sustainability

MDPI

Article

Hourly $PM_{2.5}$ Forecasting for Sustainable Urban Air Quality Management Through Multiresolution Learning and Environmental Context

Shulei Cheng [1], Xin Wen [2], Jiaying Xiao [3], Wei Fan [4] and Chong Zhou [3,*]

[1] School of Public Administration, Southwestern University of Finance and Economics, Chengdu 611130, China; chengsl@swufe.edu.cn
[2] School of Law, Southwestern University of Finance and Economics, Chengdu 611130, China; 120030106002@smail.swufe.edu.cn
[3] School of Finance, Southwestern University of Finance and Economics, Chengdu 611130, China; 42304230@smail.swufe.edu.cn
[4] School of Economics, Xinjiang College of Science & Technology, Korla 841000, China; 2026628@xjust.edu.cn
* Correspondence: 124020204002@smail.swufe.edu.cn

Abstract

Reliable hourly forecasts of fine particulate matter ($PM_{2.5}$) can help cities anticipate deteriorating air quality and prepare measures to limit exposure. A central modeling challenge is to combine rapid pollution fluctuations with meteorological and ecological observations that are updated less frequently. This study develops an adaptive frequency gated network (AFGN) that integrates hourly pollution histories, daily meteorological records, and satellite ecological indicators. Parallel temporal convolutions and recurrent encoders represent concentration changes at multiple resolutions, while attention and signed gating incorporate environmental conditions into the forecast. The framework is evaluated using observations from Beijing, Shanghai, Chengdu, and Wuhan during 2023 and 2024. AFGN achieves lower mean prediction error than the gated recurrent unit (GRU) baseline at 1, 6, and 12 h. At 6 h, mean root mean square error (RMSE) decreases from 13.498 to 12.526 $\mu g/m^3$, corresponding to an absolute reduction of 0.971 $\mu g/m^3$ and a relative reduction of 7.20%. These results demonstrate the value of combining multiresolution pollution representations with environmental context for short-range forecasting. The approach provides concentration estimates that could inform the timing of pollution surveillance and exposure advisories as part of sustainable urban air quality management.

Keywords: sustainable urban management; hourly $PM_{2.5}$ forecasting; multiresolution learning; meteorological conditions; satellite ecological indicators

Check for updates

Academic Editor: Jianming Cai

Received: 4 September 2026
Revised: 13 October 2026
Accepted: 14 October 2026
Published: 16 October 2026

Copyright: © 2026 by the authors. 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

Reducing exposure to fine particulate matter ($PM_{2.5}$) is an important objective of sustainable urban development because of its substantial health burden [1,2]. To curb the deleterious health effects of excessive $PM_{2.5}$, emerging economies have instituted stringent environmental laws and administrative regulations as a primary policy instrument. Long-term emission reduction, urban ecological planning, and strict judicial implementation address the sources and environmental conditions of pollution, whereas short-term forecasts provide time to prepare for deteriorating air quality. Blue and green infrastructure (BGI), including vegetation and water bodies, contributes to this urban environment and to wider ecosystem services [3–5]. Vegetation can remove particles by deposition and modify

local temperature and airflow [6–8], although its effects on near-surface concentrations depend on canopy structure, surrounding buildings and ventilation [9–12]. Reliable forecasts must therefore describe concentration changes within a variable environmental setting, rather than assume that a particular ecological state always corresponds to lower pollution.

How can hourly $PM_{2.5}$ forecasts represent rapid concentration changes while incorporating meteorological and ecological observations collected at longer intervals? Recent pollution records contain short fluctuations superimposed on more persistent accumulation and dispersal. A useful representation must distinguish these patterns while retaining their dependence on the surrounding environment. Meteorology affects pollutant transport and accumulation [13], and winter concentrations and extreme pollution in Chinese cities are closely associated with meteorological conditions [14,15]. Satellite observations add information on vegetation, surface moisture and temperature, although surface thermal conditions partly covary with air temperature [16]. Their relevance to ecological planning [17,18] therefore raises a specific forecasting question: do these observations add predictive information beyond recent pollution and weather?

$PM_{2.5}$ forecasting methods increasingly represent nonlinear temporal dependence, spatial relationships, and variation at multiple scales [19–22]. Integrating environmental observations requires a complementary distinction between rapidly changing sequences and more persistent states. In our dataset, $PM_{2.5}$ is observed hourly, weather is summarized daily, and satellite products are updated daily or at longer intervals. The normalized difference vegetation index (NDVI), normalized difference water index (NDWI), leaf area index (LAI), and land surface temperature (LST) can consequently describe the environmental setting of several successive forecasts. We encode these indicators separately from pollution dynamics and use them to adjust the combination of temporal features. This design accommodates their different update frequencies; whether the ecological inputs improve prediction is assessed through an otherwise identical model that omits them.

The proposed adaptive frequency gated network (AFGN) combines these two forms of representation. Parallel convolutions capture pollution changes over different receptive fields within a historical window, and recurrent encoders learn their temporal evolution. Meteorological and ecological encodings guide attention across the temporal features, followed by a signed residual adjustment. We evaluate the complete model against recurrent, linear, and attention-based alternatives in Beijing, Shanghai, Chengdu, and Wuhan over forecast horizons of 1–48 h. A separate input ablation measures the contribution of satellite information with the architecture held constant. This distinction is central to the study: the performance of a forecasting framework, the predictive value of an ecological input, and the effect of a physical BGI intervention are different questions. Our experiments address the first two to inform short-term urban air quality management; they do not estimate pollution reductions caused by BGI construction.

2. Literature Review

2.1. Meteorological and Ecological Influences on Urban $PM_{2.5}$

Studies of urban vegetation and air quality have moved from estimating particle removal to examining how deposition interacts with dispersion. Estimates of removal by trees and reviews of urban cooling demonstrate that vegetation can alter pollutant removal and local thermal conditions [6–8]. Research on urban form, street canyons and roadside barriers further shows that these effects depend on the arrangement of vegetation relative to buildings and emission sources [9–12]. Increased deposition does not necessarily imply a lower concentration at every location if airflow is also restricted. For forecasting, this literature suggests that ecological observations should be interpreted

together with meteorological conditions, rather than assigned a uniformly beneficial effect on predicted concentration.

Satellite observations provide spatially extensive descriptions of the urban environment, but they do not directly measure particle removal or pollutant transport. NDVI and LAI characterize vegetation and canopy conditions, NDWI reflects water or surface moisture, and LST describes the surface thermal environment. The association between surface and air temperature [16] also indicates that satellite and meteorological predictors may contain overlapping information. Moreover, city averages and intermittent retrievals cannot resolve the street-level processes examined in vegetation and dispersion studies. Research on ecosystem services and nature-based planning establishes why ecological conditions matter for urban management [4,5,17,18]. Their role in hourly forecasting remains a separate empirical question, requiring a comparison that changes the ecological inputs while retaining the pollution history, weather information, and network structure.

2.2. From Temporal Dependence to Multivariable Forecasting

Early statistical approaches to $PM_{2.5}$ forecasting described concentration persistence and lagged relationships, whereas subsequent nonlinear methods allowed more flexible responses to pollution and meteorological predictors [23]. Recurrent neural networks extended this development by learning directly from ordered observations. Long short-term memory (LSTM) and gated recurrent unit (GRU) retain or update historical information through gated states [24,25], making them natural baselines for hourly concentration sequences. However, their recurrent structure does not explicitly distinguish short fluctuations from slower changes. Graph-based extensions address spatial dependence: combining graph convolution with LSTM allows a forecast to draw on relationships among monitoring locations [19,26]. These methods can incorporate transport-related information from neighboring sites, but their performance also depends on the availability and spatial coverage of those observations. For city-specific forecasting, temporal representation and the use of local environmental covariates remain important design choices.

Attention mechanisms allow models to relate observations at different positions in a sequence without relying solely on recurrent state updates [27]. Informer improves the efficiency of attention over long inputs, while Autoformer combines decomposition and autocorrelation to represent recurring patterns [28,29]. PatchTST groups adjacent observations into patches, iTransformer applies attention across variables, and TimesNet represents variation within and between periods [30–32]. Together, these approaches offer different ways to organize temporal and multivariable dependence. For environmental forecasting, a further consideration is how covariates enter the prediction. Temporal Fusion Transformers (TFT) use variable selection and gating to combine heterogeneous inputs with recurrent and attention-based representations [33]. At the simpler end of the spectrum, the decomposition-based linear model (DLinear) separates a sequence into trend and residual components and applies linear prediction to each [34]. These contrasting designs motivate testing whether AFGN's multiresolution encoding and environmental fusion improve on both a parsimonious decomposition model and established attention-based alternatives.

Recent air quality studies combine temporal learning with observations from multiple sites, numerical forecasts, or data assimilation [35–37]. These approaches extend the information available for prediction, whereas their forecasting architectures determine how that information is used. Zhou et al. [23] also identify datasets and evaluation procedures as important considerations when comparing methods. For the present task, two questions therefore need to be distinguished: how effectively a model represents the available pollution and environmental observations, and whether satellite information adds predictive value when pollution history and meteorology are already included. Baseline comparisons

address the first question; an otherwise identical network with ecological inputs excluded addresses the second.

2.3. Multiscale Representation and Environmental Information Fusion

Signal decomposition separates a sequence into components before forecasting, allowing slower variation and short fluctuations to be modeled differently. Empirical mode decomposition (EMD) was developed for nonlinear, nonstationary signals; ensemble and noise-assisted extensions, including complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), address difficulties in extracting its components [38–40]. Variational mode decomposition (VMD) instead estimates modes under a bandwidth constraint [41]. $PM_{2.5}$ studies have combined EMD, wavelet-packet decomposition, or VMD with subsequent prediction models [20–22]. For example, the adaptive variational mode decomposition-based PMNet (AVMD-PMNet) combines adaptive VMD, multivariate temporal graph learning and GRU-based fusion for 12 h prediction [22]. These studies motivate representing heterogeneous temporal variation explicitly. They also leave open how to choose a representation that remains useful when meteorological conditions change, and ecological observations update infrequently.

Other models learn temporal structure as part of the forecasting network. The frequency enhanced decomposed Transformer (FEDformer) combines seasonal-trend decomposition with frequency-enhanced attention, and TimesNet organizes variation around multiple periods [32,42]. Decomposition has also been combined with recurrent learning in industrial forecasting [43], although results in that domain do not establish its effectiveness for $PM_{2.5}$. Here, we couple learned temporal representations with environmental states: AFGN optimizes parallel convolutions jointly with the forecast objective, then uses meteorological and ecological encodings to weight and adjust their outputs. The approach allows the use of temporal features to depend on the conditions surrounding a forecast. Its learned components are not constrained to form distinct physical frequency bands, and their contribution is evaluated through component ablations.

The construction of temporal features must also respect the order in which observations become available. Fitting a data-adaptive decomposition to the complete series before evaluation can allow later observations to influence earlier forecasts. AFGN instead constructs its convolutional representations within each historical input window. Imputation and scaling are fitted to training data, and environmental products are assigned according to their availability at the forecast origin. Combined with chronological evaluation [44], these restrictions prevent future observations from entering the prediction through preprocessing or feature construction.

3. Data and Methodology

3.1. Study Area and Data Sources

The analysis covers four Chinese cities featuring diverse climatic conditions, environmental judicial supervision, and pollution characteristics. Beijing (39.9° N, 116.4° E) is a northern plain metropolis. Shanghai (31.2° N, 121.5° E) serves as a coastal metropolis. Chengdu (30.6° N, 104.1° E) is located within the Sichuan Basin, while Wuhan (30.6° N, 114.3° E) features a complex hydrological system of rivers and lakes. Figure 1 illustrates their geographical locations.



Figure 1. Study areas in China. Notes: The four core cities represent northern plain, coastal delta, basin, and river and lake urban settings. The inset provides the South China Sea context. Map review approval number GS (2023) 2767.

The geographical contrast among these cities is accompanied by differences in pollution distributions and meteorological conditions. Models are trained separately for each city, allowing the same forecasting design to be evaluated across four urban environments. The experiments assess prediction within each city, not transfer to a city without local training data.

Hourly city-level $PM_{2.5}$ observations cover 1 January 2023 to 31 December 2024. They originate from the China National Environmental Monitoring Centre (CNEMC) national urban air quality platform and were obtained through archived records. Target observations are not interpolated. Each 48 h input window must contain at least 90% valid $PM_{2.5}$ observations; missing predictor values are filled using preceding observations or statistics estimated from training data. Meteorological observations come from the National Oceanic and Atmospheric Administration (NOAA) Global Surface Summary of the Day (GSOD) archive. We combine stations within 150 km of each city using inverse squared-distance weights. Chen et al. [45] provide an earlier application of inverse-distance weighting to meteorological records; the GSOD source, 150 km radius, and daily resolution are the specifications used in the present study.

Inverse-distance weighting within 150 km (IDW150) uses 13, 17, 17 and 10 GSOD stations for Beijing, Shanghai, Chengdu and Wuhan, respectively. The station weight is $wi,c = 1/(di,c^2 + \varepsilon)$, where di,c is the distance between station i and city c, and ε prevents division by zero. The resulting variables represent daily meteorological conditions and are aligned with the hourly prediction times according to availability. Satellite inputs are derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) MOD13Q1 v061 for NDVI, MOD09GA v061 surface reflectance for NDWI, MCD15A3H v061 for LAI, and MOD11A1/MYD11A1 v061 for LST. After quality filtering, values are aggregated within a 30 km city buffer. At each forecast origin, the latest available retrieval is retained only within the prescribed maximum age; initial gaps use training medians. No later retrieval is used to fill an earlier input. Valid retrieval coverage before temporal alignment ranges from 70.5% to 100% across the four indicators and cities, with product-specific values in Table A1.

$$X_t = [y_{t-L+1:t} \| M_{t-L+1:t} \| B_t \| (B_t \otimes M_t) \| \tau_t] \quad (1)$$

In this expression, $y_{t-L+1:t}$ is the historical $PM_{2.5}$ sequence, $M_{t-L+1:t}$ is the meteorological covariate sequence, and B_t is the latest BGI state.

The operator $\otimes$ constructs ecological–meteorological interactions, τ_t denotes time encoding, and $\|$ denotes feature concatenation. The prediction target is the observed $PM_{2.5}$ concentration y_{t+h} at lead time h. Data are divided chronologically before imputation and scaling are fitted to the training subset. Environmental observations are assigned to a forecast only when they are available at its origin. Convolutional representations are then calculated within that forecast's historical input window.

Table 1 summarizes the original $PM_{2.5}$ observations used in the evaluation. Meteorological and satellite records are aligned to hourly forecast origins without changing their underlying observation frequency. As a separate data consistency check, the pollution records are matched by city and exact hourly timestamp to the published archive of Silver et al. [46]. The matching statistics and paired concentrations are reported in the Appendix A.

Figures 2–5 describe the temporal variation of pollution and its association with ecological conditions. These analyses motivate the distinction between rapidly changing concentrations and more persistent environmental states. Their decomposed signals are not supplied to the forecasting network. The model instead learns its temporal representations from each historical pollution window and combines them with the meteorological and satellite observations available at that time.

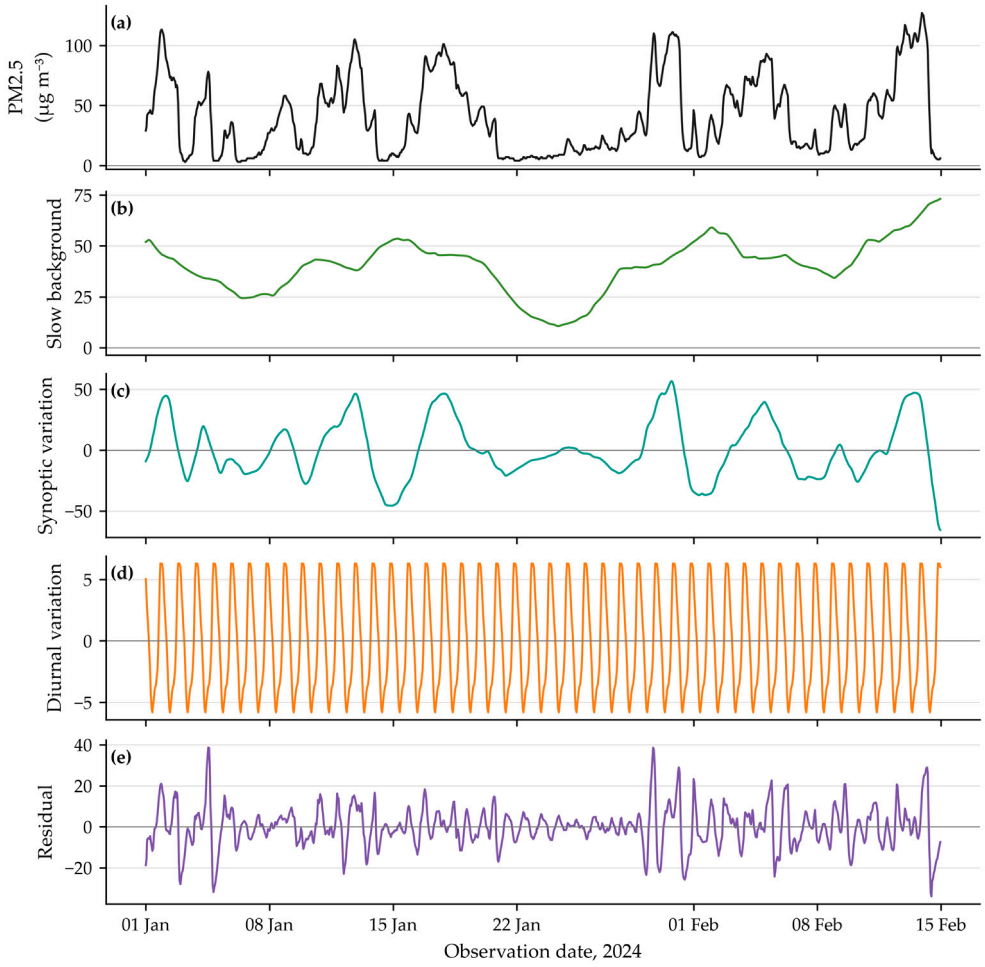


Figure 2. Descriptive multiscale decomposition of an hourly $PM_{2.5}$ segment. Subfigures show (**a**) observed concentration; (**b**) slow background; (**c**) synoptic variation; (**d**) diurnal variation; and (**e**) the residual. Notes: The decomposition is descriptive and is calculated from observed hourly $PM_{2.5}$. It illustrates temporal variation at several scales and is not the learned filter bank used by the adaptive frequency gated network (AFGN).

Table 1. Fine Particulate Matter ($PM_{2.5}$) observations in the four core cities, 2023 to 2024.

City	Valid Hours	Coverage (%)	Mean	SD	Median	P99	Range
Beijing	17,390	99.13	33.0	31.8	23	150.1	1 to 439
Shanghai	17,388	99.12	28.6	23.1	22	114.0	3 to 378
Chengdu	17,390	99.13	35.7	28.0	28	142.0	2 to 201
Wuhan	17,390	99.13	37.5	30.1	28	151.0	3 to 330

Notes: $PM_{2.5}$ is measured in $\mu g/m^3$. Coverage is the proportion of expected hourly observations retained after source-level quality control. SD denotes standard deviation, and P99 denotes the 99th percentile. The statistics describe the untrimmed observations used for the main evaluation.

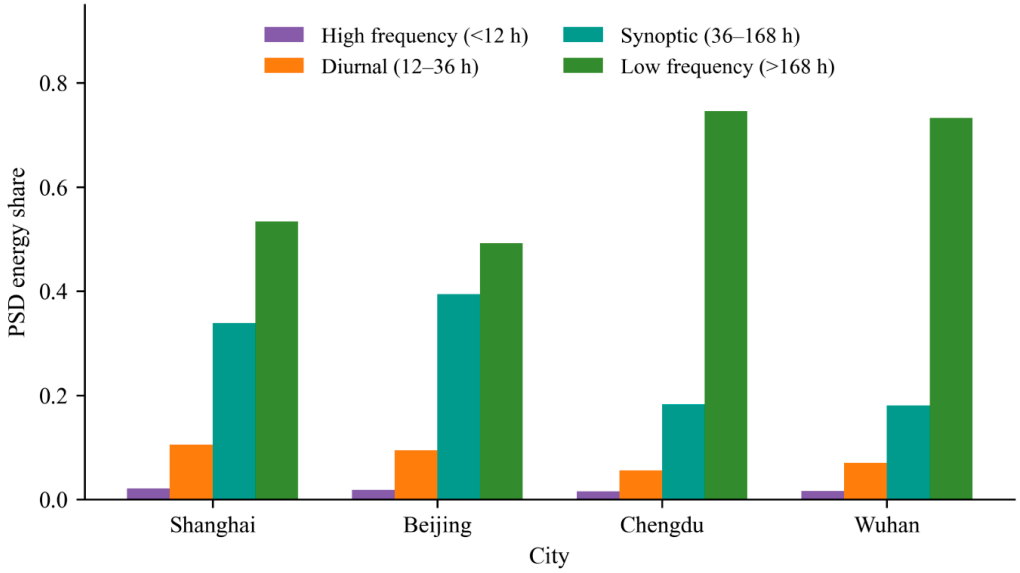


Figure 3. Normalized power spectral density (PSD) across descriptive frequency ranges in the four core cities. Notes: Power spectral density is normalized within each city. The ranges are high frequency below 12 h, diurnal from 12 to 36 h, synoptic from 36 to 168 h, and low frequency above 168 h.

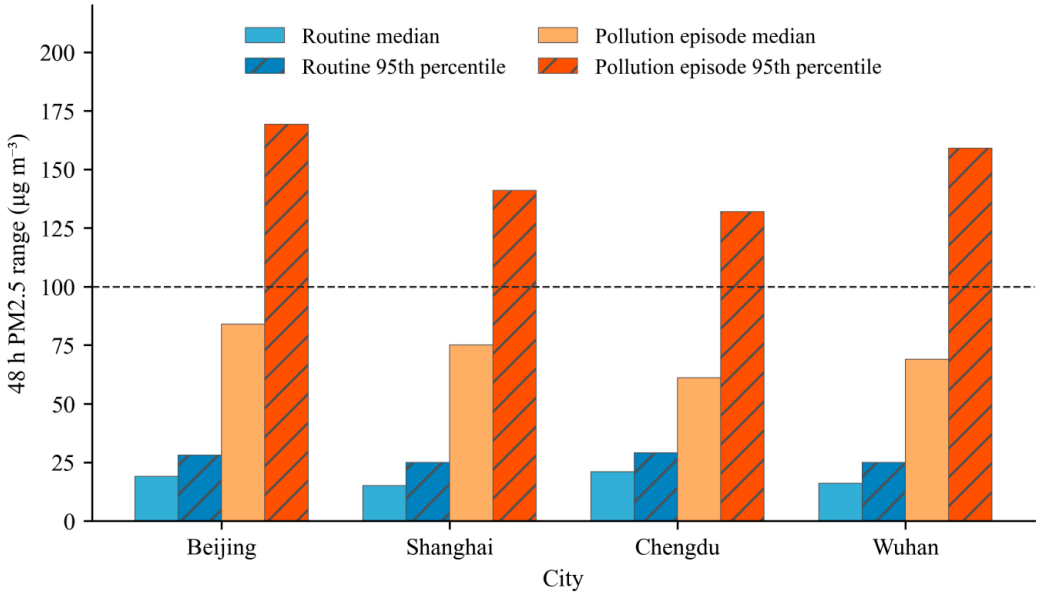


Figure 4. Temporal scale contrast between slowly updated blue and green infrastructure (BGI) indicators and hourly $PM_{2.5}$. Notes: Bars summarize 48 h $PM_{2.5}$ ranges during routine periods and pollution episodes. The horizontal dashed line marks a range of 100 $\mu g/m^3$. Satellite BGI indicators update more slowly than hourly pollution observations.

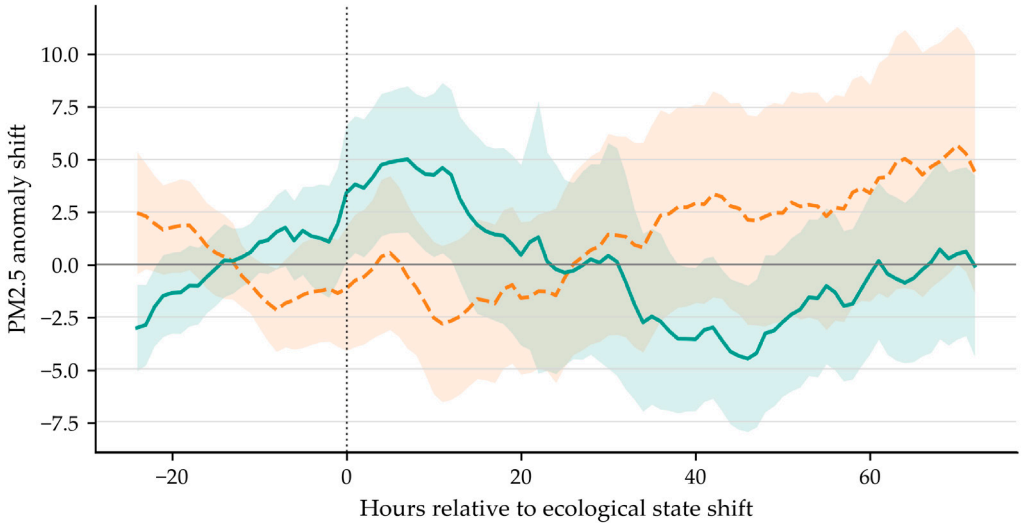


Figure 5. Event-centered association between vegetation state and $PM_{2.5}$ anomaly. Notes: Curves show $PM_{2.5}$ anomalies before and after vegetation-state transitions; shading summarizes variation across events. The comparison describes an association, not the effect of an ecological intervention.

3.2. Adaptive Frequency Gated Network

AFGN represents each 48 h $PM_{2.5}$ window through parallel temporal convolutions. One learned smoothing filter has seven coefficients; five further filters have kernel size five and dilation factors 1, 2, 4, 8, and 12. These different receptive fields allow the network to describe local fluctuations and broader changes within the window. The filters are optimized jointly with the prediction objective, so their outputs are learned temporal representations rather than predefined atmospheric frequency bands:

$$b_t^{(1)} = \sum_{j=-3}^{3} \mathrm{h}_j p_{t-j}, b_t^{(k)} = p_t - \sum_{j=-2}^{2} g_j^{(k)} p_{t-d\mathrm{k}j}, \mathrm{k} = 2, \ldots, \mathrm{K} \quad (2)$$

$$d_k \in \{1, 2, 4, 8, 12\}, \mathrm{K} = 6 \quad (3)$$

where p_t is the $PM_{2.5}$ input at time t, h_j and $g_j^{(k)}$ are learned filter coefficients, $b_t^{(k)}$ denotes *component* k, d_k is the dilation factor, and K is the number of components. The validation search considered $K \in 3, 4, 5, 6, 8, 10$, and the evaluated configuration fixes $K = 6$. Each component is projected to a common hidden dimension before recurrent encoding and gated fusion.

A single-layer bidirectional gated recurrent unit (BiGRU) encodes the sequence produced by each filter [25]. Both directions remain inside the observed historical window; neither uses the future target. A separate GRU encodes the complete predictor sequence. The network thus retains the original multivariable history alongside its resolution-specific pollution representations before combining them with the environmental states.

$$h_k = W_o \left[\overrightarrow{h}_k \,\|\, \overleftarrow{h}_k \right] \in \mathrm{R}^H \quad (4)$$

where h_k is the latent representation of *component* k, $\|$ denotes vector concatenation, W_o is the learned output projection, and H is the shared hidden dimension.

BGI features at the final time step are encoded through a two-layer perceptron with Gaussian error linear unit (GELU) activation and layer normalization (LayerNorm) [47]:

$$h_{bgi} = LayerNorm(W_2 GELU(W_1[NDVI, NDWI, LAI, LST] + b_1) + b_2) \quad (5)$$

A parallel pathway encodes the meteorological context aligned to the forecast origin as $h_{wx} \in R^H$. Here, W_1, W_2, b_1, and b_2 are learnable parameters.

For each temporal representation, attention is calculated from its encoded history together with the ecological and meteorological states. Softmax normalizes the scores into nonnegative weights. This operation determines how strongly each representation contributes to the combined pollution feature; the signed ecological adjustment is applied subsequently through the hyperbolic tangent (Tanh) gate:

$$\alpha_k = LeakyReLU(W_a[h_k \parallel h_{bgi} \parallel h_{wx}]), A_k = \frac{exp(\alpha_k)}{\sum_{j=1}^{K} exp(\alpha_j)} \quad (6)$$

where the leaky rectified linear unit (LeakyReLU) supplies the activation, α_k is the unnormalized score, A_k is its normalized attention weight, and W_a is a learnable projection. The representation obtained through attention is $h_{scale} = \sum_{k=1}^{K} A_k h_k$. A signed gate with Tanh activation then provides bidirectional modulation.

$$g = \text{tabh}(W_g[h_{scale} \parallel h_{bgi}] + b_g \quad (7)$$

where W_g and b_g are learnable gate parameters, and g is the signed gate vector. The Tanh activation permits either a positive or negative adjustment. A residual connection with a learnable coefficient, initialized at 0.1, limits the initial contribution of the ecological state:

$$\tilde{h}_{scale} = h_{scale} + \beta \cdot g \odot h_{bgi} \quad (8)$$

where β is the learnable residual coefficient and $\odot$ denotes elementwise multiplication. A separate GRU encodes the complete input into h_{ctx}. The combined context and resolution representations enter a multilayer perceptron (MLP) with two layers:

$$h_{ctx} = GRU(X)[-1], \hat{y}_{t+h} = MLP\left([h_{ctx} \parallel \tilde{h}_{scale}]\right) \quad (9)$$

Root mean square error (RMSE) and mean absolute error (MAE) quantify errors on the observed concentration scale. Here, N is the number of test observations, y_i is the observed concentration, and $\hat{y}_i$ is its prediction:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}, MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \quad (10)$$

The coefficient of determination (R^2) measures explained variance, where $\bar{y}$ is the mean observed concentration:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad (11)$$

For pollution event detection, F1 score combines precision and recall:

$$e_t(\theta) = I(y_t \geq \theta, \hat{e}_t(\theta) = I(\hat{y}_t \geq \theta), \theta \in \{35, 75\} \quad (12)$$

where θ is the $PM_{2.5}$ threshold in $\mu g/m^3$ and $I(\cdot)$ is the indicator function. Event forecasts are defined separately at 35 and 75 $\mu g/m^3$.

$$F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \text{Precision} = \frac{TP}{TP + FP}, \text{Recall} = \frac{TP}{TP + FN} \quad (13)$$

where *TP* and *FN* denote true positive and false negative pollution event counts. These equations are used consistently across all horizons and cities. The probability of detection (POD) and false alarm ratio (FAR) are defined below:

$$POD = \frac{TP}{TP + FN}, FAR = \frac{FP}{TP + FP} \tag{14}$$

where *FP* denotes false positive event counts. The Diebold–Mariano (DM) statistic compares the loss differential between AFGN and a baseline model:

$$DM = \frac{mean(d_t)}{\sqrt{Var(d_t)/T}}, d_t = L(e_t^{AFGN}) - L(e_t^{baseline}) \tag{15}$$

where L(·) is the squared error loss and d_t represents the loss differential over the test period. A negative value indicates smaller forecast errors for AFGN. Here, e_t denotes the forecast error at test time t.

Figure 6 summarizes the sequence from data preparation to forecasting and evaluation. Quality-controlled observations are aligned by availability time, followed by chronological partitioning and training-based preprocessing. AFGN then extracts and combines temporal and environmental features to predict concentration. Model comparisons, input ablations, and concentration-specific errors evaluate different aspects of the resulting forecasts.

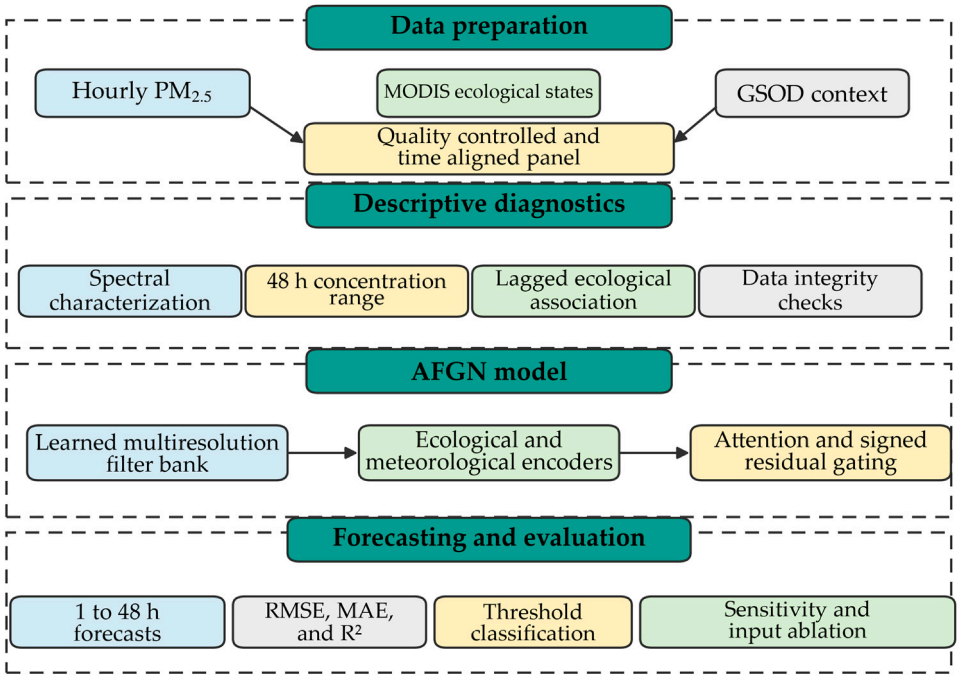


Figure 6. Methodological workflow. Notes: The workflow combines hourly $PM_{2.5}$, daily Global Surface Summary of the Day (GSOD) meteorology, and Moderate Resolution Imaging Spectroradiometer (MODIS) ecological observations. Temporal partitioning precedes fitted preprocessing. The forecasting comparisons distinguish model architecture, input information, and performance at different concentrations. Evaluation uses root mean square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2).

The ablation experiments distinguish changes to information from changes to architecture. The ecological input control retains all network modules but excludes the satellite variables and their derived features. Removing the temporal-resolution branch, by contrast, changes both the representation and the number of parameters. A third comparison disables the encoded ecological pathway. These experiments are reported separately because

an effect of removing a network component is not equivalent to an effect of withholding its input information.

The resulting architecture is shown in Figure 7. Its filters, recurrent encoders, and fusion parameters are learned from the forecasting objective. Attention weights describe the relative weighting of encoded features, while the signed gate controls a residual adjustment. Neither quantity measures particle deposition, ventilation, or the environmental effect of a BGI project. Their predictive roles are assessed through the comparisons in Section 4.

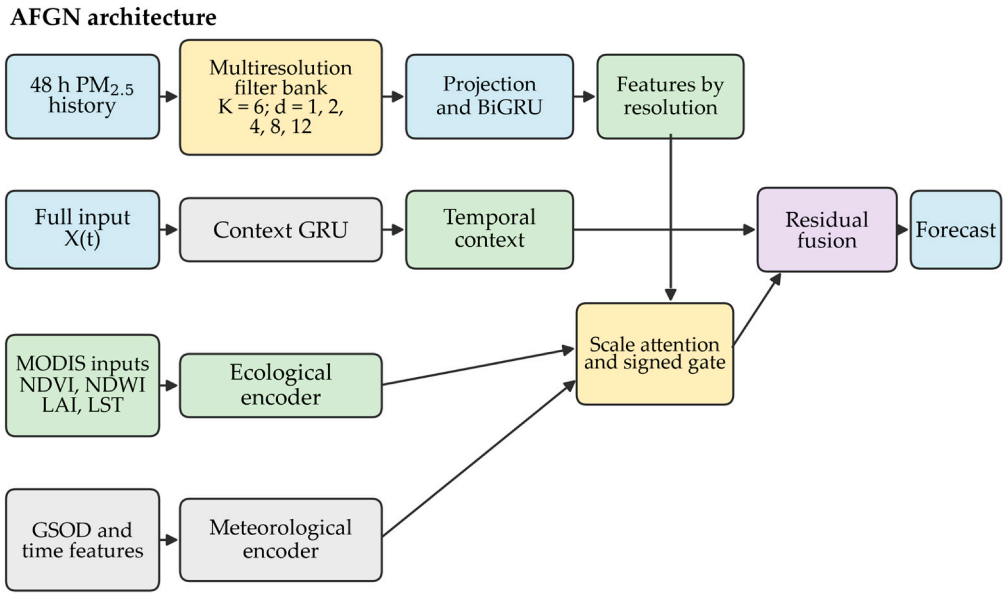


Figure 7. Architecture of AFGN. Notes: Each 48 h input window is represented at several learned temporal resolutions. The bidirectional gated recurrent unit (BiGRU) extends the gated recurrent unit (GRU). The combined attention and gate block uses nonnegative softmax attention weights followed by signed hyperbolic tangent (Tanh) residual modulation. The context encoder receives the historical predictor sequence. Satellite indicators: normalized difference vegetation index (NDVI), normalized difference water index (NDWI), leaf area index (LAI), and land surface temperature (LST).

3.3. Training and Evaluation Protocol

Model configurations are summarized in Table A2. The principal AFGN and GRU experiments use a 48 h history to predict concentrations directly at 1, 6, 12, 24 and 48 h. Eligible observations are ordered in time and assigned to training, validation, and test subsets in proportions of 70%, 15%, and 15%. Boundary windows are removed if a training or validation target reaches the start of the next subset. For the Beijing 6 h task, this leaves 11,916 training, 2549 validation, and 2555 test windows. Counts and exact boundaries differ slightly by city and horizon because the original observation mask is retained; Table A3 provides the details. Preprocessing parameters are fitted to the retained training data. Predictions are transformed back to concentration units before RMSE, MAE, and R^2 are calculated against the untrimmed observations. This ordered design preserves the temporal separation required for forecasting evaluation [44].

Standardization statistics are estimated solely on the training subset. For feature j, $\mu_{j,train}$ and $\sigma_{j,train}$ are calculated using training samples only. Each raw observation is normalized via:

$z_{t,j} = \frac{x_{t,j} - \mu_{j,train}}{\sigma_{j,train}}$. Validation and test data use these fixed statistics. AFGN is trained with the smooth L1 (SmoothL1) objective selected on validation data:

$$L_{SmoothL1} = \frac{1}{N}\sum_{i=1}^{N} \rho_{\delta}\left(\hat{y}_{i,t+h} - y_{i,t+h}\right) \quad (16)$$

where N is the number of training windows, δ is the transition parameter of SmoothL1, and $\rho_\delta(e) = \begin{cases} \frac{0.5e^2}{\delta}, |e| < \delta, \\ |e| - 0.5\delta, otherwise. \end{cases}$ RMSE, MAE, and R^2 are calculated after predictions are restored to the original concentration scale. Concentration thresholds of 35 and 75 $\mu g/m^3$ define the hourly event classifications used to calculate F1, probability of detection (POD), false alarm ratio (FAR), and missed-event frequency. Tail bias is evaluated separately. The Diebold–Mariano test [48] compares squared forecast errors for each seed.

Weight-decay candidates in validation were 0, 1×10^{-5} and 1×10^{-4}. The principal AFGN uses hidden dimension 128, dropout 0.3, batch size 512, learning rate 0.001, and weight decay 1×10^{-5}, with encoder normalization and SmoothL1 loss. AFGN, its ecological input control, and the principal GRU are trained for at most 50 epochs, with early stopping after eight epochs without improvement, adaptive moment estimation with decoupled weight decay (AdamW) optimization, and gradient clipping at 1.0. GRU and LSTM use a hidden dimension of 256 and dropout of 0.1, with batch sizes of 512 and 128, respectively. The separate LSTM, component, and capacity experiments allow up to 80 epochs. Added linear and attention baselines use Time-Series-Library implementations with mean squared error (MSE), dropout 0.1, batch size 512, a 50-epoch limit, and the same early-stopping patience. Two candidate configurations are compared on validation data for each added model family. No observed future pollution or weather is supplied to their decoders. Table A2 gives the model settings; Figure A1 shows training and validation loss trajectories; training durations are summarized in the Appendix A. A separate MSE experiment repeats AFGN and its ecological input control under the same 50-epoch limit to assess sensitivity to the loss function.

The main comparisons use five random seeds: 42, 123, 456, 789 and 2024. Models are fitted separately for each city and seed, and summary metrics give equal weight to each city. $PM_{2.5}$ inputs and targets retain their untrimmed values. Each predictor has a unique input channel; remaining missing values are filled with training medians, and scaling uses training statistics. Training uses AdamW [49], dropout [50], and early stopping [51]. SmoothL1 combines quadratic and linear penalties, following the robust-loss principle introduced by Huber [52]. Variation across seeds measures sensitivity to initialization, whereas calendar-block resampling estimates uncertainty associated with the historical test observations.

To isolate ecological information, an otherwise identical control sets 12 standardized channels to zero before the forward pass: four satellite indicators, four retrieval-age variables, and four ecological interaction terms. The network structure and training settings are unchanged. This model is termed the no-ecological-input control. An invariance check confirms that altering the excluded values cannot affect its predictions. A separate pathway control also sets the encoded ecological state to zero. Paired error intervals resample common seven-day calendar blocks across cities while retaining all five seeds. Chronological construction prevents later observations from entering earlier forecasts, but the historical test period had been examined during earlier model development. The present estimates are therefore retrospective rather than an independent prospective validation.

4. Results and Analysis

4.1. Temporal Diagnostics and Multi-Horizon Forecasting

Figures 2 and 3 show that observed $PM_{2.5}$ variation is distributed across several temporal scales. The illustrative decomposition distinguishes slow, synoptic, diurnal, and residual changes. Normalized spectra differ among cities, with more prominent synoptic variation in Beijing and Shanghai and a larger slow-variation share in Chengdu and Wuhan.

These features describe the observed series; they are not the learned outputs of AFGN's convolutional branches.

Figure 4 quantifies concentration changes within 48 h windows. Median ranges are 15–21 μg/m^3 across cities, and the 95th percentile among regular windows is approximately 25–29 μg/m^3. Windows with a range of at least 100 μg/m^3 account for 2.53–10.38% of observations, while the 95th percentile across all windows reaches 80–126 μg/m^3. Figure 5 shows that the association between vegetation-state changes and $PM_{2.5}$ anomalies changes sign over the event window. These descriptive patterns provide the context for the forecast comparisons in Table 2; they do not establish the predictive contribution of ecological inputs.

Table 2. Multi-horizon forecasting performance.

Horizon	AFGN RMSE	GRU RMSE	AFGN R^2	GRU R^2	Reduction (%)	Cities Improved
1 h	3.310	3.994	0.980	0.970	17.13	4/4
6 h	12.526	13.498	0.733	0.689	7.20	4/4
12 h	17.387	17.928	0.491	0.462	3.02	4/4
24 h	22.170	21.807	0.169	0.196	−1.67	1/4
48 h	26.163	26.047	−0.104	−0.092	−0.44	1/4

Notes: Root mean square error (RMSE) is in μg/m^3; the coefficient of determination (R^2) is dimensionless. Metrics are calculated for each city and random seed, then averaged with equal city weights over the four cities and five seeds. Positive percentage reductions indicate lower RMSE for AFGN.

AFGN has lower mean RMSE than GRU at 1, 6, and 12 h (Table 2). At 6 h, RMSE decreases from 13.498 to 12.526 μg/m^3, while mean R^2 increases from 0.689 to 0.733. The absolute RMSE reduction calculated before rounding is 0.971 μg/m^3, equivalent to 7.20%. At 24 and 48 h, GRU has lower point estimates of mean error, although the paired intervals include zero.

Table 3 gives city-specific RMSE and R^2 at 12 h for both models. Their differences are smaller than at 6 h. For the 6 h task, RMSE reductions relative to GRU are 7.19% in Beijing, 9.89% in Shanghai, 1.73% in Chengdu, and 8.75% in Wuhan. Thus, the mean reduction is not uniform across cities, with the smallest improvement in Chengdu. The concentration-specific analysis below examines another source of variation in forecast accuracy.

Table 3. Validation-selected 12 h forecasting performance by city.

City	AFGN RMSE	GRU RMSE	Reduction (%)	AFGN R^2	GRU R^2
Beijing	22.230	23.489	5.36	0.548	0.495
Shanghai	16.564	17.150	3.42	0.402	0.359
Chengdu	13.136	13.154	0.14	0.586	0.585
Wuhan	17.618	17.918	1.67	0.428	0.409

Notes: RMSE is measured in μg/m^3. Values average five fixed seeds within each city. The model settings are fixed before this protocol check, and both models are evaluated on identical target timestamps.

Average concentration error and pollution-event identification capture different aspects of forecast performance. RMSE gives greater weight to large residuals [53], but an improvement in the overall metric does not guarantee accurate prediction of the highest concentrations. Tables 4 and 5 therefore evaluate threshold classifications and concentration-specific errors using the same untrimmed observations. These comparisons hold the evaluation data fixed, unlike comparisons of published RMSE values across different cities, forecast horizons, input information, and target definitions [22,37,54–57].

Table 4. Pollution event classification at 6 h.

Threshold	Model	F1	Probability of Detection (POD)	False Alarm Ratio (FAR)	Miss Rate	Tail Bias
35	AFGN	0.812	0.798	0.170	0.202	−5.471
35	GRU	0.799	0.802	0.199	0.198	−4.546
35	Long short-term memory (LSTM)	0.779	0.771	0.207	0.229	−6.885
35	Persistence	0.804	0.800	0.193	0.200	−4.129
75	AFGN	0.573	0.559	0.355	0.441	−13.492
75	GRU	0.577	0.571	0.401	0.429	−12.935
75	LSTM	0.505	0.479	0.388	0.521	−17.419
75	Persistence	0.598	0.602	0.405	0.398	−9.811

Notes: Metrics are averaged by city and seed. Thresholds and tail bias are measured in $\mu g/m^3$. POD is the probability of detection, and FAR is the false alarm ratio. Threshold classification of a point forecast is distinct from forecasting an event probability. The long short-term memory (LSTM) model is a recurrent baseline. The F1 score is the harmonic mean of precision and recall.

Table 5. Concentration regime errors for AFGN at 6 h.

Observed $PM_{2.5}$	Unique Hours	RMSE	Mean Absolute Error (MAE)	Bias
below 35	6650	9.915	6.279	3.493
35 to below 75	2841	13.414	10.344	−2.796
75 to 150	712	24.557	19.856	−13.718
above 150	17	50.976	47.054	−47.054

Notes: Groups are defined by observed target concentration. RMSE, MAE, and bias pool observations and errors from the five runs. Counts refer to distinct city–timestamp pairs, with repeated model runs not counted as additional observations. Negative bias indicates underprediction; the highest group is strictly above 150 $\mu g/m^3$.

4.2. Episode Tracking and Pollution Event Performance

Figure 8 compares observed and predicted concentrations during a Beijing pollution accumulation episode; Figure 9 shows all matched test observations in the four cities. The forecast follows much of the increase and subsequent decline, but underestimates the observed peak of 179 $\mu g/m^3$. The scatter plots also show greater underprediction at high concentrations. Table 5 quantifies this pattern across concentration groups rather than relying on the illustrative episode alone.

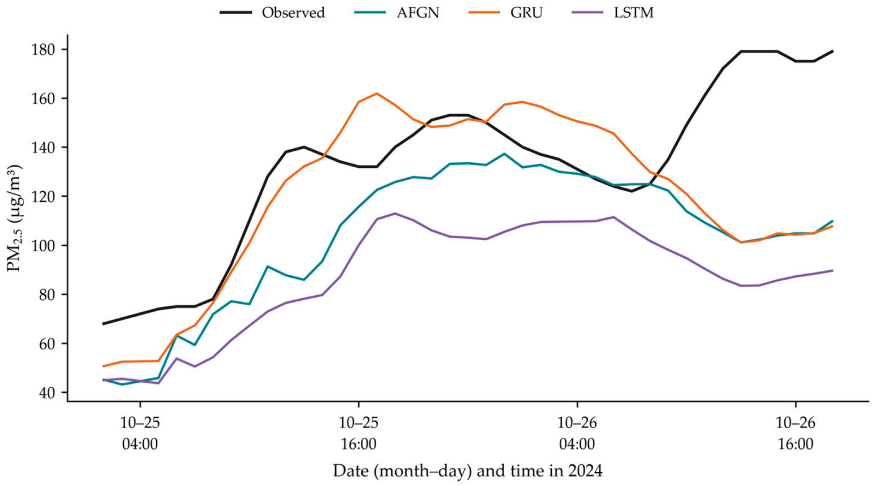


Figure 8. Observed and 6 h forecast trajectories during a selected Beijing pollution accumulation episode. Notes: The interval covers the Beijing accumulation episode from 25 October 02:00 to 26 October 18:00 in the historical test period. Curves show observed $PM_{2.5}$ and the 6 h forecasts from seed 42. This example illustrates a pollution episode and is not used to select a model.

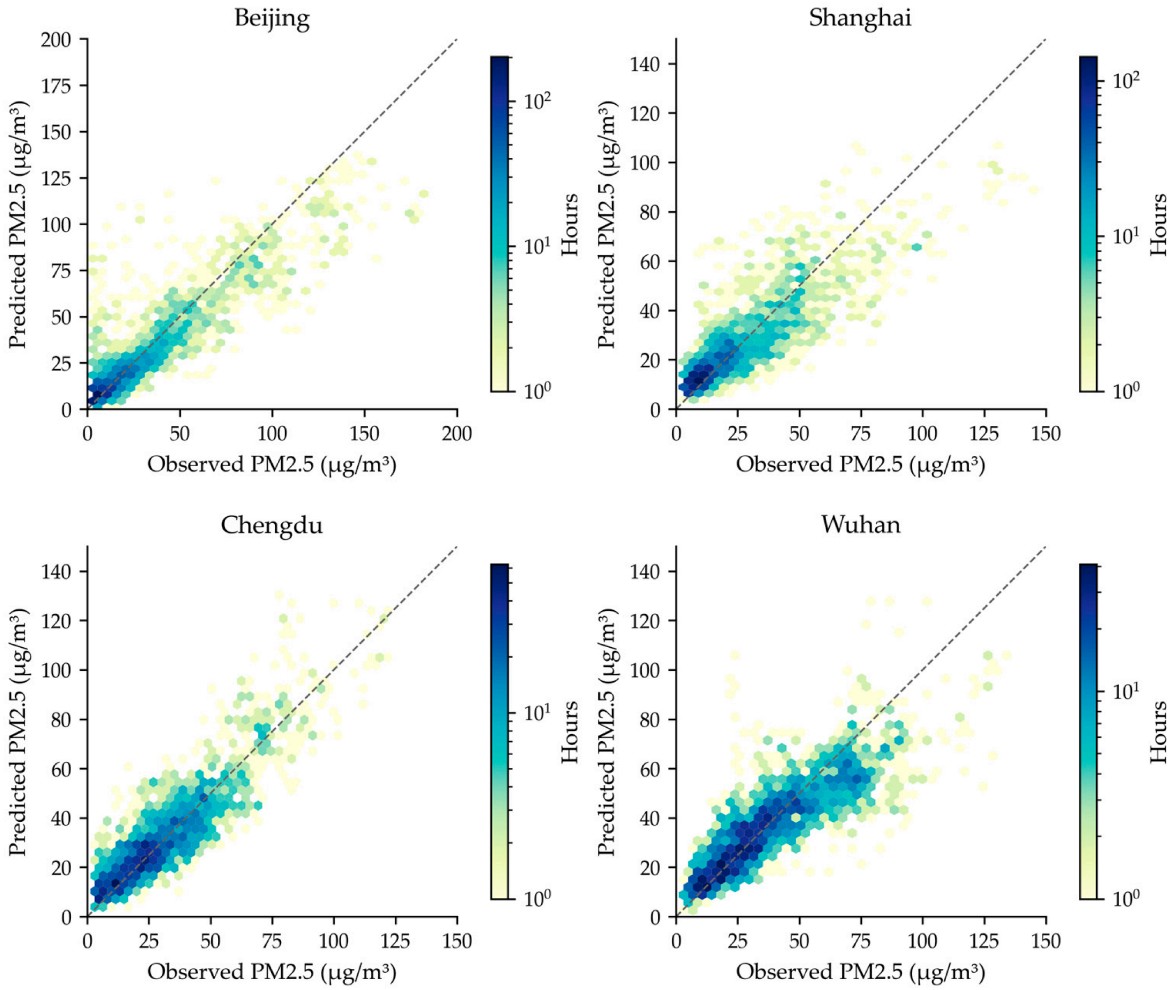


Figure 9. Observed and predicted $PM_{2.5}$ at 6 h in the four cities. Subfigures show Beijing (**upper left**), Shanghai (**upper right**), Chengdu (**lower left**), and Wuhan (**lower right**). Notes: Forecasts use AFGN with seed 42 and the original test observations. Color indicates the number of observation–prediction pairs in each hexagonal bin. The dashed diagonal denotes equality between predicted and observed concentration.

Table 4 assesses identification of threshold events in addition to the concentration fit measured by R^2. Mean F1 for AFGN is 0.812 at 35 μg/m^3 and 0.573 at 75 μg/m^3. For observations strictly above 150 μg/m^3, Table 5 reports RMSE of 50.976 μg/m^3 and mean bias of −47.054 μg/m^3. These observations comprise 17 Beijing hours in three contiguous runs. The negative bias quantifies the peak underprediction visible in Figures 8 and 9.

4.3. Comparative Performance, Sensitivity, and Component Attribution

In the 6 h comparison, mean RMSE is 12.526 μg/m^3 for AFGN and 12.840 μg/m^3 for DLinear, the closest baseline (Table 6). The paired difference is 0.314 μg/m^3, with a 95% interval of −0.030 to 0.630 μg/m^3, so the difference is not statistically significant. The other evaluated recurrent and attention-based configurations have higher mean RMSE. These results apply to the configurations selected through the validation procedures described in Section 3.3.

Table 6. Baseline comparison at 6 h.

Model	RMSE	MAE	R^2
AFGN	12.526	8.423	0.733
GRU	13.498	9.320	0.689
LSTM	14.305	9.913	0.655
Persistence	13.532	8.571	0.692
Seasonal persistence	23.657	16.101	0.062
DLinear	12.840	8.625	0.721
Informer	18.002	12.668	0.447
Autoformer	18.396	13.097	0.431
Temporal Fusion Transformers (TFT)	15.993	10.756	0.566
TimesNet	15.338	10.294	0.609

Notes: Values are averages of city-specific metrics over four cities and five seeds. RMSE and MAE are measured in $\mu g/m^3$. Persistence is deterministic. Model settings are selected using the validation procedures described in Section 3.3. The decomposition-based linear model (DLinear) is a baseline.

Table 7 supplements the mean differences with paired uncertainty intervals for AFGN and GRU [48,58]. The intervals favor AFGN at 1, 6, and 12 h, but include zero at 24 and 48 h. Each resample contains the same calendar blocks across cities and retains all five seeds, preserving temporal clustering and concurrent variation among cities. The intervals describe uncertainty within the evaluated historical period.

Table 7. Paired RMSE differences between AFGN and GRU.

Horizon	RMSE Reduction	95% Lower	95% Upper	Reduction (%)
1 h	0.684	0.026	1.480	17.13
6 h	0.971	0.705	1.404	7.20
12 h	0.541	0.107	1.141	3.02
24 h	−0.363	−0.977	0.477	−1.67
48 h	−0.116	−0.807	0.474	−0.44

Notes: The difference is GRU minus AFGN, in $\mu g/m^3$. Pointwise intervals use 2000 moving calendar block bootstrap replicates with seven-day blocks. They describe uncertainty conditional on these cities, configurations, and the historical evaluation period. They are not adjusted for the family of horizon comparisons.

Table 8 isolates the contribution of satellite ecological inputs. Removing ecological information while retaining the full network reduces mean 6 h RMSE from 12.526 to 12.199 $\mu g/m^3$. The no-ecological-input control minus full-model difference is -0.328 $\mu g/m^3$, with a 95% interval from -0.455 to -0.130 $\mu g/m^3$. Disabling the encoded ecological pathway gives the same overall ordering, and Table A4 reports the city-level results. Thus, the complete model's improvement over GRU cannot be interpreted as an overall benefit of satellite ecological inputs.

Table 8. Ecological input and pathway controls at 6 h.

Configuration	RMSE	MAE	R^2
AFGN	12.526	8.423	0.733
Ecological inputs masked	12.199	8.154	0.748
Ecological pathway removed	12.217	8.168	0.747

Notes: The no-ecological-input control sets the satellite variables and their derived features to zero after standardization, retaining the full network. The pathway control additionally disables the encoded ecological state. Models use the same five seeds and target timestamps. RMSE and MAE are in $\mu g/m^3$.

Table 9 and Figure 10 compare architectural components and model capacity. Mean RMSE is 12.594 μg/m^3 after removal of the multiresolution branch and 12.382 μg/m^3 after removal of the signed residual contribution, compared with 12.526 μg/m^3 for the complete model. A wider GRU used as a capacity control has a RMSE of 13.256 μg/m^3. Figure 10 shows how each difference varies across the five seeds. Unlike the input comparison in Table 8, branch removal also changes the architecture and parameter count.

Table 9. Component and capacity checks at 6 h.

Configuration	RMSE	MAE	Change from Full (%)
AFGN	12.526	8.423	0.00
Resolution branch removed	12.594	8.519	0.54
Signed residual removed	12.382	8.312	−1.15
GRU capacity control	13.256	9.123	5.83

Notes: RMSE and MAE are in μg/m^3. Component comparisons retain the training settings. The capacity control is a two-layer GRU with 384 hidden units, chosen to approach AFGN's parameter count without a separate hyperparameter search.

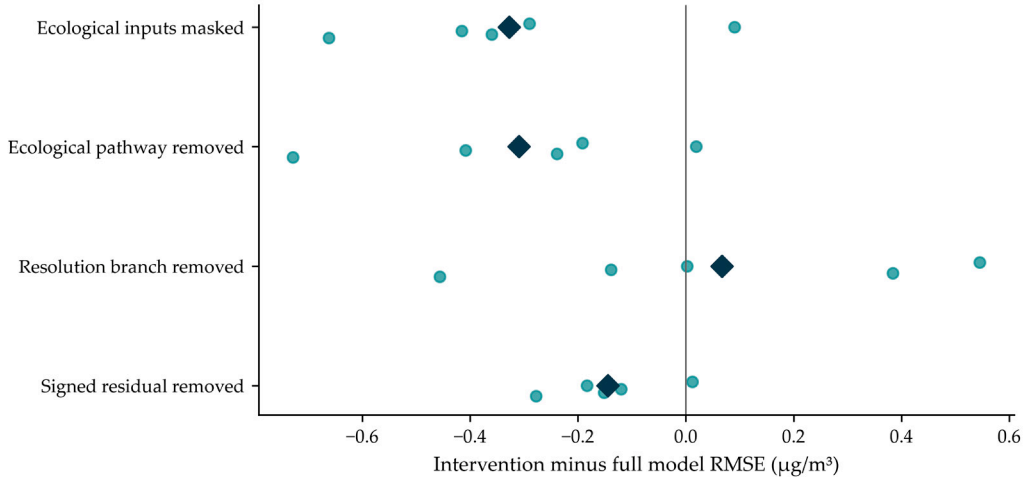


Figure 10. Changes in 6 h RMSE under input and component interventions. Notes: Points show the difference between the intervention and full model, averaged over cities for each seed. The central marker is the mean over five seeds. Positive values indicate deterioration after the intervention. The removal of a representation branch also changes model capacity.

Table 10 explores the ecological input difference under conditions observed at the forecast origin. Pollution and wind quartiles are estimated from each city's training distribution. Recent variability is the standard deviation over the preceding 24 h, requiring at least 18 valid hourly observations. Winter is defined as December to February, but the test period contains only December 2024. The reported winter groups consequently refer to that December alone. Although the group labels use information available before prediction, the subgroup analysis is retrospective; its uncertainty estimates do not provide independent confirmation of the selected conditions.

For December forecast origins in pollution quartile Q2, pooled RMSE decreases from 8.867 to 8.439 μg/m^3 across 517 city–hour observations. The reduction is 0.429 μg/m^3 or 4.83%, with an adjusted interval of 0.023–1.405 μg/m^3. The reduction remains positive for each city and seed aggregate and when individual cities, seeds, or December segments are omitted. By contrast, the Q4 pollution–Q4 wind group contains 144 h from three cities and has an interval including zero despite a positive point estimate. December alone, high pollution alone, and high recent variability do not show positive mean ecological gains. The positive result is therefore specific to the exploratory December Q2 subgroup.

Table 10. Conditional ecological input comparison at 6 h.

Condition	Hours	Full RMSE	No-Ecology RMSE	Change (%)	Adjusted Interval
All hours	10,220	12.662	12.365	−2.40	−0.477 to −0.066
Winter	2612	14.929	14.459	−3.25	−0.973 to 0.350
Origin $PM_{2.5} \geq 75$	722	28.203	27.428	−2.83	−1.773 to 0.224
Origin pollution Q4	2084	21.713	21.146	−2.68	−0.970 to 0.039
Recent variability Q4	2218	17.990	17.585	−2.30	−0.934 to 0.227
Pollution Q4 and wind Q4	144	19.856	20.465	2.97	−1.311 to 4.955
Pollution Q2 and winter	517	8.439	8.867	4.83	0.023 to 1.405

Notes: Positive differences indicate lower RMSE for the complete model. RMSE pools city–hour errors across five seeds, unlike the equal-city averages in Tables 2 and 8. Quartiles use training observations. Intervals use common seven-day blocks with nominal Bonferroni adjustment for the seven listed comparisons. Of 2000 draws, 12 contain no December subgroup observations and are excluded from the corresponding intervals. Counts refer to distinct city–hour observations. The second quartile (Q2) and fourth quartile (Q4) refer to the city-specific training distributions.

5. Discussion

5.1. Temporal Representation and Environmental Information

AFGN's improvement over GRU at 1, 6, and 12 h indicates that its temporal representation and environmental fusion are useful for short-range prediction. The similar 6 h performance of DLinear suggests that much of the predictable variation is also accessible to a simpler temporal representation. At 24 and 48 h, the AFGN advantage is no longer evident. Changes in transport, weather and emissions over these longer intervals may weaken the relationship between recent local observations and future concentrations. Neighboring observations, numerical forecasts and data assimilation offer additional information about these processes [22,36,55–57]. Their value for extending AFGN's useful forecast horizon should be evaluated on the same cities and target timestamps, rather than inferred from published errors for different tasks.

The ecological ablation shows that the complete model's advantage over GRU is not evidence of a general satellite-input benefit. Removing ecological information lowers the overall error. One possible explanation is that satellite indicators partly duplicate seasonal and meteorological information, while spatial averaging and infrequent updates limit their ability to describe rapid local changes. The positive December Q2 difference suggests a narrower possibility: ecological states may help distinguish concentration evolution within some moderate-pollution conditions. Its persistence after omitting individual cities, seeds, and time segments reduces concern that a single subset determines the sign, but all these checks reuse the same December. Independent winter observations are needed to assess recurrence. Deposition and dispersion studies [7,10] offer physical reasons to examine environmental context, yet the present ablation does not identify which mechanism produces the subgroup difference or estimate the effect of BGI construction.

Average accuracy does not capture the difficulty of forecasting severe peaks. The small number of extreme observations provides few examples from which to learn rapid accumulation, while SmoothL1 reduces the relative penalty on large standardized residuals. Replacing it with MSE makes severe-event bias less negative, from −47.054 to −44.883 $\mu g/m^3$, but raises severe-event RMSE from 50.976 to 52.430 $\mu g/m^3$ and overall RMSE from 12.526 to 12.764 $\mu g/m^3$. Thus, a change of objective alone does not resolve peak underprediction. Winter haze studies identify stagnation, ventilation, and aerosol–boundary-layer interactions as relevant processes [59,60]. Their effects cannot be separated from unobserved emission changes using the present forecast residuals. Improving peak forecasts will require more independent severe episodes and evaluation of predictors such as emission activity, boundary-layer height and vertical mixing, subject to their availability at forecast time.

5.2. Sustainable Urban Management and Scope of Application

Hourly concentration forecasts can give urban air quality management a short lead time for pollution monitoring, warning and exposure protection. Their usefulness depends on the identification of pollution episodes as well as average concentration accuracy. In this study, POD, FAR, and F1 assess classifications derived from point forecasts at thresholds of 35 and 75 $\mu g/m^3$. These thresholds define evaluation events; they are not regulatory assessments with prescribed averaging periods or calibrated event probabilities. The model uses historical meteorological observations without future numerical weather predictions or explicit emission inputs. Integrating forecasts with monitoring would therefore require evaluation of both predictive accuracy and the effectiveness of the resulting management actions.

The present evaluation covers four locally trained models and a test period mainly spanning September to December 2024. Broader seasonal coverage and validation in additional cities are needed to assess wider applicability. Daily GSOD records and intermittent MODIS retrievals characterize city-level environmental conditions, not street-level airflow or personal exposure. More frequent meteorological observations may also alter the additional information supplied by satellite indicators and warrant a matched input comparison. For BGI planning, forecasting can complement spatial monitoring and ecological assessment, but the effects of an intervention require site-specific evidence [17,18]. This distinction connects the forecasting task to sustainable urban management without treating predictive associations as causal environmental effects.

6. Conclusions

This study presents AFGN, a multiresolution framework for hourly $PM_{2.5}$ forecasting that combines pollution dynamics with meteorological and satellite ecological information. Its main methodological contribution is to represent rapid concentration changes separately from more persistent environmental conditions and combine these representations through attention and signed gating. Evaluation in Beijing, Shanghai, Chengdu, and Wuhan shows lower mean prediction error than GRU at 1, 6, and 12 h, with the clearest improvement at 6 h. The findings demonstrate the usefulness of this framework for anticipating near-term changes in urban air quality.

The controlled input comparisons distinguish the performance of the forecasting architecture from the contribution of satellite ecological information. The latter depends on the forecasting conditions and does not explain the complete model's overall improvement. A positive ecological contribution is observed for December forecasts in the second pollution quartile and remains positive when individual cities, seeds, or time segments are omitted. This finding identifies a specific condition in which ecological observations may be informative; confirmation requires independent winter data.

For sustainable urban air quality management, the study contributes a forecasting method and evidence for evaluating the information used in environmental prediction. Incorporating short-range forecasts into monitoring and early-warning systems could support timely pollution response and public exposure protection. Their planning value lies in complementing spatial and ecological assessments, rather than estimating the pollution reduction from a particular BGI intervention. Extending evaluation across seasons and cities, with greater coverage of severe pollution events, would provide a stronger basis for applying the framework in urban environmental management. Moreover, this work also delivers quantitative proof for environmental judicial cases under environmental protection laws. It supports forward legal supervision of urban blue-green ecological spaces and $PM_{2.5}$ controls.

Author Contributions: Conceptualization, S.C., W.F. and C.Z.; methodology, S.C., W.F. and C.Z.; software, C.Z.; validation, S.C., X.W., J.X. and C.Z.; formal analysis, S.C., X.W., J.X. and C.Z.; investigation, S.C., X.W., J.X. and C.Z.; data curation, C.Z.; writing, original draft preparation, S.C., X.W., J.X. and C.Z.; writing, review and editing, S.C., W.F. and C.Z.; visualization, C.Z.; supervision, S.C. and W.F.; project administration, S.C. and W.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study were derived from the following resources available in the public domain. Hourly $PM_{2.5}$ records used in this study originate from the China National Environmental Monitoring Center national urban air quality real-time reporting platform. National Oceanic and Atmospheric Administration (NOAA) Global Surface Summary of the Day (GSOD) station summaries are publicly available from the National Centers for Environmental Information at https://www.ncei.noaa.gov/. Moderate Resolution Imaging Spectroradiometer (MODIS) products are publicly available from the National Aeronautics and Space Administration (NASA) Land Processes Distributed Active Archive Center (LP DAAC) at https://lpdaac.usgs.gov/, access on 3 September 2026. Derived analysis panels aggregated by city, model configurations, and evaluation outputs are available from the corresponding author upon reasonable request.

Acknowledgments: We acknowledge CNEMC for national hourly urban air quality records, the NOAA National Centers for Environmental Information (NCEI) for GSOD station summaries, the National Aeronautics and Space Administration (NASA) Land Processes Distributed Active Archive Center (LP DAAC) for Moderate Resolution Imaging Spectroradiometer (MODIS) products, and Google Earth Engine for geospatial processing.

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

Appendix A

Tables A1–A6 provide satellite retrieval coverage, model settings, chronological data partitions, ecological input comparisons, training histories, and the external comparison of hourly $PM_{2.5}$ records. Figures A1 and A2 show the training and validation loss trajectories and the paired external observations, respectively.

Table A1. MODIS valid-retrieval coverage before temporal alignment.

City	Normalized Difference Vegetation Index (NDVI) (%)	Normalized Difference Water Index (NDWI) (%)	Leaf Area Index (LAI) (%)	Land Surface Temperature (LST) (%)
Beijing	98.6	99.0	99.4	95.5
Shanghai	96.9	92.3	97.4	84.0
Chengdu	87.5	84.3	94.7	70.5
Wuhan	100.0	91.2	97.8	86.0

Table A2. Model configurations and training settings. Loss functions are smooth L1 (SmoothL1) and mean squared error (MSE).

Model	Hidden	Dropout	Batch	Learning Rate	Loss
AFGN	128	0.3	512	0.001	SmoothL1
GRU	256	0.1	512	0.001	SmoothL1
LSTM	256	0.1	128	0.001	SmoothL1
dlinear	Not applicable	0.1	512	0.003	MSE
informer	32	0.1	512	0.001	MSE
autoformer	32	0.1	512	0.001	MSE
tft	32	0.1	512	0.001	MSE
timesnet	32	0.1	512	0.001	MSE

Table A3. Training, validation, and test windows after removal of boundary-crossing targets.

City	Lead	Train/Val/Test	Last Train Target	Last Val Target	First Test Target
Beijing	1 h	11,933/2557/2558	20 May 2024 13:00	8 September 2024 13:00	8 September 2024 14:00
Beijing	6 h	11,916/2549/2555	20 May 2024 06:00	8 September 2024 07:00	8 September 2024 13:00
Beijing	12 h	11,899/2541/2553	19 May 2024 21:00	8 September 2024 01:00	8 September 2024 13:00
Beijing	24 h	11,871/2527/2550	19 May 2024 05:00	7 September 2024 06:00	8 September 2024 06:00
Beijing	48 h	11,831/2500/2546	18 May 2024 11:00	6 September 2024 11:00	8 September 2024 11:00
Shanghai	1 h	11,932/2556/2558	20 May 2024 14:00	8 September 2024 13:00	8 September 2024 14:00
Shanghai	6 h	11,914/2549/2555	20 May 2024 06:00	8 September 2024 07:00	8 September 2024 13:00
Shanghai	12 h	11,898/2540/2553	19 May 2024 22:00	8 September 2024 01:00	8 September 2024 13:00
Shanghai	24 h	11,870/2527/2549	19 May 2024 06:00	7 September 2024 07:00	8 September 2024 07:00
Shanghai	48 h	11,829/2500/2546	18 May 2024 11:00	6 September 2024 11:00	8 September 2024 11:00
Chengdu	1 h	11,932/2557/2558	20 May 2024 13:00	8 September 2024 13:00	8 September 2024 14:00
Chengdu	6 h	11,915/2549/2555	20 May 2024 06:00	8 September 2024 07:00	8 September 2024 13:00
Chengdu	12 h	11,898/2541/2553	19 May 2024 21:00	8 September 2024 01:00	8 September 2024 13:00
Chengdu	24 h	11,870/2527/2550	19 May 2024 05:00	7 September 2024 06:00	8 September 2024 06:00
Chengdu	48 h	11,830/2500/2546	18 May 2024 11:00	6 September 2024 11:00	8 September 2024 11:00
Wuhan	1 h	11,932/2557/2558	20 May 2024 13:00	8 September 2024 13:00	8 September 2024 14:00
Wuhan	6 h	11,915/2549/2555	20 May 2024 06:00	8 September 2024 07:00	8 September 2024 13:00
Wuhan	12 h	11,898/2541/2553	19 May 2024 21:00	8 September 2024 01:00	8 September 2024 13:00
Wuhan	24 h	11,870/2527/2550	19 May 2024 05:00	7 September 2024 06:00	8 September 2024 06:00
Wuhan	48 h	11,830/2500/2546	18 May 2024 11:00	6 September 2024 11:00	8 September 2024 11:00

Table A4. City-specific ecological input control at 6 h.

City	Full RMSE	No-Ecology RMSE	Reduction	Reduction (%)
Beijing	15.205	15.015	−0.190	−1.26
Shanghai	11.739	11.460	−0.279	−2.44
Chengdu	10.238	9.537	−0.700	−7.34
Wuhan	12.923	12.782	−0.141	−1.10

Notes: Values are averages of five seeds within each city. The control excludes ecological information but retains the complete network. Positive reductions indicate lower RMSE for the full model; negative reductions indicate lower RMSE for the control. RMSE is in $\mu g/m^3$.

Table A5. Selected checkpoints and training duration in the repeated 6 h experiments.

Model	City	Selected Epoch	Total Epochs	Seeds
AFGN	Beijing	7.6	14.4	5
AFGN	Shanghai	4.4	12.4	5
AFGN	Chengdu	7.4	15.2	5
AFGN	Wuhan	4.0	12.0	5
GRU	Beijing	6.8	14.8	5
GRU	Shanghai	4.4	12.4	5
GRU	Chengdu	7.4	15.4	5
GRU	Wuhan	5.8	13.8	5
LSTM	Beijing	3.0	11.0	5
LSTM	Shanghai	1.4	9.4	5
LSTM	Chengdu	3.2	11.2	5
LSTM	Wuhan	3.6	11.2	5

Notes: Selected checkpoint and total training epochs are summarized over four cities and five random seeds. The main and ecological input comparisons use the same five seeds.

Table A6. External hourly cross-validation of the four-city modeling archive.

City	Matched Hours	Pearson R	MAE (μg/m^3)	Within 5 μg/m^3 (%)
Beijing	17,369	0.999805	0.437	99.93
Shanghai	17,367	0.999921	0.250	100.00
Chengdu	17,368	0.999226	0.921	98.49
Wuhan	17,368	0.999344	1.366	96.81

Notes: MAE denotes mean absolute difference. The within-5 μg/m^3 column gives the percentage of matched observations whose absolute difference does not exceed 5 μg/m^3. Records are matched by city and hourly timestamp against the archive of Silver et al. [46].

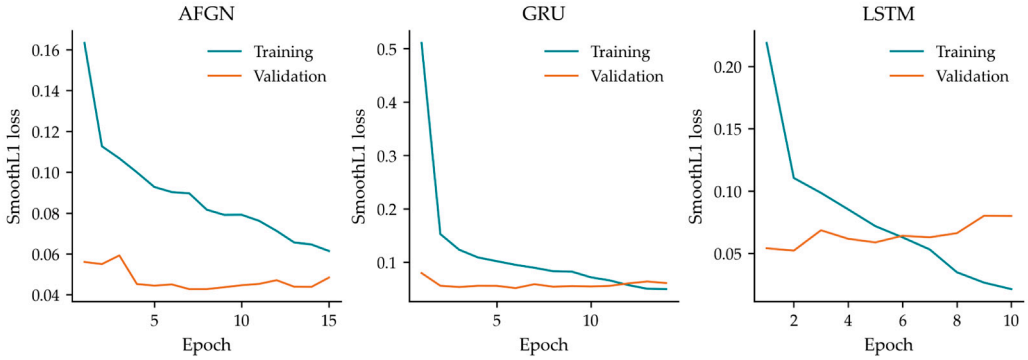


Figure A1. Beijing training and validation loss trajectories at 6 h (seed 42). The left, middle, and right panels show AFGN, GRU, and LSTM, respectively.

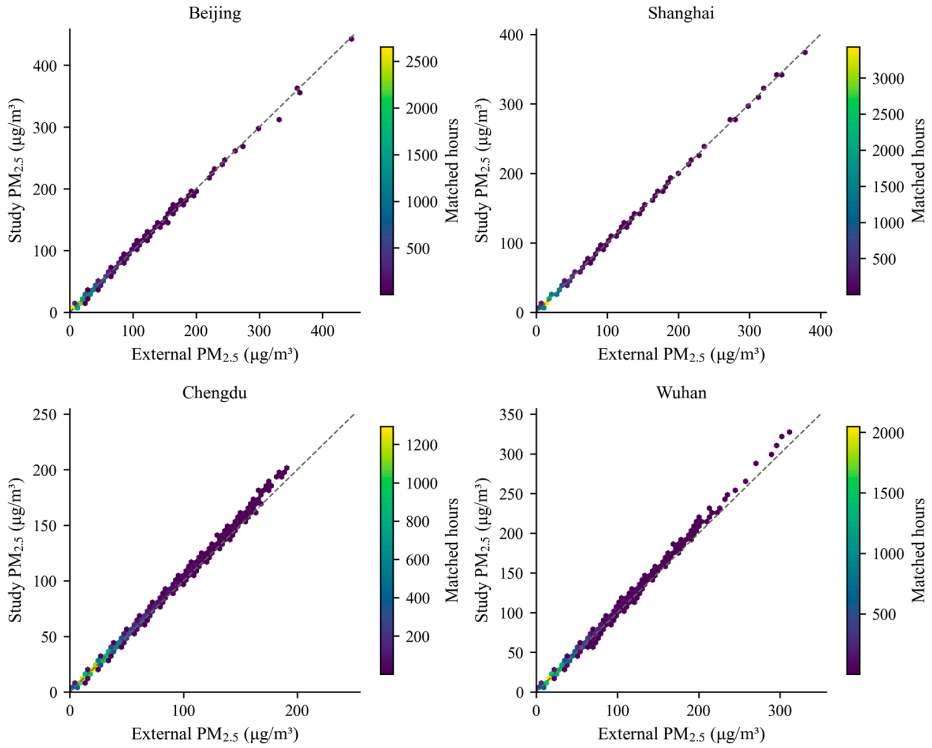


Figure A2. Hourly cross-validation against the Silver et al. archive [46]. Panels show Beijing (**upper left**), Shanghai (**upper right**), Chengdu (**lower left**), and Wuhan (**lower right**). Notes: This hex-bin plot visualises pairwise $PM_{2.5}$ records matched by city and hourly timestamp between the dataset used in our work and the open-access dataset from a representative recent publication, as shown in Reference [46]. It provides supplementary visual evidence for the reliability and robustness of the observational dataset adopted herein. Quantitative cross-validation metrics are summarised in Table A6, which serves as the primary supporting evidence for data quality assessment.

References

1. World Health Organization. *WHO Global Air Quality Guidelines: Particulate Matter ($PM_{2.5}$ and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide*; World Health Organization: Geneva, Switzerland, 2021. Available online: https://www.who.int/publications/i/item/9789240034228 (accessed on 12 October 2026).
2. GBD 2021 Risk Factors Collaborators. 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]
3. Santiago Fink, H. Human-Nature for Climate Action: Nature-Based Solutions for Urban Sustainability. *Sustainability* **2016**, *8*, 254. [CrossRef]
4. De Silva, A.; Amaratunga, D.; Haigh, R. Green and Blue Infrastructure as Nature-Based Better Preparedness Solutions for Disaster Risk Reduction: Key Policy Aspects. *Sustainability* **2022**, *14*, 16155. [CrossRef]
5. Hysa, A.; Löwe, R.; Geist, J. Ecosystem services potential is declining across European capital metropolitan areas. *Sci. Rep.* **2024**, *14*, 8903. [CrossRef] [PubMed]
6. Nowak, D.J.; Hirabayashi, S.; Bodine, A.; Hoehn, R. Modeled $PM_{2.5}$ removal by trees in ten U.S. cities and associated health effects. *Environ. Pollut.* **2013**, *178*, 395–402. [CrossRef] [PubMed]
7. Janhäll, S. Review on urban vegetation and particle air pollution: Deposition and dispersion. *Atmos. Environ.* **2015**, *105*, 130–137. [CrossRef]
8. Bowler, D.E.; Buyung-Ali, L.; Knight, T.M.; Pullin, A.S. Urban greening to cool towns and cities: A systematic review of the empirical evidence. *Landsc. Urban Plan.* **2010**, *97*, 147–155. [CrossRef]
9. Yao, M.; Smith, M.; Peng, C. Modelling the effects of vegetation and urban form on air quality in real urban environments: A systematic review of measurements, methods, and predictions. *Urban For. Urban Green.* **2025**, *105*, 128693. [CrossRef]
10. Abhijith, K.V.; Kumar, P.; Gallagher, J.; McNabola, A.; Baldauf, R.; Pilla, F.; Broderick, B.; Di Sabatino, S.; Pulvirenti, B. Air pollution abatement performances of green infrastructure in open road and built-up street canyon environments: A review. *Atmos. Environ.* **2017**, *162*, 71–86. [CrossRef]
11. Hu, M.; Ghorbany, S. Optimizing the green-grey ratio: Nonlinear effects of urban form on $PM_{2.5}$ in mid-sized cities. *Sustain. Cities Soc.* **2026**, *141*, 107293. [CrossRef]
12. Moronta-Sabad, H.; Ariño, A.H.; de la Calle-Arroyo, C.; Santos-Buitrago, R.; Santamaría, J.M.; Pons, J.J.; Elustondo, D. Evaluating the impact of roadside vegetation barriers on urban air pollution using low-cost mobile sensors. *Environ. Pollut.* **2025**, *374*, 126106. [CrossRef] [PubMed]
13. Peng, Z.; Lei, L.; Tan, Z.-M.; Zhang, M.; Ding, A.; Kou, X. Dynamics-based estimates of decline trend with fine temporal variations in China's $PM_{2.5}$ emissions. *Atmos. Chem. Phys.* **2023**, *23*, 14505–14520. [CrossRef]
14. Zhang, Y.-L.; Cao, F. Fine particulate matter ($PM_{2.5}$) in China at a city level. *Sci. Rep.* **2015**, *5*, 14884. [CrossRef] [PubMed]
15. Zhang, X.; Xiao, X.; Wang, F.; Brasseur, G.; Chen, S.; Wang, J.; Gao, M. Observed sensitivities of $PM_{2.5}$ and O_3 extremes to meteorological conditions in China and implications for the future. *Environ. Int.* **2022**, *168*, 107428. [CrossRef] [PubMed]
16. Gerçek, D.; Güven, İ.T. Assessment of Mutual Variation of Near-Surface Air Temperature, Land Surface Temperature and Driving Urban Parameters at Urban Microscale. *Sustainability* **2023**, *15*, 15710. [CrossRef]
17. Corgo, J.; Cruz, S.S.; Conceição, P. Nature-based solutions in spatial planning and policies for climate change adaptation: A literature review. *Ambio* **2024**, *53*, 1599–1617. [CrossRef] [PubMed]
18. Dunlop, T.; Khojasteh, D.; Cohen-Shacham, E.; Glamore, W.; Haghani, M.; van den Bosch, M.; Rizzi, D.; Greve, P.; Felder, S. The evolution and future of research on nature-based solutions to address societal challenges. *Commun. Earth Environ.* **2024**, *5*, 132. [CrossRef]
19. Qi, Y.; Li, Q.; Karimian, H.; Liu, D. A hybrid model for spatiotemporal forecasting of $PM_{2.5}$ based on graph convolutional neural network and long short-term memory. *Sci. Total Environ.* **2019**, *664*, 1–10. [CrossRef] [PubMed]
20. Wang, P.; Wu, Q.; Zhang, G. A hybrid prediction model for $PM_{2.5}$ concentration based on high-frequency and low-frequency IMFs with EMD decomposition. *Sci. Rep.* **2026**, *16*, 4969. [CrossRef] [PubMed]
21. Zheng, Q.; Tian, X.; Yu, Z.; Jiang, N.; Elhanashi, A.; Saponara, S.; Yu, R. Application of wavelet-packet transform driven deep learning method in $PM_{2.5}$ concentration prediction: A case study of Qingdao, China. *Sustain. Cities Soc.* **2023**, *92*, 104486. [CrossRef]
22. Pei, Y.; Huang, C.-J.; Shen, Y.; Ma, Y. An Ensemble Model with Adaptive Variational Mode Decomposition and Multivariate Temporal Graph Neural Network for $PM_{2.5}$ Concentration Forecasting. *Sustainability* **2022**, *14*, 13191. [CrossRef]
23. Zhou, S.; Wang, W.; Zhu, L.; Qiao, Q.; Kang, Y. Deep-learning architecture for $PM_{2.5}$ concentration prediction: A review. *Environ. Sci. Ecotechnol.* **2024**, *21*, 100400. [CrossRef] [PubMed]
24. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]

25. Cho, K.; van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2014; pp. 1724–1734. [CrossRef]
26. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. In Proceedings of the International Conference on Learning Representations (ICLR), Toulon, France, 24–26 April 2017. Available online: https://openreview.net/forum?id=SJU4ayYgl (accessed on 9 August 2026).
27. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems 30*; Guyon, I., von Luxburg, U., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30. Available online: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html (accessed on 12 October 2026).
28. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 11106–11115. [CrossRef]
29. Wu, H.; Xu, J.; Wang, J.; Long, M. Autoformer: Decomposition Transformers with auto-correlation for long-term series forecasting. In *Advances in Neural Information Processing Systems 34*; Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P.S., Vaughan, J.W., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2021; Volume 34, pp. 22419–22430. Available online: https://proceedings.neurips.cc/paper/2021/hash/bcc0d400288793e8bdcd7c19a8ac0c2b-Abstract.html (accessed on 12 October 2026).
30. Nie, Y.; Nguyen, N.H.; Sinthong, P.; Kalagnanam, J. A time series is worth 64 words: Long-term forecasting with Transformers. In Proceedings of the International Conference on Learning Representations (ICLR), Kigali, Rwanda, 1–5 May 2023. Available online: https://openreview.net/forum?id=Jbdc0vTOcol (accessed on 9 August 2026).
31. Liu, Y.; Hu, T.; Zhang, H.; Wu, H.; Wang, S.; Ma, L.; Long, M. iTransformer: Inverted Transformers are effective for time series forecasting. In Proceedings of the International Conference on Learning Representations (ICLR), Vienna, Austria, 7–11 May 2024. Available online: https://proceedings.iclr.cc/paper_files/paper/2024/hash/2ea18fdc667e0ef2ad82b2b4d65147ad-Abstract-Conference.html (accessed on 12 October 2026).
32. Wu, H.; Hu, T.; Liu, Y.; Zhou, H.; Wang, J.; Long, M. TimesNet: Temporal 2D-variation modeling for general time series analysis. In Proceedings of the International Conference on Learning Representations (ICLR), Kigali, Rwanda, 1–5 May 2023. Available online: https://openreview.net/forum?id=ju_Uqw384Oq (accessed on 9 August 2026).
33. Lim, B.; Arık, S.Ö.; Loeff, N.; Pfister, T. Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. *Int. J. Forecast.* **2021**, *37*, 1748–1764. [CrossRef]
34. Zeng, A.; Chen, M.; Zhang, L.; Xu, Q. Are Transformers Effective for Time Series Forecasting? *Proc. AAAI Conf. Artif. Intell.* **2023**, *37*, 11121–11128. [CrossRef]
35. Teng, M.; Li, S.; Yang, J.; Chen, J.; Fan, C.; Ding, Y. A new hybrid deep neural network for multiple sites $PM_{2.5}$ forecasting. *J. Clean. Prod.* **2024**, *473*, 143542. [CrossRef]
36. Cai, S.; Fang, F.; Peuch, V.-H.; Alexe, M.; Navon, I.M.; Wang, Y. D-DNet: A dual deep neural network framework for high-efficiency operational $PM_{2.5}$ and AOD550 forecasting with data assimilation. *J. Geophys. Res. Mach. Learn. Comput.* **2026**, *3*, e2025JH000949. [CrossRef]
37. Qiu, Y.; Feng, J.; Zhang, Z.; Zhao, X.; Li, Z.; Ma, Z.; Liu, R.; Zhu, J. Regional aerosol forecasts based on deep learning and numerical weather prediction. *npj Clim. Atmos. Sci.* **2023**, *6*, 71. [CrossRef]
38. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.-C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. A* **1998**, *454*, 903–995. [CrossRef]
39. Torres, M.E.; Colominas, M.A.; Schlotthauer, G.; Flandrin, P. A complete ensemble empirical mode decomposition with adaptive noise. In *Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, 22–27 May 2011*; IEEE: New York, NY, USA, 2011; pp. 4144–4147. [CrossRef]
40. Colominas, M.A.; Schlotthauer, G.; Torres, M.E. Improved complete ensemble EMD: A suitable tool for biomedical signal processing. *Biomed. Signal Process. Control* **2014**, *14*, 19–29. [CrossRef]
41. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [CrossRef]
42. Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; Jin, R. FEDformer: Frequency Enhanced Decomposed Transformer for Long-term Series Forecasting. In *Proceedings of the 39th International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2022*; PMLR: Hamilton, New Zealand, 2022; Volume 162, pp. 27268–27286. Available online: https://proceedings.mlr.press/v162/zhou22g.html (accessed on 12 October 2026).
43. Shen, Q.; Guo, J.; Tian, Y.; Chen, Z.-S. An EMD-based forecasting framework integrating GMM and BiLSTM for helicopter engine anomaly detection. *J. Ind. Inf. Integr.* **2025**, *48*, 101003. [CrossRef]

44. Cerqueira, V.; Torgo, L.; Mozetič, I. Evaluating time series forecasting models: An empirical study on performance estimation methods. *Mach. Learn.* **2020**, *109*, 1997–2028. [CrossRef]
45. Chen, X.; Gao, J.; Chen, L.; Khanna, M.; Gong, B.; Auffhammer, M. The spatiotemporal pattern of surface ozone and its impact on agricultural productivity in China. *Proc. Natl. Acad. Sci. Nexus* **2024**, *3*, pgad435. [CrossRef] [PubMed]
46. Silver, B.; Reddington, C.L.; Chen, Y.; Arnold, S.R. A decade of China's air quality monitoring data suggests health impacts are no longer declining. *Environ. Int.* **2025**, *197*, 109318. [CrossRef] [PubMed]
47. Xu, J.; Sun, X.; Zhang, Z.; Zhao, G.; Lin, J. Understanding and improving layer normalization. In *Advances in Neural Information Processing Systems 32*; Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2019; Volume 32. Available online: https://proceedings.neurips.cc/paper/2019/hash/2f4fe03d77724a7217006e5d16728874-Abstract.html (accessed on 12 October 2026).
48. Diebold, F.X.; Mariano, R.S. Comparing predictive accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–263. [CrossRef]
49. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. In Proceedings of the International Conference on Learning Representations (ICLR), New Orleans, LA, USA, 6–9 May 2019. Available online: https://openreview.net/forum?id=Bkg6RiCqY7 (accessed on 9 August 2026).
50. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
51. Prechelt, L. Early Stopping-But When? In *Neural Networks: Tricks of the Trade*; Orr, G.B., Müller, K.-R., Eds.; Springer: Berlin, Germany, 1998; pp. 55–69. [CrossRef] [PubMed]
52. Huber, P.J. Robust Estimation of a Location Parameter. *Ann. Math. Stat.* **1964**, *35*, 73–101. [CrossRef]
53. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)? Arguments Against Avoiding RMSE in the Literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [CrossRef]
54. Xiao, F.; Yang, M.; Fan, H.; Fan, G.; Al-Qaness, M.A.A. An improved deep learning model for predicting daily $PM_{2.5}$ concentration. *Sci. Rep.* **2020**, *10*, 20988. [CrossRef] [PubMed]
55. Zhang, Z.; Xu, J.; Zhao, X.; Cheng, S. Improving $PM_{2.5}$ and Visibility Predictions in China Using Machine Learning and Ensemble Forecasting. *J. Geophys. Res. Mach. Learn. Comput.* **2025**, *2*, e2025JH000640. [CrossRef]
56. Cai, S.; Fang, F.; Tang, X.; Zhu, J.; Wang, Y. A Hybrid Data-Driven and Data Assimilation Method for Spatiotemporal Forecasting: $PM_{2.5}$ Forecasting in China. *J. Adv. Model. Earth Syst.* **2024**, *16*, e2023MS003789. [CrossRef]
57. Gualtieri, G.; Brilli, L.; Carotenuto, F.; Cavaliere, A.; Gioli, B.; Giordano, T.; Putzolu, S.; Vagnoli, C.; Zaldei, A. Assessing Capability of Copernicus Atmosphere Monitoring Service to Forecast $PM_{2.5}$ and PM10 Hourly Concentrations in a European Air Quality Hotspot. *Atmos. Pollut. Res.* **2025**, *16*, 102567. [CrossRef]
58. Künsch, H.R. The Jackknife and the Bootstrap for General Stationary Observations. *Ann. Stat.* **1989**, *17*, 1217–1241. [CrossRef]
59. Sun, Y.; Chen, C.; Zhang, Y.; Xu, W.; Zhou, L.; Cheng, X.; Zheng, H.; Ji, D.; Li, J.; Tang, X.; et al. Rapid formation and evolution of an extreme haze episode in Northern China during winter 2015. *Sci. Rep.* **2016**, *6*, 27151. [CrossRef] [PubMed]
60. Li, J.; Han, Z.; Wu, Y.; Xiong, Z.; Xia, X.; Li, J.; Liang, L.; Zhang, R. Aerosol radiative effects and feedbacks on boundary layer meteorology and $PM_{2.5}$ chemical components during winter haze events over the Beijing-Tianjin-Hebei region. *Atmos. Chem. Phys.* **2020**, *20*, 8659–8690. [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.