

Supplementary Material

Physicochemically Motivated Decomposition for Long-Sequence Urban O₃ Forecasting: A Cross-Method Comparison

Jucheng Li¹, Chuanyu Zhou², Guangming Shi², Xinyi Liu^{1,2*}, Yu Zhan²

¹Department of Environmental Science and Engineering, Sichuan University, Chengdu, 610065, China

²College of Carbon Neutrality Future Technology, Sichuan University, Chengdu, 610065, China

*e-mail: liuxinyi1@stu.scu.edu.cn

Text S1 Technical Specifications of the NCEP-GFS Dataset and Temporal Interpolation Methodology

The Global Forecast System (GFS), maintained by the National Centers for Environmental Prediction (NCEP), is a sophisticated global numerical weather prediction (NWP) system [1]. The system generates four forecast cycles daily, initiated at 02:00, 08:00, 14:00, and 20:00 CST, utilizing the Finite-Volume Cubed-Sphere dynamical core to provide global grids at a $0.25^\circ \times 0.25^\circ$ resolution. In this study, to ensure a stable and reliable data supply for our daily 23:00 CST air quality predictions, we consistently utilize the GFS forecast cycle initialized at 08:00 CST. Two data access modes are involved in this workflow. For actual operational application, near-real-time GFS forecasts can be obtained from the official NCEP portal [2]. However, because this portal only retains a short rolling window of recent data, it cannot support retrospective data collection. Therefore, for retrospective model development and evaluation, this study used archived operational GFS forecast products from the archive system.

Regarding temporal resolution, the GFS archive provides hourly outputs for the first 120 hours, before transitioning to a 3-hour interval thereafter. To maintain a uniform 1-hourly sequence for the entire 168-hour forecast window, Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) is employed to interpolate the 3-hourly data segments [3]. PCHIP is specifically chosen for meteorological data due to its ‘shape-preserving’ characteristics. Unlike standard cubic splines or linear interpolations, PCHIP ensures that the interpolant is monotonic on intervals where the data is monotonic, and it avoids overshooting at local extrema. Mathematically, for a given set of data points (x_i, y_i) , PCHIP constructs a cubic polynomial on each subinterval $[x_i, x_{i+1}]$ such that both the function values and the first derivatives (slopes) are continuous. The slopes d_i at each point x_i are

calculated using a harmonic mean approach that considers the neighboring intervals, ensuring that the resulting curve does not oscillate beyond the range of the surrounding data.

Text S2 Detailed Architectures and Data Flow of the LSTM and Transformer Models

LSTM Architecture

We employ a deep LSTM network designed for high-dimensional time series forecasting (Text S3 and Fig. S1b), and the data flow strictly follows the modules illustrated in Fig. 2b:

1. **History Branch:** Before Feature Fusion, the raw pollutant and meteorological inputs (Poll&Meteo) from the History Branch are first passed through Time Projection (fully connected network) to obtain a high-dimensional latent sequence aligned to the prediction horizon.
2. **Future Branch:** Future meteorological forecasts are combined element-wise with encoded time features to form a time-aware future sequence (without time projection).
3. **Feature Fusion:** The latent representations from the history branch and the time-aware meteorological sequences from the Future Branch are integrated via a channel-wise concatenation.
4. **LSTM Layers:** The fused feature sequence is then fed into a multi-layer LSTM architecture, where hierarchical temporal dependencies across multiple time scales are extracted.
5. **Boost Module:** The LSTM outputs are subsequently passed through a Boost Module composed of feed-forward layers, which applies non-linear transformations to further enhance feature representation capacity.
6. **Feature Projection & Output Target:** Finally, a Feature Projection layer constructed with fully connected networks maps the latent representations to the target dimension to predict the forecast increment ΔY . The final output is obtained by adding ΔY to the initial value Y_{t0} , yielding the model's O_3 forecast.

Transformer Architecture

Complementing the recurrent structure, the Transformer architecture (Fig. 2c) leverages attention mechanisms to capture global temporal dependencies [4], and its data flow follows the encoder-decoder paradigm:

1. **History Branch & Future Branch:** Unlike the LSTM architecture, the Transformer history branch is not passed through the Time Projection layer, while the remaining processing follows the same design as in the LSTM.
2. **Encoder Layer:** The processed history sequence is fed into the encoder stack, and each layer consists of multi-head self-attention and feed-forward networks, each wrapped with Add & Norm (residual connection and layer normalization), to extract long-term patterns from past observations.

3. **Decoder Layer & Cross-Attention:** The decoder stack first employs masked multi-head self-attention to prevent future leakage. Crucially, the subsequent multi-head attention sub-layer serves as cross-attention, where the decoder utilizes its future state as Queries (Q) to attend to the encoder's historical outputs (acting as Keys K and Values V), enabling the decoder to retrieve relevant historical context for forecasting (Text S3 and Fig. S1c, S1d).
4. **Feature Projection & Output Target:** The feature projection and output reconstruction are identical to the LSTM architecture.

Text S3 Detailed Architectures of Neural Network Modules

Time Feature Encoder (Fig. S1a)

To effectively capture the multi-scale temporal dependencies inherent in air quality data, the Time Feature Encoder (Fig. S1a) processes timestamp sequences spanning the historical input window (T_{history}) up to the future 168-hour forecast horizon (T_{future}). The architecture operates through three parallel streams, each dedicated to specific time components:

1. **Phase Encoding:** This stream targets continuous cyclical patterns by processing time components such as the Hour of Day (0–23) and the Day of Week (0–6), which are mapped into continuous cyclical coordinates using sine and cosine functions, which accurately preserve the periodicity of daily and weekly pollutant variations.
2. **Categorical Embeddings:** This stream handles discrete temporal identifiers to capture non-linear contextual dependencies. It processes the Month of Year (1–12), Day of Month (1–31), and Hour of Day (0–23).
3. These discrete inputs are transformed into high-dimensional dense vectors using learnable embedding layers, enabling the model to distinguish between seasonal shifts (Months), monthly trends (Days), and daily variations (Hours).
4. **Positional Encoding:** This stream incorporates absolute sequence position information (e.g., time steps from 1 to T) to ensure that the model can distinguish the chronological order of the input time series throughout the entire observation-to-forecast period [4].
5. Finally, these heterogeneous features are integrated through a Feature Fusion layer, forming a unified, learnable time representation that captures both periodic and non-periodic temporal dependencies.

Long Short-Term Memory (LSTM) Unit (Fig. S1b)

The LSTM unit (Fig. S1b) is designed to mitigate the vanishing gradient problem in long sequences by relying on a core mechanism of three gating structures [5]: The Forget Gate determines which information to discard from the previous cell state. The Input Gate controls the update of new information into the cell state. The Output Gate regulates the information flow to

the next hidden state. This architecture enables the model to selectively retain historical air quality patterns over extended periods while filtering out irrelevant noise.

Scaled Dot-Product Attention (Fig. S1c)

As the fundamental mechanism of the Transformer, the Scaled Dot-Product Attention (Fig. S1c) computes the relevance between elements using the classic formula [4]:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (S.1)$$

Where Q (Query), K (Key), and V (Value) are the input matrices. The attention scores are derived from the dot product of the queries and keys (QK^T), scaled by the inverse square root of the key dimension (d_k) to prevent gradient vanishing in the SoftMax function. An optional Mask operation is applied to prevent the model from attending to future positions. Finally, the normalized weights are applied to the values (V) to generate the output.

Multi-Head Attention (Fig. S1d)

To capture information from different representation subspaces, the Multi-Head Attention mechanism (Fig. S1d) projects Q , K , and V into multiple parallel heads. Each head performs Scaled Dot-Product Attention independently, allowing the model to focus on various aspects of the meteorological and pollutant interactions simultaneously. The outputs of all heads are then concatenated and linearly projected to generate the final context-aware feature representation.

Text S4 Computational Environment and Hardware Specifications

The deep learning models were constructed, trained, and evaluated using the Ubuntu 24.04.3 LTS operating system. We leveraged the PyTorch framework (version 1.13.1+cu117) to build the deep learning architectures and conduct model predictions. The computations were performed on a high-performance computing server with the following specifications: dual AMD EPYC 7763 64-Core Processors (totaling 128 physical cores and 256 threads), eight NVIDIA GeForce RTX 4080 SUPER GPUs (16 GB VRAM each), and 2.0 TiB of RAM.

Text S5 Calculation of Maximum Daily 8-hour Average (MDA8) O₃

The MDA8 O₃ concentration is widely used as a standard metric for air quality assessment, which is computed from hourly O₃ observations using a strict two-step procedure:

Running 8-hour Averages: The 8-hour moving average concentrations are calculated for each time step t by averaging the hourly concentrations within the interval $[t, t+7]$:

$$A_t = \frac{1}{N_v} \sum_{i=0}^7 C_{t+i} \cdot I(C_{t+i} \text{ is valid}) \quad (S.2)$$

where C denotes the hourly concentration, and valid averages require at least 6 valid hourly readings ($N_v \geq 6$) within the window.

Daily Maximum Selection: The MDA8 for day d is determined as the maximum of the 24 possible rolling averages starting within that calendar day:

$$\text{MDA8}_d = \max\{A_t \mid t \in_d\} \quad (\text{S. 3})$$

Text S6 Formula of the statistical metrics

We assessed model performance using four complementary metrics: the coefficient of determination (R^2), the Pearson correlation coefficient (R), the root mean square error (RMSE), and the index of agreement (IA). Here, R^2 is **not** interpreted as R squared. Instead, R^2 summarizes how well the predictions reproduce the observed variability (i.e., overall fit), whereas R captures the strength and direction of the linear association between predicted and observed values. RMSE reports the typical magnitude of prediction errors in the original units, serving as an accuracy measure. IA provides a dimensionless, bounded agreement score (closer to 1 indicates better agreement), reflecting how closely predictions match observations while penalizing large deviations.

$$R^2 = \frac{\sum_{i=1}^N (x_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (\text{S. 4})$$

$$R = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (\text{S. 5})$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (x_i - y_i)^2}{N}} \quad (\text{S. 6})$$

$$\text{IA} = 1 - \frac{\sum_{i=1}^N (x_i - y_i)^2}{\sum_{i=1}^N (|x_i - \bar{y}| + |y_i - \bar{y}|)^2} \quad (\text{S. 7})$$

x_i : Predicted value for the i -th observation,

y_i : True value for the i -th observation,

$\bar{y}$: Mean of all observations,

N : Number of observations

Text S7 Descriptive Statistics

During 2019 to 2024, the station-based data indicated that Shanghai ($69.0 \pm 38.9 \mu\text{g}/\text{m}^3$) and Beijing ($63.5 \pm 47.4 \mu\text{g}/\text{m}^3$) experienced higher levels of O_3 pollution, compared to Chengdu ($57.0 \pm 44.5 \mu\text{g}/\text{m}^3$) and Guangzhou ($56.0 \pm 46.1 \mu\text{g}/\text{m}^3$) (Table S3) [6]. O_3 concentrations across these cities generally exhibited a dip-and-rebound pattern, reaching a relative minimum in 2021 before rebounding significantly in 2022 [7], especially in Beijing and Shanghai (Table S4). These four cities exhibited distinct seasonal O_3 cycles: Beijing and Chengdu concentrations peaked in summer and reached their lowest levels in winter. In contrast, Shanghai showed elevated levels in both

spring and summer, while Guangzhou exhibited a unique peak during the autumn months (Table S4). The spatiotemporal variations of O₃ were closely linked to meteorological drivers, as solar radiation (DSWRF, R : 0.59-0.71) and temperature (R : 0.32-0.65) emerged as primary positive drivers across all cities, underscoring the dominance of photochemical production. Conversely, O₃ levels were generally suppressed by relative humidity (R : -0.59 to -0.16) and surface pressure (R : -0.43 to 0.03). In Guangzhou, relative humidity exhibited the strongest negative association with O₃ variability (Pearson correlation coefficient R = -0.59), whereas in Beijing and Chengdu, O₃ variations are more strongly linked to temperature and solar radiation. Regarding wind influence, O₃ concentrations in Beijing and Chengdu showed moderate correlations with wind components (e.g., VW in Beijing, R = 0.29), reflecting the role of regional transport, while wind correlations were generally weaker in Shanghai and Guangzhou (Fig. S2).

Text S8 Process-oriented evaluation of the extracted diurnal O₃ component

To further substantiate the ‘physicochemically motivated’ interpretation, we performed a process-oriented evaluation of the extracted diurnal O₃ component from STL, GAM, CEEMDAN, and VMD. First, we examined temporal consistency using the absolute deviation between extracted diurnal peaks and meteorological driver peaks (Table S9), together with the peak-time bias relative to observed daily O₃ peaks (Table S10) [8]. For alignment with meteorological drivers, STL and GAM generally show slightly smaller deviations across cities, with mean deviations of about 1.60-2.39 h against Tem peaks and 2.31-3.65 h against USWRF peaks. By comparison, CEEMDAN and VMD tend to be slightly larger. For peak-time bias (Table S10), VMD shows the smallest deviations in most cities, with mean biases of about 1.00-1.51 h, whereas STL and GAM generally exhibit larger biases across the four cities.

Second, we examined whether the extracted diurnal component preserved regime-dependent variability under changing meteorological conditions using a local diurnal isolation index (DII) [9], defined as the fraction of wavelet power within the 20-28 h band relative to the total power in the 2-72 h range (Table S11). Across the four cities, CEEMDAN shows the strongest regime sensitivity, with frontal-to-stagnant DII ratios of 1.27 in Beijing and 1.12 in Chengdu, while VMD shows weaker but still detectable responses (1.04 and 1.01, respectively). In contrast, STL and GAM remain close to 1.00, indicating much weaker sensitivity to regime-dependent redistribution of diurnal-scale power.

Third, we evaluated the chemical relevance of the extracted diurnal component by relating its daily amplitude to the daily amplitude of NO₂ during 2019-2023 (Table S12) [10]. STL and GAM exhibit negligible day-to-day amplitude variation due to their structurally constrained decomposition, and therefore, the corresponding correlations with NO₂ amplitude cannot be computed. In contrast, CEEMDAN and VMD better preserve amplitude modulation associated with precursor variability, leading to stronger positive correlations. Between these two,

CEEMDAN performs slightly better overall, reaching 0.54 and 0.51 in Chengdu and Guangzhou, compared with 0.52 and 0.50 for VMD. In Shanghai, CEEMDAN also remains higher (0.32 vs. 0.25), whereas Beijing shows near-zero correlations for both methods. These results provide chemical support for the physicochemical interpretation of the extracted diurnal component.

Text S9 Comparison with Chemical Transport Models (CTMs)

To evaluate the performance of the proposed data-driven framework against traditional chemical transport models (CTMs), we compared ΔY -Transformer-VMD with the operational Goddard Earth Observing System Composition Forecast (GEOS-CF) [11] product in Chengdu for the full year of 2024 (Table S14). GEOS-CF is a widely used global atmospheric chemistry-climate model that provides daily forecasts of ozone and other trace gases by simulating emissions, transport, chemical reactions, and deposition processes across the troposphere. Chengdu was selected as a meteorologically complex basin city, where stagnant conditions, episodic ventilation, and boundary-layer variability make O₃ forecasting challenging, providing a meaningful test case.

The comparison was performed at the daily scale for forecast Days 1, 3, and 5. The results show that ΔY -Transformer-VMD consistently outperforms GEOS-CF: R^2 increases by 0.19, 0.16, and 0.16, and RMSE decreases by 9.2, 6.9, and 6.0 $\mu\text{g}/\text{m}^3$, respectively. This demonstrates the added value of the proposed framework in improving forecast accuracy. These findings highlight the superior predictive performance of the proposed data-driven framework compared to traditional CTMs.

Text S10 Uncertainty Quantification

We conducted a 2.5th-97.5th quantile (q2.5-q97.5) prediction interval analysis to evaluate the uncertainty of the O₃ forecasts, where the 5th and 95th quantiles denote the lower and upper bounds of the predicted distribution, forming a 95% prediction interval [12]. The Prediction Interval Coverage Probability (PICP) measures the fraction of observations within this interval, and the Mean Prediction Interval Width (MPIW) indicates its average width. For the four cities, PICP values are 0.169 (Beijing), 0.245 (Chengdu), 0.143 (Guangzhou), and 0.349 (Shanghai), with corresponding MPIW values of 7.82, 12.44, 7.06, and 19.86 $\mu\text{g}/\text{m}^3$ (Table S15). These results indicate that robust uncertainty quantification remains challenging with the current deterministic framework.

Text S11 Analysis of Spatial Heterogeneity

The spatial heterogeneity in optimal decomposition-model combinations likely reflects differences in climatology, emission structures, and dominant temporal variability across cities. These factors influence both the separability of temporal scales in the ozone signal and the suitability of different learning architectures.

In Beijing, ozone variability is strongly influenced by pronounced seasonal cycles, regional transport, and synoptic-scale meteorological patterns over the North China Plain [13]. These

processes produce relatively clear, large-scale temporal structures and long-range dependencies in the time series. Consequently, the GAM-Y-Transformer combination performs well, as GAM effectively captures smooth long-term trends while the Transformer architecture is capable of modeling long-range temporal dependencies.

In Chengdu, located within the Sichuan Basin, basin topography often traps pollutants and enhances interactions among local emissions, boundary-layer dynamics, and regional circulation [14]. These conditions generate complex multi-scale variability in ozone concentrations. Under such conditions, VMD-based decomposition helps separate overlapping temporal components, while the ΔY -Transformer formulation captures incremental changes more effectively in highly variable environments.

In Shanghai, coastal meteorology introduces multiple interacting temporal processes, including sea-land breeze circulations and synoptic transport from surrounding regions [15]. These mechanisms generate relatively clear multi-scale temporal structures. Methods such as VMD, which explicitly isolate different frequency modes, combined with Transformer-based models, can effectively utilize this multi-scale information.

By contrast, Guangzhou experiences a humid subtropical climate with frequent cloud variability, high humidity, and strong photochemical sensitivity to meteorological fluctuations [16]. These factors introduce stronger short-term disturbances and reduce the stability of decomposed components. Under such conditions, smoother decomposition methods such as STL, combined with lighter architectures like LSTM, tend to provide more stable representations.

Table S1. Summary of models and data settings for air quality forecasting studies

City	Target Pollutant	Model	Decomposition	Input Variables	Subsequence Count	Max Forecast Horizon
Beijing[17]	PM _{2.5}	BiLSTM	VMD	PM _{2.5}	8	1 hour (hourly)
Changchun/Beijing/Shanghai/Guangzhou/Shenzhen/Chongqing[18]	AQI	TimesNet-iTransformer-CPO	STL	PM _{2.5} , PM ₁₀	3 (trend, seasonal, residual)	15 days (daily)
Harbin/Cho	PM _{2.5}	SVR-GWO	CEEMD	PM _{2.5}	3 IMFs + residual	1 day

ngqing[19]					(4 total)	(daily)
Beijing/Dalian[20]	PM _{2.5}	RNN-LSTM	VMD / CEEMD / CEEMDAN / ICEEMDAN	PM _{2.5}	10 IMFs → 4 reconstructed sequences, Hybrid: Range Entropy	1 day (daily)
Beijing/Tangshan[21]	AQI	TimeGAN- ARIMA- PSO-VMD- CNN_BiLSTM- XGBoost	PSO-VMD	Six Polls, AQI, Tem, Pres, RH, WS, ARIMA error as auxiliary input	4 (Tangshan), 5 (Beijing), Hybrid: Envelope entropy (fitness for PSO- VMD)	1 hour (hourly)
Guangzhou /Lanzhou[22]	PM _{2.5}	PSR- LSSVM	EEMD	PM _{2.5}	5 IMFs + residual	1 day (daily)
Shanghai/Nanjing/Lianyungang/Beijing/Guangzhou[23]	AQI	Hybrid- FTS	CEEMD / EMD	AQI	Not stated (EMD/CEEMD + SE recombination), Hybrid: Sample Entropy	24 hours (hourly)
Beijing/Baoding[24]	AQI	LSTM	VMD	AQI	8 BIMFs → 5 recombined subseries, Hybrid: Sample Entropy	1 day (daily)
Beijing[25]	PM _{2.5}	CNN-GRU	EMD	PM _{2.5}	IMFs grouped into 3 sequences, Hybrid: CNN- based merge	24 hours (hourly)
Warsaw[26]	PM ₁₀	eSNN- Stacking- Bagging	CEEMDAN	PM ₁₀ , Tem, TP, RH, WS, TCC, DOW	5 CEEMDAN components	1 day (daily)
Beijing/Shanghai/Guangzhou[27]	AQI	RGTCN	STL + VMD (NSGA-II)	Six Polls, AQI, Tem, DP, RH, TP, WS, WG, WD, Pres, Flux	Not stated (VMD K optimized by NSGA-II), SecDec: STL	3 hours (hourly, multi- step)

					trend/season → VMD residual	
Beijing[28]	PM _{2.5}	DeepTCN	CEEMDAN	Six Polls, WS, WD, Tem, TP, Pres, RH, Wx, Time	15 IMFs + residual (16 total)	3 hours (hourly, 1–3 h)
Beijing/Tianjin/Baoding[29]	PM _{2.5}	LSTM	CEEMDAN + AVMD	PM _{2.5}	10 IMFs + residual, IMF1 → 8 AVMD modes, SecDec: IMF1 re-decomposition, Hybrid: Sample Entropy	1 day (daily)
South Korea[30]	PM ₁₀	LSTM	STL	CO ₂ , PM ₁₀ , Tem, RH, VOCs, Time	3 STL components (residual removed, trend+seasonal used)	1 hour (hourly)
Jing-Jin-Ji[31]	PM _{2.5}	HALN	None	PM _{2.5} , PM ₁₀ , SO ₂ , TCC, Vis, RH, WS	N/A	120 hours (hourly)
Baoding[32]	Six Polls	LSTM	CEEMDAN	Six Polls	11 IMFs → 6 groups (recombined), Hybrid: Sample + Fuzzy Entropy	6 hours (6-hourly)
Beijing[33]	AQI	AOA-DBGRU-MDN	CEEMDAN	Six Polls, AQI, Tem, RH, Pres, WS, WD, TP	8, Hybrid: Approximate Entropy, Fuzzy Entropy	1 hour (hourly)
Jinan/Shanghai/Harbin[34]	Six polls and AQI	MOSSA-ENN	MCEEMDAN (CEEMDAN + VMD)	Six Polls, AQI	8 (VMD modes), SecDec: Yes (VMD applied to high-frequency IMFs produced by CEEMDAN)	3 days (daily, 1–3-step)
Beijing[35]	O ₃	Transformer	STL	Six Polls, AQI, Tem, DP, WS, WD, Pres	3 (trend, seasonal, remainder)	168 hours (hourly)

Note: Table S1 summarizes representative decomposition–ensemble air-quality forecasting studies. STL = Seasonal–Trend decomposition using LOESS,

EMD/EEMD/CEEMD/CEEMDAN (and ICEEMDAN/MCEEMDAN) = empirical mode decomposition families producing multiple IMFs, VMD = variational mode decomposition with preset mode number K . ‘Subsequence count’ reports the number of IMFs/modes and/or the number of reconstructed (recombined) subseries after grouping. Entropy metrics (e.g., Sample/Range/Fuzzy/Envelope entropy) are used for component grouping/selection. Metaheuristics (PSO, GWO, AOA, MOSSA, NSGA-II, etc.) denote optimization of decomposition parameters and/or model hyperparameters. ‘Sub-models tuned separately’ indicates component-wise tuning. ‘Six Polls’ refers to $PM_{2.5}$, PM_{10} , SO_2 , NO_2 , CO , and O_3 , meteorological inputs (e.g., Tem, RH, Pres, wind, TP, PBLH), and time features (e.g., DOW) are included as stated. ‘Decomposition Strategy & Leakage Risk’ column assesses data leakage based on decomposition timing. ‘Separately applied’ effectively isolates the test set (low risk), whereas ‘Full then split’ leaks future information (high risk). Studies labeled ‘Not stated’ are assumed to use full-sequence decomposition by default and are thus classified as high risk.

Table S2. Hyperparameter search spaces and architectural configurations for LSTM and Transformer models optimized via the Optuna framework

Hyperparameter	LSTM	Transformer
Input Window (T_{history})	24, 48, 96, 168	24, 48, 96, 168
Hidden Size	32, 64, 128	32, 64, 128
LSTM Layers	1, 2	-
Transformer Encoder Layers	-	1, 2, 3
Transformer Decoder Layers	-	1, 2, 3
Attention Heads	-	1, 2, 4
Dropout Rate	0, 0.1, 0.2, 0.3	0, 0.1, 0.2, 0.3
Batch Size	128, 256, 512, 640, 768	128, 256, 512
Learning Rate	1e-3, 5e-4, 1e-4, 5e-5, 1e-5	1e-3, 5e-4, 1e-4, 5e-5, 1e-5
Optimizer	Adam, AdamW	Adam, AdamW
Weight Decay	1e-3, 1e-4, 1e-5, 1e-6	1e-3, 1e-4, 1e-5, 1e-6

Table S3. Descriptive statistics of station-based air pollutant data for the four cities ($\mu\text{g}/\text{m}^3$)

Variable	City	Max	Min	Mean	SD	Median	IQR
PM _{2.5}	CD	210.1	0.6	37.9	26.9	31	30.6
PM ₁₀	CD	442.5	1.4	59	36.9	50.4	45.4
NO ₂	CD	130.1	1.9	30.8	16.7	28	23.1
O ₃	CD	271.7	1	57	44.5	45.9	52.6
CO	CD	1.9	0.2	0.7	0.2	0.6	0.2
SO ₂	CD	48.5	1.4	5.1	2.4	4.7	2.8
PM _{2.5}	BJ	297	0	34.6	34	23.9	34.4
PM ₁₀	BJ	499	0	60.7	50.2	48.8	53.1
NO ₂	BJ	138.6	1	26.6	17.5	22.4	24.1
O ₃	BJ	325.3	0.1	63.5	47.4	54.8	54.9
CO	BJ	7.9	0	0.6	0.3	0.5	0.4
SO ₂	BJ	48.7	1	3.3	2.1	2.6	0.8
PM _{2.5}	SH	229.3	1.7	29.5	22.8	22.5	22.8
PM ₁₀	SH	500	0	44.3	31.7	36.2	29.7
NO ₂	SH	171.8	3.6	33.6	19.8	28	22.4
O ₃	SH	321.6	1.3	69	38.9	65.1	47.9
CO	SH	2.3	0.3	0.7	0.2	0.6	0.2
SO ₂	SH	30	3.1	6.5	2	6.2	2.2
PM _{2.5}	GZ	138.3	0.7	23.6	14.3	20.5	18.7
PM ₁₀	GZ	239.1	0.9	42.7	24.9	37	30.5
NO ₂	GZ	182.7	4.7	32.5	18.8	27.9	20.4
O ₃	GZ	327.9	2	56	46.1	42.9	58
CO	GZ	2	0.4	0.7	0.2	0.7	0.2
SO ₂	GZ	35.4	3.4	6.6	1.8	6.1	1.9

Table S4. Yearly changes of O₃ for the four cities (μg/m³)

City	2019	2020	2021	2022	2023	2024
Chengdu	50.6±42.2	55.2±44.9	52±42.2	60.5±46.2	61±44.9	62.9±45.1
Beijing	62.8±53.5	62.7±47.7	57.6±40.8	66.8±47.2	66.2±47.8	64.8±46
Shanghai	69.5±41.8	68.5±38.7	62.9±36.9	74±38.9	70.5±39.4	68.9±36.7
Guangzhou	54.4±51.8	54.3±46.7	56.5±45.5	57.6±48	56.9±43.2	56.3±40.2

Table S5. Seasonal changes of O₃ for the four cities (μg/m³)

City	Fall	Spring	Summer	Winter	All
Chengdu	43.4±34.4	68.4±45.2	82.8±50	32.9±24.9	57±44.5
Beijing	45.3±39.7	74.3±42.2	96.5±53.3	37.1±23.1	63.5±47.4
Shanghai	65.7±34.8	83.8±37.9	77±44.8	49.2±26.5	69±38.9
Guangzhou	69±49.8	53±46.1	54.4±47.8	47.6±36.7	56±46.1

Table S6. Quantitative comparison of decomposition performance across different methods

Method	Residual Standard Deviation ($\mu\text{g}/\text{m}^3$)	Orthogonality Index (IO)
STL	23.1	0.1127
GAM	28.69	0.0002
CEEMDAN	36.38	0.0238
VMD	9.46	0.006

Table S7. VMD Decomposition Metrics Summary Table

Component	Beijing	Chengdu	Guangzhou	Shanghai
<i>K2</i> -IMF1	8758.6 (0.338)	8758.6 (0.327)	8758.6 (0.433)	8758.6 (0.455)
<i>K2</i> -IMF2	24.0 (0.508)	24.0 (0.505)	24.0 (0.435)	24.0 (0.667)
<i>K2</i> -Resi	12.0 (1.944)	12.0 (1.854)	12.0 (1.781)	12.0 (1.913)
<i>K8</i> -MP-LowFreq	8758.6 (0.662)	8758.6 (0.622)	8758.6 (0.794)	8758.6 (0.770)
<i>K8</i> -MP-Diurnal	24.0 (0.344)	24.0 (0.369)	24.0 (0.395)	24.0 (0.533)
<i>K8</i> -MP-HighFreq	12.0 (1.513)	12.0 (1.739)	12.0 (1.236)	12.0 (1.883)
<i>K8</i> -MSE-LowFreq	8758.6 (1.316)	8758.6 (1.476)	24.0 (1.469)	24.0 (1.649)
<i>K8</i> -MSE-MidFreq	12.0 (1.339)	8.0 (1.556)	12.0 (0.768)	12.0 (1.857)
<i>K8</i> -MSE -HighFreq	6.0 (1.852)	6.0 (1.813)	6.0 (1.691)	6.0 (1.774)

Note: Values are presented as the Dominant Period [h] followed by Sample Entropy in parentheses. *K2* and *K8* denote Variational Mode Decomposition (VMD) configurations with mode numbers $K = 2$ and $K = 8$, respectively. MP refers to the Merging based on Period strategy, while MSE refers to the Merging based on Sample Entropy strategy used to reconstruct the sub-sequences.

Table S8. CEEMDAN Decomposition Metrics Summary Table

Component	Beijing	Chengdu	Guangzhou	Shanghai
MP-LowFreq	8758.6 (0.469)	8758.6 (0.495)	8758.6 (0.549)	8758.6 (0.599)
MP-Diurnal	24.0 (1.261)	24.0 (1.088)	24.0 (0.890)	24.0 (1.242)
MP-HighFreq	6.0 (1.567)	12.0 (1.212)	12.0 (0.933)	12.0 (1.289)
MSE -LowFreq	8758.6 (0.020)	8758.6 (0.012)	8758.6 (0.011)	8758.6 (0.015)
MSE -MidFreq	181.7 (0.604)	331.8 (0.628)	24.0 (1.506)	782.0 (0.643)
MSE -HighFreq	24.0 (1.266)	24.0 (1.317)	12.0 (1.726)	24.0 (1.455)

Note: Values are presented as the Dominant Period [h] followed by Sample Entropy in parentheses. MP refers to the Merging based on Period strategy, while MSE refers to the Merging based on Sample Entropy strategy used to reconstruct the CEEMDAN components.

Table S9. Phase alignment of O₃ diurnal component peaks with meteorological drivers (hours)

City	STL	GAM	CEEMDAN	VMD
Beijing vs Tem	2.38 ± 1.41	2.38 ± 1.41	2.39 ± 2.25	2.42 ± 1.93
Beijing vs USWRF	3.65 ± 1.60	3.65 ± 1.60	3.56 ± 2.24	3.54 ± 1.95
Chengdu vs Tem	2.39 ± 1.68	2.39 ± 1.68	2.96 ± 2.34	2.89 ± 2.04
Chengdu vs USWRF	2.31 ± 1.58	2.31 ± 1.58	2.76 ± 2.33	2.56 ± 2.08
Guangzhou vs Tem	1.60 ± 1.31	1.60 ± 1.31	2.13 ± 1.93	1.89 ± 1.51
Guangzhou vs USWRF	2.40 ± 1.62	2.40 ± 1.62	2.83 ± 2.09	2.63 ± 1.82
Shanghai vs Tem	1.98 ± 1.72	1.98 ± 1.72	2.70 ± 2.75	2.54 ± 2.30
Shanghai vs USWRF	2.61 ± 1.68	2.61 ± 1.68	3.27 ± 2.60	3.13 ± 2.30

Note: ‘Tem’ denotes temperature, and ‘USWRF’ denotes upward shortwave radiation flux. Phase alignment is defined as the absolute time difference between the daily peaks of the extracted O₃ component and the corresponding meteorological driver.

Table S10. Absolute deviation of O₃ diurnal component peaks from observed daily O₃ peaks
(Peak-time bias, hours)

City	STL	GAM	CEEMDAN	VMD
Beijing	1.68 ± 2.13	1.68 ± 2.13	0.83 ± 1.69	1.20 ± 1.72
Chengdu	1.48 ± 2.18	1.48 ± 2.18	1.23 ± 1.91	1.16 ± 1.78
Guangzhou	1.37 ± 1.74	1.37 ± 1.74	1.26 ± 1.97	1.00 ± 1.49
Shanghai	2.07 ± 2.78	2.07 ± 2.78	1.62 ± 2.24	1.51 ± 2.03

Table S11. Regime-specific local diurnal isolation index (DII) of the extracted diurnal component across four cities.

City	Method	Frontal DII	Stagnant DII	Ratio (Frontal/Stagnant)
Beijing	STL	0.75	0.75	1.00
Beijing	GAM	0.75	0.75	1.00
Beijing	CEEMDAN	0.64	0.51	1.27
Beijing	VMD	0.74	0.72	1.04
Chengdu	STL	0.68	0.68	1.00
Chengdu	GAM	0.69	0.69	1.00
Chengdu	CEEMDAN	0.67	0.59	1.12
Chengdu	VMD	0.73	0.73	1.01
Guangzhou	STL	0.69	0.69	1.00
Guangzhou	GAM	0.70	0.70	1.00
Guangzhou	CEEMDAN	0.70	0.72	0.98
Guangzhou	VMD	0.76	0.77	0.99
Shanghai	STL	0.69	0.69	1.00
Shanghai	GAM	0.69	0.69	1.00
Shanghai	CEEMDAN	0.51	0.62	0.82
Shanghai	VMD	0.65	0.73	0.88

Table S12. Spearman correlations between O₃ diurnal amplitude and NO₂ amplitude for four decomposition methods

City	STL	GAM	CEEMDAN	VMD
Beijing	/	/	0.06	-0.02
Chengdu	/	/	0.54	0.52
Guangzhou	/	/	0.51	0.5
Shanghai	/	/	0.32	0.25

Note: ‘/’ indicates not applicable. The diurnal components extracted by STL and GAM have fixed amplitudes by construction, so their day-to-day amplitudes do not vary and the corresponding Spearman correlations cannot be calculated.

Table S13. Spectral power of O₃ time series at 12 h, 24 h, 168 h, and 8640 h across four cities (2019-2023)

City	Power-12h	Power-24h	Power-168h	Power-8640h
Beijing	4.34E+05	1.15E+07	1.01E+03	1.46E+07
Chengdu	1.51E+06	1.19E+07	9.24E+03	9.74E+06
Guangzhou	2.24E+06	1.88E+07	2.82E+04	7.25E+05
Shanghai	8.81E+05	7.42E+06	3.79E+03	3.62E+06

Table S14. Comparison of Daily O₃ Forecast Performance Between GEOS-CF and ΔY -Transformer-VMD in Chengdu for 2024

City	Model	Forecast Day	R^2	RMSE
Chengdu	GEOS-CF	1	0.64	29.2
Chengdu	ΔY -Transformer-VMD	1	0.83	20.0
Chengdu	GEOS-CF	3	0.62	29.9
Chengdu	ΔY -Transformer-VMD	3	0.78	23.0
Chengdu	GEOS-CF	5	0.53	33.0
Chengdu	ΔY -Transformer-VMD	5	0.69	27.0

Table S15. Prediction interval coverage probability (PICP) and mean prediction interval width (MPIW) for the four cities

City	PICP	MPIW
Beijing	0.17	7.82
Chengdu	0.24	12.44
Guangzhou	0.14	7.06
Shanghai	0.35	19.86

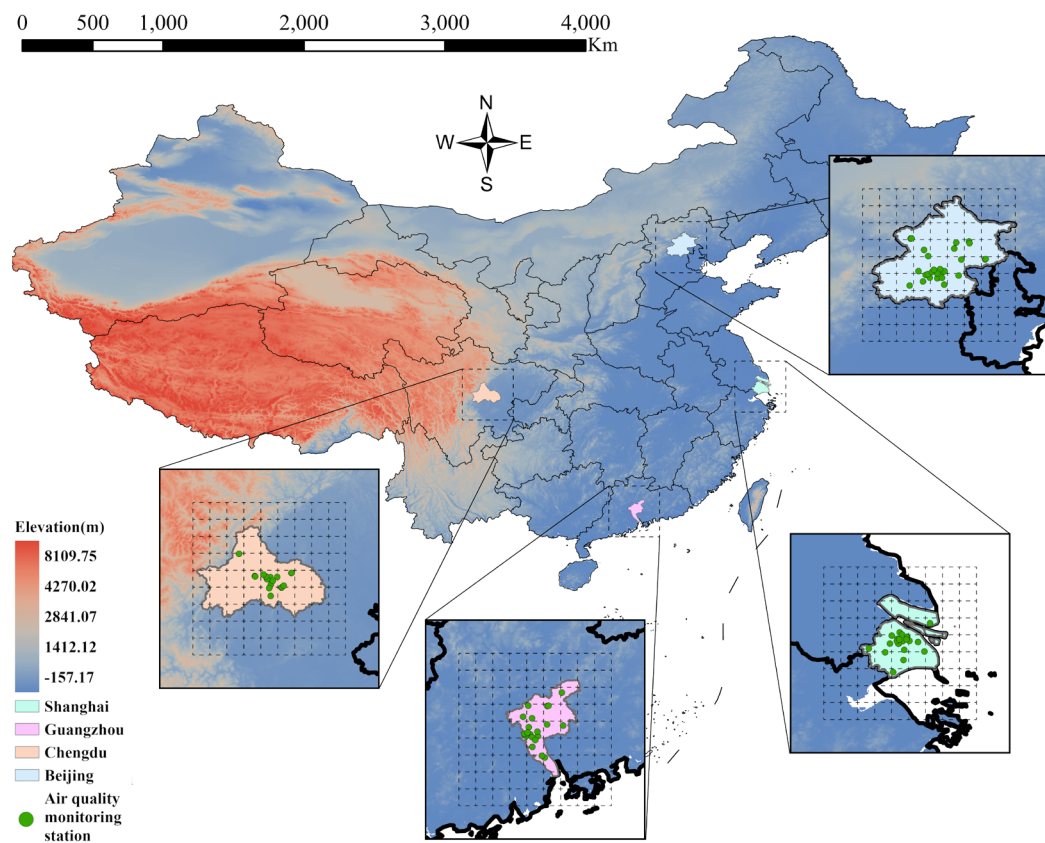
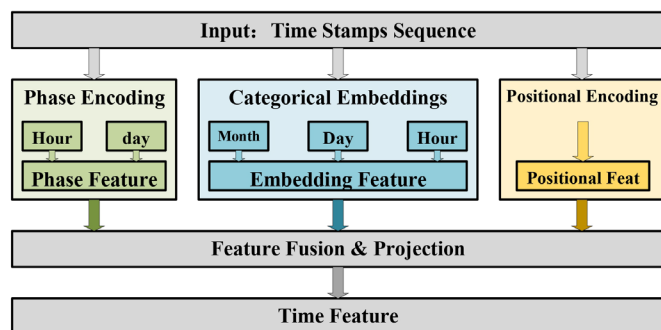
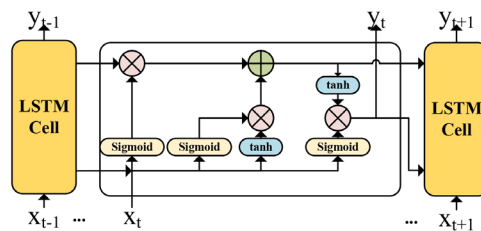


Fig. S1. The study domains and the coverage of regional forecast meteorological data for each domain. The dashed rectangles indicate the selected 9×9 meteorological grid with a spatial resolution of 0.25° . The number of air quality monitoring stations in each city is 28 in Beijing, 19 in Chengdu, 21 in Guangzhou, and 22 in Shanghai.

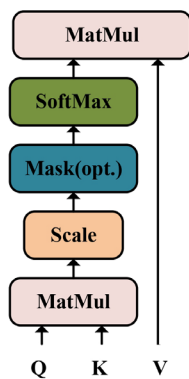
(a) Time Feature Encoder



(b) LSTM



(c) Scaled Dot-Product Attention



(d) Multi-Head Attention

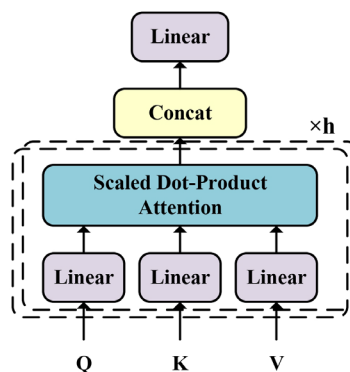


Fig. S2. Schematic diagram of the core neural network modules. (a) Time Feature Encoder, which integrates phase encoding, categorical embeddings, and positional encoding, (b) The internal structure of the LSTM unit, (c) Scaled Dot-Product Attention mechanism, (d) Multi-Head Attention architecture.

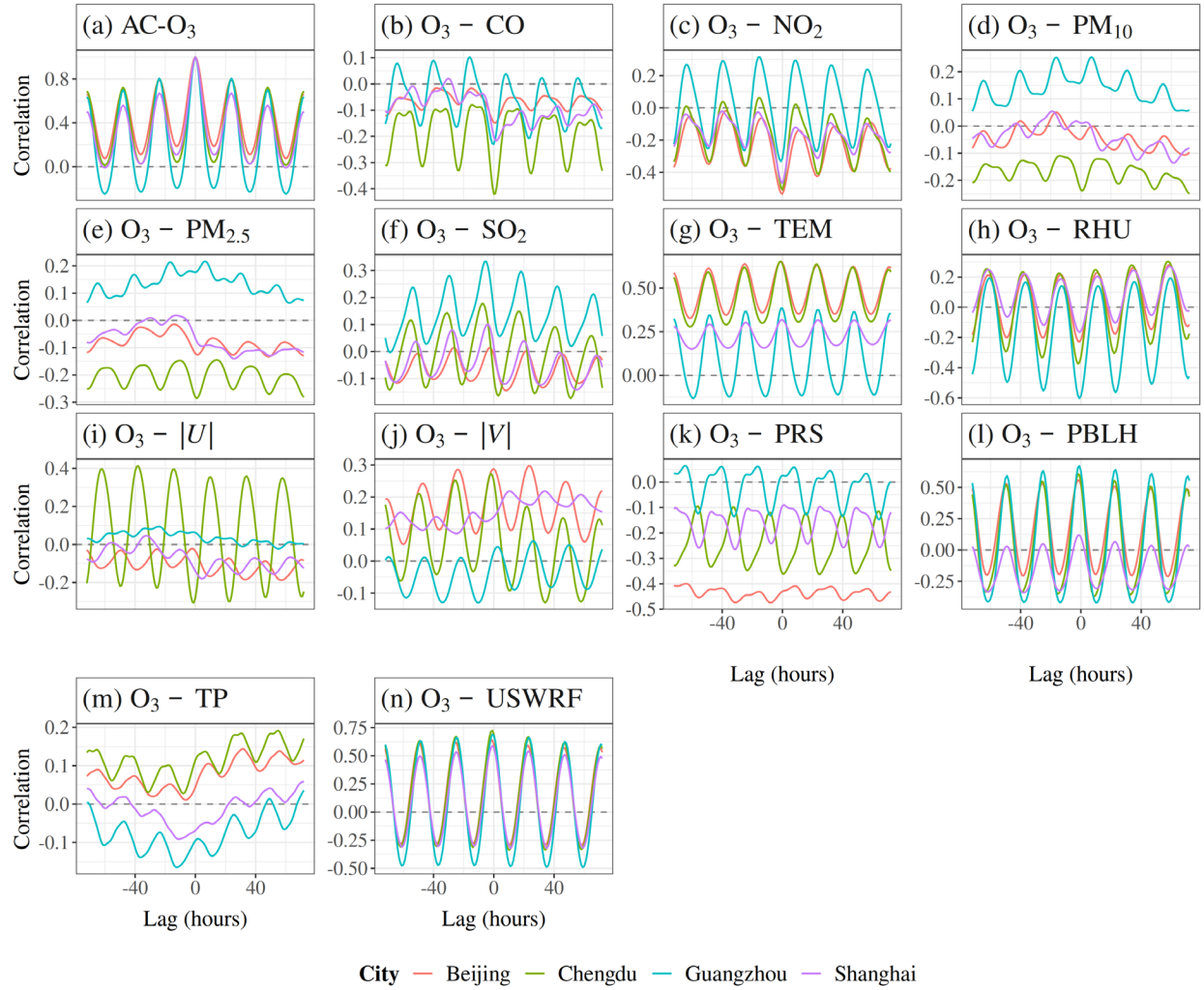


Fig. S3. Time-lagged cross-correlation analysis between O_3 and meteorological/pollutant variables across four cities. (a) Autocorrelation function (ACF) of the O_3 series. (b)–(n) Cross-correlation functions (CCF) between O_3 and exogenous variables, including pollutants (CO, NO_2 , PM_{10} , $PM_{2.5}$, SO_2) and meteorological factors (TEM: temperature, RHU: relative humidity, $|U|/|V|$: wind components, PRS pressure, PBLH: planetary boundary layer height, TP: total precipitation, USWRF: up short-wave radiation flux). The x-axis represents the time lag (in hours), where negative values indicate the correlation with historical observations. The distinct colored curves correspond to Beijing, Chengdu, Guangzhou, and Shanghai, illustrating regional variations in temporal dependencies.

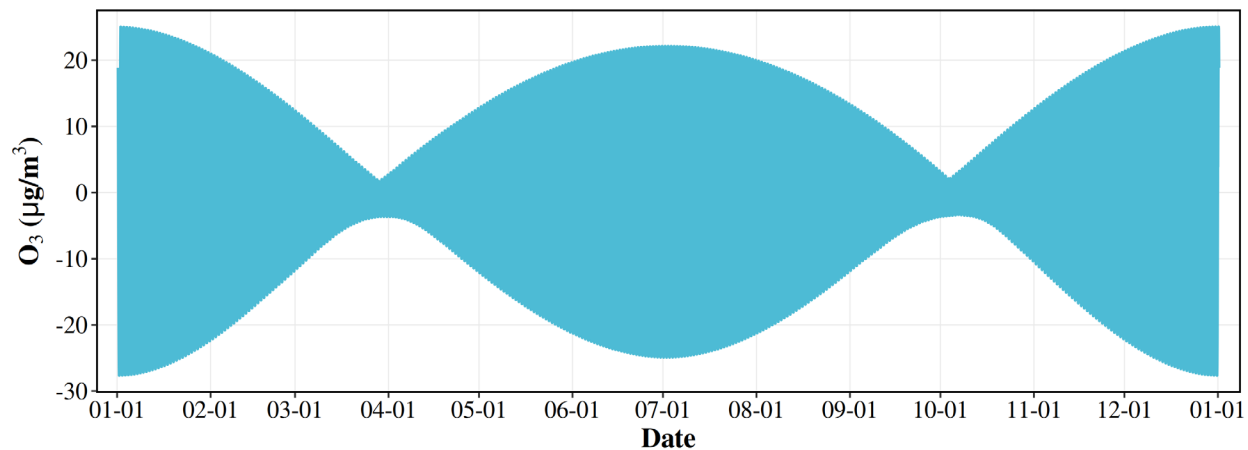


Fig. S4. Visualization of the interaction component in the GAM framework. The plot illustrates the coupled effect of seasonal (day of year) and diurnal (hour of day) variations on O₃ concentrations. The vertical spread at any given date represents the magnitude of the diurnal fluctuation (i.e., the range between the daily maximum and minimum contribution) estimated by the model.

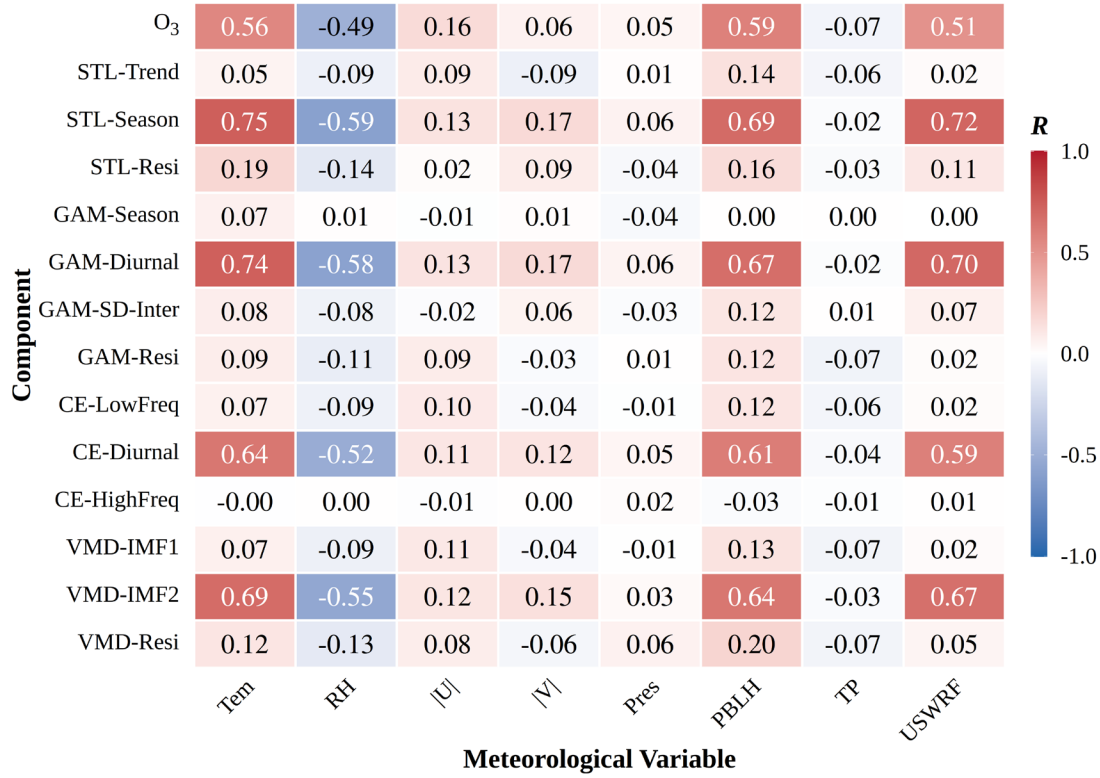


Fig. S5. Pearson-correlation heatmap between decomposed O₃ components and meteorological drivers. The heatmap reports Pearson correlation coefficients (R) between the original surface O₃ and the decomposed sub-sequences obtained from STL, GAM, CEEMDAN (CE), and VMD. Meteorological predictors (x-axis) include Temperature (Tem), Relative Humidity (RH), zonal/meridional wind components ($|U|$, $|V|$), Surface Pressure (Pres), Planetary Boundary Layer Height (PBLH), Total Precipitation (TP), and Downward Shortwave Radiation (USWRF). On the y-axis, CE-LowFreq/CE-Diurnal/CE-HighFreq denote period-grouped CEEMDAN components. GAM-SD-Inter denotes the Seasonal–Diurnal (SD) interaction term (i.e., the seasonal modulation of the diurnal cycle). Colors indicate the magnitude and sign of R (red positive, blue negative), and the numerical values are annotated in each cell.

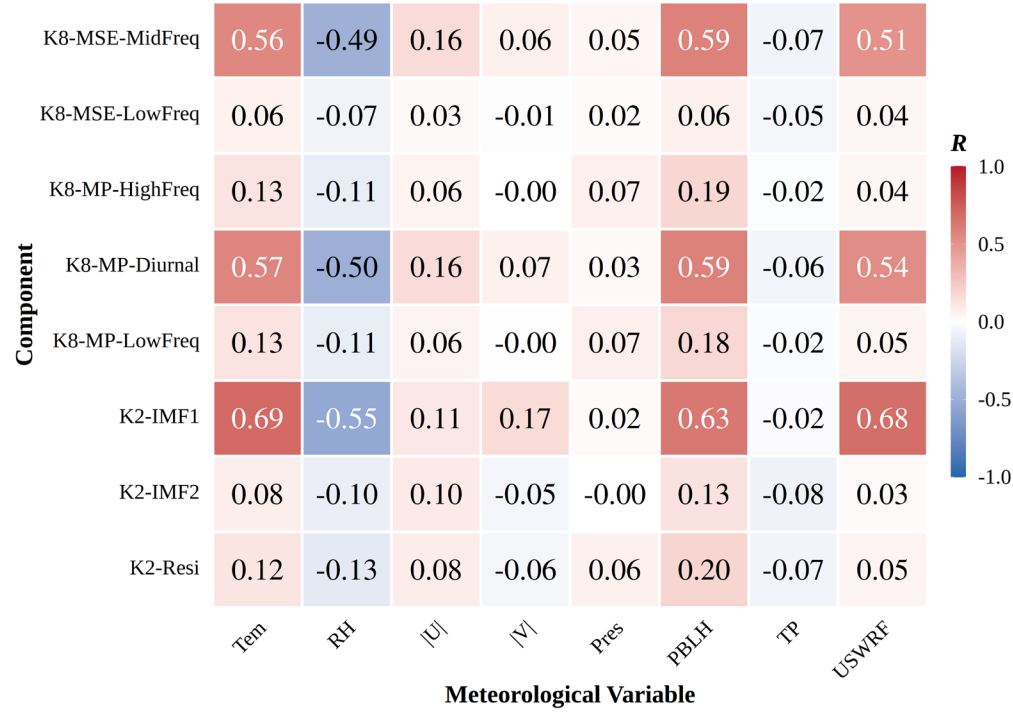


Fig. S6. Heatmap of Pearson correlation coefficients (R) between meteorological variables and O_3 components extracted via Variational Mode Decomposition (VMD) in Beijing. The prefixes $K8$ and $K2$ denote the preset number of decomposed modes (K) configured in the VMD algorithm (i.e., $K = 8$ and $K = 2$). ‘MP’ and ‘MSE’ indicate that the VMD sub-modes were subsequently merged based on Mean Period and Multiscale Entropy, respectively, while ‘Resi’ represents the residual component.

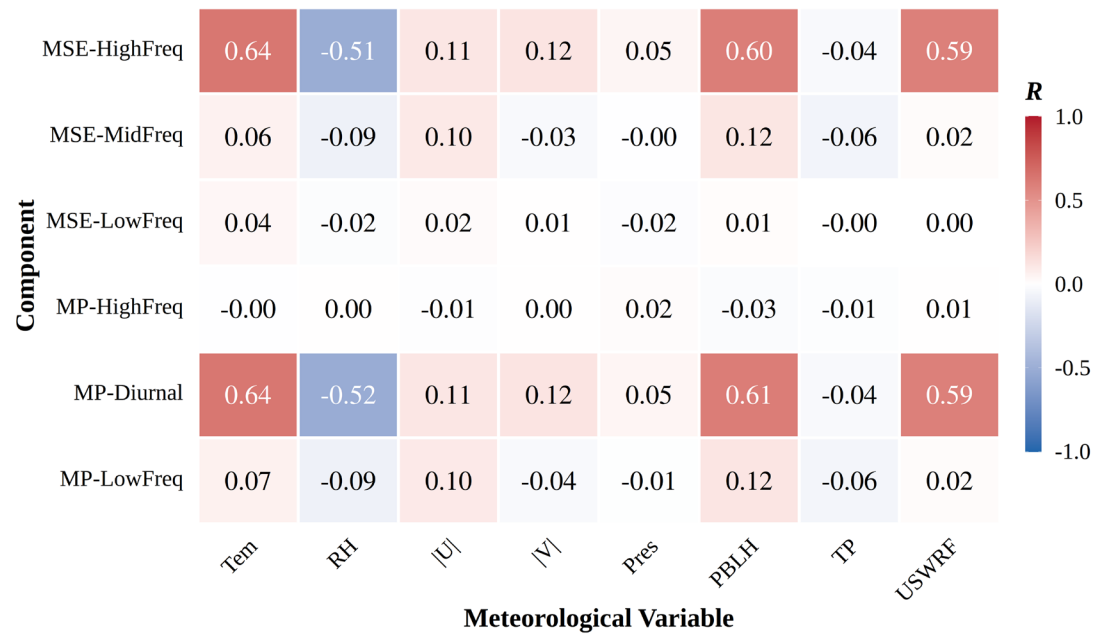


Fig. S7. Heatmap of Pearson correlation coefficients (R) between the original O_3 series and its decomposed/merged components by VMD and the meteorological variables (Beijing). Here, MP and MSE denote component groups merged based on period (MP) and sample entropy (MSE), respectively.

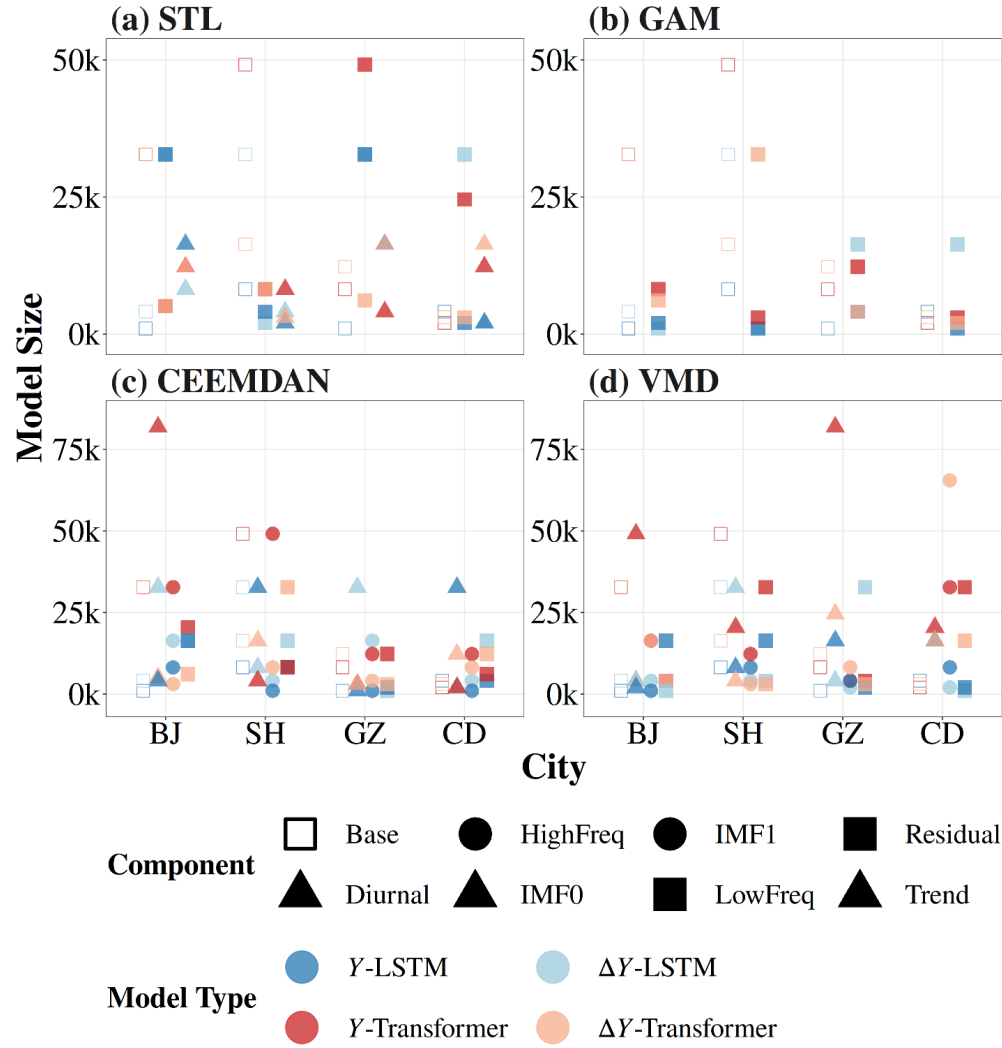


Fig. S8. Model size comparison across decomposition methods and cities. Model sizes (number of trainable parameters) for component-wise predictors under four decomposition schemes-(a) STL, (b) GAM, (c) CEEMDAN, and (d) VMD-across Beijing, Shanghai, Guangzhou, and Chengdu. Marker shapes denote decomposed components, and colors indicate model variants (Y-/ΔY-based LSTM and Transformer).

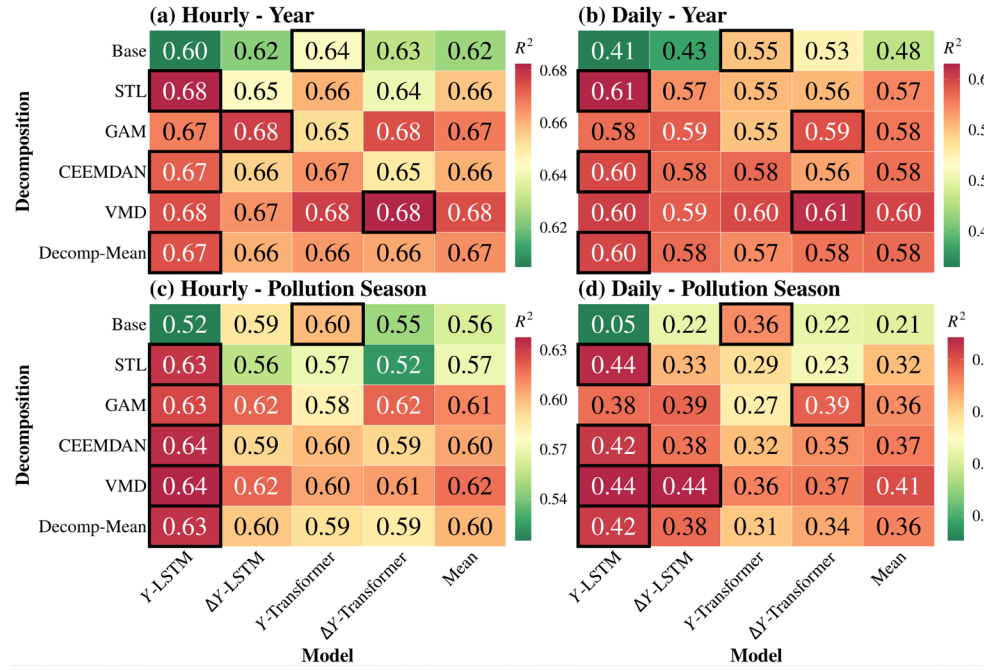


Fig. S9. City-averaged R^2 across four evaluation scenarios. The heatmap layout, row/column configurations, and evaluation settings are identical to Fig. 3, but for the Coefficient of Determination R^2 . Black outlines indicate the best-performing (highest-value) configuration for each decomposition strategy.

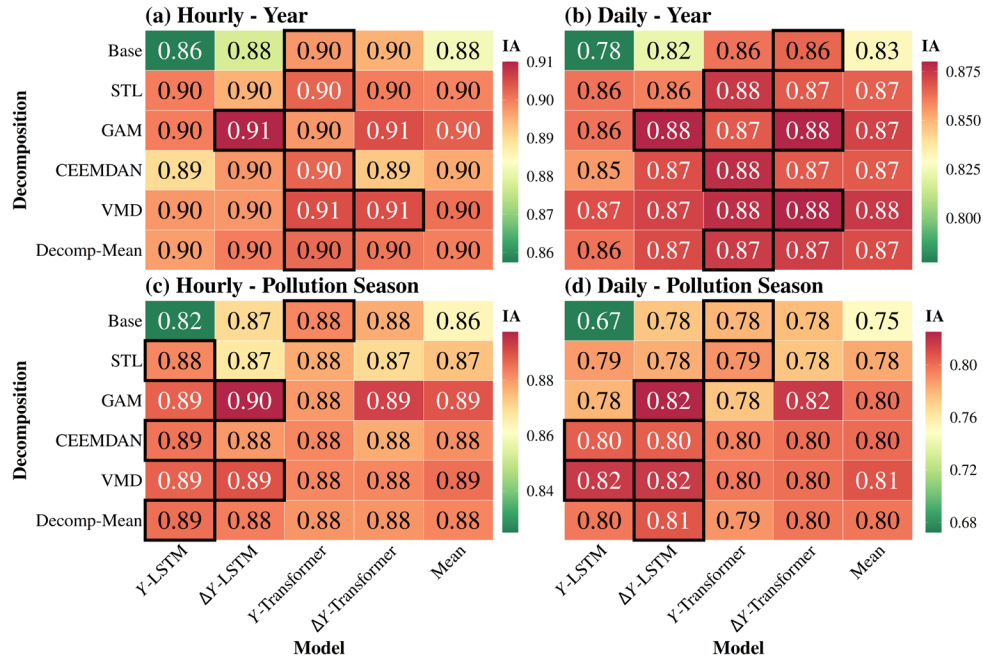


Fig. S10. City-averaged IA across four evaluation scenarios. The heatmap layout, row/column configurations, and evaluation settings are identical to Fig. 3, but for the Coefficient of Determination IA. Black outlines indicate the best-performing (highest-value) configuration for each decomposition strategy.

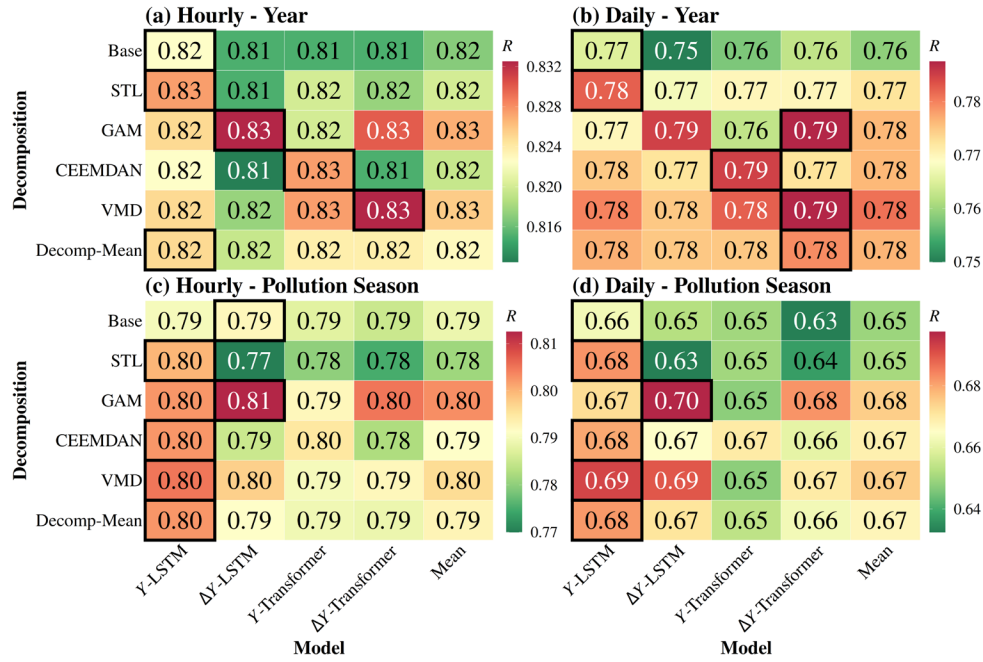


Fig. S11. City-averaged R across four evaluation scenarios. The heatmap layout, row/column configurations, and evaluation settings are identical to Fig. 3, but for the Coefficient of Determination R . Black outlines indicate the best-performing (highest-value) configuration for each decomposition strategy.

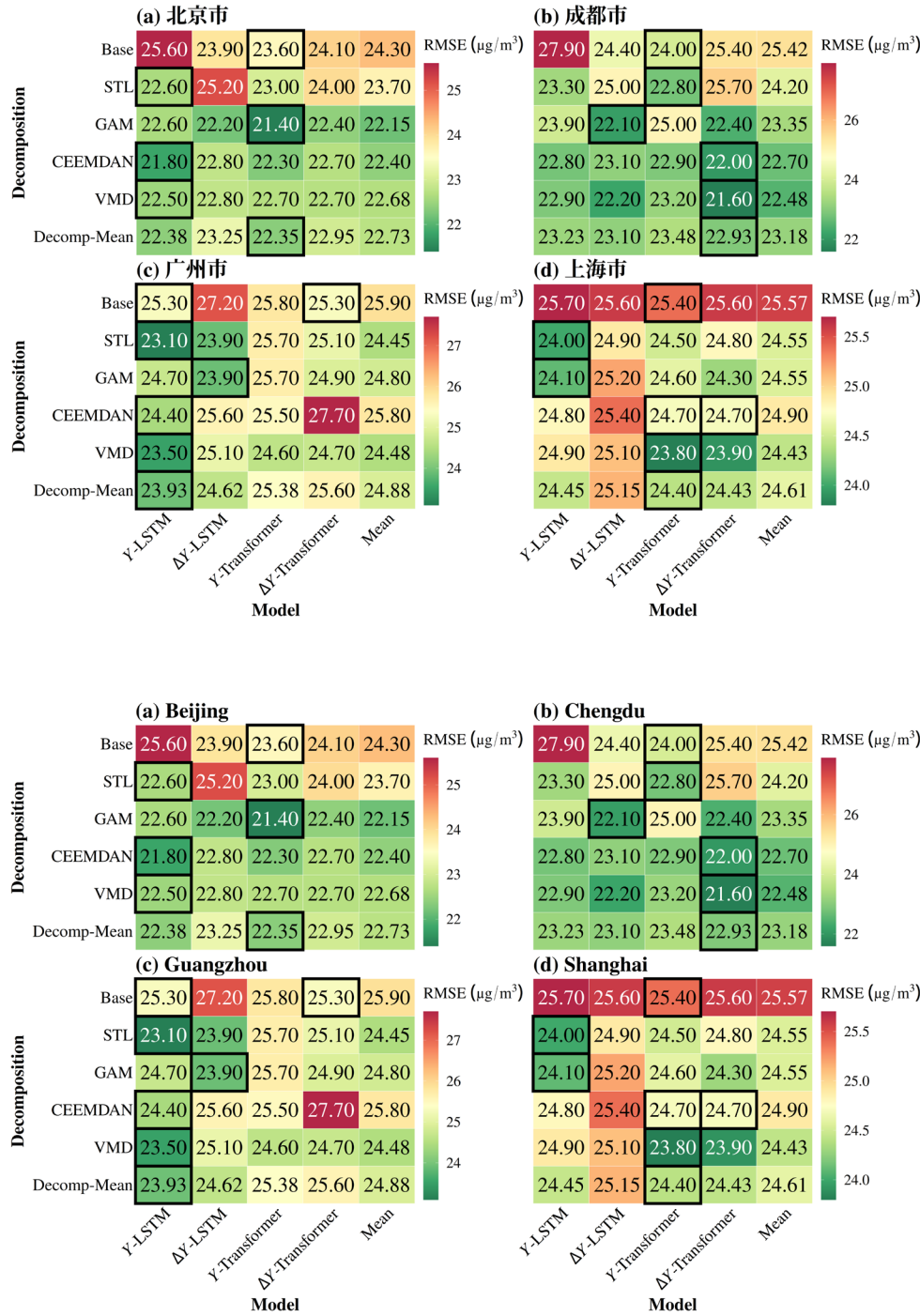


Fig. S12. RMSE performance heatmaps for the Hourly–Year evaluation scenario across four cities. Panels (a–d) display the results for Beijing, Chengdu, Guangzhou, and Shanghai, respectively. Rows represent decomposition strategies (including the mean performance ‘Decomp-Mean’), and columns denote model backbones (with ‘Mean’ indicating the row-wise average). Black outlines highlight the optimal model configuration (lowest RMSE) within each decomposition strategy.

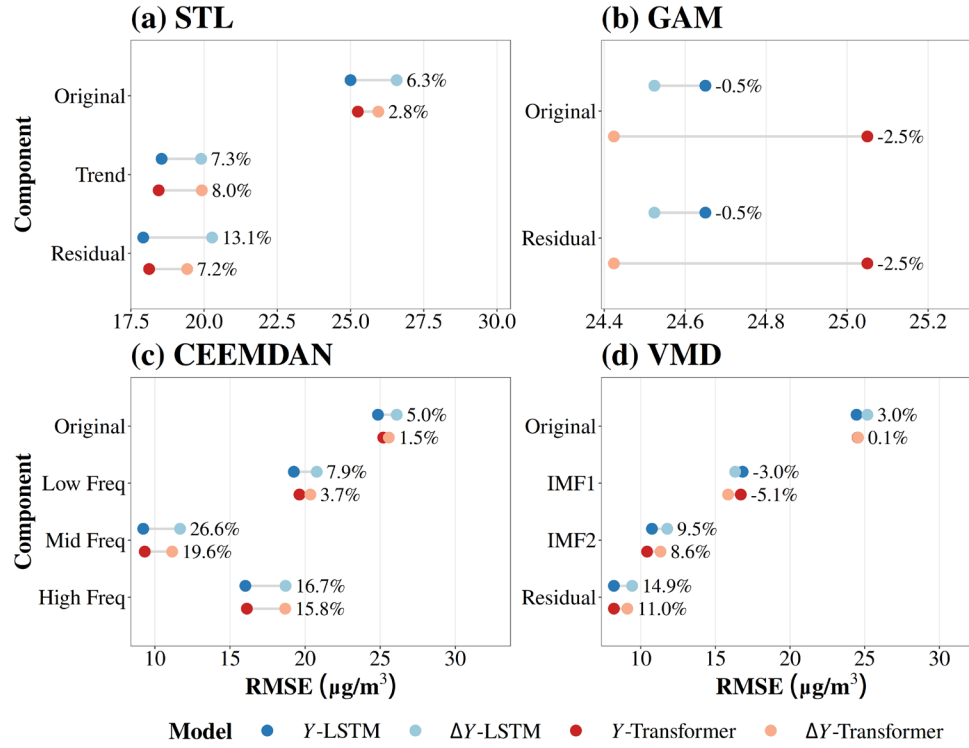


Fig. S13. Dumbbell plots of RMSE values comparing the prediction accuracy of specific decomposition sub-sequences under different model-input configurations. For each backbone (LSTM or Transformer), the two endpoints of a dumbbell correspond to the Y -based and ΔY -based inputs on the same sub-sequence/component. The annotated percentage indicates the relative RMSE change of ΔY compared with Y under the same backbone and sub-sequence

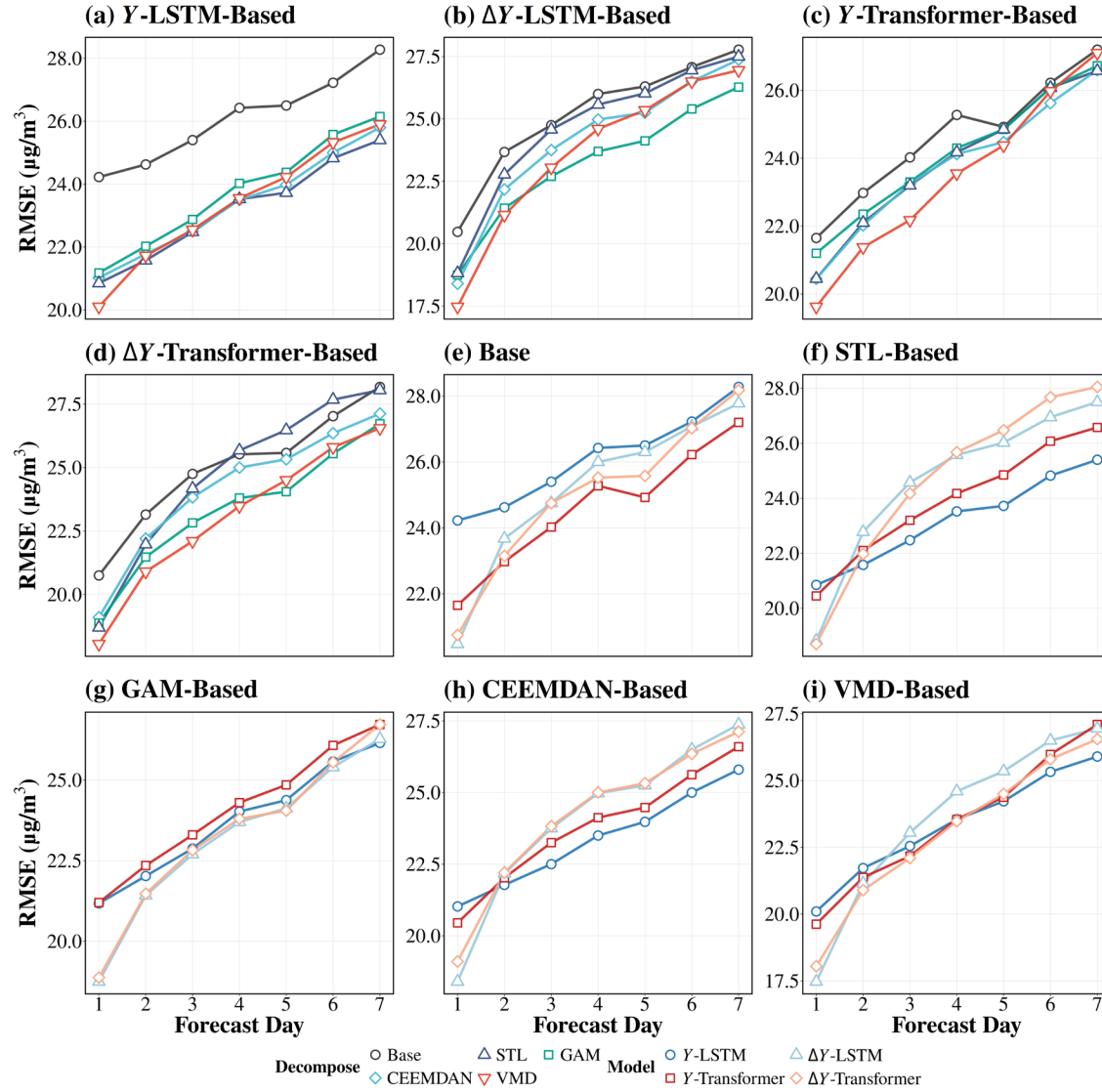


Fig. S14. Comprehensive evaluation of daily RMSE trends across 1–7 day forecast horizons under different decomposition-model combinations. (a)–(d) Impact of decomposition methods: Comparison of five decomposition strategies (Base, STL, GAM, CEEMDAN, VMD) under fixed model backbones: (a) Y-LSTM, (b) ΔY -LSTM, (c) Y-Transformer, and (d) ΔY -Transformer. (e)–(i) Impact of forecasting strategies: Comparison of four model architectures (Y-LSTM, ΔY -LSTM, Y-Transformer, and ΔY -Transformer) within specific decomposition frameworks: (e) Baseline (undecomposed), (f) STL, (g) GAM, (h) CEEMDAN, and (i) VMD.

Reference

1. NATIONAL CENTERS FOR ENVIRONMENTAL PREDICTION. NCEP GFS 0.25 Degree Global Forecast Grids Historical Archive. <https://rda.ucar.edu/datasets/ds084.1/>, (Accessed in **2025/10/1**).
2. NATIONAL CENTERS FOR ENVIRONMENTAL PREDICTION. NCEP GFS 0.25 Degree Global Forecast Grids Operational Data. <https://nomads.ncep.noaa.gov/pub/data/nccf/com/gfs/prod/>, (Accessed in **2025/10/1**).
3. RABBATH C., CORRIVEAU D. J. D. T. A comparison of piecewise cubic Hermite interpolating polynomials, cubic splines and piecewise linear functions for the approximation of projectile aerodynamics. **15**, (5), 741, **2019**.
4. VASWANI A., SHAZEER N., PARMAR N., USZKOREIT J., JONES L., GOMEZ A. N., KAISER L., POLOSUKHIN I. Attention Is All You Need. Advances in Neural Information Processing Systems 30 (NIPS 2017). 5998, **2017**.
5. GREFF K., SRIVASTAVA R. K., KOUTNÍK J., STEUNEBRINK B. R., SCHMIDHUBER J. LSTM: A Search Space Odyssey. IEEE Transactions on Neural Networks and Learning Systems. **28**, (10), 2222, **2017**.
6. LI K., JACOB D. J., LIAO H., QIU Y., SHEN L., ZHAI S., BATES K. H., SULPRIZIO M. P., SONG S., LU X. J. P. O. T. N. A. O. S. Ozone pollution in the North China Plain spreading into the late-winter haze season. **118**, (10), e2015797118, **2021**.
7. WANG L., CHEN B., OUYANG J., MU Y., ZHEN L., YANG L., XU W., TANG L. J. E. S., ECOTECHNOLOGY Causal-inference machine learning reveals the drivers of China's 2022 ozone rebound. **24**, 100524, **2025**.
8. SOLAZZO E., BIANCONI R., HOGREFE C., CURCI G., TUCCELLA P., ALYUZ U., BALZARINI A., BARO R., BELLASIO R., BIESER J., BRANDT J., CHRISTENSEN J. H., COLETTE A., FRANCIS X., FRASER A., VIVANCO M. G., JIMENEZ-GUERRERO P., IM U., MANDERS A., NOPMONGCOL U., KITWIROON N., PIROVANO G., POZZOLI L., PRANK M., SOKHI R. S., UNAL A., YARWOOD G., GALMARINI S. Evaluation and error apportionment of an ensemble of atmospheric chemistry transport modeling systems: multivariable temporal and spatial breakdown. Atmospheric Chemistry and Physics. **17**, (4), 3001, **2017**.
9. TORRENCE C., COMPO G. P. A practical guide to wavelet analysis. Bulletin of the American Meteorological society. **79**, (1), 61, **1998**.
10. LI K., JACOB D. J., LIAO H., SHEN L., ZHANG Q., BATES K. H. Anthropogenic drivers of 2013–2017 trends in summer surface ozone in China. Proceedings of the National Academy of Sciences. **116**, (2), 422, **2019**.

11. KELLER C. A., KNOWLAND K. E., DUNCAN B. N., LIU J., ANDERSON D. C., DAS S., LUCCHESI R. A., LUNDGREN E. W., NICELY J. M., NIELSEN E. Description of the NASA GEOS composition forecast modeling system GEOS - CF v1. 0. *Journal of Advances in Modeling Earth Systems*. **13**, (4), e2020MS002413, **2021**.
12. KHOSRAVI A., NAHAVANDI S., CREIGHTON D., ATIYA A. F. Comprehensive review of neural network-based prediction intervals and new advances. *IEEE Transactions on Neural Networks and Learning Systems*. **22**, (9), 1341, **2011**.
13. CHEN Z., CHEN D., ZHAO C., KWAN M.-P., CAI J., ZHUANG Y., ZHAO B., WANG X., CHEN B., YANG J. Influence of meteorological conditions on PM_{2.5} concentrations across China: A review of methodology and mechanism. *Environment international*. **139**, 105558, **2020**.
14. NING G., WANG S., MA M., NI C., SHANG Z., WANG J., LI J. Characteristics of air pollution in different zones of Sichuan Basin, China. *Science of the Total Environment*. **612**, 975, **2018**.
15. LI L., CHEN C., HUANG C., HUANG H., ZHANG G., WANG Y., WANG H., LOU S., QIAO L., ZHOU M. Process analysis of regional ozone formation over the Yangtze River Delta, China using the Community Multi-scale Air Quality modeling system. *Atmospheric Chemistry Physics*. **12**, (22), 10971, **2012**.
16. LI L., QIAN J., OU C.-Q., ZHOU Y.-X., GUO C., GUO Y. Spatial and temporal analysis of Air Pollution Index and its timescale-dependent relationship with meteorological factors in Guangzhou, China, 2001–2011. *Environmental Pollution*. **190**, 75, **2014**.
17. ZHANG Z., ZENG Y., YAN K. A hybrid deep learning technology for PM_{2.5} air quality forecasting. *Environmental Science and Pollution Research*. **28**, (29), 39409, **2021**.
18. MA G., XU K., ZHANG Y., ZHANG L., CHEN Z. A novel deep learning model for air quality index prediction integrating time series decomposition and intelligent optimization. *Results in Engineering*. **27**, 106078, **2025**.
19. NIU M., WANG Y., SUN S., LI Y. A novel hybrid decomposition-and-ensemble model based on CEEMD and GWO for short-term PM_{2.5} concentration forecasting. *Atmospheric Environment*. **134**, 168, **2016**.
20. ZHAO L., LI Z., QU L. A novel machine learning-based artificial intelligence method for predicting the air pollution index PM_{2.5}. *Journal of Cleaner Production*. **468**, 143042, **2024**.
21. WANG K., YANG T., KONG S., LI M. Air quality index prediction through TimeGAN data recovery and PSO-optimized VMD-deep learning framework. *Applied Soft Computing*. **170**, 112626, **2025**.
22. NIU M., GAN K., SUN S., LI F. Application of decomposition-ensemble learning

paradigm with phase space reconstruction for day-ahead PM 2.5 concentration forecasting. *Journal of Environmental Management*. **196**, 110, **2017**.

23. DONG K., RAN P., FAN Q., LIU X., LI W. Comparison of improved hybrid FTS models for forecasting the urban air quality index. *Journal of Cleaner Production*. **428**, 139234, **2023**.
24. WU Q., LIN H. Daily urban air quality index forecasting based on variational mode decomposition, sample entropy and LSTM neural network. *Sustainable Cities and Society*. **50**, 101657, **2019**.
25. JIN X.-B., YANG N.-X., WANG X.-Y., BAI Y.-T., SU T.-L., KONG J.-L. Deep Hybrid Model Based on EMD with Classification by Frequency Characteristics for Long-Term Air Quality Prediction. *Mathematics*. **8**, (2), 214, **2020**.
26. MACIĄG P. S., BEMBENIK R., PIEKARZEWICZ A., DEL SER J., LOBO J. L., KASABOV N. K. Effective air pollution prediction by combining time series decomposition with stacking and bagging ensembles of evolving spiking neural networks. *Environmental Modelling & Software*. **170**, 105851, **2023**.
27. HU K., CHE J., XIA W., XU Y., LI Y. Enhancing multi-step air quality prediction with deep learning using residual neural network and adaptive decomposition-based multi-objective optimization. *Expert Systems with Applications*. **273**, 126969, **2025**.
28. JIANG F., ZHANG C., SUN S., SUN J. Forecasting hourly PM 2.5 based on deep temporal convolutional neural network and decomposition method. *Applied Soft Computing*. **113**, 107988, **2021**.
29. SUN W., XU Z. A hybrid Daily PM2.5 concentration prediction model based on secondary decomposition algorithm, mode recombination technique and deep learning. *Stochastic Environmental Research and Risk Assessment*. **36**, (4), 1143, **2022**.
30. YIN H., JIN D., HONG H., MOON J., GU Y. H. IAQ-STL-ML: A novel indoor air quality prediction pipeline using meta-learning framework with STL decomposition. *Environmental Technology & Innovation*. **38**, 104107, **2025**.
31. ZHENG H., ZHAO J., ZHU J., YE Z., DENG F. Long-term urban air quality prediction with hierarchical attention loop network. *Sustainable Cities and Society*. **118**, 106010, **2025**.
32. NIU C., NIU Z., QU Z., WEI L., LI Y. Research and Application of the Mode Decomposition-Recombination Technique Based on Sample-Fuzzy Entropy and K-Means for Air Pollution Forecasting. *Frontiers in Environmental Science*. **10**, 941405, **2022**.
33. CHE J., HU K., XIA W., XU Y., LI Y. Short-term air quality prediction using point and interval deep learning systems coupled with multi-factor decomposition and data-driven tree compression. *Applied Soft Computing*. **166**, 112191, **2024**.

34. HAO Y., TIAN C. The study and application of a novel hybrid system for air quality early-warning. *Applied Soft Computing*. **74**, 729, **2019**.
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. *Journal of Environmental Management*. **366**, (000), 11, **2024**.