

Original Research

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

Received: 30 January 2026

Accepted: 07 April 2026

Abstract

Surface ozone (O₃) forms via nonlinear chemistry and is highly non-stationary, complicating 168-h forecasts. Many models omit key physical drivers, mix time scales, and degrade quickly with lead time. We propose a decomposition-ensemble framework that fuses surface observations with NWP predictors (e.g., solar radiation, PBL height). To prevent data leakage, decomposition (STL, GAM, CEEMDAN, VMD) was performed exclusively on the training and validation data to isolate the annual, diurnal, and residual components. We then train LSTM and Transformer models on each component using both direct and incremental (ΔY) learning. Across Beijing, Shanghai, Guangzhou, and Chengdu, decomposition lowers the error bound and curbs horizon-wise error growth. Among the tested strategies, VMD ($K = 2$) and period-based merging (for CEEMDAN) robustly isolate the 24-h diurnal mode, limit cross-scale leakage, and provide the best balance of separability, interpretability, and forecast skill. The results further reveal pronounced heterogeneity in optimal model architectures across temporal scales and city-specific regimes, indicating that a single universal model is insufficient to jointly optimize short-term accuracy and long-term stability. Overall, this study provides a scenario-adaptive forecasting paradigm that offers practical guidance for O₃ pollution early warning and decision support.

Keywords: ozone forecasting, time-series decomposition, deep learning, transformer and LSTM, long-term modeling

Introduction

Air pollution, particularly in the form of elevated surface ozone (O₃) and particulate matter concentrations, has become a critical global environmental challenge [1]. Among these pollutants, surface O₃ has drawn

increasing concern due to its strong oxidizing capacity, broad spatial extent, and substantial impacts on human health, ecosystems, and climate, with persistent exposure exacerbating cardiopulmonary diseases and reducing agricultural productivity. Accurate and reliable O₃ forecasting is essential for supporting air quality management, emission control strategies, and public health protection through timely early warnings [2]. Existing O₃ forecasting approaches can be broadly

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

Tel.: +86 28 85996656

Fax: +86 28 85405613

 0000-0002-8692-0603

categorized into deterministic numerical models, statistical methods, and artificial intelligence-based data-driven techniques [3]. Deterministic chemical transport models, such as the Weather Research and Forecasting model coupled with Chemistry (WRF-Chem) and the Community Multiscale Air Quality model (CMAQ), explicitly represent atmospheric dynamics and photochemical processes but are often limited by high computational costs and uncertainties in emission inventories and chemical parameterizations. In contrast, traditional statistical methods, including autoregressive integrated moving average models and multiple linear regression, are computationally efficient yet generally inadequate for capturing the strong nonlinearity, non-stationarity, and intermittency that characterize surface O_3 time series [4].

In recent years, artificial intelligence and hybrid modeling approaches have gained increasing popularity due to their powerful nonlinear and high-order interaction learning capabilities [5]. Deep learning architectures such as Long Short-Term Memory networks (LSTM) [6], Gated Recurrent Units (GRU) [7], Convolutional Neural Networks (CNN) [8], and Transformer models [9] have been adopted for O_3 forecasting. To further enhance predictive performance, the decomposition-ensemble paradigm has become a mainstream strategy [10]. In this framework, the original O_3 series is decomposed into multiple subsequences using techniques such as Empirical Mode Decomposition (EMD) [11, 34], Ensemble Empirical Mode Decomposition (EEMD) [12], Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) [13], Variational Mode Decomposition (VMD) [14], or Seasonal-Trend decomposition using Loess (STL) [15]. Two distinct strategies are generally adopted: modeling each component separately with subsequent aggregation, or integrating the decomposed subsequences directly into the input feature space [11]. To reduce computational complexity, many studies further employ entropy-based indicators [33], including Sample Entropy (SE) [16] or Range Entropy (RE) [17], to reconstruct decomposed components, while intelligent optimization algorithms are frequently used to tune model hyperparameters.

Despite their popularity, existing decomposition-based O_3 forecasting studies still suffer from notable methodological shortcomings (Table S1). Many decomposition-recombination frameworks prioritize optimization metrics and use entropy-based recombination that merges subsequences solely by statistical similarity, thereby overlooking O_3 diurnal/seasonal temporal structure and potentially mixing signals with different physicochemical meanings [16]. Mixing meteorology-driven high-frequency and emission/seasonal low-frequency components obscures O_3 mechanisms and reduces interpretability. Distinct-frequency components are often modeled with uniform architectures without component-specific tuning, limiting the exploitation of decomposed O_3 information.

Decomposing before splitting causes leakage and inflates accuracy beyond real-world performance. Additionally, although forecasting the increment relative to the initial state (ΔY) [18] has proven effective for non-stationary series, existing decomposition-based studies predominantly rely on direct absolute value forecasting, largely overlooking the stability benefits of forecasting the increment (ΔY) rather than the absolute values, a strategy that anchors predictions to the most recent local state.

To address the above limitations, this study aims to develop a comparative framework designed to benchmark the predictive capabilities of various decomposition methods, specifically focusing on their ability to integrate physicochemical understanding with data-driven modeling. The proposed approach is specifically designed to address forecasting challenges over extended lead times up to 168 h by improving the representation of slowly varying background conditions and mitigating error accumulation as the forecasting horizon increases. In addition, a physically informed decomposition and modeling strategy is adopted that explicitly respects the intrinsic diurnal and seasonal structure of O_3 variability, thereby avoiding the mixing of components with distinct physicochemical meanings. Crucially, we perform decomposition only once on the combined training and validation set (not separately), keeping the test set fully isolated and thereby preventing data leakage. In order to explicitly evaluate the efficacy of residual learning strategies within decomposition frameworks, we systematically implement both direct and incremental predictive paradigms across LSTM and Transformer architectures (specifically $Y/\Delta Y$ -LSTM and $Y/\Delta Y$ -Transformer) to identify the optimal target representation for each decomposed signal component. By combining component-wise modeling with strictly time-ordered evaluation, this study improves predictive skill and physicochemical interpretability, offering more robust guidance for air quality forecasting and early warning.

Materials and Methods

Study Region and Data Collection

To evaluate the effectiveness and generalizability of the proposed methodology, we selected four representative Chinese megacities: Beijing, Shanghai, Guangzhou, and Chengdu (Fig. S1), as case studies [19]. These cities represent key regions for China's atmospheric environment and economic development, each facing distinct pollution challenges. This study integrates ground-based air pollutant observations with numerical weather forecasting data spanning the period from 2019 to 2024. Surface air quality observations are obtained from the China National Environmental Monitoring Centre (CNEMC) [20], which provides site-level concentrations of six major criteria pollutants:

PM_{2.5}, PM₁₀, O₃, CO, NO₂, and SO₂. Meteorological data are sourced from the NCEP Global Forecast System (GFS) 0.25° Global Forecast Grids (Supplementary Text S1) [21], which provides operational forecast products issued at 00:00 UTC (08:00 Beijing Time). These forecasts, archived for reference, were used for training, validation, and the 2024 test set, ensuring that the model only accesses information available in real forecasting conditions rather than perfect hindsight reanalysis.

To ensure a robust representation of urban air quality, city-level pollutant concentrations are derived by averaging valid observations from all monitoring stations within each city. The geographic centroid of these monitoring stations is calculated as the mean latitude and longitude of all stations, and this centroid is used as the reference location for meteorological data extraction. Considering that regional meteorological conditions better capture atmospheric transport processes and spatial heterogeneity than single-point measurements [22], we delineate a 9×9 grid domain (2.25°×2.25°) centered on this centroid to derive grid-averaged values. Following previous studies [23], we selected 2 m temperature (Tem), 2 m relative humidity (RH), 10 m wind vectors (U and V), surface pressure (SP), planetary boundary layer height (PBLH), total precipitation (TP), and up short-wave radiation flux (USWRF) as predictor variables (Supplementary Text S7). For the treatment of missing values, Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) is applied to fill pollutant records and to interpolate the GFS meteorological data from its 3-hourly historical resolution to a uniform 1-hourly interval (Supplementary Text S1). We consistently utilize the GFS forecast products initialized at 08:00 (CST), covering forecast lead times from f_1 to f_{191} . This extended horizon ensures that for any forecasting baseline hour x within the daily operational window, the required 168-h future meteorological forcing can be precisely extracted from the sequence f_{x-8} to f_{x+160} .

Predictive Model Construction

Leveraging the Long Short-Term Memory (LSTM) and Transformer architectures, we constructed two predictive models to learn the mapping from meteorological and pollutant predictors to the target O₃ series and its decomposed subsequence. To effectively integrate historical observations with future meteorological forcing, both models adopt a dual-branch input structure consisting of a History Branch and a Future Branch (Fig. 1), which is explicitly designed to respect the causal nature of atmospheric forecasting:

1. **History Branch:** Takes as input the pollutant concentrations, meteorological observations, and time stamps from the previous history window.
2. **Future Branch:** Inputs GFS meteorological forecasts for the next 168 h and the corresponding timestamps.

To capture temporal dependencies in O₃ variability, including diurnal and seasonal cycles, we design a Time Feature Encoder that processes timestamps via phase encoding, categorical embeddings, and positional encoding (Fig. S2a)). Instead of simple concatenation, the encoded time features are combined via element-wise addition with the meteorological and pollutant features (or their latent projections).

Within this framework, the two architectures serve complementary roles in capturing temporal dynamics (Supplementary Text S2):

1. **LSTM-based Model:** Exploits the recurrent structure to extract hierarchical temporal dependencies and local sequential patterns across multiple time scales.
2. **Transformer-based Model:** Leverages the encoder-decoder paradigm with self-attention mechanisms to capture global temporal dependencies and long-range interactions that may be overlooked by recurrent networks.

The detailed mathematical formulations and implementation details of the proposed encoding strategies are provided in Supplementary Text S3 and Fig. S2.

Sequence Decomposition Methods

Given the complex and non-stationary nature of ambient O₃ concentrations driven by periodic physicochemical cycles and meteorological variability, we adopt a decomposition-ensemble framework that decomposes the original time series into relatively stable subsequences for independent modeling and subsequent integration (Fig. 1). Ambient O₃ concentrations exhibit complex variability, primarily governed by robust physicochemical cycles (seasonal and diurnal) and further modulated by evolving meteorological conditions. To address the challenges posed by highly non-stationary signals, we explore multiple decomposition strategies, including additive approaches (STL), interaction-aware formulations (GAM), and adaptive signal processing techniques (VMD and CEEMDAN), which systematically capture long-term trends (seasonal variability), diurnal variability, and their time-varying evolution. Using hourly O₃ observations from 2019 to 2024, the sequence decomposition is strictly confined to the training (2019-2022) and validation (2023) subsets to generate target variables, effectively isolating the test set (2024) to prevent any information leakage from future data. The specific mathematical principles and implementation details for each employed decomposition approach are elaborated below.

Seasonal-Trend Decomposition Using LOESS (STL)

STL is a robust time series decomposition method based on LOESS smoothing, and in this study its parameters are carefully configured to explicitly isolate the seasonal cycle, diurnal cycle, and residual variability,

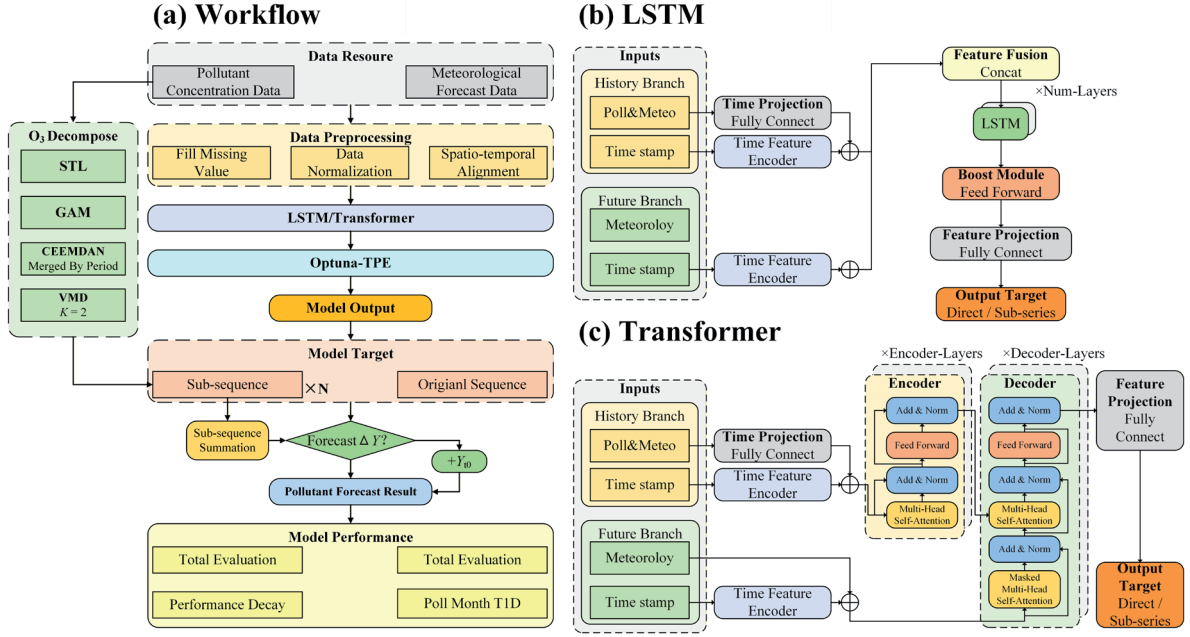


Fig. 1. Schematic Illustration of the Proposed Deep Learning Frameworks for Air Quality Forecasting. This figure presents the schematic illustration of the proposed deep learning frameworks for air quality forecasting. a) Workflow: Overview of the framework, including pollutant concentration data and meteorological forecast data inputs, data preprocessing steps (e.g., missing value imputation, normalization, spatiotemporal alignment), the use of LSTM/Transformer models, and Optuna-TPE for hyperparameter tuning. b) LSTM: Architecture of the LSTM model, comprising history and future branches, time feature encoding, feature fusion, and output target (direct or sub-sequence). c) Transformer: Architecture of the Transformer model, including encoder and decoder layers with multi-head self-attention and output projection to forecast pollutant levels.

so that each component can be associated with a distinct physicochemical mechanism [24, 50]. STL assumes an additive model, decomposing the original time series $Y(t)$ into three components:

$$Y(t) = T(t) + S(t) + R(t) \quad (1)$$

In our implementation, we set the seasonal period to 24, which assigns the STL seasonal component $S(t)$ to represent the diurnal cycle. A periodic seasonal window ($s.window = \text{periodic}$) is further applied to enforce a stable daily pattern, ensuring that long-term variability, particularly the annual seasonality, is captured by the STL trend component $T(t)$. Robust fitting is used so that the residual component $R(t)$ primarily reflects stochastic noise while minimizing the influence of short-lived outliers.

Generalized Additive Model (GAM)

The Generalized Additive Model (GAM) provides a flexible statistical framework that can capture non-linear interactions across multiple temporal scales [25]. To explicitly model the seasonal modulation of the O_3 diurnal profile, we express O_3 concentration series as a function of the day-of-year (x_{doy} , 1-366) and hour-of-day (x_{hour} , 1-24), utilizing cyclic cubic splines to model the periodic variations:

$$Y(t) = f_1(x_{\text{doy}}) + f_2(x_{\text{hour}}) + f_3(x_{\text{doy}}, x_{\text{hour}}) + \varepsilon(t) \quad (2)$$

Where $f_1(x_{\text{doy}})$ captures the long-term annual seasonality, while $f_2(x_{\text{hour}})$ represents the baseline diurnal cycle. The interaction term $f_3(x_{\text{doy}}, x_{\text{hour}})$ accounts for the synergistic effects (e.g., the seasonal variation in diurnal peak amplitude). The residual $\varepsilon(t)$ captures the unexplained stochastic noise that cannot be explained by these deterministic periodic components.

Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN)

CEEMDAN improves upon EMD by adding adaptive white-noise realizations to alleviate mode mixing, but it often produces a large number of intrinsic mode functions (IMFs) [26]. Although prior studies often use metrics such as sample entropy (SE) to group IMFs, these data-driven approaches may undesirably conflate high-frequency physical oscillations with instrumental or stochastic noise [27]. To explicitly isolate the targeted annual and diurnal cycles, we implement a physics-informed merging strategy where IMFs are aggregated based on their dominant periods estimated from Power Spectral Density (PSD) [28]:

1. High-frequency (Period < 23.5 h): Sum of high-frequency IMFs, defined as the Stochastic Residual.

2. Diurnal (23.5 h $\leq$ Period $\leq$ 24.5 h): Sum of IMFs with a dominant period near 24 h, intended to represent the Diurnal Cycle.

3. Low-frequency (Period $>$ 24.5 h): Sum of low-frequency IMFs, representing the low-frequency trend component spanning seasonal to annual scales.

$$Y_{target}(t) = \sum_{i \in \text{Group}} IMF_i(t) \quad (3)$$

Where group refers to the PSD-based period band (HF/Diurnal/LF) defined above and the corresponding set of IMF indices used for component reconstruction. This aggregation strategy effectively transforms raw mathematical modes into physically interpretable ones, thereby simplifying the subsequence learning task.

Variational Mode Decomposition (VMD)

Variational Mode Decomposition (VMD) is an adaptive signal-processing technique that decomposes a time series into band-limited modes via constrained variational optimization [29]. Assuming the signal consists of K discrete modes u_k with corresponding center frequencies ω_k , VMD minimizes the sum of their estimated bandwidths:

$$\min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|^2 \right\} \quad (4)$$

where $u_k(t)$ denotes the k -th band-limited mode and ω_k is its center angular frequency, K is the number of modes, and t is time. $\delta(t)$ is the Dirac delta (representing a unit impulse at $t = 0$), and $\frac{1}{\pi t}$ represents the Hilbert-transform kernel. In signal processing, convolution with this kernel effectively shifts the phase of the signal by 90°, providing the imaginary component necessary to construct the analytic signal, and $e^{-j\omega_k t}$ shifts the mode to baseband.

To align with our physics-informed assumption that the deterministic component comprises a seasonal cycle and a diurnal cycle, we intentionally set $K = 2$. This configuration constrains the algorithm to extract two dominant oscillatory components:

1. Model1 (IMF1): Captures the low-frequency trend, including annual seasonality.

2. Model2 (IMF2): Captures the diurnal-scale oscillation.

Accordingly, any remaining variability not captured by these two modes is defined as the stochastic residual:

$$r(t) = Y(t) - (\text{IMF1}(t) + \text{IMF2}(t)) \quad (5)$$

This approach ensures that $r(t)$ aggregates high-frequency fluctuations, measurement noise, and other unexplained variability, thereby separating

the deterministic components from *stochastic* variability.

Model Training and Implementation Details

We adopt a strictly chronological data-splitting strategy to evaluate model performance and enable a rigorous assessment of temporal generalization. Data from 2019-2022 are used for training, 2023 for validation and hyperparameter tuning, and 2024 as an independent test set. All meteorological and pollutant variables are Z-score normalized using training-set statistics only, which are then applied to the validation/test sets to avoid leakage. We used mean squared error (MSE) as the objective loss function. To promote stable convergence and prevent overfitting, we employed early stopping with a patience of 30 epochs, based on a sliding moving average of the validation MSE to reduce sensitivity to stochastic fluctuations. Hyperparameter optimization was conducted using the Tree-structured Parzen Estimator (TPE), a Bayesian optimization algorithm as implemented in Optuna [30]. Compared with conventional grid search, this method more efficiently explores the hyperparameter space by sequentially evaluating promising configurations for key hyperparameters, including optimizer type (Adam vs. AdamW), learning rate, and model architectural dimensions (Table S2). In this study, all models were implemented in PyTorch and trained on a high-performance Linux cluster. Detailed information on software versions and hardware specifications (CPUs, GPUs, and memory) is provided in Supplementary Text S4. Although the decomposition and multi-subsequence training strategy introduces additional preprocessing and training cost, the computational burden remains manageable for routine automated deployment. In our computing environment, baseline training required 1.5 ± 0.3 h, while decomposition-based frameworks required about 2-5 h. After training, all methods were computationally efficient in inference, generating one full year of predictions within 1 min.

Model Performance Evaluation

We employ a two-stage evaluation framework that verifies the validity of signal decomposition prior to assessing forecasting skill. Decomposition quality is assessed by inspecting component structures for mode mixing/leakage and quantified using residual standard deviation (unexplained energy) and the orthogonality index (inter-component redundancy). Physical consistency is evaluated by relating decomposed components to key meteorological drivers and by comparing alternative recombination strategies to verify the preservation of interpretable meteorology- O_3 relationships. Guided by these diagnostics, we adopt a parsimonious decomposition scheme, favoring period-based merging (CEEMDAN) and a low-order VMD setting ($K = 2$) to balance interpretability

and computational efficiency. Differences in optimal model size across evaluation scenarios indicate that distinct settings require tailored configurations, highlighting the importance of automated optimization. Robustness is quantified using statistical metrics averaged across four megacities and four benchmark scenarios defined by temporal resolution and seasonal regime. Lead-time-dependent behavior is characterized over the full 168-h horizon by analyzing RMSE trajectories across Days 1-7, complemented by episode-based diagnostics (monthly performance and Top-1 days) and time-series comparisons during high-volatility episodes to evaluate temporal coherence and phase alignment between forecasts and observations. Details of the evaluation metrics and the calculation of MDA8 from hourly concentrations are provided in Supplementary Text S5 and S6.

Results and Discussion

Analysis of Sequence Decomposition Characteristics

Among the four decomposition methods, O_3 variability is partitioned into markedly different temporal-scale components, as illustrated for Beijing (Fig. 2). STL and GAM delineate broad seasonal and diurnal structures; however, their residuals retain considerable high-frequency variance and intermittently clustered fluctuations, suggesting incomplete separation of structured variability. In particular, the periodic constraints in STL and GAM yield highly regular diurnal patterns (and, for GAM, an additional seasonal component), while non-stationary variability is left in the residual term. Notably, within GAM, this tensor-product interaction term allows the model to capture the seasonal modulation of the diurnal O_3 cycle's amplitude, a real-world phenomenon driven by changes in day length and solar zenith angle representing non-linear coupling between annual-scale conditions and daily O_3 cycles (Fig. S4) [31]. CEEMDAN is widely regarded as more adaptive than classical EMD, yet it can still be affected by residual mode mixing. Consistent with this limitation, our reconstructed components exhibit partial leakage of variability between the low-frequency and diurnal terms, while the high-frequency series retains intermittent spikes. By contrast, VMD yields a compact set of frequency-localized components, including a low-frequency mode capturing long-term variability, a distinct diurnal-scale mode, and a lower-amplitude residual relative to the original series [29].

The patterns in Fig. 2 are further supported by quantitative metrics used to assess decomposition quality (Table S6). Using the residual standard deviation as a proxy for unmodeled information, VMD achieves the lowest residual variance ($9.46 \mu\text{g}/\text{m}^3$), substantially outperforming STL ($23.10 \mu\text{g}/\text{m}^3$), GAM ($28.69 \mu\text{g}/\text{m}^3$), and CEEMDAN ($36.38 \mu\text{g}/\text{m}^3$), suggesting that VMD

removes a larger fraction of structured variability from the residual. In addition, the Orthogonality Index (IO), which measures redundancy among decomposed components, provides complementary evidence in favor of VMD [32]. VMD yields a low IO value ($\text{IO} = 0.0060$), compared with STL ($\text{IO} = 0.1127$) and CEEMDAN ($\text{IO} = 0.0238$), indicating a near-orthogonal decomposition with limited overlap among components. Although GAM yields a near-zero IO ($\text{IO} = 0.0002$), this primarily reflects the orthogonality implied by its additive regression form rather than strong separation of stochastic variability, consistent with its comparatively large residual dispersion. Taken together, these metrics suggest that VMD better balances residual reduction with reduced component redundancy, yielding components that are more physically interpretable.

The decomposed subsequences fall into three regimes: low-frequency, diurnal, and high-frequency residuals, each reflecting distinct physicochemical controls.

1. The low-frequency (annual) component is weakly correlated with meteorology ($|R| < 0.1$), suggesting slower anthropogenic/emissions influences and persistence-based predictability (Fig. S5).

2. The diurnal component correlates strongly with solar radiation and temperature ($R > 0.6$), indicating that it captures the recurrent daytime photochemical enhancement and boundary-layer evolution that jointly shape the 24-h O_3 cycle, and is therefore more meteorology-driven in predictability. The high-frequency residuals show weak correlations with meteorological variables ($|R| < 0.2$) and predominantly reflect stochastic variability.

Thus, decomposition quality largely depends on isolating the diurnal component that carries most meteorology-linked predictability. While STL and GAM yield slightly higher correlations with radiation ($R \approx 0.72$ for USWRF) than VMD ($R \approx 0.67$), this may be partly attributed to their rigid, imposed 24-h periodicity, which closely follows the mean solar cycle but may not capture non-stationary features (Supplementary Text S8). In contrast, VMD allows subtle non-stationary deviations in the diurnal cycle to be retained in the diurnal mode rather than being absorbed into the residual. By reducing leakage of diurnal variability to the residual, VMD yields a more realistic O_3 representation and is consistent with improved interpretability and predictive performance [1].

Optimization of Decomposition Strategy and Model Configuration

Direct VMD ($K = 2$) and period-based merging isolate a meteorology-coupled diurnal O_3 component, whereas entropy-based recombination does not [1]. Spectral metrics (Table S7) show that VMD ($K = 2$) with period-based merging consistently preserves the 24-h diurnal mode. By contrast, entropy-based merging often misallocates the 24-h signal into the low-frequency

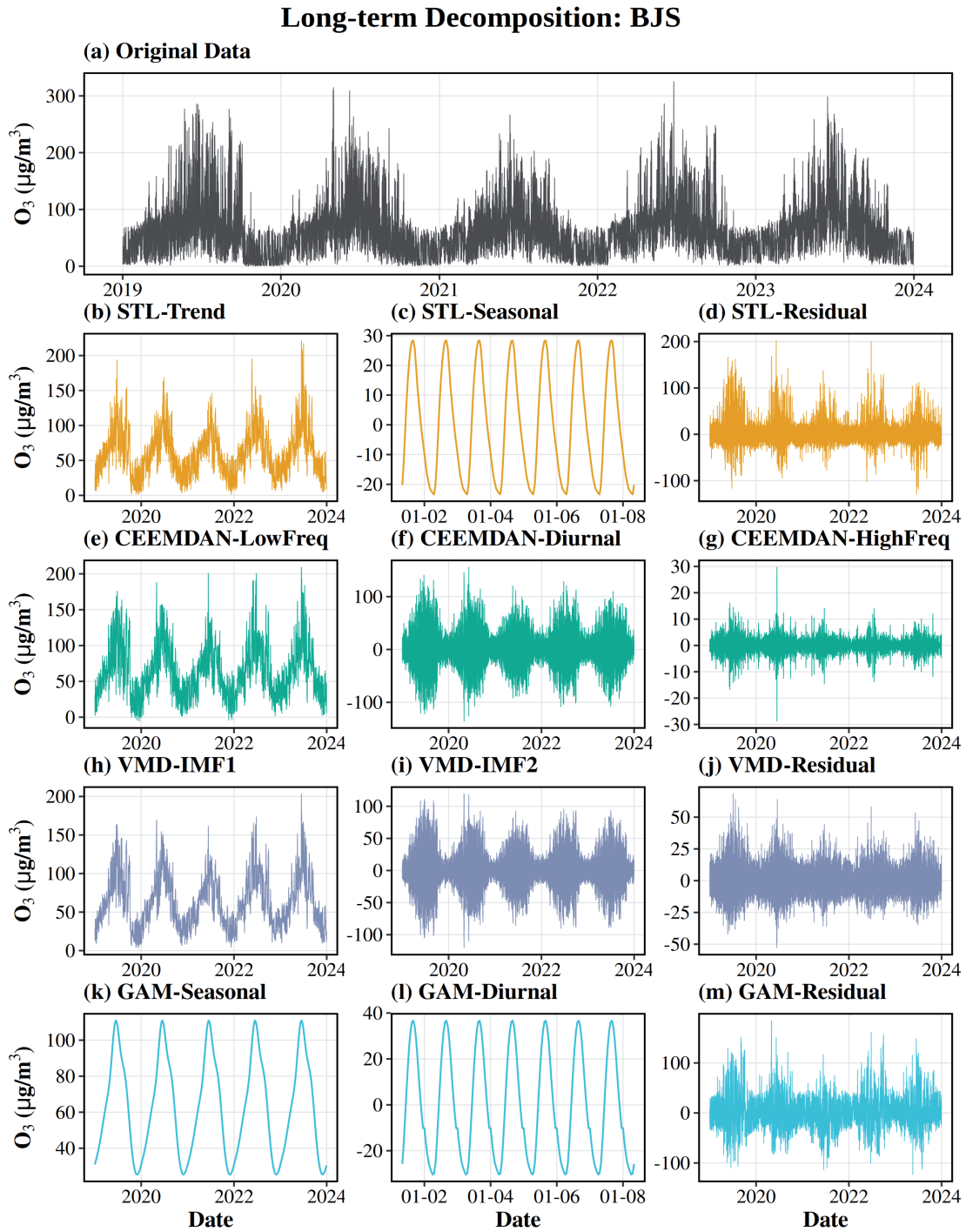


Fig. 2. Comparison of time-series decomposition methods for surface ozone (O_3) concentrations (2019-2023). a) Original observed time series. (b-d) STL decomposition results, displaying the Trend, Seasonal (daily), and Residual components. (e-g) GAM decomposition results, illustrating the Seasonal (annual), Diurnal (daily), and Residuals. (h-j) CEEMDAN decomposition showing selected Intrinsic Mode Functions (IMFs) representing the trend, diurnal cycle, and high-frequency oscillations. (k-m) VMD decomposition showing the Low-Frequency Trend (IMF1), Diurnal Component (IMF2), and High-Frequency Residuals.

term in Guangzhou and Shanghai. This likely reflects the limitation of sample entropy as a purely statistical metric: in these coastal cities, cloud variability and sea breezes introduce strong high-frequency disturbances, making entropy-based grouping more likely to split physically related diurnal components. Entropy-based recombination shows only moderate correlation with

radiation ($R \approx 0.54$), whereas the diurnal components from VMD ($K = 2$) or $K = 8$ with period-based merging correlate more strongly with both temperature ($R \approx 0.69$) and solar radiation ($R \approx 0.67$), comparable to the period-based reconstruction (Fig. S6). This indicates that $K = 2$ successfully captures the main meteorology-linked O_3 cycles (Table S13), reducing cross-scale leakage

and allowing clearer attribution across timescales, all without requiring an additional decompose-and-merge step, which makes it computationally efficient for operational forecasting. For CEEMDAN, where the number of modes cannot be prespecified and merging is required (Table S8, Fig. S7), period-based merging likewise outperforms entropy-based recombination in isolating the meteorology-coupled diurnal signal.

Optimal model capacity differs substantially between the baseline O_3 forecasting task and the decomposed subsequences, highlighting strong heterogeneity in parameter requirements across frequency components and geographic settings (Fig. S8) [35]. Whereas the baseline model favors a relatively uniform architecture to accommodate coupled signals, the decomposed components require more specialized parameterizations that reflect their distinct physical and statistical structure. Clear frequency dependence is observed: the high-frequency VMD IMF1 component, which captures rapid fluctuations, consistently favors larger capacity (mean parameter metric $\approx 20,096$), whereas the low-frequency VMD IMF2 component, dominated by smooth background evolution, can be represented with a more compact architecture (mean $\approx 12,032$). Optimal configurations also vary markedly across regions. For instance, the preferred size for VMD IMF2 reaches 27,136 parameters in Chengdu but drops to 4,608 in Guangzhou, a nearly sixfold difference that likely reflects location-specific atmospheric and environmental regimes. Collectively, these results indicate that a static, one-size-fits-all architecture is insufficient for decomposed O_3 forecasting. Instead, scenario-specific model adaptation appears necessary to fully leverage the predictive information embedded in heterogeneous spatiotemporal O_3 components.

Forecast Performance

Using the 2024 dataset, we assess robustness by averaging RMSE across four representative cities under four forecasting scenarios that span Hourly vs. Daily resolution and Annual vs. Pollution-season conditions (Fig. 3). This design systematically tests sensitivity to rising task difficulty driven by temporal aggregation and seasonal regime shifts. As shown in Fig. 3(a-d), baseline models and decomposition-integrated frameworks exhibit markedly different error behaviors. Baseline performance is highly sensitive to both temporal granularity and seasonality. RMSE rises sharply from $25.30 \mu\text{g}/\text{m}^3$ in the Hourly-Annual case (Fig. 3a) to $36.85 \mu\text{g}/\text{m}^3$ in the Daily-Pollution case (Fig. 3d), indicating substantial degradation under compounded smoothing and pollution-season volatility [36]. By contrast, decomposition-based frameworks remain stable across all scenarios. Even the weakest decomposition route consistently outperforms the strongest baseline model, suggesting that decomposition changes the error-growth pattern as forecasting complexity increases [37]. Overall, decomposition delivers an “error-dampening”

effect that mitigates uncertainty from temporal aggregation and seasonal transitions, bringing diverse model architectures into a narrower, more stable RMSE range.

Signal decomposition lowers the attainable error level of O_3 forecasting across all scenarios, clearly separating baseline and decomposition-based performance regimes. Notably, the highest RMSE values observed within decomposition-based frameworks are $23.95 \mu\text{g}/\text{m}^3$ (Fig. 3a), $26.65 \mu\text{g}/\text{m}^3$ (Fig. 3b), $25.54 \mu\text{g}/\text{m}^3$ (Fig. 3c), and $28.32 \mu\text{g}/\text{m}^3$ (Fig. 3d), all of which remain substantially lower than the minimum RMSE achieved by the corresponding baseline models, namely $24.87 \mu\text{g}/\text{m}^3$, $27.60 \mu\text{g}/\text{m}^3$, $26.77 \mu\text{g}/\text{m}^3$, and $33.77 \mu\text{g}/\text{m}^3$, respectively [38]. This “ceiling-below-floor” pattern indicates that decomposition is pivotal for surpassing the accuracy limits of direct forecasting [39]. Consistently, R^2 and the Index of Agreement improve under decomposition-based modeling across scenarios (Figs. S9 and S10). Interestingly, the relatively limited variation in R across scenarios (Fig. S11) suggests that decomposition does not fundamentally modify the intrinsic temporal dependency structure of the O_3 time series [40]. Instead, its primary contribution lies in improving global scale alignment and amplitude calibration, thereby reducing systematic bias while largely preserving phase relationships and relative variability. This indicates that decomposition acts as a structural regularization mechanism, enhancing predictability without distorting the underlying dynamical signatures of O_3 variability. The present study provides deterministic point forecasts, and robust uncertainty quantification remains challenging (Supplementary Text S10 and Table S15).

Beyond the universal gains from decomposition, approaching the accuracy ceiling depends on matching signal characteristics to model architecture. A larger backbone, such as a Transformer, does not automatically improve performance after decomposition [18]. Instead, clear pairings emerge: STL and CEEMDAN work best with Y -LSTM to preserve temporal continuity, whereas GAM aligns better with ΔY -LSTM. VMD is scenario dependent, coupling most effectively with ΔY -Transformer in annual forecasts to capture long-horizon structure (Fig. 3a and b)) but adapting better with Y -LSTM during the volatile pollution season (Fig. 3c and d)). This divergence is further reinforced by inter-city variability and by the distinct preferences of decomposed subsequences (Fig. S12). Optimal decomposition-model combinations vary across cities: GAM- Y -Transformer in Beijing, VMD- ΔY -Transformer in Chengdu, VMD-based Transformer models in Shanghai, and STL- Y -LSTM in Guangzhou. These differences align with city-specific climatological and emission characteristics, such as large-scale temporal structure in Beijing [41], basin effects in Chengdu [42], multi-scale separability in Shanghai [43], and humid subtropical conditions in Guangzhou favoring smoother models (Supplementary Text S11) [44]. To further

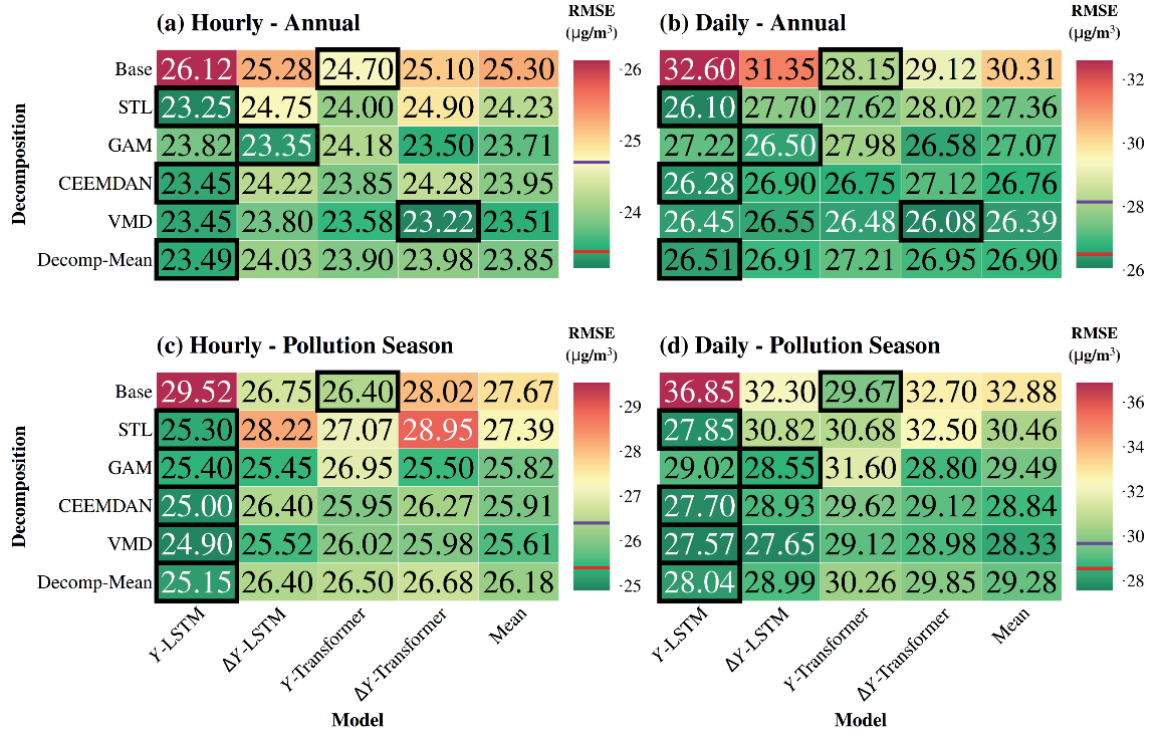


Fig. 3. City-averaged RMSE across four evaluation scenarios for decomposition-integrated forecasting frameworks. Panels (a-d) report the mean RMSE across four cities under Hourly-Year, Daily-Year, Hourly-Pollution Season, and Daily-Pollution Season settings. Rows correspond to decomposition strategies (Base, STL, GAM, CEEMDAN, VMD, and Decomp-Mean), and columns denote model backbones (Y-LSTM, ΔY -LSTM, Y-Transformer, ΔY -Transformer), with the last column showing the row-wise mean. Black outlines mark the best-performing (lowest-RMSE) configuration within each decomposition strategy (i.e., the optimal model-decomposition pairing for that row). Cell colors encode RMSE magnitude (green lower, red higher). The purple line in the legend indicates the best-performing baseline configuration, while the red line indicates the worst-performing decomposition-based configuration.

investigate the architectural affinities at the sub-series level, a detailed component-wise breakdown reveals that most components exhibit stronger adaptability to Y-based models, whereas high-frequency elements such as IMF1 in VMD and the granular residuals in GAM require ΔY -based architectures to effectively resolve fine-scale fluctuations (Fig. S13) [45]. Although VMD-based configurations typically rank first, their advantage over other decomposition routes can be marginal. For example, in the Hourly-Annual scenario, the gap between the theoretical SOTA model (VMD- ΔY -Transformer, 23.22 $\mu\text{g}/\text{m}^3$) and the cost-effective leader (STL-Y-LSTM, 23.25 $\mu\text{g}/\text{m}^3$) is only 0.03 $\mu\text{g}/\text{m}^3$. This implies that capacity, differencing, and attention depth should be tuned per component and region to avoid overfitting smooth modes or underfitting variability [46]. From an operational perspective, LSTM-based models are generally lighter, with training times of 1.2 ± 0.2 h, compared to 1.7 ± 0.3 h for Transformer-based models, with only modest differences overall.

Divergent Optimal Architectures for Short-Versus Long-Term Forecasting

The comparative results indicate that the optimal architecture differs between short-term and medium-

to long-term O_3 forecasts [47]. Using the mean RMSE in Fig. 4a), we select the five best decomposition-based models together with the baseline for detailed comparison. Performance is strongly horizon dependent: VMD- ΔY -Transformer achieves the overall best results, with its advantage concentrated in Days 1-3. Leveraging the Y-LSTM backbone due to its robust average performance with decomposition techniques, the comparative evaluation (Fig. 4b)) reveals that VMD yields the lowest initial error on Day 1. Within both VMD and STL-based settings, the incremental strategy (ΔY) consistently reduces Day-1 RMSE relative to direct Y forecasting [48]. Accordingly, the short-term edge of VMD- ΔY -Transformer is likely driven by the combined ability of VMD and ΔY to mitigate initial-state error likely arises from VMD's cleaner separation of high-frequency variability, enabling the Transformer's attention mechanism to capture rapid changes [49]. In contrast, STL-Y-LSTM benefits from STL's smoother low-frequency structure, which aligns with the LSTM's recurrent memory for tracking slow, persistent background evolution, explaining its advantage at longer lead times (Days 5-7). This illustrates how the interaction between decomposition characteristics and model architecture drives horizon-dependent forecasting performance. The more detailed RMSE trajectories

corresponding to these analyses are systematically visualized in Fig. S14.

Performance Evaluation during Representative Pollution Episodes

Using the day-by-day RMSE rankings for Beijing, Chengdu, Guangzhou, and Shanghai in the T1D task (Fig. 5), we quantify dominance stability via the “number of daily bests” defined as the count of days on which a model attains the lowest RMSE. Across 123 evaluation days, VMD- ΔY -Transformer (30 daily bests) and GAM- ΔY -LSTM (28) jointly contribute 47.2% of all wins, indicating strong robustness and transferability across cities and dates [51]. STL-Y-LSTM and CEEMDAN-Y-LSTM (18 days each) form a second tier, suggesting that decomposition/denoising strategies can be competitive in specific periods but with slightly weaker dominance stability. VMD-Y-LSTM (15 days) and the baseline Base-Y-Transformer (14 days) rank third, implying greater sensitivity to temporal variations when structural decomposition is absent or limited. Notably, T1D reflects the frequency of winning rather than the magnitude of improvement. Thus, similar win counts may still correspond to different RMSE margins, because this metric reflects how often a model ranks

first across dates rather than the size of its improvement on any given day [52].

Across cities, the identity of the optimal model varies markedly, indicating strong spatial heterogeneity in model suitability [53]. For Beijing and Chengdu, most T1D wins are delivered by GAM- ΔY -LSTM, VMD- ΔY -Transformer, and CEEMDAN-Y-LSTM, suggesting that ΔY forecasting and decomposition help stabilize performance under more complex, non-stationary intraday and interday variability [54]. Guangzhou, on the other hand, presents a characteristic of “methodological divergence”: the advantages of VMD- ΔY -Trans and STL-Y-LSTM are more concentrated, while the baseline model also secured a certain number of best days (8 days), implying that the sequence structure in this city during some periods is closer to baseline assumptions, with narrowed gaps between models and advantages relying more on specific pollution processes and meteorological driving stages. In Shanghai, the optimal model switches more frequently, indicating that performance advantages are more episodic and difficult to sustain with a single configuration over a long horizon. Overall, these cross-city patterns indicate that the “best model” is not fixed and motivate city-aware modeling, stratified evaluation by season or weather regime, and adaptive ensemble strategies to improve stability across stages [55, 56].

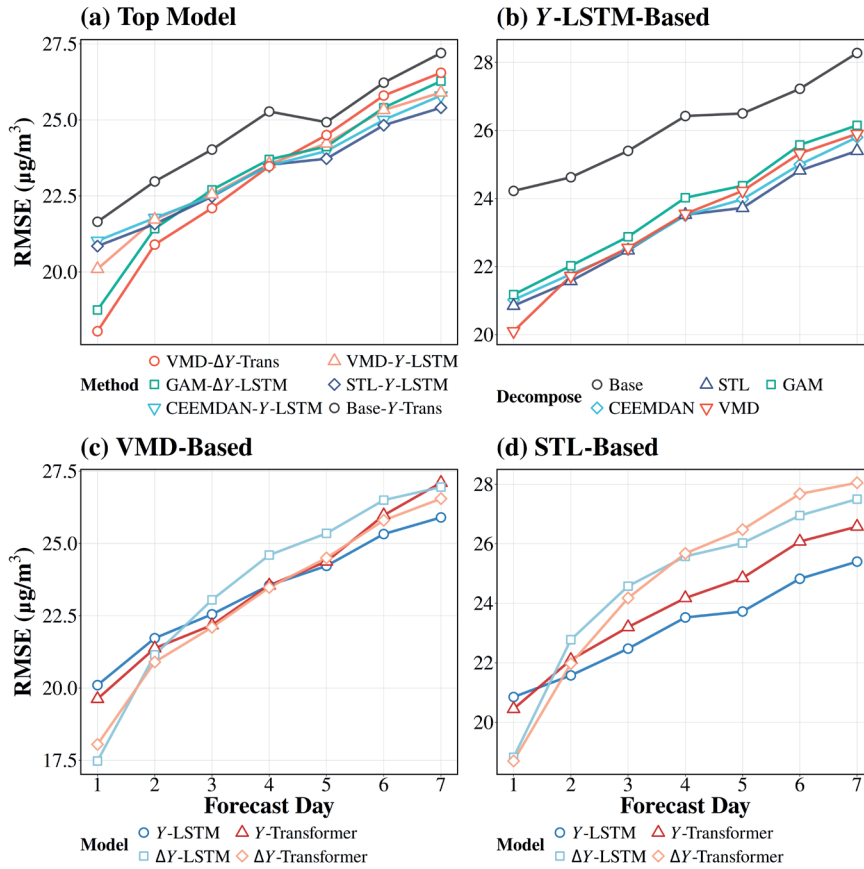


Fig. 4. RMSE across forecast horizons (1-7 days) averaged over four representative cities. a) Top-performing methods versus the baseline, b) performance of Y-LSTM-based frameworks across decomposition choices, c) performance of VMD-based frameworks across model choices, d) performance of STL-based frameworks across model choices.

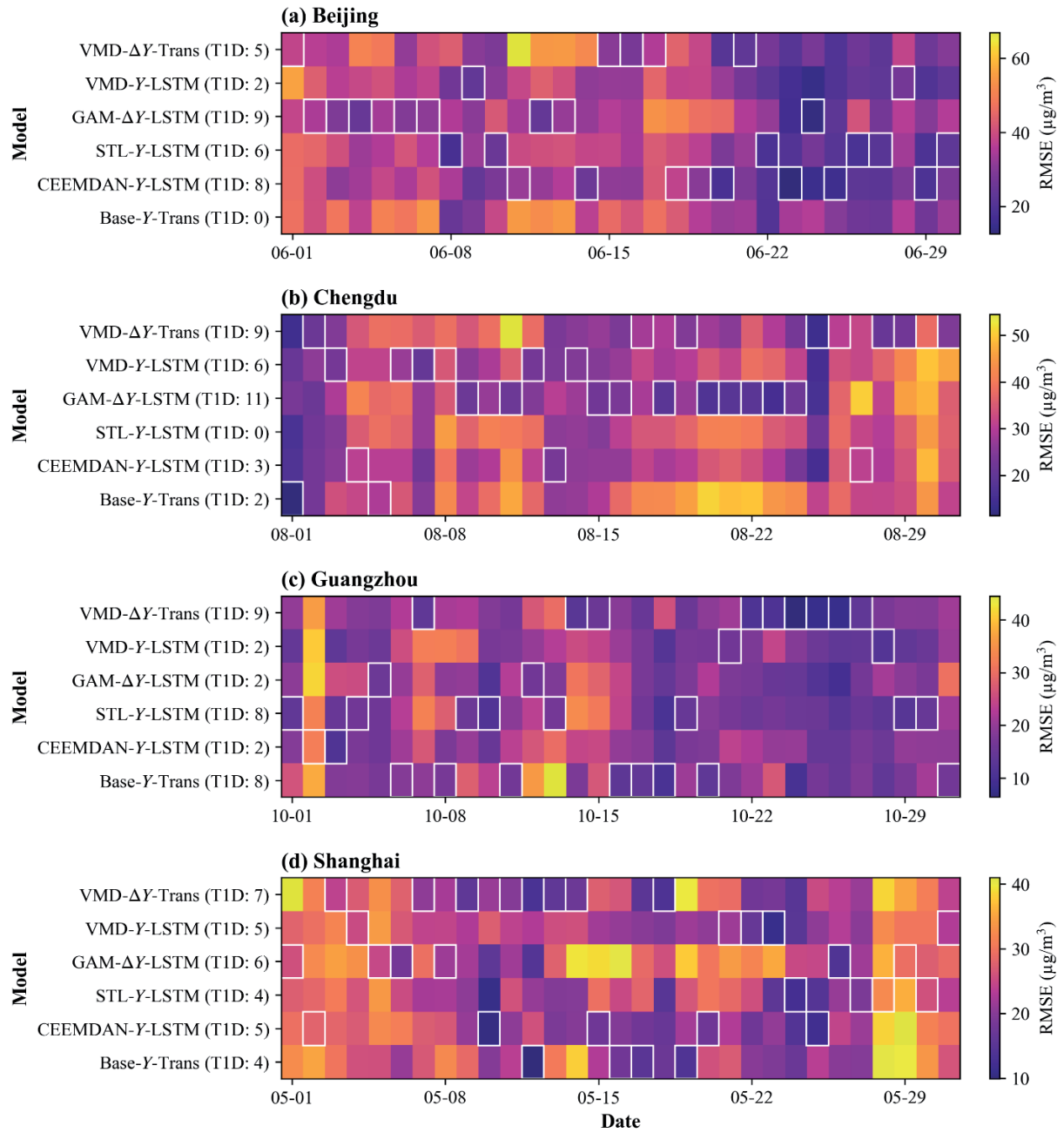


Fig. 5. The evaluation of daily forecasting performance across representative high-ozone months. The performance comparison of decomposition-integrated forecasting frameworks during representative high-ozone months. Panels (a-d) show the daily RMSE for Beijing (June), Chengdu (August), Guangzhou (October), and Shanghai (May). For each city, T1D (Top-1 Days) indicates the number of days within the corresponding high-pollution month on which a method ranks 1st (i.e., achieves the lowest RMSE among all compared methods). Here, Trans denotes the Transformer-based model. White boxes mark the days when the corresponding method attains the best performance (lowest RMSE) among all methods.

Limitations and Future Work

This study has several limitations that motivate future work. Meteorological forcing is represented using a fixed 9×9 neighborhood average for each city, which can smooth horizontal gradients relevant to transport and stagnation and does not explicitly capture vertical structure. Future meteorology over the 168-h horizon is taken from a single NWP product, so potential

gains from multi-product complementarity (e.g., bias characteristics, spatial resolution, and regime-dependent skill) remain unexplored. Our benchmark focuses on external decomposition methods (STL, GAM, CEEMDAN, VMD) coupled with LSTM/Transformer backbones and does not include decomposition-integrated Transformer variants such as Autoformer or FEDformer. We also evaluate decomposition methods independently rather than integrating multiple strategies to exploit complementary mechanisms. Future work

could (i) replace fixed-area averages with spatial encoders over gridded, multi-level meteorological fields to better represent transport and vertical structure, (ii) incorporate multi-source NWP inputs, (iii) benchmark decomposition-integrated Transformer variants such as Autoformer and FEDformer, and (iv) explore hybrid or ensemble decomposition schemes that integrate complementary mechanisms.

Conclusions

This study develops a decomposition-ensemble framework for 168-h O_3 forecasting and shows that resolving time–frequency structure is critical for improving skill across lead times. By separating multi-scale signals, the framework better captures meteorology-coupled diurnal variability and reduces cross-scale interference. Parsimonious VMD ($K = 2$) and period-based merging for CEEMDAN provide the best overall trade-off among spectral separation, forecastability, and computational efficiency. Optimal capacity and architecture vary across components and city regimes, which supports automated, component-wise hyperparameter tuning rather than one-size-fits-all designs. Across four scenarios, decomposition-based models consistently beat the undecomposed baseline and show “error dampening” with slower error growth as horizon increases. The best decomposition-architecture pairing is regime dependent: VMD- ΔY -Transformer is strongest for short-term (1-3 day) forecasts, while STL- Y -LSTM is more stable and cost-effective at medium to long lead times (5-7 days). Cross-city comparisons identify VMD- ΔY -Transformer and GAM- ΔY -LSTM as the most robust overall, and the observed heterogeneity motivates scenario-adaptive forecasting that selects models by horizon and regime for decision support.

Acknowledgments

The authors gratefully acknowledge the China National Environmental Monitoring Centre (CNEMC) for providing ground-based surface air quality observations, and the National Centers for Environmental Prediction (NCEP) for making the Global Forecast System (GFS) 0.25° Global Forecast Grids Historical Archive (ds084.1) publicly available. These datasets provide essential support for the development and evaluation of the proposed forecasting framework.

AI Usage Disclosure Statement

During the preparation of this manuscript, the authors used ChatGPT (OpenAI) solely for language editing to improve the clarity and readability of the

English text. The tool was not used for data generation, data analysis, model development or training, figure preparation, result calculation, scientific interpretation, or the generation of conclusions. All AI-assisted language revisions were carefully reviewed and revised by the authors. The authors take full responsibility for the accuracy, integrity, and originality of the manuscript.

Conflict of Interest

The authors declare no conflict of interest.

References

- LIU H., MI X., LI Y. Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM. *Energy Conversion and Management*, **159**, 54, **2018**.
- SHOJAEI BARJOEE S., FAROKHI H., ABADI S., SEDGHI S. Investigation of morphological and mineralogy characteristics of dust around the lead and zinc mine in Mehdiabad, Yazd, Iran, Using the SEM and XRD analyzes. *Journal of Air Pollution and Health*, **5** (2), 121, **2020**.
- SHOJAEI BARJOEE S., NIKBAKHT M., MALVERDI E., ABADI Z., NAGHDI M. Modeling the Consequences of Benzene Leakage from Tank using ALOHA in Tar Refining Industrial of Kerman, Iran. *Pollution*, **7** (1), 217, **2021**.
- 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 Climate Atmospheric Science*, **6** (1), 71, **2023**.
- SHOJAEI BARJOEE S., RODIONOV V., BABKIN R., TUMANOV M., SHOJAEI B. A fuzzy set-based hybrid SWARA-CoCoSo-William Fine framework for safety risk assessment in a ceramic granule preparation unit. *Proceedings of the Indian National Science Academy*, **2025**.
- PANDE C.B., KUSHWAHA N.L., ALAWI O.A., SAMMEN S.S., SIDEK L.M., YASEEN Z.M., PAL S.C., KATIPOGLU O.M. Daily scale air quality index forecasting using bidirectional recurrent neural networks: Case study of Delhi, India. *Environmental Pollution*, **351**, 124040, **2024**.
- 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**.
- SALMAN A.K., CHOI Y., SINGH D., KAYASTHA S.G., DIMRI R., PARK J. Temporal CNN-based 72-h ozone forecasting in South Korea: Explainability and uncertainty quantification. *Atmospheric Environment*, **343**, 120987, **2025**.
- 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**, 121883, **2024**.
- ZAINI N.A., EAN L.W., AHMED A.N., ABDUL MALEK M., CHOW M.F. $PM_{2.5}$ forecasting for an urban area based

- on deep learning and decomposition method. *Scientific Reports*, **12** (1), 17565, **2022**.
11. 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**.
12. HUANG M.-L., CHAMNISAMPAN N., KE Y.-R. A Deep Learning Model Integrating EEMD and GRU for Air Quality Index Forecasting. *Atmosphere*, **16** (9), 1095, **2025**.
13. 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**.
14. 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**.
15. 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**.
16. 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**.
17. 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**.
18. ZENG A., CHEN M., ZHANG L., XU Q. Are transformers effective for time series forecasting? *Proceedings of the AAAI Conference on Artificial Intelligence*, **37** (9), 11121, **2023**.
19. SHOJAEI BARJOEE S., REZAEI N., ABADI S., MALVERDI E., NIKBAKHT M., NAGHDI M. Multivariate and Statistical Approaches for the Source Apportionment and Evaluation of Trace Elements Pollution at Mining Areas (Case study: Mehdi Abad Pb/Zn mine). *Advances in Environmental Technology*, **6** (1), 19, **2020**.
20. CHINA NATIONAL ENVIRONMENTAL MONITORING CENTRE. China National Environmental Monitoring Centre Official Website. 2025/10/1. Available online: <https://www.cnemc.cn/> (Accessed in 2025/10/1).
21. NATIONAL CENTERS FOR ENVIRONMENTAL PREDICTION. NCEP GFS 0.25 Degree Global Forecast Grids Historical Archive. Available online: <https://rda.ucar.edu/datasets/ds084.1/> (Accessed in 2025/10/1).
22. LIU X., PU X., LU C., ZHANG H., LI T., GRIENEISEN M.L., LI J., MA N., YAN C., ZHAN Y., YANG F. Interpretable regional meteorological feature extraction enhances deep learning for extended 120-h $PM_{2.5}$ forecasting. *Journal of Cleaner Production*, **483**, 144287, **2024**.
23. ZHANG A., FU T.M., FENG X., GUO J., LIU C., CHEN J., MO J., ZHANG X., WANG X., WU W., HOU Y., YANG H., LU C. Deep learning-based ensemble forecasts and predictability assessments for surface ozone pollution. *Geophysical Research Letters*, **50** (8), e2022GL102611, **2023**.
24. CLEVELAND R.B., CLEVELAND W.S., MCRAE J.E., TERPENNING I. STL: A seasonal-trend decomposition. *Journal of Official Statistics*, **6** (1), 3, **1990**.
25. HASTIE T., TIBSHIRANI R. Generalized additive models. *Statistical Science*, **1** (3), 297, **1986**.
26. RILLING G., FLANDRIN P., GONCALVES P. On empirical mode decomposition and its algorithms. *Proceedings of the IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP 2003)*, **2003**.
27. RICHMAN J.S., MOORMAN J.R. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, **278** (6), H2039, **2000**.
28. YOUNGWORTH R.N., GALLAGHER B.B., STAMPER B.L. An overview of power spectral density (PSD) calculations. *Optical Manufacturing and Testing VI*, **5869**, **2005**.
29. DRAGOMIRETSKIY K., ZOSSO D. Variational mode decomposition. *IEEE Transactions on Signal Processing*, **62** (3), 531, **2013**.
30. AKIBA T., SANO S., YANASE T., OHTA T., KOYAMA M. Optuna: A Next-Generation Hyperparameter Optimization Framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623, **2019**.
31. SHOJAEI BARJOEE S., AZIZI M., KOUHKAN M., ALIPOURFARD I., BAYAT A., SHAHBAZ Y.H., BADIIEFAR A., LATIF M.T. The Impacts and Analysis of Individual and Social Risks of the Stochastic Emission of Benzene from Floating-Roof Tanks Using Response Surface Analysis and MPACT Model. *Archives of Environmental Contamination and Toxicology*, **84** (3), 347, **2023**.
32. 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. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, **454** (1971), 903, **1998**.
33. CHEN W., WANG Z., XIE H., YU W. Characterization of surface EMG signal based on fuzzy entropy. *IEEE Transactions on Neural Systems Rehabilitation Engineering*, **15** (2), 266, **2007**.
34. JÁNOSI I.M., MÜLLER R. Empirical mode decomposition and correlation properties of long daily ozone records. *Physical Review E-Statistical, Nonlinear, Soft Matter Physics*, **71** (5), 056126, **2005**.
35. ZHANG G.P. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, **50**, 159, **2003**.
36. CHEN B., YANG X., XU J. Spatio-temporal variation and influencing factors of ozone pollution in Beijing. *Atmosphere*, **13** (2), 359, **2022**.
37. KOURENTZES N., PETROPOULOS F., TRAPERO J.R. Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting*, **30** (2), 291, **2014**.
38. ULLAH S., CHEN X., HAN H., WU J., DONG J., LIU R., DING W., LIU M., LI Q., QI H., HUANG Y., YU P.L. A novel hybrid ensemble approach for wind speed forecasting with dual-stage decomposition strategy using optimized GRU and transformer models. *Energy*, **329**, 136739, **2025**.
39. SHOJAEI BARJOEE S., GENDLER S. Sustainable illumination: Experimental and simulation analysis of illumination for workers wellbeing in the workplace. *Heliyon*, **10** (24), e40745, **2024**.
40. WILLMOTT C.J. On the validation of models. *Physical Geography*, **2** (2), 184, **1981**.
41. CHEN Z., CHEN D., ZHAO C., KWAN M.-P., CAI J., ZHUANG Y., ZHAO B., WANG X., CHEN B., YANG

- J., LI B., HE B., GAO B., WANG K., XU B. Influence of meteorological conditions on $PM_{2.5}$ concentrations across China: A review of methodology and mechanism. *Environment International*, **139**, 105558, **2020**.
42. 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**.
 43. LI L., CHEN C., HUANG C., HUANG H., ZHANG G., WANG Y., WANG H., LOU S., QIAO L., ZHOU M., CHEN M.H., CHEN D.G., FU J.S., JANG C.J 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**.
 44. 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**.
 45. LIU Z., CHENG M., LI Z., HUANG Z., LIU Q., XIE Y., CHEN E. Adaptive normalization for non-stationary time series forecasting: A temporal slice perspective. *Advances in Neural Information Processing Systems*, **36**, 14273, **2023**.
 46. WOLPERT D.H., MACREADY W.G. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, **1** (1), 67, **1997**.
 47. BEN TAIEB S., BONTEMPI G., ATIYA A.F., SORJAMAA A. A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. *Expert Systems with Applications*, **39** (8), 7067, **2012**.
 48. 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**.
 49. HE K., ZHANG X., REN S., SUN J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770, **2016**.
 50. THEODOSIOU M. Forecasting monthly and quarterly time series using STL decomposition. *International Journal of Forecasting*, **27** (4), 1178, **2011**.
 51. CERQUEIRA V., TORGO L., MOZETIČ I. Evaluating time series forecasting models: An empirical study on performance estimation methods. *Machine Learning*, **109** (11), 1997, **2020**.
 52. ARMSTRONG J.S., COLLOPY F. Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, **8** (1), 69, **1992**.
 53. WANG T., XUE L., BRIMBLECOMBE P., LAM Y.F., LI L., ZHANG L. Ozone pollution in China: A review of concentrations, meteorological influences, chemical precursors, and effects. *Science of the Total Environment*, **575**, 1582, **2017**.
 54. ZHAO S., YU Y., YIN D., QIN D., HE J., DONG atial patterns and temporal variations of six criteria air pollutants during 2015 to 2017 in the city clusters of Sichuan Basin, China. *Science of the Total Environment*, **624**, 540, **2018**.
 55. WOLPERT D.H. The lack of a priori distinctions between learning algorithms. *Neural Computation*, **8** (7), 1341, **1996**.
 56. REN Y., SUGANTHAN P.N., SRIKANTH N. A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods. *IEEE Transactions on Sustainable Energy*, **6** (1), 236, **2014**.

Supplementary Material

link to file: <https://www.pjoes.com/SuppFile/220417/1/>