zhang, G. Dynamic graph neural network with adaptive edge attributes for air quality predictions. *arXiv* **2023**, arXiv:2302.09977. [CrossRef]
38. Chen, Q.; Ding, R.; Mo, X.; Li, H.; Xie, L.; Yang, J. An adaptive adjacency matrix-based graph convolutional recurrent network for air quality prediction. *Sci. Rep.* **2024**, *14*, 4408. [CrossRef] [PubMed]
39. Muñoz-Sabater, J.; Dutra, E.; Agustí-Panareda, A.; Albergel, C.; Arduini, G.; Balsamo, G.; Boussetta, S.; Choulga, M.; Harrigan, S.; Hersbach, H.; et al. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data* **2021**, *13*, 4349–4383. [CrossRef]
40. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2006.
41. Goovaerts, P. *Geostatistics for Natural Resources Evaluation*; Oxford University Press: New York, NY, USA, 1997.
42. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph attention networks. In Proceedings of the International Conference on Learning Representations (ICLR), Vancouver, BC, Canada, 30 April–3 May 2018.
43. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, Long Beach, CA, USA, 4–9 December, 2017; Volume 30, pp. 5998–6008.
44. Kafi, F.; Yousefi, E.; Ehteram, M.; Ashrafi, K. Stabilized Long Short Term Memory (SLSTM) model: A new variant of the LSTM model for predicting ozone concentration data. *Earth Sci. Inform.* **2025**, *18*, 311. [CrossRef]
45. Cherukat, S.; Niharika, G.; Raghavendran, S.; Bajpai, I.K.; Lalmohan, K.S. Comparative analysis of LSTM, GRU, and random forest regression for air quality prediction. In *NIELIT's International Conference on Communication, Electronics and Digital Technologies*; Springer: Berlin/Heidelberg, Germany, 2025; pp. 327–341.
46. Sun, L.; Lan, Y.; Liang, X.; Sun, X.; Nie, H.; Su, Y.; He, Y.; Wang, J.; Xia, D. The spatio-temporal prediction of ozone in Zhuhai based on graph convolutional memory network. *Acta Sci. Nat. Univ. SunYatseni* **2024**, *63*, 48–59.
47. Niresi, K.F.; Zhao, M.; Bissig, H.; Baumann, H.; Fink, O. Spatial-temporal graph attention fuser for calibration in IoT air pollution monitoring systems. In Proceedings of the 2023 IEEE SENSORS, Vienna, Austria, 29 October–1 November 2023; pp. 1–4.
48. Hickman, S.H.; Griffiths, P.T.; Nowack, P.J.; Archibald, A.T. Short-term forecasting of ozone air pollution across Europe with transformers. *Environ. Data Sci.* **2023**, *2*, e43. [CrossRef]
49. Yang, T.; Li, S.; Chen, B. Spatial and temporal prediction of ozone concentration in the Pearl River Delta region based on a dynamic graph convolutional network. *J. Atmos.-Sol.-Terr. Phys.* **2025**, *273*, 106559. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.