

PolluVCCT: Multi-Scale Hybrid Learning for Robust Air Pollution Forecasting Across Diverse Climate Zones

Yuhe Wu

Dongbei University of Finance and Economics
Dalian, Liaoning, China
wyhttkx727@gmail.com

Xinyue Su

Dongbei University of Finance and Economics
Dalian, Liaoning, China
xinyuesu656@gmail.com

Yuran Chen

Dongbei University of Finance and Economics
Dalian, Liaoning, China
yuranchen06@gmail.com

Zhuang Liu

Dongbei University of Finance and Economics
Dalian, Liaoning, China
liuzhuang@dufe.edu.cn

Abstract

Reliable air pollution forecasting remains a major hurdle for environmental science, as conventional models are ill-equipped to handle the noisy, non-linear, and multi-scale nature of atmospheric data. This limitation severely hampers effective public health warnings and policy interventions. To overcome this, we introduce PolluVCCT, a hybrid deep learning framework designed for robust and long-range forecasting. At its core, PolluVCCT employs a hierarchical decoupling strategy. A cascaded VMD-CEEMDAN signal processing pipeline first disentangles complex time-series signals, which a parallel CNN-Transformer architecture then leverages to capture local emission patterns and long-range dependencies. Experiments on a benchmark of seven diverse climate zones confirm PolluVCCT's dual strengths. It pioneers a new level of cross-regional robustness, maintaining a MAPE variance below 3% across starkly different environments, from UV-intensive plateaus to inversion-prone basins. Our work provides a powerful, deployable tool for more reliable public health warnings and proactive environmental policy.

CCS Concepts

• **Applied computing;**

Keywords

Air pollution forecasting, Hybrid deep learning, Cross-Climate Generalization, Multi-Scale Decomposition

ACM Reference Format:

Yuhe Wu, Yuran Chen, Xinyue Su, and Zhuang Liu. 2025. PolluVCCT: Multi-Scale Hybrid Learning for Robust Air Pollution Forecasting Across Diverse

All authors contributed equally to this work

The open-source implementation is available at <https://github.com/Kzczc/KDD2025-PolluVCCT>.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
KDD,

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-XXXX-X/2025/08
<https://doi.org/XXXXXXX.XXXXXXX>

Climate Zones. In *Proceedings of August 03–07, 2025 (KDD)*. ACM, New York, NY, USA, 8 pages. <https://doi.org/XXXXXXX.XXXXXXX>

1 Introduction

The acceleration of industrialization and urbanization has made air pollution one of the most pressing environmental challenges worldwide, with profound implications for public health, ecosystem stability, and socioeconomic development[32]. According to the World Health Organization (WHO), more than seven million premature deaths every year are attributable to air-pollution-related illnesses such as cardiovascular diseases, respiratory disorders, and lung cancer[6]. Furthermore, air pollution threatens crop yields, degrades water quality, and weakens ecosystem biodiversity, but also carries latent risks[17], such as depressive symptoms [14, 20], leading to substantial labor losses and declines in productivity[7, 9, 33]. In response, governments and international organizations have promulgated a wide range of environmental regulations and have continuously improved air-quality monitoring and forecasting systems. Nevertheless, existing forecasting techniques still struggle to meet practical requirements owing to the inherent noise, strong non-stationarity, and complex multi-scale dynamics of pollutant time series[10, 12].

Against this backdrop, we propose PolluVCCT—a hybrid signal-decomposition and deep learning framework that couples variational Mode Decomposition (VMD) and complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) with a parallel CNN-Transformer architecture. VMD first isolates high-frequency noise and stochastic fluctuations, after which CEEMDAN conducts a second-stage decomposition of the residual intrinsic mode function (IMF). Subsequently, a convolutional neural network extracts local spatio-temporal patterns, while a Transformer equipped with a global self-attention mechanism captures long-range dependencies; multi-scale features are then fused to yield the final prediction. By hierarchically disentangling features and integrating multiple scales, PolluVCCT simultaneously mitigates data non-stationarity and balances local patterns with global trends.

Our contributions are threefold:

- (1) We introduce an adaptive multi-stage decomposition strategy. By using sample entropy to guide a VMD-CEEMDAN cascade, our method precisely decouples complex pollution signals from noise.

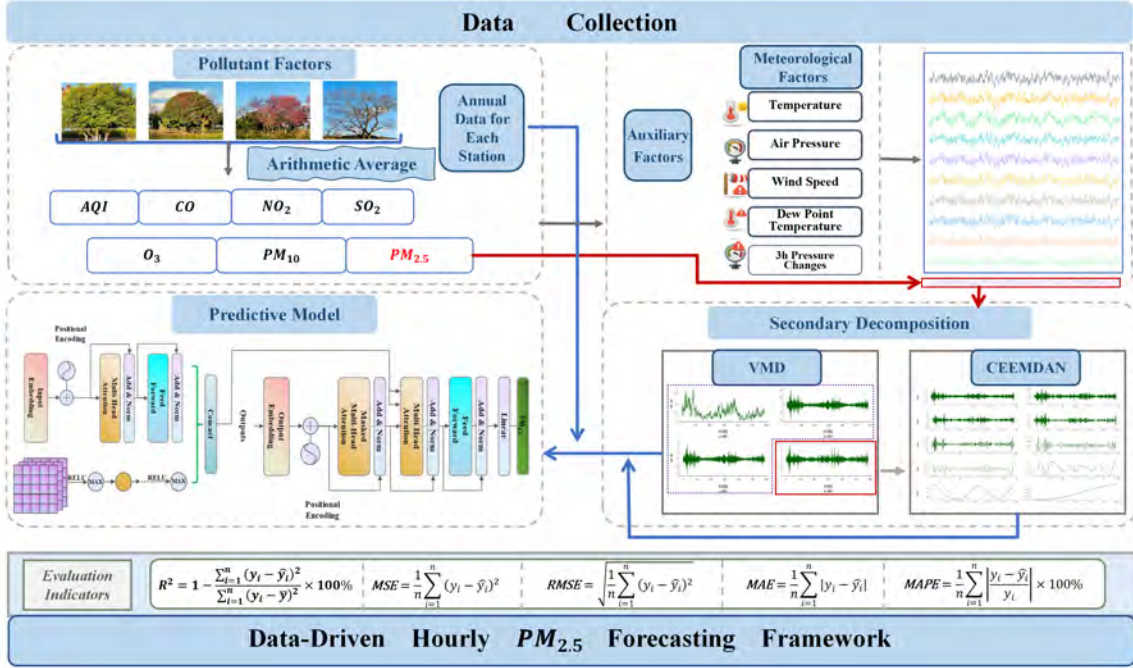


Figure 1: Multi-Source Data-Driven Forecasting Framework of the PolluVCCT Model

- (2) We design a parallel fusion architecture where dilated causal convolutions capture local emission spikes and a Transformer models long-range transport, effectively unifying multi-scale spatio-temporal dependencies.
- (3) We conduct extensive experiments on a new cross-climate benchmark we built covering multiple climate zones, on which our model achieves unprecedented long-term accuracy and generalization across diverse climates.

2 Related Works

Early research on air-quality forecasting was dominated by statistical regression and Eulerian-Lagrangian numerical simulation models that explicitly solved atmospheric transport and chemical-reaction equations[3, 35]. While physically interpretable, these approaches require dense, high-quality input data and often struggle to accommodate the strong non-linearity and multi-factor coupling characteristic of real-world pollution processes[2, 19, 22, 28]. To overcome these shortcomings, data-driven methods based on deep learning have gained traction. Qin et al.[18] introduced an autoencoder-enhanced multi-layer LSTM that improved temporal representation of pollutant series. Hardini et al.[13] demonstrated that convolutional neural network (CNN) can extract visual cues from low-cost mobile imagery, enabling multi-modal air-quality prediction in resource-constrained settings. Spatial-temporal hybrids further broadened this line of work: Xu et al.[23] and Zhang et al.[29] fused CNN and LSTM modules into unified frameworks, effectively capturing local spatial patterns and long-range temporal dependencies, while Guo et al.[11] showed that stacking DenseNet with ConvLSTM yields substantial gains for regional PM_{2.5} concentration forecasting.

Parallel to architectural innovations, signal-processing hybrids have been explored to mitigate the non-stationarity of pollutant series[1, 8, 16, 26, 27]. Zhang et al.[31] combined VMD with bidirectional LSTM to model forward and backward temporal relationships. More recently, Cao et al.[5] integrated LSTM, Temporal Convolutional Networks, and Transformers within a meta-heuristic optimization loop, achieving greater robustness under highly volatile conditions. Despite these advances, two critical issues remain: (1) pollutant data often contain high-frequency noise and exhibit pronounced multi-scale dynamics that simple pre-processing cannot fully disentangle[4, 15, 21, 34]; and (2) most models show limited generalization when deployed across disparate climate zones or emission regimes[24, 25, 30].

The proposed PolluVCCT framework addresses both gaps by cascading VMD and CEEMDAN to isolate stable intrinsic modes and by employing a parallel CNN-Transformer backbone to reconcile local receptive-field learning with global self-attention. Empirical results confirm that this design enhances cross-regional adaptability and predictive accuracy, positioning PolluVCCT as a compelling successor to existing deep-learning and hybrid paradigms.

3 Methodology

3.1 Problem Definition

Urban air quality prediction is a key task in the field of environmental monitoring and public health. The aim of this study is to construct a high-precision model using multi-source time-series data to predict PM_{2.5} concentrations in a single urban area. Given the historical PM_{2.5} and related environmental variable observations to date, the goal is to predict future PM_{2.5} concentrations over a short-term forecast horizon.

The inputs may include a variety of data sources such as ground monitoring station readings and meteorological sensors. These data sources typically have different temporal resolutions and noise characteristics. Accurate prediction requires dealing with the complex, nonlinear, and multi-scale dynamics of urban PM2.5 time-series data. Therefore, we first preprocess and time-align the input signals, and then employ a modeling framework that captures both short-term variations and long-range dependencies.

3.2 Model Architecture

The proposed model consists of two main modules: signal decomposition and parallel feature extraction and forecasting. The detailed process is illustrated in Algorithm 1 as follow.

Algorithm 1 PolluVCCT Framework

```

1: Require:
2:    $X_{raw}$ : Raw time-series signal
3:    $K, \alpha$ : VMD hyperparameters
4:    $I, \sigma_{noise}$ : CEEMDAN hyperparameters
5:    $\Theta$ : Model hyperparameters
6: Ensure:
7:    $\hat{Y}$ : The forecasted sequence
8: – Stage 1: Adaptive Signal Decomposition –
9:    $\{u_k\}_{k=1}^K \leftarrow \text{VMD-Decompose}(X_{raw}, K, \alpha)$   $\triangleright$  Decompose into  $K$  modes (IMFs)
10:   $E \leftarrow [\text{SampleEntropy}(u_1), \dots, \text{SampleEntropy}(u_K)]$   $\triangleright$  Compute entropy for each IMF
11:   $m \leftarrow \text{argmax}(E)$   $\triangleright$  Find index of the most complex IMF
12:   $\text{IMF}_m \leftarrow u_m$ 
13:   $\{\text{IMF}'_j\}_{j=1}^L \leftarrow \text{CEEMDAN-Decompose}(\text{IMF}_m, I, \sigma_{noise})$   $\triangleright$  Further decompose the complex IMF
14:   $X_{proc} \leftarrow \text{ReassembleComponents}(\{u_k\}, m, \{\text{IMF}'_j\})$   $\triangleright$  Assemble the final component set
15: – Stage 2: Parallel Feature Learning –
16:   $Y^{CNN} \leftarrow \text{CNN-Branch}(X_{proc}, \Theta)$   $\triangleright$  Extract local features
17:   $Y^{Trans} \leftarrow \text{Transformer-Encoder}(X_{proc}, \Theta)$   $\triangleright$  Extract global dependencies
18: – Stage 3: Multi-scale Feature Fusion –
19:   $\tilde{Y}^{CNN} \leftarrow \text{Interpolate}(Y^{CNN})$   $\triangleright$  Align temporal dimensions
20:   $Y^{Fusion} \leftarrow \tilde{Y}^{CNN} + Y^{Trans}$   $\triangleright$  Fuse features via element-wise addition
21: – Stage 4: Prediction Generation –
22:   $\hat{Y} \leftarrow \text{Transformer-Decoder}(Y^{Fusion}, Y^{Trans})$   $\triangleright$  Generate forecast via attention
23: return  $\hat{Y}$ 

```

3.2.1 Signal Decomposition. In the signal decomposition module, each input time series is first decomposed using VMD. VMD decomposes a complex signal into a predefined number of IMFs plus a residual trend. It achieves this by solving a variational optimization problem in which each IMF is constrained to be a narrow-band component around a certain center frequency:

$$\min_{\{u_k\}, \{\omega_k\}} \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \quad (1)$$

$$\text{subject to } \sum_{k=1}^K u_k(t) = f(t).$$

$u_k(t)$ is the k -th mode function extracted from the signal $f(t)$. ω_k is the central frequency corresponding to $u_k(t)$. $f(t)$ is the original input signal to be decomposed.

We use the augmented Lagrangian method ADMM with Lagrangian multipliers for optimization:

$$\mathcal{L}(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-i\omega_k t} \right\|_2^2$$

$$+ \left\| f(t) - \sum_{k=1}^K u_k(t) \right\|_2^2 + \langle \lambda(t), f(t) - \sum_{k=1}^K u_k(t) \rangle. \quad (2)$$

$\lambda(t)$ is the Lagrange multiplier used to constrain that the sum of all modal functions must equal the original signal. α is a weighting parameter used in the augmented Lagrangian formulation.

The result of VMD is a set of IMFs that capture oscillatory components of the signal at different frequency bands, effectively removing high-frequency noise and revealing the multi-scale structure of the data.

After VMD, we compute the sample entropy of each IMF to quantify its complexity (irregularity). Sample entropy is higher for more unpredictable or complex sequences. We select the IMF with the highest sample entropy as the target for further decomposition, since this component likely contains the most irregular fluctuations that may suffer from mode mixing. This selected mode is then decomposed by CEEMDAN to obtain cleaner intrinsic modes.

In CEEMDAN, Gaussian white noise is added to the target signal to create an ensemble of noisy copies:

$$Y_k(t) = \text{IMF}_m(t) + w_k(t) \quad (k = 1, 2, \dots, I). \quad (3)$$

Each noisy signal is decomposed by standard Empirical Mode Decomposition (EMD), and the first IMF from each is averaged to form a refined mode:

$$\text{IMF}_j = \frac{1}{I} \sum_{k=1}^I E_1(Y_k(t)) \quad (4)$$

This refined mode is subtracted from the residual, and the process repeats iteratively: at each iteration, noise is added to the current residual, EMD extracts its first IMF, and the average of these IMFs becomes the next mode:

$$r_j(t) = r_{j-1}(t) - \text{IMF}_j. \quad (5)$$

Iterations continue until the residual becomes essentially monotonic (no remaining extrema). The outcome is a set of refined IMFs and a final residual for the selected mode:

$$\text{IMF}_m(t) = \sum_{i=1}^L \text{IMF}_i + r_L(t). \quad (6)$$

Overall, the two-stage decomposition produces multiple intrinsic components for each original signal, with reduced noise and mode mixing. All decomposed components are then collected and reconstructed into a new input sequence. Concretely, each original signal yields several component time series. We assemble these components as separate channels of a multi-channel time series. This reconstructed sequence retains the key multi-scale features of the original data, with each channel corresponding to one intrinsic mode or trend. This multi-channel sequence serves as the input to the parallel feature extraction module.

3.2.2 Parallel Feature Extraction and Forecasting. After data preprocessing and VMD-CEEMDAN decomposition, the raw signals are input to the CNN and Transformer modules for parallel processing. The two modules extract temporal features from different scales to enhance the characterization ability through local convolution operation and global self-attention mechanism, respectively which are implemented as follows:

(1) The convolutional neural network performs local feature extraction on the input sequence $X \in \mathbb{R}^{L \times d}$ through N convolutional kernels of size $k \times d$ and step size s , and outputs the feature matrix

$Y^{CNN} \in \mathbb{R}^{L' \times N}$. The operation of each convolutional kernel at the temporal position is:

$$y_i(t) = \sum_{j=1}^k W_i(j) \cdot x(t+j-1) + b_i. \quad (7)$$

The outputs of all the convolution kernels are further spliced into a localized feature matrix:

$$Y^{CNN} \in \mathbb{R}^{L' \times N}. \quad (8)$$

(2) The Transformer module performs hierarchical feature abstraction of the input sequence X through non-linear transformations, generating a high-dimensional embedding matrix $Z \in \mathbb{R}^{L \times d_{model}}$. It captures complex global dependencies via multi-head attention, enabling dynamic feature integration and context-aware representation enhancement. The global feature matrix is obtained after further multi-head output splicing and mapping:

$$Y^{Transformer} = \text{Concat}(\text{Attention}_1, \dots, \text{Attention}_h) W^O \in \mathbb{R}^{L \times d_{model}}. \quad (9)$$

After the outputs of the two modules are interpolated and aligned with the fully connected layer to adjust the dimensionality, the feature fusion is accomplished by element-wise addition. The process of feature fusion is as follows:

First, the CNN output $Y^{CNN} \in \mathbb{R}^{L' \times N}$ is linearly interpolated to extend the time dimension to L to obtain $\tilde{Y}^{CNN} \in \mathbb{R}^{L \times N}$, and then the feature dimension is adjusted to d_{model} by the fully connected layer:

$$\tilde{Y}_{adj}^{CNN} = \tilde{Y}^{CNN} W_{adj} \in \mathbb{R}^{L \times d_{model}}. \quad (10)$$

Then, in the feature fusion stage, the local feature matrix \tilde{Y}_{adj}^{CNN} of the CNN adjusted by interpolation and fully connected layer is summed element by element with the global feature matrix $Y^{Transformer}$ output by the Transformer module

$$Y^{Fusion} = \tilde{Y}_{adj}^{CNN} + Y^{Transformer} \in \mathbb{R}^{L \times d_{model}}. \quad (11)$$

The fused features are then passed to a Transformer decoder to generate the forecasted $PM_{2.5}$ sequence. The Transformer decoder uses masked self-attention to refine the output sequence and encoder-decoder attention to incorporate the fused encoder features. In the encoder-decoder attention, each output position attends to the fused feature representation, aligning the forecast with the input patterns. A final linear projection layer then maps the decoder outputs to the predicted $PM_{2.5}$ concentration values for each future time step. In this way, the model produces a multi-step prediction of future $PM_{2.5}$ levels. Overall, this architecture adaptively learns complex nonlinear temporal patterns and integrates multi-scale information from the input signals, enabling accurate short-term $PM_{2.5}$ forecasting.

4 Experiment and Analysis

4.1 Experimental preparation and design

4.1.1 Research Objects. Seven representative cities were selected for analysis in this study, as shown in Table 1: Shenyang, Beijing, Hangzhou, Nanning, Xi'an, Changsha and Kunming. These cities have a wide geographic distribution, covering a wide range of topographic features from plateaus, plains, and basins and most climate types. According to the national ambient air quality status

in 2024 published by the Ministry of Ecology and Environment, the compliance rate of ambient air quality in Kunming is 99.7%, which is ranked as the first among 168 key cities in China. In this context, Kunming was selected as the main city of the experiment for the research study.

4.1.2 Data sources. In this study, data on the five most common components of the atmosphere that have a significant impact on air quality: $PM_{2.5}$, CO , NO_2 , SO_2 , PM_{10} , and the air quality index (AQI) were obtained from the real-time national urban air quality dissemination platform of the China Environmental Monitoring General Station. In addition, meteorological information of the same day was obtained from the public weather website <https://rp5.ru/> as an auxiliary factor. In this study, the ratio of 7:3 is used to divide the training set and test set, and the time periods are allocated based on the time distribution characteristics of the fully populated data.

4.1.3 Correlation Analysis. Pearson's correlation coefficient was employed to quantify the linear association between any two variables. Given paired observations $\{(x_i, y_i)\}_{i=1}^n$, the coefficient is defined as

$$\rho_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (12)$$

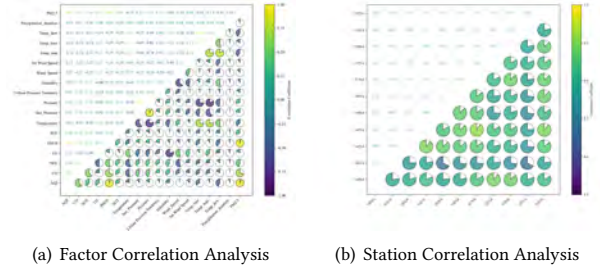


Figure 2: Correlation Analysis

(1) **Factor- $PM_{2.5}$ Correlation.** To elucidate the influence of potential predictors, Pearson correlation coefficients were computed between $PM_{2.5}$ concentrations and eighteen candidate factors. As illustrated in 2(a), temperature, barometric pressure, wind speed, dew-point temperature, and the three-hour pressure tendency exhibit the most pronounced correlations. These variables therefore constitute the core feature set for subsequent forecasting experiments.

(2) **Inter-Station Correlation.** Air-quality dynamics are modulated by local geography and micro-meteorological conditions. We therefore analysed pairwise correlations among the eleven monitoring stations in Kunming (2(b)). All coefficients exceed 0.65, indicating highly consistent temporal patterns across sites. Such spatial consistency likely arises from shared emission sources and homogeneous synoptic conditions within the metropolitan area.

Given the strong inter-station similarity, station-wise $PM_{2.5}$ readings were averaged to obtain a single representative series for each city. This aggregation alleviates computational burden without compromising the fidelity of regional pollution trends.

Table 1: Analysis of the Selected Research Area

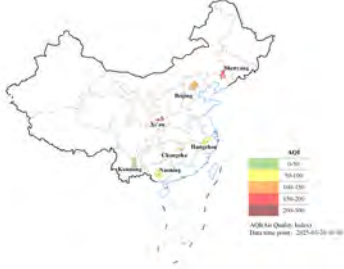
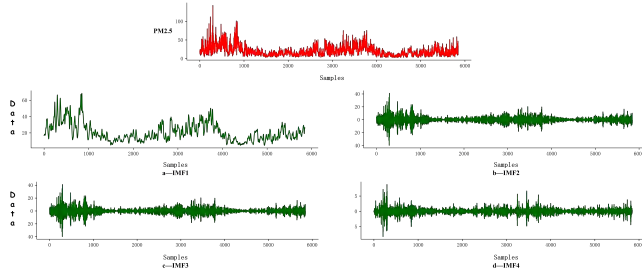
Display	City	Region (China)	Climate Zone	Salient Characteristics
	Shenyang	Northeast	Temperate monsoon	Heavy-industry base; severe winter cold; high coal-fired heating demand
	Beijing	North	Temperate monsoon	National capital; dense population; strong regional transport and policy control
	Hangzhou	East	Sub-tropical monsoon	Core YRD economy; large industrial & transport load; frequent stagnation episodes
	Nanning	South	Tropical monsoon	Rapidly urbanizing hot-humid city; influenced by trans-boundary ASEAN pollutants
	Xi'an	Northwest	Temperate continental	Basin terrain restricts mixing; spring dust events plus local coal/industrial sources
	Changsha	Central	Sub-tropical monsoon	Central-region growth pole; diversified sources with rising industrial VOC load
	Kunming	Southwest	Plateau-mountain monsoon	High-altitude city with low background PM but elevated ozone potential

Table 2: Center Frequencies under Different Parameters K

K	Center Frequency/HZ							
1	0.000353	-	-	-	-	-	-	-
2	0.000351	0.247038	-	-	-	-	-	-
3	0.000338	0.123904	0.248938	-	-	-	-	-
4	0.000338	0.123899	0.248464	0.323582	-	-	-	-
5	0.000332	0.122688	0.150275	0.249144	0.327691	-	-	-
6	0.000199	0.031398	0.124395	0.247767	0.286665	0.366451	-	-
7	0.000195	0.030684	0.123273	0.151879	0.24854	0.29072	0.368709	-
8	0.000195	0.030655	0.123278	0.151836	0.248478	0.288736	0.361697	0.452133

**Figure 3: Result of VMD Decomposition**

4.2 Analysis of VMD Decomposition Results

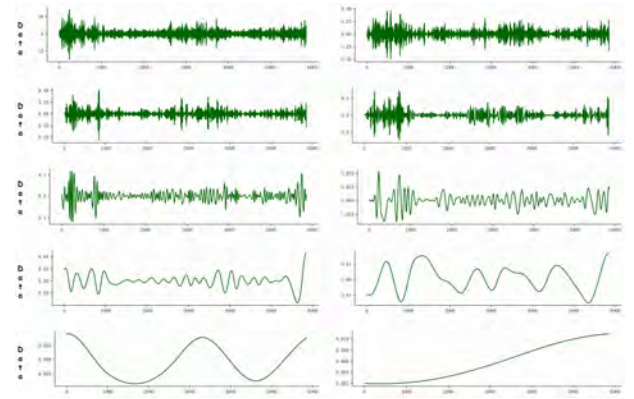
The original $PM_{2.5}$ concentration series exhibits strong volatility. To reduce the impact of this volatility on prediction results, this study first employed VMD to decompose the $PM_{2.5}$ concentration. Considering that the value of the VMD parameter K significantly affects model complexity, this study performed decompositions by setting K to 1, 2, 3, ..., 8, respectively. The center frequencies for each K value were recorded, as shown in Table 2. It was observed that when $K=4$, central modes with similar frequencies appeared, indicating an over-decomposition phenomenon. Therefore, $K=4$ and $\alpha = 800$ were selected as the VMD parameters. Using these parameters, VMD decomposed the original series into 4 IMF components, each with a distinct frequency. Figure 3 illustrates the results of the VMD decomposition. Compared to the original series, all decomposed IMF components showed a certain degree of reduction in both volatility and complexity. However, it was also noted that as the decomposition frequency of the components increased, their volatility also tended to increase.

To address this issue, the Sample Entropy (SampEn) of each component series was calculated to evaluate its complexity. This helped identify the component with the highest instability. A secondary

decomposition was then performed on this most unstable component to enhance data effectiveness and improve prediction accuracy. Table 3 presents the results of the Sample Entropy calculations.

Table 3: Calculation of Sample Entropy

IMF	Sample Entropy Value
IMF_1	0.150190988
IMF_2	0.501608364
IMF_3	0.532750198
IMF_4	0.796696871

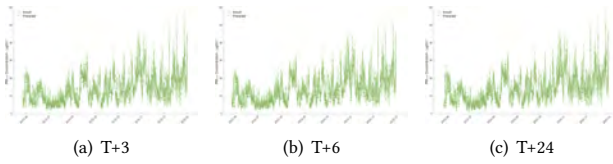
**Figure 4: Result of CEEMDAN Decomposition**

Observation reveals that IMF_4 exhibits the highest complexity. Therefore, CEEMDAN is employed for its secondary decomposition. The standard deviation threshold is set to 0.005, and the number of iterations to 200, while other settings adopt their default values. The results are illustrated in Figure 4. The CEEMDAN algorithm decomposes this IMF_4 sequence into 10 IMF components, labeled U1 to U10, with distinct frequencies. A comparison with the aforementioned IMF_1 sequence reveals that the frequencies and complexities of these new components progressively decrease, exhibiting more stable periodicity. Modes U1–U5 possess higher frequencies and undergo relatively rapid changes. Their fluctuations primarily manifest as short-term, high-frequency random components, which can significantly reflect the local details and random disturbances of the original sequence. Conversely, modes U6–U10 have lower

Table 4: Performance of Deep Learning Models on Kunming City Data

Model	T+3					T+6					T+24				
	R^2	MSE	RMSE	MAE	MAPE	R^2	MSE	RMSE	MAE	MAPE	R^2	MSE	RMSE	MAE	MAPE
XGBOOST	66.30	22.74	4.77	3.72	40.58	56.41	29.37	5.42	4.21	45.55	39.49	40.58	6.37	5.01	53.39
CEEMDAN-CNN-Transformer	91.11	6.33	2.51	1.96	13.99	86.47	9.65	3.10	2.40	15.57	82.70	12.35	3.51	2.82	18.80
CNN-Transformer	80.27	13.30	3.65	2.73	23.42	78.79	14.31	3.78	2.92	26.67	74.10	17.50	4.18	3.23	27.04
Transformer	77.89	14.92	3.86	2.88	24.53	70.13	20.15	4.49	3.45	31.03	66.72	22.48	4.74	3.68	31.67
Ours	95.88	2.78	1.66	1.23	9.99	94.06	4.01	2.00	1.55	11.76	93.66	4.28	2.06	1.57	13.72

frequencies, change more slowly, and display weaker volatility. Specifically, U6–U8 might capture seasonal variations or cyclical trends, whereas U9–U10 are more likely to represent the overall evolutionary trend of the sequence.

**Figure 5: Model Performance in Predicting Kunming at Different Time Horizons**

In this study, the XGBoost, CEEMDAN-CNN-Transformer, CNN-Transformer, and Transformer models were employed for comparison. This comparison aimed to validate the predictive performance of the proposed PolluVCCT model. All models were applied to prediction tasks at three time horizons: $T + 3$, $T + 6$, and $T + 24$. Five evaluation metrics were used to measure the performance differences among these models. The final results are presented in Table 4. Observing the PolluVCCT model individually in Figure 5, at the $T + 3$ horizon, it achieved an R^2 of 95.88%, an MSE of 2.78, an RMSE of 1.66, an MAE of 1.23, and a MAPE of 9.99%. As the prediction sequence length increased, the model’s accuracy experienced a slight decline. However, the R^2 value consistently remained above 90%. Overall, the PolluVCCT model demonstrated substantially superior predictive performance compared to the other models in this experimental task. Regarding model architecture comparisons, ablation study results indicated that each component contributed synergistically to enhancing predictive performance. When the VMD preprocessing module was removed, the RMSE for $T + 24$ prediction increased by 70.39%. This confirms the advantage of the VMD algorithm in reducing the non-stationarity of the original sequence. The MAE of the CNN-Transformer combination model at the $T + 6$ stage increased by 87.74% compared to the complete PolluVCCT model. This reveals the importance of the multi-scale feature fusion mechanism for medium-to-long-term predictions. In contrast, the standalone Transformer architecture exhibited significant performance degradation in $T + 24$ prediction, with its MAPE value reaching 31.67%. This validates its limitations in modeling long-term temporal dependencies. The hybrid architecture proposed in this study effectively mitigates this deficiency by incorporating a CNN for local feature extraction.

Compared to traditional machine learning methods, the XGBoost model performed significantly worse across all three prediction horizons. Particularly for the $T + 24$ task, its R^2 value was only 39.49%, and its MAPE exceeded 53%. The PolluVCCT model achieved an

81.6% lower MAPE than XGBoost. This indicates that the proposed composite model significantly outperforms traditional machine learning models in this specific task.

To validate the generalization ability of the model and the reliability of the experimental conclusions, this study conducted identical experimental tasks on $PM_{2.5}$ data from other regions. The environmental conditions and parameter settings were kept consistent throughout these experiments.

The specific results are presented in Table 5. As shown, among the six cities, the PolluVCCT model demonstrated significantly superior predictive performance compared to the other models in most cases. The comparative performance against other models was largely consistent with the results observed for Kunming city. These findings indicate that the model proposed in this study consistently exhibits superior performance in $PM_{2.5}$ prediction tasks across different cities and even different regions. This demonstrates the model’s excellent generalization capability. Consequently, it holds potential for future application in predicting $PM_{2.5}$ concentrations in other urban areas.

5 Conclusion and Future Work

This study proposes the PolluVCCT, a hybrid decomposition and deep learning framework, aiming to effectively address the issues of high noise levels, significant nonlinearity, and inherent multi-scale characteristics in $PM_{2.5}$ concentration time series. A large number of experiments conducted in seven cities in China, each with a distinct climate type and representing a representative sample, have shown that PolluVCCT consistently outperformed state-of-the-art baselines in terms of accuracy, robustness, and cross-regional generalization. Specifically, its parallel CNN-Transformer backbone successfully balanced local feature extraction with global trend modeling, while the two-stage VMD-CEEMDAN decomposition mitigated non-stationarity and enhanced sensitivity to both abrupt local variations and long-range dynamics. Consequently, compared with the best competing methods, PolluVCCT reduced RMSE by nearly 3% and increased R^2 by 4%–10%.

While PolluVCCT shows strong empirical performance, three extensions warrant exploration: First, incorporating heterogeneous data sources—for example, satellite imagery, traffic flow, or meteorological forecasts—may further enrich the model’s representation capacity. Second, pursuing model compression and architecture-level pruning could facilitate deployment on resource-constrained edge devices. Third, cross-regional transfer-learning strategies merit investigation to alleviate data scarcity in under-monitored areas. Advancing these directions could expand PolluVCCT’s applicability in environmental science and sustainable development, strengthening its role in evidence-based air quality governance.

References

- [1] Erik Altman, Jovan Blanuša, Luc Von Niederhäusern, Béni Egressy, Andreea Anghel, and Kubilay Atas. 2023. Realistic synthetic financial transactions for anti-money laundering models. *Advances in Neural Information Processing Systems*, 36, 29851–29874.
- [2] Lu Bai, Jianzhou Wang, Xuejiao Ma, and Haiyan Lu. 2018. Air pollution forecasts: an overview. *International journal of environmental research and public health*, 15, 4, 780.
- [3] Marleen Balvert. 2024. Iterative rule extension for logic analysis of data: an milp-based heuristic to derive interpretable binary classifiers from large data sets. *INFORMS Journal on Computing*, 36, 3, 723–741. doi:10.1287/IJOC.2021.0284.
- [4] Abdellatif Bekkar, Badr Hssina, Samira Douzi, and Khadija Douzi. 2021. Air-pollution prediction in smart city, deep learning approach. *Journal of big Data*, 8, 1–21.
- [5] Junqian Cao, Huan Li, and Xinyue Mo. 2025. Ensemble forecasting of ozone concentration based on deep learning and metaheuristic algorithms. *Computer Measurement & Control*, 1–18.
- [6] Zhongping Cui, Xiaomin Liu, Shuang Lu, and Yu Liu. 2024. Dynamic comprehensive evaluation of the development level of china's green and low carbon circular economy under the double carbon target. *Polish Journal of Environmental Studies*, 33, 1.
- [7] Rui Dong, Raymond Fisman, Yongxiang Wang, and Nianhang Xu. 2021. Air pollution, affect, and forecasting bias: evidence from chinese financial analysts. *Journal of Financial Economics*, 139, 3, 971–984.
- [8] Mingjiang Duan, Tongya Zheng, Yang Gao, Gang Wang, Zunlei Feng, and Xinyu Wang. 2024. Dga-gnn: dynamic grouping aggregation gnn for fraud detection. In *Proceedings of the AAAI conference on artificial intelligence*. Vol. 38, 11820–11828.
- [9] Yueyi Feng, Miao Ning, Wenbo Xue, Miaomiao Cheng, and Yu Lei. 2023. Developing china's roadmap for air quality improvement: a review on technology development and future prospects. *Journal of Environmental Sciences*, 123, 510–521.
- [10] Ihsane Gryech, Chaimae Asaad, Mounir Ghogho, and Abdellatif Kobbane. 2024. Applications of machine learning & internet of things for outdoor air pollution monitoring and prediction: a systematic literature review. *Engineering Applications of Artificial Intelligence*, 137, 109182.
- [11] Kailin Guo, Ruiju Zhang, Jian Wang, Haibo Li, Dong Li, Cai Chen, and Hua Zhong. 2022. Regional pm_{2.5} concentration prediction combining densenet and convlstm. *Bulletin of Surveying and Mapping*, 12, 55.
- [12] N Srinivasa Gupta, Yashvi Mohta, Khyati Heda, Raahil Armaan, B Valarmathi, and G Arulkumar. 2023. Prediction of air quality index using machine learning techniques: a comparative analysis. *Journal of Environmental and Public Health*, 2023, 1, 4916267.
- [13] Marviola Hardini, Mochamad Heru Riza Chakim, Lena Magdalena, Hiroshi Kenta, Ageng Setiani Rafika, and Dwi Julianingsih. 2023. Image-based air quality prediction using convolutional neural networks and machine learning. *Aptisi Transactions on Technopreneurship (ATT)*, 5, 1Sp, 109–123.
- [14] Jennifer Jie Li, Massimo Massa, Hong Zhang, and Jian Zhang. 2021. Air pollution, behavioral bias, and the disposition effect in china. *Journal of Financial Economics*, 142, 2, 641–673.
- [15] Kaidi Li et al. 2024. Sefraud: graph-based self-explainable fraud detection via interpretative mask learning. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 5329–5338.
- [16] Yun-Chia Liang, Yona Maimury, Angela Hsiang-Ling Chen, and Josue Rodolfo Cuevas Juarez. 2020. Machine learning-based prediction of air quality. *applied sciences*, 10, 24, 9151.
- [17] Tanisha Madan, Shrddha Sagar, and Deepali Virmani. 2020. Air quality prediction using machine learning algorithms—a review. In *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*. IEEE, 140–145.
- [18] Dongming Qin, Zhijun Ding, Yupeng Jin, and Qin Zhao. 2019. Air pollutant concentration prediction based on autoencoder networks. *Journal of Tongji University (Natural Science Edition)*, 47, 5, 681–687.
- [19] Nikhil Sharma, Avinash Kumar Agarwal, Peter Eastwood, Tarun Gupta, and Akhilendra Pratap Singh. 2018. Introduction to air pollution and its control. *Air Pollution and Control*, 3–7.
- [20] Fei Wang et al. 2025. Short-term exposure to pm_{2.5} and high pollution events on depressive symptoms among adolescents. *Journal of Hazardous Materials*, 492, 138131.
- [21] Qianyu Wang, Wei-Tek Tsai, and Tianyu Shi. 2025. Graphalm: active learning for detecting money laundering transactions on blockchain networks. *IEEE Network*, 39, 2, 294–303.
- [22] Shanshan Wang and Erick Delage. 2024. A column generation scheme for distributionally robust multi-item newsvendor problems. *INFORMS Journal on Computing*, 36, 3, 849–867. doi:10.1287/IJOC.2022.0010.
- [23] Ailan Xu, Zaifeng Zhang, Qiang Sun, Yanmin Zhu, Xiaoyan Peng, and Xiangxiang Yu. 2021. Construction of an ai air quality forecasting system based on deep learning. *China Environmental Monitoring*, 37, 2, 89–95.
- [24] Vibha Yadav, Amit Kumar Yadav, Vedant Singh, and Tej Singh. 2024. Artificial neural network an innovative approach in air pollutant prediction for environmental applications: a review. *Results in Engineering*, 102305.
- [25] Yuning You, Tianlong Chen, Zhangyang Wang, and Yang Shen. 2023. Graph domain adaptation via theory-grounded spectral regularization. In *The eleventh international conference on learning representations*.
- [26] Dazhou Yu, Xiaoyun Gong, Yun Li, Meikang Qiu, and Liang Zhao. 2024. Self-consistent deep geometric learning for heterogeneous multi-source spatial point data prediction. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 4001–4011.
- [27] Qian Yu, Hong-wu Yuan, Zhao-long Liu, and Guo-ming Xu. 2024. Spatial weighting emd-lstm based approach for short-term pm_{2.5} prediction research. *Atmospheric Pollution Research*, 15, 10, 102256.
- [28] Xin Yu, Rongye Shi, Pu Feng, Yongkai Tian, Simin Li, Shuhao Liao, and Wenjun Wu. 2024. Leveraging partial symmetry for multi-agent reinforcement learning. In *AAAI 2024*, 17583–17590. doi:10.1609/AAAI.V38I16.29709.
- [29] Bo Zhang, Yunjie Lu, Dongming Qin, and Guojian Zou. 2022. A convolutional autoencoder deep-learning model for joint multi-site air pollution prediction. *Acta Electronica Sinica*, 50, 6, 1410–1427.
- [30] Bo Zhang, Yi Rong, Ruihan Yong, Dongming Qin, Maozhen Li, Guojian Zou, and Jianguo Pan. 2022. Deep learning for air pollutant concentration prediction: a review. *Atmospheric Environment*, 290, 119347.
- [31] Zhendong Zhang, Yongkang Zeng, and Ke Yan. 2021. A hybrid deep learning technology for pm_{2.5} air quality forecasting. *Environmental Science and Pollution Research*, 28, 39409–39422.
- [32] Zhenhua Zhang, Guoxing Zhang, and Lili Li. 2023. The spatial impact of atmospheric environmental policy on public health based on the mediation effect of air pollution in china. *Environmental Science and Pollution Research*, 30, 55, 116584–116600.
- [33] Bin Zhao, Shuxiao Wang, and Jiming Hao. 2024. Challenges and perspectives of air pollution control in china. *Frontiers of Environmental Science & Engineering*, 18, 6, 68.
- [34] Qinghe Zheng et al. 2024. Application of complete ensemble empirical mode decomposition based multi-stream informer (ceemd-msi) in pm_{2.5} concentration long-term prediction. *Expert Systems with Applications*, 245, 123008.
- [35] Shiyun Zhou, Wei Wang, Long Zhu, Qi Qiao, and Yulin Kang. 2024. Deep-learning architecture for pm_{2.5} concentration prediction: a review. *Environmental Science and Ecotechnology*, 21, 100400.

Appendix

A Results of Generalization Experimental

Table 5: Performance of Deep Learning Models on Cities’ Data

City	Model	T+3					T+6					T+24				
		R^2	MSE	RMSE	MAE	MAPE	R^2	MSE	RMSE	MAE	MAPE	R^2	MSE	RMSE	MAE	MAPE
Beijing	XGBOOST	50.46	333.76	18.27	14.43	152.93	45.85	365.01	19.11	15.06	159.22	26.89	494.32	22.23	17.32	185.97
	CEEMDAN-CNN-Transformer	90.95	81.94	9.05	6.99	56.62	90.09	89.76	9.47	7.85	80.53	84.44	141.01	11.87	9.02	92.00
	CNN-Transformer	85.08	100.51	10.03	7.55	60.06	84.16	107.12	10.35	7.34	55.93	83.05	111.01	10.54	7.59	58.94
	Transformer	80.35	132.40	11.51	8.03	57.92	79.35	139.64	11.82	8.31	64.33	74.11	174.48	13.21	9.37	74.16
	Ours	93.98	41.05	6.41	4.72	36.04	90.12	67.38	8.21	6.40	54.39	88.58	78.04	8.83	6.86	56.68
Hangzhou	XGBOOST	66.04	97.09	9.85	7.93	61.88	57.42	121.65	11.03	8.71	66.99	33.20	190.43	13.80	10.65	80.74
	CEEMDAN-CNN-Transformer	91.46	30.58	5.53	4.17	25.35	89.64	37.10	6.09	4.83	28.45	70.76	20.89	4.57	3.97	33.77
	CNN-Transformer	81.05	54.17	7.36	4.93	27.85	79.24	59.44	7.71	5.77	37.65	78.99	60.15	7.75	5.74	32.94
	Transformer	71.05	82.78	9.10	6.90	40.14	61.12	111.29	10.55	7.86	48.36	59.41	116.07	10.77	8.30	53.71
	Ours	97.95	5.84	2.42	1.82	10.71	96.97	8.64	2.94	2.36	14.40	95.15	13.89	3.73	3.06	19.79
Nanning	XGBOOST	69.29	101.41	10.07	7.71	48.12	64.08	118.16	10.87	8.27	51.11	52.22	155.64	12.48	9.49	58.89
	CEEMDAN-CNN-Transformer	92.53	27.31	5.23	4.07	18.34	90.32	35.37	5.95	4.73	23.16	89.90	36.92	6.08	4.47	17.30
	CNN-Transformer	87.53	41.17	6.42	4.61	23.56	86.53	44.50	6.67	4.92	27.65	85.66	47.43	6.89	4.98	30.66
	Transformer	84.66	50.64	7.12	5.31	32.59	76.92	76.26	8.73	6.55	37.94	69.43	101.10	10.06	7.34	45.41
	Ours	97.58	7.96	2.82	2.23	13.98	96.55	11.39	3.37	2.51	14.83	96.23	12.45	3.53	2.55	13.78
Shenyang	XGBOOST	55.22	143.87	11.99	10.01	76.78	44.15	179.43	13.40	11.06	83.70	34.88	209.60	14.48	11.03	75.90
	CEEMDAN-CNN-Transformer	85.75	57.12	7.56	5.88	33.45	64.12	143.99	12.00	10.40	58.92	61.75	153.73	12.40	10.19	49.54
	CNN-Transformer	69.30	98.62	9.93	7.67	44.28	68.23	102.06	10.10	8.18	51.99	67.84	103.54	10.18	7.95	45.93
	Transformer	68.63	100.76	10.04	7.87	45.61	67.04	105.87	10.29	7.89	44.84	59.93	154.42	12.42	9.93	56.48
	Ours	93.99	19.30	4.39	3.23	18.45	92.73	23.34	4.83	3.71	21.71	90.09	31.91	5.65	4.66	29.15
Xi'an	XGBOOST	61.76	251.29	15.85	13.97	75.06	56.81	283.58	16.84	14.68	77.98	40.10	393.27	19.83	16.74	86.90
	CEEMDAN-CNN-Transformer	89.73	82.97	9.11	6.98	21.56	89.12	87.88	9.37	7.44	24.65	81.70	147.89	12.16	9.30	25.12
	CNN-Transformer	87.18	84.26	9.18	6.94	28.47	82.14	117.46	10.84	7.86	28.76	79.24	136.98	11.70	8.62	31.34
	Transformer	77.73	146.43	12.10	8.70	29.80	76.94	151.55	12.31	9.13	33.66	65.62	226.78	15.06	11.12	41.44
	Ours	96.22	24.80	4.98	3.73	13.16	95.33	30.70	5.54	4.27	16.88	93.00	46.19	6.80	5.73	29.60
Changsha	XGBOOST	61.03	364.25	19.09	15.29	80.57	56.54	405.89	20.15	16.01	83.38	42.70	533.27	23.09	17.72	90.68
	CEEMDAN-CNN-Transformer	87.19	133.85	11.57	8.82	29.46	86.28	143.40	11.97	8.89	25.68	75.37	257.57	16.05	11.96	31.95
	CNN-Transformer	81.94	168.84	12.99	9.14	37.84	80.77	179.80	13.41	10.30	53.15	72.21	260.44	16.14	11.76	51.42
	Transformer	76.09	224.07	14.97	10.34	42.67	67.53	303.56	17.42	11.75	46.57	62.76	348.07	18.66	13.02	51.72
	Ours	95.85	38.75	6.23	5.05	29.00	95.21	44.70	6.69	5.45	22.50	94.34	53.03	7.28	5.08	22.21