frontiers | Frontiers in Environmental Science

TYPE Original Research
PUBLISHED 17 March 2025
DOI 10.3389/fenvs.2025.1549209

OPEN ACCESS

EDITED BY
Chen Siyu,
Lanzhou University, China

REVIEWED BY
Chunsong Lu,
Nanjing University of Information Science and Technology, China
Yang Yang,
Nanjing University of Information Science and Technology, China

*CORRESPONDENCE
Xiaolong Chen,
chen_xiaolong123@126.com
Hongfeng Zhang,
hfengzhang@mpu.edu.mo

RECEIVED 24 December 2024
ACCEPTED 13 February 2025
PUBLISHED 17 March 2025

CITATION
Qin Z, Wei B, Gao C, Chen X, Zhang H and In Wong CU (2025) SFDformer: a frequency-based sparse decomposition transformer for air pollution time series prediction.
Front. Environ. Sci. 13:1549209.
doi: 10.3389/fenvs.2025.1549209

SFDformer: a frequency-based sparse decomposition transformer for air pollution time series prediction

Zhenkai Qin [1,2], Baozhong Wei[1], Caifeng Gao[1], Xiaolong Chen[3]*, Hongfeng Zhang[3]* and Cora Un In Wong[3]

[1]School of Information Technology, Guangxi Police College, Nanning, China, [2]School of Computer Science and Artificial Intelligence, Southwest Jiaotong University, Chengdu, China, [3]Faculty of Humanities and Social Sciences, Macao Polytechnic University, Macao, China

Introduction: With the rapid advancement of industrialization and the prevalent occurrence of haze weather, $PM_{2.5}$ contamination has emerged asa significant threat to public health and environmental sustainability. The concentration of $PM_{2.5}$ exhibits intricate dynamic attributes and is profoundly correlated with meteorological conditions as well as the concentrations of other pollutants, thereby substantially augmenting the complexity of predictive endeavors.

Methods: A novel predictive methodology has been developed. It integrates time seriesfrequency domain analysis with the decomposition of deep learning models. This approach facilitates the capture of interdependencies among high - dimensional features through time series decomposition, employs Fourier Transform to mitigate noise interference, and incorporates sparse attention mechanisms to selectively filter critical frequency components, thereby enhancing time - dependent modeling. Importantly, this technique effectively reduces computational complexity from $O(L^2)$ to $O(L \log L)$.

Results: Empirical findings substantiate that this methodology yields notably superior predictive accuracy relative to conventional models across a diverse array of real-world datasets.

Discussion: This advancement not only offers an efficacious resolution for $PM_{2.5}$ prediction tasks but also paves the way for innovative research and application prospects in the realm of complex time series modeling.

KEYWORDS

frequency, time series forecasting, air pollution, transformer, sparse attention mechanism

1 Introduction

Over the past several decades, the rapid pace of industrialization has precipitated the frequent occurrence of smog, thereby intensifying environmental pollution. Fine particulate matter ($PM_{2.5}$), characterized by particles with a diameter of 2.5 μm or less, has emerged as a pivotal pollutant that poses considerable risks to human health. As indicated by the World Health Organization (WHO), nearly 90% of the global populace inhales air that surpasses its quality standards, rendering $PM_{2.5}$ a primary contributor to respiratory ailments (Ailshire and Crimmins, 2014; Pöschl, 2005). Additionally, short-term exposure to $PM_{2.5}$ (spanning

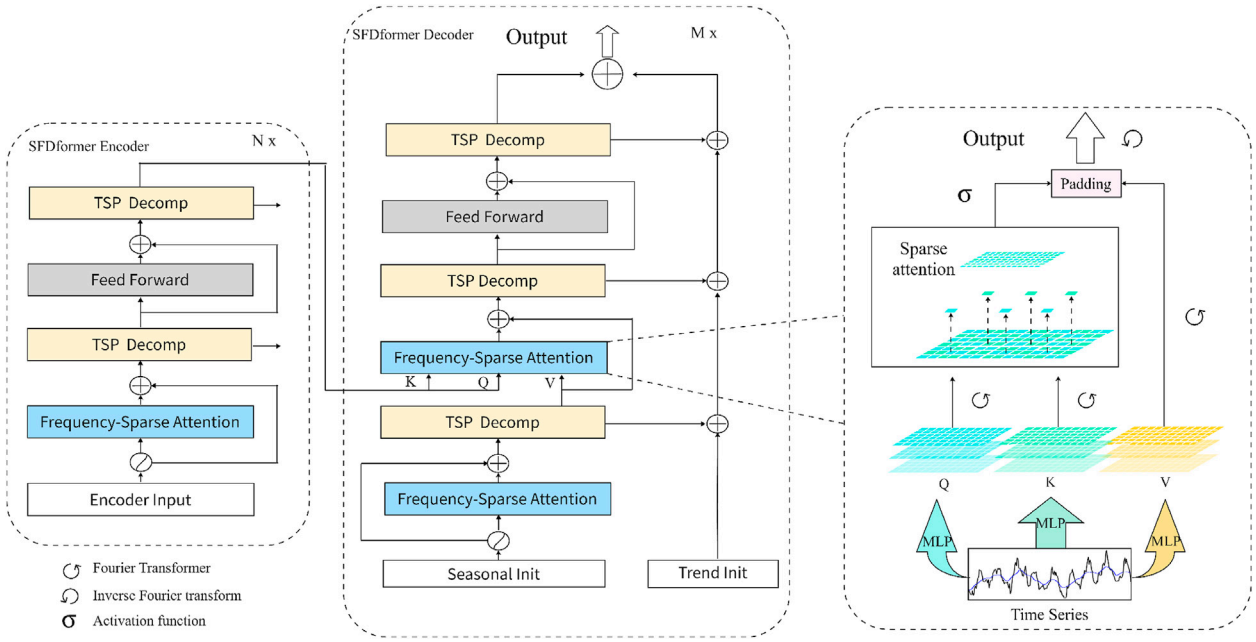

FIGURE 1
Schematic overview of the proposed SFDformer method. Initially, the Sparse Frequency Decomposition Attention module (Frequency-Sparse Attention, blue block) is designed to perform frequency transformation and reduce model parameters by leveraging Fourier transform and sparse attention mechanisms. More specifically, the Fourier transform converts time-domain data into the frequency domain to reduce the impact of noise, while sparse attention is employed to filter the critical frequency weight matrices. After that, the time series pooling decomposition (TSP Decomp, yellow block) method is utilized to extract seasonal and trend patterns from the input time series data.

from hours to weeks) has been associated with cardiovascular-related mortality and other health sequelae (Du et al., 2016). Beyond its ramifications for public health, deteriorating air quality imposes substantial economic burdens. A report by the Organization for Economic Cooperation and Development (OECD) (Lanzi, 2016) underscores that air pollution could lead to global GDP losses of up to 1%.

Developing an efficient air pollution monitoring and prediction system is, consequently, imperative for safeguarding human health and alleviating economic losses. Nonetheless, the formation mechanism of $PM_{2.5}$ is exceptionally intricate(Lu et al., 2021), encompassing complex interactions among various external pollutants, which markedly complicates the prediction process. Furthermore, air quality data exhibit a strong temporal dependence, constituting a prototypical time-series dataset with distinct periodic features. Predicting $PM_{2.5}$ concentrations constitutes a formidable task, necessitating the incorporation of meteorological factors (e.g., precipitation and temperature) and historical data (e.g., PM_{10}, SO_2) into time-series modeling. Extensive research has shown that these factors are highly correlated and have complex relationships in the formation of air pollution (Rakholia et al., 2024; He et al., 2017; Luo et al., 2020; Neiburger, 1969). Consequently, effectively discerning these complex interactions and integrating them into pollutant prediction models has emerged as a pivotal aspect in comprehending pollution mechanisms and improving prediction accuracy (Deng et al., 2024). To tackle the dynamic variations in pollutant concentrations and their intricate feature relationships, a plethora of modeling approaches have been suggested. Conventional statistical methods were extensively utilized in the initial phases of air quality prediction research. These methods predominantly depend on historical data for model training, employing frequently used techniques such as Autoregressive Moving Average (ARMA) (Liu and Yang, 2021) models and Autoregressive Integrated Moving Average (ARIMA) (Liu and Yang, 2021) models. However, as the volume and complexity of data have escalated, these methods have encountered difficulties in meeting the practical demands of real-time forecasting of pollutant concentrations due to prolonged training times and limited scalability.

The advent of deep learning technologies has led to the emergence of Transformer-based deep learning models as innovative solutions for tackling complex problems and enhancing performance. These models are particularly efficacious as they account for the temporal correlations inherent in pollutant concentration sequences. To date, deep learning models have demonstrated state-of-the-art capabilities in time-series prediction tasks. By capitalizing on the neural networks' ability to extract temporal features from time-series data, the precision of pollutant concentration predictions can be substantially improved. Empirical studies on air pollutant prediction have shown that deep learning models surpass traditional methods, including classical machine learning algorithms, by more effectively capturing high-dimensional feature dependencies and temporal patterns (Panneerselvam and Thiagarajan 2024). Nevertheless, conventional Transformer models encounter several challenges, particularly their substantial computational cost, which is especially significant when dealing with large-scale environmental

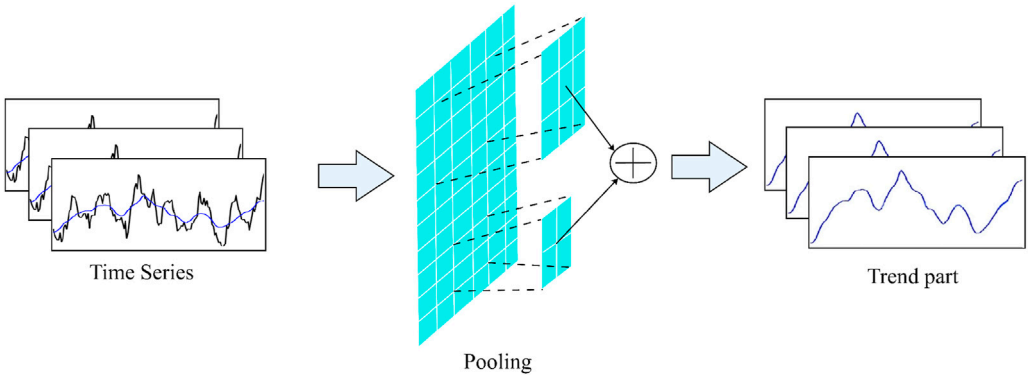

FIGURE 2
Time series pooling decomposition module.

datasets. The temporal continuity, dynamic fluctuations, and complex intercorrelations within pollutant concentration time-series data further complicate accurate prediction and analysis. Moreover, challenges such as noise, nonlinearity, and high-dimensional complexity inherent in environmental big data pose considerable obstacles for extracting temporal correlation information between pollutant concentrations and meteorological factors (Chen et al., 2024).

To tackle these challenges, this study introduces an end-to-end framework named Sparse Frequency Decomposition Transformer (SFDformer) for predicting the time series of pollutant concentrations. Figure 1 illustrates an overview of the proposed method. This approach uses time-series decomposition to capture the interdependencies among high-dimensional features and employs Fourier Transform to convert the data into the frequency domain, effectively reducing noise interference. The SFDformer integrates a sparse attention mechanism that selectively allocates weights to key frequency components, reducing the computational complexity from quadratic to linear time complexity. This design enhances computational efficiency while accurately extracting crucial features, providing a more accurate and efficient solution for forecasting pollutant concentrations. In summary, the main contributions of this paper are as follows.

- By fully considering the temporal dependencies in the time domain and the characteristic information in the frequency domain, a dual-domain modeling approach is used to accurately extract the complex correlation features between pollutant concentrations and meteorological data.
- We have introduced a frequency sparse attention mechanism based on Fourier transform, which combines sparse attention with Fourier transform to reduce the computational cost of self-attention layers and the impact of noise during the prediction process.
- In the issue of air pollution prediction, extensive experiments on eight real datasets have demonstrated the practicality and feasibility of the proposed model in $PM_{2.5}$ concentration forecasting. Furthermore, the results obtained in this work outperform other deep learning models reported in the literature.

2 Related work

The prediction of air pollutant concentrations is currently accomplished through two primary methodologies: physicochemical approaches and data-driven approaches. Physicochemical approaches entail the simulation and analysis of the physical and chemical processes that regulate air pollutants, employing fundamental physical and chemical principles to forecast pollutant behavior across diverse spatial and temporal scales (Thongthammachart et al., 2021; Kang et al., 2018; Hofman et al., 2022). Although these approaches can yield high prediction accuracy, they typically necessitate intricate model configurations and extensive parameter tuning, which may result in limited model generalization and diminished robustness in practical applications (Wang et al., 2020).

Emergence of meteorological stations and analogous monitoring devices, air quality monitoring stations, and meteorological satellites has enabled the gathering of data on air pollutant concentrations and meteorological conditions. This data provides strong support for research on air quality prediction (Gu et al., 2021; Kim et al., 2022). Data-driven methodologies have been increasingly employed to forecast air pollutant concentrations. In the nascent stages of air pollution prediction research, conventional machine learning models such as ARIMA and SARIMA were extensively utilized. These models forecast pollutant concentrations by examining the historical trends and periodic characteristics of time series data (Marvin et al., 2022). While these methods excel in modeling stationary time series and capturing short-term dependencies, they exhibit notable limitations when addressing complex nonlinear relationships and long-term sequence dependencies (Zhou et al., 2018). Specifically, the omission of high-frequency information in traditional machine models results in the loss of critical data, thereby constraining prediction accuracy and applicability. Furthermore, these methods encounter difficulties in leveraging multidimensional data (such as meteorological features and concentrations of other pollutants) to delineate more

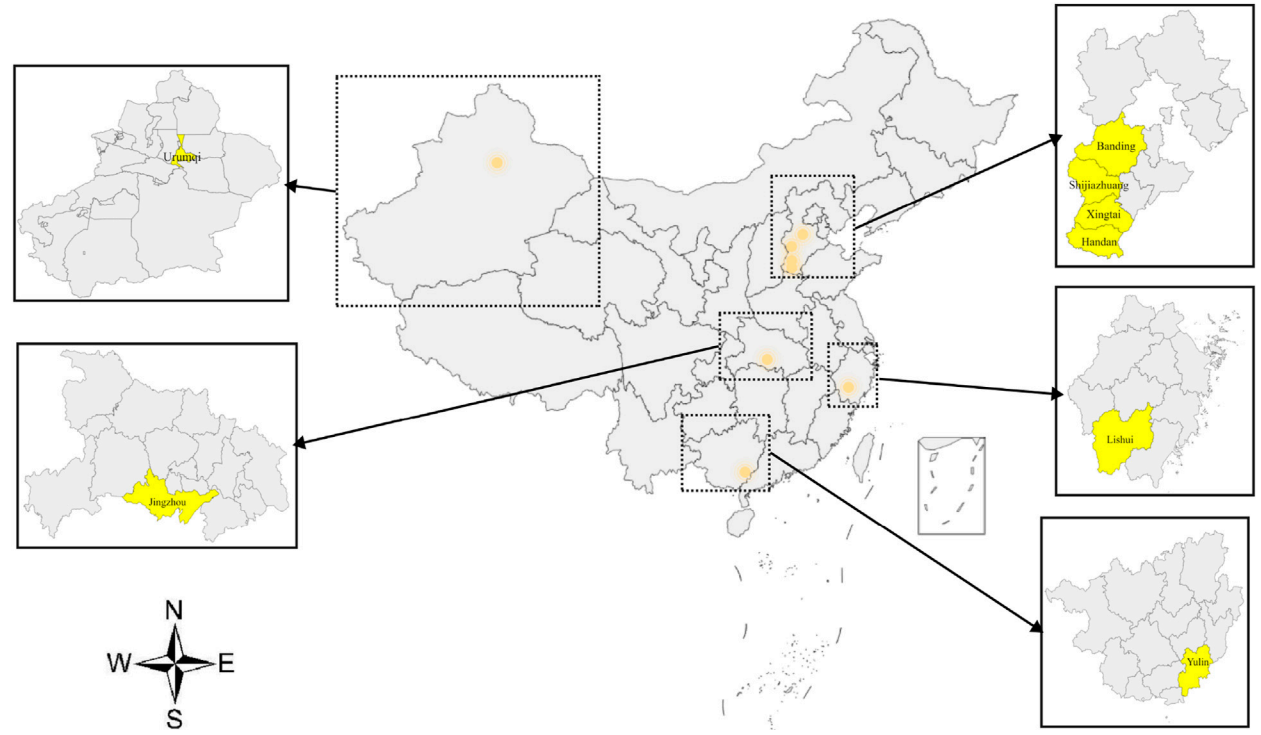

FIGURE 3
Geographical locations of cities.

comprehensive pollutant characteristics (Tagliabue et al., 2021). With advancements in data scale and computational power, machine learning methodologies have progressively emerged as more versatile options. Models such as Support Vector Regression (SVR), Random Forest (RF), and Multi-Layer Perceptron (MLP) have gained widespread adoption due to their efficacy in managing nonlinear relationships (Haq and Ahmad Khan, 2022; Rybarczyk and Zalakeviciute, 2018). These methodologies demonstrate superior predictive performance compared to traditional statistical methods by utilizing multidimensional data for modeling (Ma X. et al., 2023; Pan et al., 2023). However, they depend on manually crafted feature engineering, and their capacity to model the interdependencies of other multidimensional data influencing air pollutant concentrations remains limited (Zaini et al., 2022). Nonetheless, these methodologies have furnished valuable insights into air pollution prediction and established a foundation for investigating hybrid models that integrate traditional methods with deep learning technologies (Kshirsagar and Shah, 2022; Méndez et al., 2023).

The rapid advancement of deep learning technologies has led to significant breakthroughs in their application to time series forecasting, especially in the realm of air pollution prediction. In comparison to traditional statistical methods and classical machine learning techniques, deep learning models exhibit considerable advantages due to their robust ability to model non-linearity and precisely capture temporal dependencies. Recurrent Neural Networks (RNNs) and their sophisticated variants, such as Long Short-Term Memory Networks [LSTMs Han et al. (2023)] and Gated Recurrent Units (GRUs), have been extensively utilized to process time series data (Espinosa et al., 2021). These models adeptly capture long-term dependencies through memory units, effectively mitigating the challenges of vanishing and exploding gradients (Athira et al., 2018; Faraji et al., 2022). Nonetheless, individual models still possess certain limitations in modeling high-dimensional features (Sarkar et al., 2022). To further enhance the performance of air pollution time series forecasting, researchers have devised hybrid architectures, such as LSTM-CNN (Ghimire et al., 2019), LSTM-RNN (Ozcanli et al., 2020), and CNN-LSTM-RNN (Ko and Jung, 2022). These models amalgamate the strengths of distinct neural networks: LSTM-CNN extracts intricate features via CNNs while LSTM captures temporal dependencies, rendering it suitable for managing complex time series data; LSTM-RNN integrates RNN's capability to handle short-term dependencies with LSTM's capacity to capture long-term trends, making it ideal for data exhibiting both short-term fluctuations and long-term patterns; CNN-LSTM-RNN consolidates the advantages of CNNs, LSTMs, and RNNs, enabling it to process more intricate air pollution data scenarios. Despite these hybrid models demonstrating substantial performance improvements, they are accompanied by several limitations, such as elevated model complexity, extended training times, substantial hardware resource demands, and difficulties in hyperparameter tuning, which escalate optimization costs. Furthermore, the intricacy of these models often results in overfitting, particularly when data is limited or of inferior quality (Wang et al., 2022; Yuan et al., 2020).

To address these challenges, Transformer-based models have demonstrated exceptional performance in tackling the intricacies of

TABLE 1 Characteristic indicators of air pollution time prediction dataset. We utilize Indicator to represent various features within the air pollution dataset. The Indicator Definition elucidates the meaning of each feature, while the Corresponding Characteristics describe the specific attributes associated with these features.

Indicator	Indicator definition	Corresponding characteristics
Date	Air Quality Measurement Day	Temporal Characteristics
Quality Level	Air Quality Index Assessment	Assessment Characteristic
AQI	Pollutant Composite Index	Quality Characteristic
Daily AQI Ranking	Air Quality Trend	
PM_{10}	Air Pollutants	Harmful Substances in the Air
SO_2		
NO_2		
CO		
O_3		
$PM_{2.5}$		
Temperature	Meteorological Factors	Meteorological Characteristic
Wind speed		
Precipitation		

feature modeling, primarily due to their attention mechanism (Zhang and Zhang, 2023). However, conventional Transformer models typically exhibit high computational complexity and are susceptible to noise when managing high-dimensional dependencies (Guo and Mao, 2023). To alleviate these issues, researchers have introduced sparse attention mechanisms that concentrate on crucial dependencies, substantially reducing computational complexity to linear levels while maintaining robust global modeling capabilities (Al-qaness et al., 2023; Ma Z. et al., 2023). Considering that air pollutant concentrations frequently display significant seasonal variations influenced by meteorological factors, integrating time series decomposition and autocorrelation mechanisms can aid the model in better grasping the complex interdependencies among various features in the time series. Furthermore, frequency-domain enhancement techniques have substantially improved the overall performance and efficiency of the models by diminishing noise interference in long-term dependencies (Zeng et al., 2023). Inspired by these advancements, we propose the SFDformer method. This approach employs time series decomposition techniques to segregate the data into seasonal and trend components, effectively capturing factors such as air pollution, which are subject to seasonal fluctuations and trend variations. By employing Fourier transforms to transform time-domain data into frequency-domain data, we mitigate noise interference. The sparse attention mechanism further prioritizes essential frequency components and assigns them higher weights, enabling the model to capture critical short-term alterations while preserving vital long-term traits. This enhancement not only significantly boosts computational efficiency but also bolsters the model's stability and robustness in capturing the dependencies between high-dimensional features of air pollution concentrations, offering a more efficient and precise solution for intricate air pollution forecasting tasks.

3 Methodology

3.1 Background

The air pollution forecasting problem can be defined in a rolling prediction setting, where the future air quality over a given time horizon is predicted based on historical observations within a fixed-size window. At each time point t, the input sequence $\mathcal{X}^t = \{x_1^t, \ldots, x_{L_x}^t\}$ consists of observed values across multiple feature dimensions. The output sequence $\mathcal{Y}^t = \{y_1^t, \ldots, y_{L_y}^t\}$ predicts air quality indicators, such as concentrations of $PM_{2.5}$, PM_{10}, NO_2, *etc.*, over several future time points. This setup enables the model to predict multiple pollutants simultaneously, making it highly suitable for air quality monitoring and management applications. By providing such predictions, relevant authorities can take proactive measures to mitigate the impact of air pollution, thereby enhancing the quality of life for urban residents.

3.2 Time series pooling decomposition module

In real-world air pollution time series data, intricate seasonal patterns often intertwine with trend components, making them difficult to disentangle. Traditional fixed-window average pooling methods struggle to effectively capture such diverse temporal characteristics. To address this challenge, as depicted in Figure 2, we introduce a Time Series Pooling Decomposition Module (TSP Decomp), meticulously designed to tackle the complexities inherent in environmental time series forecasting.

This module incorporates a variety of average pooling filters with differing window sizes, allowing for the adaptable extraction of multiple trend components from the input signal. Furthermore, a dynamic weighting mechanism, based on the attributes of the input

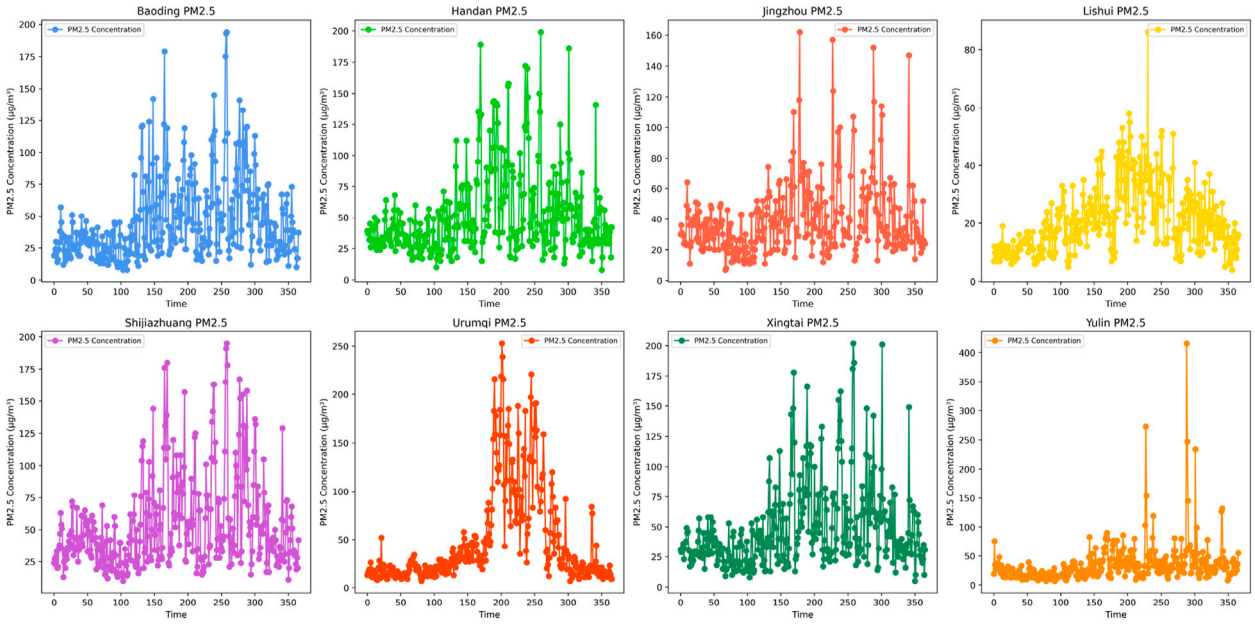

FIGURE 4
$PM_{2.5}$ concentrations in various cities.

data, combines these trend components into a comprehensive final trend depiction. As shown in Equations 1, 2:

$$X_{\text{trend}} = \text{Softmax}(T(x)) \cdot P(x) \tag{1}$$

$$X_{\text{season}} = \mathcal{X} - X_{\text{trend}} \tag{2}$$

In these two formulas, $\mathbf{P}(\cdot)$ denotes a set of average pooling filters, crafted to capture trends across diverse temporal scales. Furthermore, $\text{Softmax}(\mathbf{T}(x))$ acts as a data-dependent weight allocation function, effectively combining these identified trends into a cohesive final trend representation.

3.3 The mutual conversion between the time domain and the frequency domain

In the scholarly domain of air pollution time series forecasting, the Discrete Fourier Transform (DFT) and its counterpart, the Inverse Discrete Fourier Transform (IDFT), are instrumental in scrutinizing complex periodicity and trend variation patterns. This is accomplished by enabling the transition of time series data between the temporal and frequency domains. The DFT decomposes the time series into long-term trends and periodic components, which facilitates the identification of significant periodic features and the elimination of random noise. Subsequently, the IDFT reconstructs the processed signal back into the time domain.

For a time series $x[n] \in R^N$ with a specific length, the DFT is given by Equation 3 as follows:

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-i\frac{2\pi}{N}kn}, \quad k = 0, 1, \ldots, N-1 \tag{3}$$

The IDFT uses Equation 4 to restore the frequency-domain data to the time domain:

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] \cdot e^{i\frac{2\pi}{N}kn}, \quad n = 0, 1, \ldots, N-1 \tag{4}$$

In DFT, determines series length and frequency resolution, indexes frequency components, and $e^{-i\frac{2\pi}{N}kn}$ extracts sinusoidal elements, enabling frequency-domain decomposition. In IDFT, these parameters reconstruct the time-domain signal, with providing amplitude and phase, and $e^{i\frac{2\pi}{N}kn}$ synthesizing the signal. Together, DFT and IDFT support feature extraction, periodic pattern recognition, trend analysis, and noise reduction. Additionally, represents frequency-domain coefficients, where low frequencies indicate trends, and high frequencies reflect noise or rapid fluctuations.

3.4 Frequency-sparse attention mechanism with fourier transform

3.4.1 Traditional attention mechanisms with quadratic complexity

The conventional attention mechanisms utilize three inputs: Q (the query), K (the key), and V (the value) matrices. These mechanisms compute scaled dot-product attention. This is determined by Equation 5:

$$Attention(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{5}$$

In the formulas, the matrices $Q \in R^{L_q \times d}$, $K \in R^{L_k \times d}$, and $V \in R^{L_v \times d}$ are defined, with d representing the dimensionality of the input data. When examining the traditional attention

TABLE 2 Multivariate results with different prediction lengths $O \in \{12, 36, 58, 96\}$ for eight different datasets when $I = 96$. MSE Reduction refers to the percentage decrease in MSE of SFDformer compared to other models. The best average results are in bold, while the second-best results are underlined.

Model		SFDformer			FiLM			Autoformer			Informer			Reformer			Pyraformer		
Metric		MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE
Baoding	**12**	**0.294**	**0.368**	**0.542**	0.397	0.440	0.630	0.390	0.448	0.624	0.321	0.395	0.567	0.310	0.388	0.557	0.299	0.373	0.547
	36	**0.328**	**0.376**	**0.573**	0.341	0.387	0.584	0.474	0.502	0.688	0.352	0.421	0.593	0.366	0.420	0.605	0.338	0.406	0.581
	58	**0.325**	**0.401**	**0.570**	0.449	0.448	0.670	0.429	0.465	0.655	0.361	0.416	0.601	0.446	0.495	0.668	0.328	0.405	0.573
	96	**0.362**	**0.442**	**0.602**	0.493	0.475	0.702	0.420	0.463	0.648	0.372	0.447	0.610	0.518	0.599	0.720	0.510	0.575	0.714
Handan	**12**	**0.402**	**0.424**	**0.634**	0.521	0.515	0.722	0.470	0.475	0.686	0.406	0.437	0.637	0.403	0.433	0.635	0.405	0.432	0.636
	36	**0.421**	**0.436**	**0.649**	0.447	0.446	0.669	0.475	0.489	0.689	0.427	0.444	0.653	0.465	0.492	0.682	0.469	0.483	0.685
	58	**0.463**	**0.487**	**0.680**	0.564	0.502	0.751	0.510	0.514	0.714	0.507	0.509	0.712	0.550	0.557	0.742	0.471	0.492	0.686
	96	**0.537**	**0.514**	**0.733**	0.618	0.534	0.786	0.542	0.516	0.736	0.745	0.662	0.863	0.649	0.650	0.806	0.619	0.586	0.787
Shijiazhuang	**12**	0.329	0.388	0.574	0.412	0.450	0.642	0.403	0.455	0.655	0.353	0.429	0.594	**0.327**	0.407	0.572	0.338	**0.394**	0.581
	36	0.372	0.402	0.610	**0.352**	**0.400**	**0.593**	0.444	0.491	0.666	0.427	0.475	0.653	0.373	0.439	0.611	0.382	0.446	0.618
	58	**0.368**	**0.445**	**0.607**	0.452	0.452	0.672	0.413	0.456	0.643	0.515	0.530	0.718	0.466	0.501	0.683	0.375	0.447	0.612
	96	0.470	**0.465**	0.686	0.496	0.477	0.704	**0.441**	0.472	**0.664**	0.592	0.581	0.709	0.541	0.604	0.736	0.560	0.602	0.748
Xingtai	**12**	**0.376**	**0.422**	**0.613**	0.487	0.486	0.698	0.429	0.461	0.655	0.382	0.424	0.618	0.398	0.447	0.631	0.408	0.434	0.639
	36	**0.417**	**0.426**	**0.646**	0.421	0.431	0.649	0.481	0.493	0.694	0.452	0.480	0.672	0.443	0.480	0.666	0.488	0.508	0.699
	58	**0.436**	**0.439**	**0.660**	0.527	0.481	0.726	0.462	0.470	0.680	0.443	0.464	0.666	0.499	0.524	0.706	0.451	0.485	0.672
	96	**0.503**	**0.489**	**0.709**	0.571	0.508	0.756	0.511	0.497	0.715	0.691	0.638	0.831	0.584	0.610	0.764	0.651	0.632	0.807
Yulin	**12**	**0.764**	**0.517**	**0.874**	1.102	0.703	1.050	0.861	0.581	0.928	0.773	0.516	0.879	0.812	0.522	0.901	0.768	0.523	0.876
	36	**0.758**	**0.517**	**0.871**	0.910	0.565	0.954	0.762	0.565	0.873	0.813	0.525	0.902	0.832	0.534	0.912	0.818	0.530	0.904
	58	**0.757**	**0.503**	**0.870**	1.076	0.625	1.037	0.763	0.579	0.873	0.788	0.559	0.888	0.890	0.596	0.943	0.796	0.519	0.892
	96	**0.771**	**0.524**	**0.878**	1.149	0.664	1.072	0.962	0.655	0.981	0.784	0.533	0.885	0.890	0.606	0.943	0.822	0.541	0.907
Lishui	**12**	**0.532**	**0.488**	**0.729**	0.565	0.537	0.752	0.551	0.523	0.742	0.645	0.549	0.803	0.570	0.500	0.755	0.540	0.530	0.735
	36	**0.556**	**0.532**	**0.746**	0.587	0.543	0.766	0.599	0.541	0.774	0.645	0.568	0.803	0.667	0.587	0.817	0.616	0.580	0.785
	58	**0.588**	0.550	**0.767**	0.786	0.622	0.887	0.601	**0.533**	0.775	0.625	0.573	0.791	0.650	0.568	0.806	0.666	0.601	0.816
	96	**0.644**	**0.565**	**0.802**	0.852	0.653	0.923	0.786	0.678	0.887	0.658	0.566	0.811	0.679	0.589	0.824	0.770	0.650	0.877
Urumqi	**12**	**0.290**	**0.379**	**0.539**	0.385	0.431	0.620	0.347	0.439	0.589	0.347	0.422	0.589	0.304	0.389	0.551	0.395	0.463	0.628

(Continued on following page)

10.3389/fenvs.2025.1549209

TABLE 2 (*Continued*) Multivariate results with different prediction lengths $O \in A12, 36, 58, 96A$ for eight different datasets when $I = 96$. MSE Reduction refers to the percentage decrease in MSE of SFDformer compared to other models. The best average results are in bold, while the second-best results are underlined.

Model		SFDformer			FiLM			Autoformer			Informer			Reformer			Pyraformer		
Metric		MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE
	36	**0.320**	**0.410**	**0.566**	0.408	0.450	0.639	0.394	0.462	0.628	0.335	0.414	0.579	0.333	0.415	0.577	0.437	0.486	0.661
	58	**0.336**	**0.415**	**0.580**	0.661	0.601	0.813	0.493	0.525	0.702	0.364	0.435	0.603	0.364	0.444	0.603	0.576	0.570	0.759
	96	**0.353**	**0.423**	**0.594**	0.812	0.664	0.901	0.511	0.541	0.715	0.376	0.434	0.613	0.415	0.484	0.644	0.789	0.670	0.888
Jingzhou	**12**	**0.661**	**0.514**	**0.813**	0.844	0.605	0.919	0.761	0.577	0.872	0.712	0.552	0.844	0.678	0.529	0.823	0.794	0.587	0.891
	36	**0.727**	**0.556**	**0.853**	0.806	0.585	0.898	0.859	0.63	0.927	0.782	0.593	0.884	0.742	0.569	0.861	0.887	0.642	0.942
	58	**0.770**	**0.578**	**0.877**	1.069	0.701	1.034	0.896	0.645	0.947	0.895	0.648	0.946	0.823	0.610	0.907	0.982	0.695	0.991
	96	**0.862**	**0.629**	**0.928**	1.194	0.74	1.093	0.969	0.674	0.984	0.988	0.683	0.994	0.878	0.644	0.937	0.996	0.749	0.998
Count		29	29	29	1	1	1	1	1	1	0	0	0	1	0	1	0	1	0
MSE reduction		———			22.46			12.58			9.97			9.92			14.14		

10.3389/fenvs.2025.1549209

TABLE 3 Univariate results with different prediction lengths $O \in \{12, 36, 58, 96\}$ for eight different datasets when $I = 96$. MAE Reduction refers to the percentage decrease in MAE of SFDformer compared to other models. The best average results are in bold, while the second-best results are in underlined.

Model		SFDformer			FEDformer			Autoformer			Informer			Reformer			LightTS		
Metric		MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE
Baoding	**12**	**0.220**	**0.335**	**0.469**	0.244	0.339	0.494	0.305	0.406	0.552	0.254	0.391	0.504	0.225	0.351	0.474	0.256	0.404	0.506
	36	**0.208**	**0.315**	**0.456**	0.219	0.325	0.468	0.299	0.390	0.547	0.212	0.333	0.460	0.213	0.317	0.462	0.237	0.391	0.487
	58	**0.223**	**0.335**	**0.472**	0.244	0.339	0.494	0.305	0.406	0.552	0.255	0.391	0.505	0.225	0.350	0.474	0.256	0.404	0.506
	96	**0.212**	**0.330**	**0.460**	0.215	0.333	0.464	0.294	0.388	0.542	0.246	0.363	0.496	0.244	0.360	0.494	0.360	0.499	0.600
Handan	**12**	**0.314**	**0.401**	**0.560**	0.319	0.415	0.565	0.328	0.404	0.573	0.374	0.457	0.612	0.344	0.410	0.587	0.386	0.479	0.621
	36	**0.304**	**0.387**	**0.551**	0.325	0.412	0.570	0.312	0.393	0.559	0.333	0.403	0.577	0.326	0.443	0.571	0.357	0.466	0.597
	58	**0.313**	**0.402**	**0.559**	0.319	0.415	0.565	0.328	0.404	0.573	0.370	0.452	0.608	0.345	0.411	0.587	0.386	0.479	0.621
	96	**0.329**	**0.386**	**0.574**	0.335	0.391	0.579	0.355	0.405	0.596	0.346	0.436	0.588	0.350	0.443	0.592	0.520	0.586	0.721
Shijiazhuang	**12**	**0.181**	**0.313**	**0.425**	0.216	0.320	0.465	0.247	0.359	0.497	0.185	0.325	0.430	0.189	0.324	0.435	0.227	0.385	0.476
	36	**0.171**	**0.293**	**0.414**	0.199	0.312	0.446	0.256	0.370	0.506	0.174	0.298	0.417	0.185	0.314	0.430	0.214	0.375	0.463
	58	**0.182**	**0.314**	**0.427**	0.216	0.320	0.465	0.247	0.359	0.497	0.188	0.329	0.434	0.189	0.323	0.435	0.227	0.385	0.476
	96	**0.192**	**0.310**	**0.438**	0.209	0.321	0.457	0.229	0.356	0.479	0.202	0.341	0.449	0.210	0.345	0.458	0.317	0.467	0.563
Xingtai	**12**	**0.231**	**0.342**	**0.481**	0.258	0.344	0.508	0.279	0.382	0.528	0.242	0.350	0.492	0.235	0.345	0.485	0.288	0.419	0.537
	36	**0.218**	**0.332**	**0.467**	0.244	0.341	0.494	0.269	0.375	0.519	0.224	0.338	0.473	0.236	0.347	0.486	0.265	0.403	0.515
	58	**0.230**	**0.340**	**0.480**	0.258	0.344	0.508	0.279	0.382	0.528	0.242	0.351	0.492	0.234	0.343	0.484	0.288	0.419	0.537
	96	**0.235**	**0.347**	**0.485**	0.261	0.353	0.511	0.270	0.381	0.520	0.242	0.356	0.492	0.247	0.362	0.497	0.391	0.509	0.625
Yulin	**12**	**1.034**	**0.545**	**1.017**	1.079	0.581	1.039	1.134	0.597	1.065	1.042	0.553	1.021	1.100	0.551	1.049	1.298	0.689	1.139
	36	**1.057**	**0.539**	**1.028**	1.107	0.592	1.052	1.091	0.582	1.045	1.063	0.544	1.031	1.102	0.591	1.050	1.192	0.641	1.092
	58	**1.039**	**0.547**	**1.019**	1.079	0.581	1.039	1.133	0.596	1.064	1.041	0.551	1.020	1.101	0.551	1.049	1.298	0.689	1.139
	96	1.079	0.572	1.039	1.135	0.606	1.065	1.137	0.595	1.066	**1.069**	**0.546**	**1.034**	1.114	0.574	1.055	1.472	0.763	1.213
Lishui	**12**	**0.164**	**0.315**	**0.405**	0.202	0.353	0.449	0.178	0.321	0.422	0.172	0.323	0.415	0.174	0.318	0.417	0.220	0.368	0.469
	36	**0.169**	**0.317**	**0.411**	0.192	0.337	0.438	0.190	0.340	0.436	0.197	0.347	0.444	0.215	0.351	0.464	0.265	0.418	0.515
	58	**0.182**	**0.333**	**0.427**	0.245	0.379	0.495	0.201	0.336	0.448	0.219	0.362	0.468	0.237	0.388	0.487	0.331	0.473	0.575
	96	**0.252**	**0.397**	**0.502**	0.291	0.413	0.539	0.299	0.424	0.547	0.310	0.463	0.557	0.411	0.560	0.641	0.486	0.593	0.697
Urumqi	**12**	**0.155**	**0.240**	**0.394**	0.254	0.323	0.504	0.225	0.345	0.474	0.193	0.301	0.439	0.187	0.255	0.432	0.221	0.314	0.470

(Continued on following page)

10.3389/fenvs.2025.1549209

10.3389/fenvs.2025.1549209

TABLE 3 (*Continued*) Univariate results with different prediction lengths $O \in A12, 36, 58, 96A$ for eight different datasets when $I = 96$. MAE Reduction refers to the percentage decrease in MAE of SFDformer compared to other models. The best average results are in bold, while the second-best results are in underlined.

Model		SFDformer			FEDformer			Autoformer			Informer			Reformer			LightTS		
Metric		MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE
	36	**0.180**	**0.248**	**0.424**	0.349	0.378	0.591	0.296	0.364	0.544	0.237	0.341	0.487	0.197	0.296	0.444	0.316	0.404	0.562
	58	**0.237**	**0.326**	**0.487**	0.409	0.420	0.640	0.327	0.391	0.572	0.291	0.357	0.539	0.288	0.380	0.537	0.375	0.427	0.612
	96	**0.288**	**0.333**	**0.537**	0.614	0.550	0.784	0.444	0.465	0.666	0.359	0.419	0.599	0.349	0.457	0.591	0.647	0.686	0.804
Jingzhou	**12**	**0.381**	**0.434**	**0.617**	0.421	0.456	0.649	0.414	0.454	0.643	0.530	0.536	0.728	0.445	0.484	0.667	0.439	0.513	0.663
	36	**0.396**	**0.439**	**0.629**	0.472	0.482	0.687	0.473	0.493	0.688	0.489	0.503	0.699	0.415	0.476	0.644	0.533	0.581	0.730
	58	**0.426**	**0.458**	**0.653**	0.525	0.517	0.725	0.489	0.496	0.699	0.522	0.546	0.722	0.517	0.547	0.719	0.611	0.632	0.782
	96	**0.485**	**0.477**	**0.696**	0.562	0.531	0.750	0.545	0.533	0.738	0.914	0.791	0.956	0.497	0.548	0.705	0.804	0.755	0.897
Count		31	31	31	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
MAE reduction		———			8.54			11.04			10.40			8.48			25.06		

TABLE 4 Ablation study results with different prediction lengths $O \in \{12, 36, 58, 96\}$ for eight different datasets when $I = 96$. MSE Reduction refers to the percentage decrease in MSE of SFDformer compared to other models. The best average results are shown in bold, and the second-best in underlined.

Model		SFDformer			SFDformerV1			SFDformerV2			Informer			Reformer			Transformer		
Self-att		SFDAtt			AutoAtt			AutoAtt			ProbAtt			ReAtt			FullAtt		
Cross-att		SFDAtt			SFDAtt			AutoAtt			ProbAtt			ReAtt			FullAtt		
Metric		MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE
Baoding	**12**	**0.294**	**0.368**	**0.542**	0.337	0.406	0.581	0.362	0.445	0.602	0.321	0.395	0.567	0.310	0.388	0.557	0.303	0.379	0.550
	36	**0.328**	**0.376**	**0.573**	0.371	0.437	0.609	0.353	0.428	0.594	0.352	0.421	0.593	0.366	0.420	0.605	0.323	0.389	0.568
	58	**0.325**	**0.401**	**0.570**	0.396	0.446	0.629	0.410	0.484	0.640	0.361	0.416	0.601	0.446	0.495	0.668	0.333	0.390	0.577
	96	**0.362**	**0.442**	**0.602**	0.390	0.443	0.624	0.436	0.488	0.660	0.372	0.447	0.610	0.518	0.599	0.720	0.368	0.446	0.607
Handan	**12**	**0.402**	**0.424**	**0.634**	0.444	0.468	0.666	0.465	0.485	0.682	0.406	0.437	0.637	0.403	0.433	0.635	0.451	0.478	0.672
	36	**0.421**	**0.436**	**0.649**	0.480	0.491	0.693	0.496	0.499	0.704	0.427	0.444	0.653	0.465	0.492	0.682	0.494	0.503	0.703
	58	**0.463**	**0.487**	**0.680**	0.508	0.511	0.713	0.546	0.541	0.739	0.507	0.509	0.712	0.550	0.557	0.742	0.518	0.523	0.720
	96	0.537	**0.514**	0.733	**0.533**	0.540	**0.730**	0.536	0.532	0.732	0.745	0.662	0.863	0.649	0.650	0.806	0.542	0.545	0.736
ShiJiazhuang	**12**	0.329	**0.388**	0.574	0.447	0.475	0.669	0.463	0.477	0.680	0.353	0.429	0.594	**0.327**	0.402	**0.572**	0.332	0.390	0.573
	36	**0.372**	**0.402**	**0.610**	0.468	0.482	0.684	0.473	0.490	0.688	0.427	0.475	0.653	0.373	0.439	0.611	0.376	0.419	0.613
	58	**0.368**	**0.445**	**0.607**	0.512	0.512	0.716	0.524	0.528	0.724	0.515	0.530	0.718	0.466	0.501	0.683	0.400	0.454	0.632
	96	**0.470**	**0.465**	**0.686**	0.532	0.505	0.729	0.512	0.517	0.716	0.592	0.581	0.769	0.541	0.604	0.736	0.515	0.541	0.718
Xingtai	**12**	**0.376**	**0.422**	**0.613**	0.428	0.460	0.783	0.445	0.469	0.667	0.382	0.424	0.618	0.398	0.447	0.631	0.379	0.423	0.616
	36	**0.417**	**0.426**	**0.646**	0.471	0.489	0.804	0.444	0.478	0.666	0.452	0.480	0.672	0.443	0.48	0.666	0.431	0.433	0.657
	58	**0.436**	**0.439**	**0.660**	0.498	0.509	0.813	0.534	0.534	0.731	0.443	0.464	0.666	0.499	0.524	0.706	0.448	0.463	0.669
	96	**0.503**	**0.489**	**0.709**	0.523	0.529	0.842	0.531	0.536	0.729	0.691	0.638	0.831	0.584	0.610	0.764	0.519	0.498	0.720
Yulin	**12**	**0.764**	0.517	**0.874**	0.806	0.551	0.898	0.835	0.584	0.914	0.773	**0.516**	0.879	0.812	0.522	0.901	0.784	0.542	0.885
	36	**0.758**	**0.517**	**0.871**	0.917	0.592	0.958	0.919	0.599	0.959	0.813	0.525	0.902	0.832	0.534	0.912	0.902	0.581	0.950
	58	**0.757**	**0.503**	**0.870**	0.947	0.608	0.973	0.941	0.609	0.970	0.788	0.559	0.888	0.890	0.596	0.943	0.929	0.582	0.964
	96	**0.771**	**0.524**	**0.878**	0.972	0.626	0.986	0.959	0.658	0.979	0.784	0.533	0.885	0.890	0.606	0.943	0.938	0.632	0.969
Lishui	**12**	**0.532**	**0.488**	**0.729**	0.542	0.512	0.736	0.554	0.519	0.744	0.645	0.549	0.803	0.570	0.500	0.755	0.584	0.510	0.764
	36	**0.556**	**0.532**	**0.746**	0.572	0.557	0.756	0.581	0.564	0.762	0.645	0.568	0.803	0.667	0.587	0.817	0.650	0.563	0.806
	58	**0.588**	0.550	**0.767**	0.613	0.551	0.783	0.624	0.58	0.790	0.625	0.573	0.791	0.660	0.568	0.806	0.617	**0.537**	0.785

(Continued on following page)

10.3389/fenvs.2025.1549209

TABLE 4 (*Continued*) Ablation study results with different prediction lengths $O \in A12, 36, 58, 96A$ for eight different datasets when $I = 96$. MSE Reduction refers to the percentage decrease in MSE of SFDformer compared to other models. The best average results are shown in bold, and the second-best in underlined.

Model		SFDformer			SFDformerV1			SFDformerV2			Informer			Reformer			Transformer		
Self-att		SFDAtt			AutoAtt			AutoAtt			ProbAtt			ReAtt			FullAtt		
Cross-att		SFDAtt			SFDAtt			AutoAtt			ProbAtt			ReAtt			FullAtt		
Metric		MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE
	96	**0.644**	**0.565**	**0.802**	0.682	0.578	0.826	0.793	0.661	0.891	0.658	0.566	0.811	0.679	0.589	0.824	0.675	0.577	0.822
Urumqi	**12**	**0.29**	**0.379**	**0.539**	0.316	0.394	0.562	0.324	0.405	0.569	0.347	0.422	0.589	0.304	0.389	0.551	0.334	0.411	0.578
	36	**0.320**	**0.410**	**0.566**	0.327	0.421	0.572	0.332	0.438	0.576	0.335	0.414	0.579	0.333	0.415	0.577	0.326	0.417	0.571
	58	**0.336**	**0.415**	**0.580**	0.343	0.436	0.586	0.349	0.443	0.591	0.364	0.435	0.603	0.364	0.444	0.603	0.356	0.424	0.597
	96	**0.353**	**0.423**	**0.594**	0.444	0.465	0.666	0.457	0.473	0.676	0.376	0.434	0.613	0.415	0.484	0.644	0.373	0.442	0.611
Jingzhou	**12**	**0.661**	**0.514**	**0.813**	0.749	0.568	0.865	0.777	0.587	0.881	0.712	0.552	0.844	0.678	0.529	0.823	0.859	0.625	0.927
	36	**0.727**	**0.556**	**0.853**	0.804	0.599	0.897	0.801	0.597	0.895	0.782	0.593	0.884	0.742	0.569	0.861	0.850	0.624	0.922
	58	**0.770**	**0.578**	**0.877**	0.97	0.675	0.985	0.982	0.685	0.991	0.895	0.648	0.946	0.823	0.610	0.907	0.885	0.653	0.941
	96	**0.862**	**0.629**	**0.928**	1.038	0.698	1.019	1.056	0.704	1.028	0.988	0.683	0.994	0.878	0.644	0.937	1.133	0.768	1.064
Count		30	30	30	1	0	1	0	0	0	0	0	0	1	0	1	0	0	0
MSE reduction		———			12.44			14.46			9.97			9.92			10.21		

10.3389/fenvs.2025.1549209

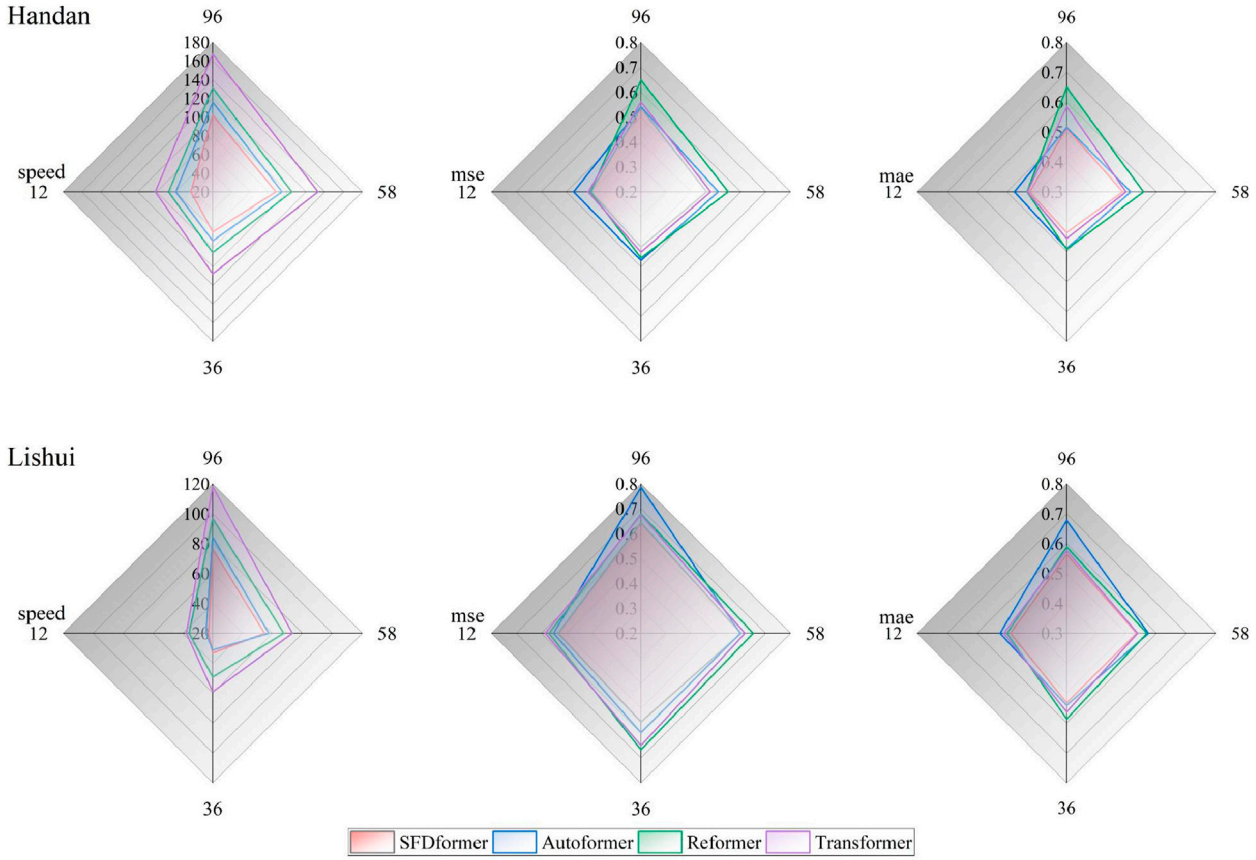

FIGURE 5
Experiment for evaluating computational efficiency and performance. The input length was fixed at $I = 96$, with prediction lengths set to $O \in \{12, 36, 58, 96\}$. Computational efficiency was measured by the time (in seconds) required for each model to complete one hundred epochs. Performance was assessed using Mean Squared Error (MSE) and Mean Absolute Error (MAE) as the key metrics.

mechanisms, particular attention is paid to the distribution of attention for the i-th query, referred to as q_i. This distribution is calculated using an asymmetric kernel smoother, which yields the attention associated with the i-th query, as shown in Equation 6:

$$Attention(q_i, K, V) = \sum_j \frac{k(q_i, k_j)}{\sum_j k(q_i, k_j)} v_j - E_{p(k_j|q_i)}[v_j] \quad (6)$$

The probability $p(k_j|q_i)$ is calculated as $\frac{k(q_i,k_j)}{\sum_j k(q_i,k_j)}$, where $k(q_i, k_j)$ represents the asymmetric exponential kernel, expressed mathematically as $\exp(\frac{q_i k_j^T}{\sqrt{d}})$. This computation entails quadratic dot-product operations, resulting in a computational complexity of $O(L_q L_k)$. This complexity poses a considerable challenge in terms of memory usage for models designed to improve predictive performance.

3.4.2 Query sparsity measurement

In the traditional attention mechanisms, the attention distribution $p(K_i|Q_i)$ for ith query is represented as a weighted aggregation over all keys. High dot products between queries and keys lead to uneven attention distributions, potentially reducing the significance of individual values. In order to tackle this issue, a mechanism grounded in KL divergence is proposed to assess the resemblance between the attention distribution and a predefined baseline. The degree of similarity is determined by using Equation 7:

$$KL(Q\|p) = -\ln\left(\frac{1}{L_n}\sum_{j=1}^{L_n} e^{\frac{Q_i K_j^T}{\sqrt{d}}}\right) + \ln L_K - \frac{1}{L_n}\sum_{j=1}^{L_n}\frac{Q_i K_j^T}{\sqrt{d}} \quad (7)$$

In the above formula, L_n represents the number of keys, $Q_i K_j^T$ indicates the dot product between the query and the key, and d is the dimensionality of the features. The distillation measure, denoted as $M(Q_i, K)$, is defined by Equation 8 as follows:

$$M(Q_i, K) = -\ln\left(\sum_{j=1}^{L_n} e^{\frac{Q_i K_j^T}{\sqrt{d}}}\right) + \frac{1}{L_n}\sum_{j=1}^{L_n}\frac{Q_i K_j^T}{\sqrt{d}} \quad (8)$$

A higher $M(Q_i, K)$ value indicates a more diverse attention distribution for the ith query, potentially focusing on dominant query-key pairs in the tail of the self-attention output. This approach allows the model to prioritize influential query-key pairs, thereby enhancing the overall effectiveness of the knowledge extraction process.

3.4.3 Frequency-sparse attention mechanism

We apply the Discrete Fourier Transform (DFT) to transform the queries q, keys k, and values v. Subsequently, we execute a

TABLE 5 Comparison of accuracy and efficiency metrics for different methods.

Methods		SFDformer	Autoformer	Informer	LogTrans	Transformer	LSTM
Training	**Time**	$\mathcal{O}(L \log L)$	$\mathcal{O}(L \log L)$	$\mathcal{O}(L \log L)$	$\mathcal{O}(L \log L)$	$\mathcal{O}(L^2)$	$\mathcal{O}(L)$
	Memory	$\mathcal{O}(L \log L)$	$\mathcal{O}(L \log L)$	$\mathcal{O}(L \log L)$	$\mathcal{O}(L^2)$	$\mathcal{O}(L^2)$	$\mathcal{O}(L)$
Testing	**Steps**	1	1	1	1	L	L

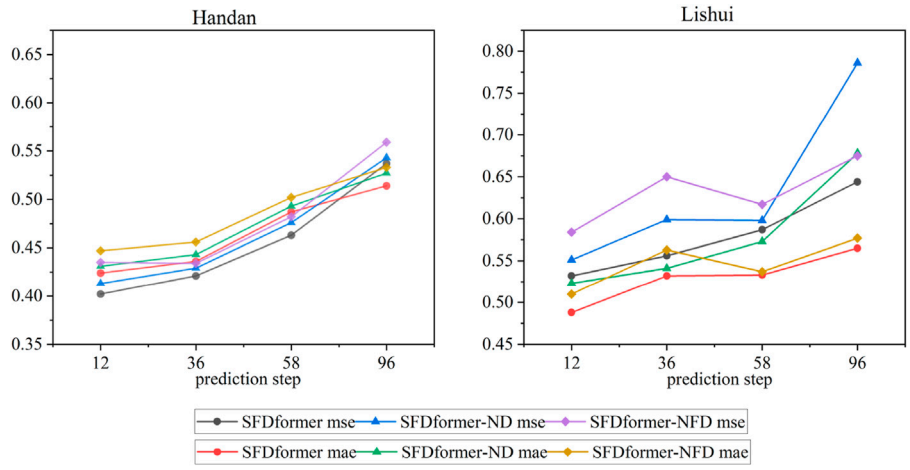

FIGURE 6
$PM_{2.5}$ Concentration Prediction Results for Handan and Xintai Datasets. The prediction results of the SFDformer model are indicated by a red line, the SFDformer-ND model is represented by a blue line, and the performance of the SFDformer-NFD model is depicted by a yellow line.

comparable attention mechanism in the frequency domain by choosing the Top u weight matrix patterns. The versions of the queries, keys, and values after the DFT transformation are represented as $\tilde{Q} \in C^{M\times D}$, $\tilde{K} \in C^{M\times D}$, and $\tilde{V} \in C^{M\times D}$. The Frequency Sparse Attention Mechanism incorporating Fourier Transform(SFD) is outlined as follows in Equations 9–12:

$$\tilde{Q} = \mathrm{Top}_u(F(q)) \tag{9}$$

$$\tilde{K} = \mathrm{Top}_u(F(k)) \tag{10}$$

$$\tilde{V} = \mathrm{Top}_u(F(v)) \tag{11}$$

$$\mathrm{SFDAttention}(q,k,v) = F^{-1}\left(\mathrm{Padding}\left(\sigma\left(\tilde{Q}\cdot\tilde{K}^T\right)\cdot\tilde{V}\right)\right) \tag{12}$$

In the above formula, σ represents an activation function. We utilize either softmax or tanh as the activation function, since their convergence performance differs among various datasets. Let Y be defined as $Y = \sigma(\tilde{Q}\cdot\tilde{K}^T)\cdot\tilde{V}$, where $Y \in C^{M\times D}$. The structure of the Frequency Sparse Attention Mechanism with Fourier Transform (SFD) is depicted in Figure 1.

At the screening frequency, it is sufficient to randomly sample $u = L_K \ln L_Q$ dot-product pairs for the computation of $M(Q_i, K)$, with the remaining pairs being effectively filled with zeros. From these sampled pairs, the sparse Top_u is selected as Q. The maximum operator in $M(Q_i, K)$ exhibits reduced sensitivity to zero values, thereby ensuring numerical stability. In practical applications, the input lengths of queries and keys are typically equivalent in self-attention computations, i.e., $L_Q = L_K = L$. Consequently, the overall time complexity and space complexity of the SFDAttention mechanism are $O(L \log L)$.

4 Experiment

4.1 Data description

This research employed historical data on pollutant concentrations and meteorological conditions, gathered from monitoring stations situated in eight different cities throughout China. The dataset spans the timeframe from 28 October 2013, to 31 May2021. The experimental data in this study is based on a city-level perspective, where daily sample data for each city is represented as a one-dimensional feature vector, with feature elements consisting of pollutants and meteorological factors. The eight selected cities are Baoding, Handan, Jingzhou, Shijiazhuang, Xingtai, Yulin, Lishui, Urumqi,Jingzhou are among the selected cities, each exhibiting unique economic development characteristics within China. These cities are strategically positioned across various geographical regions of the country(see Figure 3). In the analysis, six distinct types of pollutants were considered, alongside three indicators for evaluating pollution levels and three meteorological factors that influence pollutant concentrations. each of which has distinctive characteristics in terms of economic development in

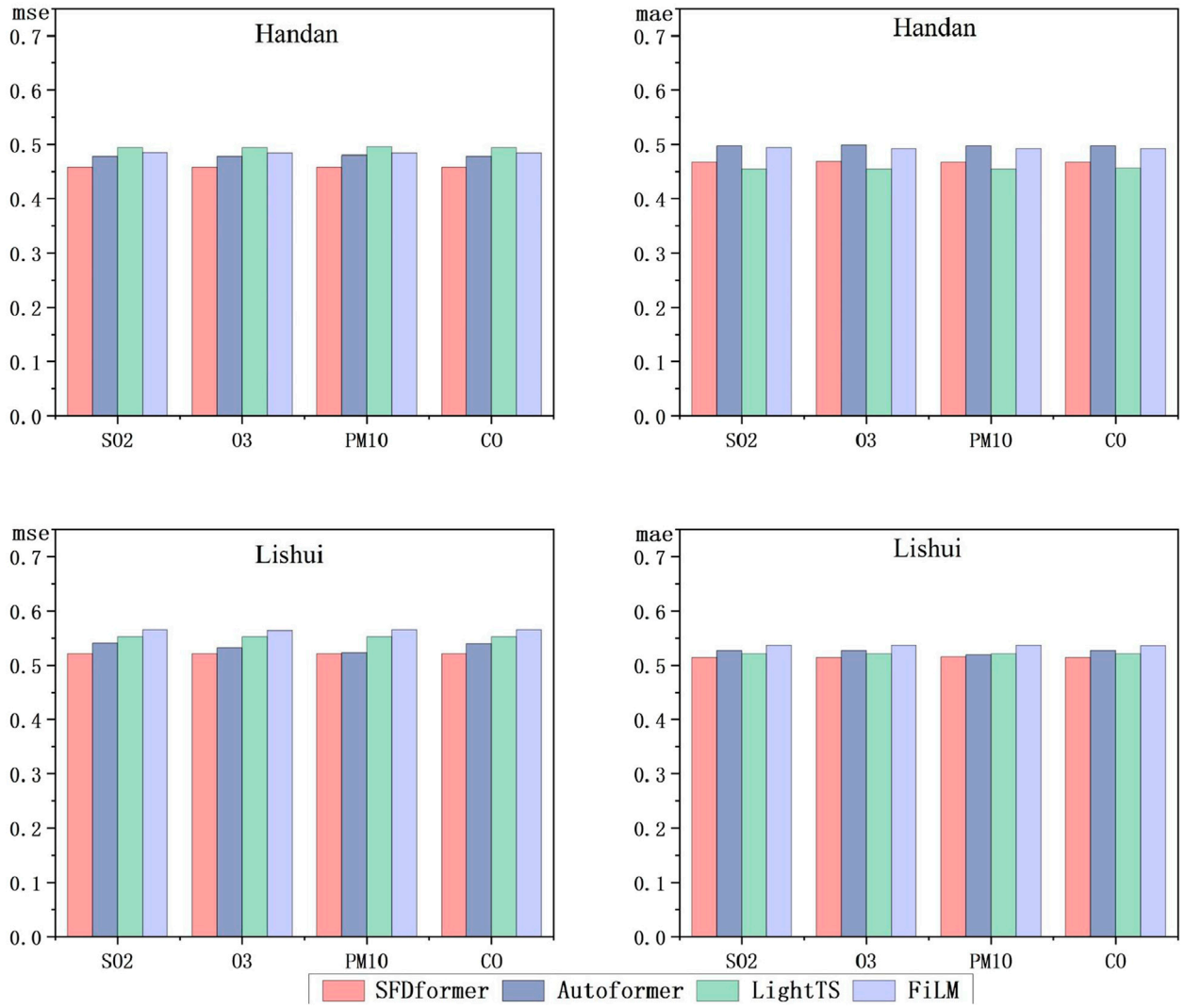

FIGURE 7
Prediction results for SO_2, O_3, PM_{10}, and CO concentrations. The red areas represent the SFDformer model, the blue areas represent the Autoformer model, the green areas represent the LightTS model, and the purple areas represent the Film model.

China. The selected cities are strategically located across diverse geographical regions within the nation, each presenting distinct pollution characteristics. (refer to Table 1 for details): air quality grade, AQI index, daily AQI ranking, O_3, PM_{10}, So_2, No_2, CO, $PM_{2.5}$, Temperature, Wind speed, and Precipitation. In Figure 4, we present the daily $PM_{2.5}$ concentration for each city segment in the dataset from 28 October 2020, to 31 May 2021. In accordance with established procedures, the entirety of the compiled datasets was methodically divided into training, validation, and test subsets, arranged sequentially over time, and following a specified allocation ratio of 7:1:2 (Hua et al., 2019).

4.2 Implementation details

Harnessing the benefits of Transformer architectures in managing time series information, we integrated residual connections into our model, embedding them within decomposition blocks (Yu et al., 2024). These blocks incorporate functionalities like moving averages, which assist in evening out periodic oscillations and highlighting long-term tendencies within the time series data. As a result, residual connections significantly improve the model's ability to perceive and assimilate complex patterns inherent in time series, thereby significantly boosting its proficiency in long-term projections. To further enhance the self-attention mechanism, we subjected the input features to nonlinear transformations and dimensional alterations *via* a Multi-Layer Perceptron (MLP), resulting in innovative feature renditions. This tactic allows the model to more precisely detect intricate patterns and profound interconnections embedded in the time series information, ultimately refining its overall predictive capabilities. Our training methodology utilizes L2 loss along with the ADAM optimizer (Kingma and Ba, 2015), initiated with a learning rate of 0.0001 and a batch size of 32. The attention factor is established at 3, and weight decay is set to 0.1. Training concludes prematurely after 10 epochs. Every experiment was

replicated thrice and executed using PyTorch (Paszke et al., 2019), facilitated on a solitary NVIDIA Tesla V100 32 GB GPU (Markidis et al., 2018).

In this study, we utilize Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) as three essential criteria to assess the predictive accuracy of the SFDformer model. The detailed explanations for calculating these indicators are provided in Equations 13–15:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{13}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{14}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{15}$$

Where y_i represents the actual observed value, $\hat{y}_i$ is the predicted value from the model, and n is the total number of data points, these metrics allow us to intuitively evaluate the accuracy of the model's predictions. Lower values of Mean Squared Error (MSE) and Mean Absolute Error (MAE) indicate that the predicted values are closer to the actual values, suggesting better predictive performance. Additionally, a lower Root Mean Squared Error (RMSE) implies a better model fit to the data, indicating more reliable prediction results.

We evaluated seven baseline methods for comparative analysis. In the multivariate setting, we selected four Transformer-based models: Autoformer (Wu et al., 2021), Informer (Zhou et al., 2021), Reformer (Kitaev et al., 2020), and Pyraformer Liu et al. (2022), in addition to one model based on linear networks: FiLM (Zhou et al., 2022a). For the univariate setting, we considered more competitive baselines: FEDformer (Zhou et al., 2022b), and a model based on MLP: LightTS (Campos et al., 2023).

4.3 Main results

4.3.1 Multivariate results

Multivariate analysis involves the simultaneous consideration of multiple time series to examine the interrelationships and influences among them.In the multivariate settings, we conducted experiments using eight different datasets. The results indicate that SFDformer consistently achieved state-of-the-art performance across most baseline and prediction horizon configurations (see Table 2). Specifically, under the input-96-predict-58 (The model utilizes 96 historical data points to forecast 58 future data points) configuration, SFDformer reduces the MSE by 0.9% in Baoding (0.328 → 0.325), 1.6% in Handan (0.471 → 0.463), 1.8% in Shijiazhuang (0.375 → 0.368), 1.5% in Xingtai (0.443 → 0.436), 0.7% in Yulin (0.763 → 0.757), 2.1% in Lishui (0.601 → 0.588), 7.6% in Urumqi (0.364 → 0.336), and 6.4% in Jingzhou (0.823 →0.770) compared to previous state-of-the-art results. Overall, in this configuration, the average MSE reduction for SFDformer is 22.46%. Furthermore, on the Shijiazhuang dataset, SFDformer did not exhibit optimal performance in the input-96-predict-12 and input-96-predict-36 settings. However, its performance improves as the prediction horizon extends. This improvement can be attributed to the relatively minor impact of noise in short-term forecasting, whereas long-term forecasting is more influenced by the intricate temporal patterns inherent in real-world time series, demonstrating SFDformer's ability to better handle complex temporal patterns.

4.3.2 Univariate results

Univariate analysis predicts future values based solely on the historical data of a single time series. We showcase the univariate outcomes for eight illustrative datasets, as depicted in Table 3. In contrast with numerous baseline models, SFDformer achieves cutting-edge performance in prediction tasks. Notably, under the input-96-predict-58 setup, our model diminishes the mean absolute error (MAE) on the Baoding dataset by 1.1% (0.339 → 0.335). Regarding the Handan dataset, the model lowers the MAE by 0.4% (0.404 → 0.402), Regarding the Shijiazhuang dataset, the model lowers the MAE by 1.8% (0.320 → 0.314), Regarding the Xingtai dataset, the model lowers the MAE by 1.1% (0.344 → 0.340), Regarding the Yulin dataset, the model lowers the MAE by 0.7% (0.551 → 0.547), Regarding the Lishui dataset, the model lowers the MAE by 0.8% (0.336 → 0.333), Regarding the Urumqi dataset, the model lowers the MAE by 8% (0.357 → 0.326), Regarding the Jingzhou dataset, the model lowers the MAE by 7% (0.496 → 0.458). Moreover, as the prediction timeline extends, the model's proficiency stays consistent, underscoring its resilience in forecasting $PM_{2.5}$ air pollution concentration levels.

4.3.3 Ablation research

This study assesses the impact of the Sparse Frequency Domain Attention (SFDA) module on model performance *via* an ablation experiment. Three variants of SFDformer were tested: KEDformer, which entirely substitutes both the self-attention and cross-attention mechanisms with SFDA; SFDformerV1, which replaces only the self-attention mechanism with SFDA while maintaining the cross-correlation attention mechanism; and SFDformerV2, which employs self-correlation attention to manage both mechanisms. The experiments were conducted on eight datasets, as illustrated in Table 4. SFDformer exhibited performance enhancements in 90 out of 96 test cases. Importantly, the SFDformer integrated with the SFDA module consistently demonstrated improvements across all cases, corroborating the effectiveness of SFDA in substituting traditional attention mechanisms and significantly improving the model's performance.

5 Discusion

5.1 Efficiency analysis and performance analysis

The present study comprehensively evaluates the impact of various self-attention mechanisms on model performance and computational efficiency, with a detailed analysis of the trade-offs between these two aspects (see Figure 5). To further verify the model's generalization capability across regions with different levels of air pollution, two distinct locations were selected: Handan, situated in northern China and characterized by relatively severe air pollution, and Lishui, located in eastern China with relatively

mild air pollution. The SFDformer model stands out from other models by integrating Fourier transform and sparse attention techniques into its attention mechanism, thereby significantly enhancing prediction accuracy. Compared with traditional Transformer models, SFDformer effectively mitigates the inherent quadratic complexity of conventional attention mechanisms, leading to a substantial improvement in operational efficiency. This feature makes SFDformer particularly well-suited for handling large-scale time series datasets, such as those used in air pollution forecasting tasks.

5.2 Computation efficiency

In the multivariate setting and with the current optimal implementation of all methods, SFDformer has achieved a significant enhancement in computational efficiency compared to conventional Transformer models. This improvement effectively addresses the challenges associated with the quadratic time complexity $O(L^2)$ and memory usage $O(L^2)$ inherent to standard self-attention mechanisms. By employing sparse attention and the discrete Fourier transform, SFDformer reduces both the time complexity and memory usage to $O(L \log L)$, thereby enhancing the model's capability to handle real-world scenarios of air pollutant concentration prediction. During the testing phase, SFDformer completes predictions in a single step, in contrast to traditional models that require $O(L)$ steps, thereby substantially increasing its efficiency. As demonstrated in Table 5, SFDformer strikes a superior balance between computational efficiency and predictive accuracy, rendering it a practical solution for air pollutant concentration prediction tasks in resource-constrained environments.

5.3 Performance impact of time series decomposition and frequency transformation

To explore the effectiveness of time series decomposition and Fourier transform techniques, we conducted experimental studies using datasets from Handan and Lishui, two regions with significantly different levels of air pollution. As illustrated in Figure 6, the SFDformer model integrates both techniques, whereas the SFDformer-NF model excludes the Fourier transform step, and the SFDformer-NFD model omits both techniques. The experimental results elucidate that the SFDformer model surpasses the other two models, with performance enhancements stemming from several pivotal factors. Primarily, the time series decomposition technique enables the model to directly emulate the seasonal variations in air pollutant concentrations, thereby more accurately capturing periodic patterns and significantly improving the model's ability to make predictions based on historical data. Secondly, the application of the Fourier transform allows the model to discern and accentuate crucial features in the data while mitigating noise interference, ensuring the model concentrates on the most pertinent information during predictions. These findings substantiate the efficacy of time series decomposition and Fourier transform techniques in improving model performance. This version adheres to the standards for scientific articles, employing clear and precise language.

5.4 Generalization and predictive insights of the model on pollutant levels

In this study, the SFDformer model demonstrated remarkable precision in predicting $PM_{2.5}$ concentrations. Industrial development is one of the sources of various air pollutants and is also a key factor contributing to air pollution. To further evaluate the generalization ability of this model, we selected two regions with different industrial characteristics for experiments: Handan, a city with significant heavy industrial development, and Lishui, a region dominated by light industrial activities. We applied the SFDformer model to predict the concentrations of additional pollutants, including PM_{10}, carbon monoxide, sulfur dioxide, and ozone. As shown in Figure 7, the SFDformer model exhibited remarkable proficiency across these diverse pollutant prediction tasks. The experimental results clearly indicate that the SFDformer model outperforms alternative models in terms of generalization capability.

6 Conclusion

The rapid advancement of deep learning technologies has led to their widespread adoption across both academia and industry. This paper presents a novel framework, SFDformer, which seamlessly integrates time series decomposition, Fourier transform, and sparse attention mechanisms. Through the employment of time series decomposition, SFDformer adeptly captures the seasonal fluctuations and long-term trends of $PM_{2.5}$ concentrations, elucidating the interplay between short-term variations and long-term patterns. The fusion of Fourier transform and sparse attention mechanisms not only substantially reduces computational complexity from quadratic to linear, thereby significantly enhancing computational efficiency, but also effectively mitigates noise interference from air pollution features during the prediction process. This dual mechanism's design minimizes the impact of noise on prediction outcomes, enabling the model to better adapt to the temporal dynamics of the real world, which is pivotal for the accurate forecasting of $PM_{2.5}$ concentrations, a critical air pollution indicator.

In future research, we will focus on enhancing the adaptability of SFDformer to diverse datasets, especially those with irregular patterns. We are confident that through further optimization and expansion, SFDformer will achieve even more remarkable results in the highly challenging field of air pollution time-series forecasting. In summary, SFDformer has made significant breakthroughs in addressing the complexities of air pollution time-series forecasting. This achievement not only demonstrates its strong effectiveness but also highlights its great potential and broad application prospects in this critical field.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and

accession number(s) can be found in the article/supplementary material.

Author contributions

ZQ: Conceptualization, Methodology, Formal Analysis, Investigation, Writing–original draft. BW: Methodology, Software, Investigation, Data curation, Writing–original draft, Visualization. CG: Investigation, Methodology, Validation, Writing–original draft. XC: Conceptualization, Investigation, Resources, Writing–review and editing, Supervision, Project administration. HZ: Conceptualization, Validation, Formal analysis, Resources, Writing–review and editing, Supervision, Project administration. CUIW: Validation, Resources, Writing–review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

References

Ailshire, J. A., and Crimmins, E. M. (2014). Fine particulate matter air pollution and cognitive function among older us adults. *Am. J. Epidemiol.* 180, 359–366. doi:10.1093/aje/kwu155

Al-qaness, M. A., Dahou, A., Ewees, A. A., Abualigah, L., Huai, J., Abd Elaziz, M., et al. (2023). Resinformer: residual transformer-based artificial time-series forecasting model for pm2. 5 concentration in three major Chinese cities. *Mathematics* 11, 476. doi:10.3390/math11020476

Athira, V., Geetha, P., Vinayakumar, R., and Soman, K. (2018). Deepairnet: Applying recurrent networks for air quality prediction. *Procedia Comput. Sci.* 132, 1394–1403. doi:10.1016/j.procs.2018.05.068

Campos, D., Zhang, M., Yang, B., Kieu, T., Guo, C., and Jensen, C. S. (2023). Lightts: Lightweight time series classification with adaptive ensemble distillation. *Proc. ACM Manag. Data* 1, 1–27. doi:10.1145/3589316

Chen, S., He, L., Shen, S., Zhang, Y., and Ma, W. (2024). Improving air quality prediction via self-supervision masked air modeling. *Atmosphere* 15, 856. doi:10.3390/atmos15070856

Deng, Y., Zhi, P., Zhu, W., Sang, T., and Li, Y. (2024). Prediction of pm2. 5 concentration based on bayesian optimization random forest, in *2024 43rd Chinese Control conference (CCC) (IEEE)*, 8507–8511.

Du, Y., Xu, X., Chu, M., Guo, Y., and Wang, J. (2016). Air particulate matter and cardiovascular disease: the epidemiological, biomedical and clinical evidence. *J. Thorac. Dis.* 8, E8–E19. doi:10.3978/j.issn.2072-1439.2015.11.37

Espinosa, R., Palma, J., Jiménez, F., Kamińska, J., Sciavicco, G., and Lucena-Sánchez, E. (2021). A time series forecasting based multi-criteria methodology for air quality prediction. *Appl. Soft Comput.* 113, 107850. doi:10.1016/j.asoc.2021.107850

Faraji, M., Nadi, S., Ghaffarpasand, O., Homayoni, S., and Downey, K. (2022). An integrated 3d cnn-gru deep learning method for short-term prediction of pm2. 5 concentration in urban environment. *Sci. Total Environ.* 834, 155324. doi:10.1016/j.scitotenv.2022.155324

Ghimire, S., Deo, R. C., Raj, N., and Mi, J. (2019). Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. *Appl. Energy* 253, 113541. doi:10.1016/j.apenergy.2019.113541

Gu, J., Yang, B., Brauer, M., and Zhang, K. M. (2021). Enhancing the evaluation and interpretability of data-driven air quality models. *Atmos. Environ.* 246, 118125. doi:10.1016/j.atmosenv.2020.118125

Guo, Y., and Mao, Z. (2023). Long-term prediction model for nox emission based on lstm–transformer. *Electronics* 12, 3929. doi:10.3390/electronics12183929

Han, J., Lin, H., and Qin, Z. (2023). Prediction and comparison of in-vehicle co2 concentration based on arima and lstm models. *Appl. Sci.* 13, 10858. doi:10.3390/app131910858

Haq, M. A., and Ahmad Khan, R. (2022). Smotednn: a novel model for air pollution forecasting and aqi classification. *Comput. Mater. & Continua* 71, 1403–1425. doi:10.32604/cmc.2022.021968

He, J., Gong, S., Yu, Y., Yu, L., Wu, L., Mao, H., et al. (2017). Air pollution characteristics and their relation to meteorological conditions during 2014–2015 in major Chinese cities. *Environ. Pollut.* 223, 484–496. doi:10.1016/j.envpol.2017.01.050

Hofman, J., Do, T. H., Qin, X., Bonet, E. R., Philips, W., Deligiannis, N., et al. (2022). Spatiotemporal air quality inference of low-cost sensor data: evidence from multiple sensor testbeds. *Environ. Model. & Softw.* 149, 105306. doi:10.1016/j.envsoft.2022.105306

Hua, Y., Zhao, Z., Li, R., Chen, X., Liu, Z., and Zhang, H. (2019). Deep learning with long short-term memory for time series prediction. *IEEE Commun. Mag.* 57, 114–119. doi:10.1109/mcom.2019.1800155

Kang, G. K., Gao, J. Z., Chiao, S., Lu, S., and Xie, G. (2018). Air quality prediction: big data and machine learning approaches. *Int. J. Environ. Sci. Dev.* 9, 8–16. doi:10.18178/ijesd.2018.9.1.1066

Kim, M. K., Cremers, B., Liu, J., Zhang, J., and Wang, J. (2022). Prediction and correlation analysis of ventilation performance in a residential building using artificial neural network models based on data-driven analysis. *Sustain. Cities Soc.* 83, 103981. doi:10.1016/j.scs.2022.103981

Kingma, D. P., and Ba, J. (2015). Adam: a method for stochastic optimization 3rd international conference on learning representations, in *ICLR 2015-Conference Track Proceedings*, 1.

Kitaev, N., Kaiser, Ł., and Levskaya, A. (2020). Reformer: the efficient transformer. arXiv preprint arXiv:2001.04451.

Ko, K. K., and Jung, E. S. (2022). Improving air pollution prediction system through multimodal deep learning model optimization. *Appl. Sci.* 12, 10405. doi:10.3390/app122010405

Kshirsagar, A., and Shah, M. (2022). Anatomization of air quality prediction using neural networks, regression and hybrid models. *J. Clean. Prod.* 369, 133383. doi:10.1016/j.jclepro.2022.133383

Lanzi, E. (2016). *The economic consequences of outdoor air pollution*. OECD.

Liu, H., and Yang, R. (2021). A spatial multi-resolution multi-objective data-driven ensemble model for multi-step air quality index forecasting based on real-time decomposition. *Comput. Industry* 125, 103387. doi:10.1016/j.compind.2020.103387

Liu, S., Yu, H., Liao, C., Li, J., Lin, W., Liu, A. X., et al. (2022). *Pyraformer: low-complexity pyramidal attention for long-range time series modeling and forecasting*. ICLR.

Lu, D., Mao, W., Xiao, W., and Zhang, L. (2021). Non-linear response of pm2. 5 pollution to land use change in China. *Remote Sens.* 13, 1612. doi:10.3390/rs13091612

Luo, H., Han, Y., Cheng, X., Lu, C., and Wu, Y. (2020). Spatiotemporal variations in particulate matter and air quality over China: National, regional and urban scales. *Atmosphere* 12, 43. doi:10.3390/atmos12010043

Ma, X., Chen, T., Ge, R., Xv, F., Cui, C., and Li, J. (2023a). Prediction of pm2. 5 concentration using spatiotemporal data with machine learning models. *Atmosphere* 14, 1517. doi:10.3390/atmos14101517

Ma, Z., Luo, W., Jiang, J., Wang, B., Ma, Z., Lin, J., et al. (2023b). Spatial and temporal characteristics analysis and prediction model of pm2. 5 concentration based on spatiotemporal-informer model. *Plos one* 18, e0287423. doi:10.1371/journal.pone.0287423

Markidis, S., Der Chien, S. W., Laure, E., Peng, I. B., and Vetter, J. S. (2018). Nvidia tensor core programmability, performance & precision, in 2018 IEEE international parallel and distributed processing symposium workshops (IPDPSW) *(IEEE)*, 522–531.

Marvin, D., Nespoli, L., Strepparava, D., and Medici, V. (2022). A data-driven approach to forecasting ground-level ozone concentration. *Int. J. Forecast.* 38, 970–987. doi:10.1016/j.ijforecast.2021.07.008

Méndez, M., Merayo, M. G., and Núñez, M. (2023). Machine learning algorithms to forecast air quality: a survey. *Artif. Intell. Rev.* 56, 10031–10066. doi:10.1007/s10462-023-10424-4

Neiburger, M. (1969). The role of meteorology in the study and control of air pollution. *Bull. Am. Meteorological Soc.* 50, 957–966. doi:10.1175/1520-0477-50.12.957

Ozcanli, A. K., Yaprakdal, F., and Baysal, M. (2020). Deep learning methods and applications for electrical power systems: a comprehensive review. *Int. J. Energy Res.* 44, 7136–7157. doi:10.1002/er.5331

Pan, Q., Harrou, F., and Sun, Y. (2023). A comparison of machine learning methods for ozone pollution prediction. *J. Big Data* 10, 63. doi:10.1186/s40537-023-00748-x

Panneerselvam, V., and Thiagarajan, R. (2024). Toward accurate multi-region air quality prediction: integrating transformer-based deep learning and crossover boosted dynamic arithmetic optimization (cdao). *Signal Image Video Process* 18, 4145–4156. doi:10.1007/s11760-024-03061-z

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). Pytorch: an imperative style, high-performance deep learning library. *Adv. neural Inf. Process. Syst.* 32. doi:10.48550/arxiv.1912.01703

Pöschl, U. (2005). Atmospheric aerosols: composition, transformation, climate and health effects. *Angew. Chem. Int. Ed.* 44, 7520–7540. doi:10.1002/anie.200501122

Rakholia, R., Le, Q., Vu, K., Ho, B. Q., and Carbajo, R. S. (2024). Accurate pm2. 5 urban air pollution forecasting using multivariate ensemble learning accounting for evolving target distributions. *Chemosphere* 364, 143097. doi:10.1016/j.chemosphere.2024.143097

Rybarczyk, Y., and Zalakeviciute, R. (2018). Machine learning approaches for outdoor air quality modelling: a systematic review. *Appl. Sci.* 8, 2570. doi:10.3390/app8122570

Sarkar, N., Gupta, R., Keserwani, P. K., and Govil, M. C. (2022). Air quality index prediction using an effective hybrid deep learning model. *Environ. Pollut.* 315, 120404. doi:10.1016/j.envpol.2022.120404

Tagliabue, L. C., Cecconi, F. R., Rinaldi, S., and Ciribini, A. L. C. (2021). Data driven indoor air quality prediction in educational facilities based on iot network. *Energy Build.* 236, 110782. doi:10.1016/j.enbuild.2021.110782

Thongthammachart, T., Araki, S., Shimadera, H., Eto, S., Matsuo, T., and Kondo, A. (2021). An integrated model combining random forests and wrf/cmaq model for high accuracy spatiotemporal pm2. 5 predictions in the kansai region of Japan. *Atmos. Environ.* 262, 118620. doi:10.1016/j.atmosenv.2021.118620

Wang, X., Ahmad, I., Javeed, D., Zaidi, S. A., Alotaibi, F. M., Ghoneim, M. E., et al. (2022). Intelligent hybrid deep learning model for breast cancer detection. *Electronics* 11, 2767. doi:10.3390/electronics11172767

Wang, Y., Wang, H., and Zhang, S. (2020). Prediction of daily pm2. 5 concentration in China using data-driven ordinary differential equations. *Appl. Math. Comput.* 375, 125088. doi:10.1016/j.amc.2020.125088

Wu, H., Xu, J., Wang, J., and Long, M. (2021). Autoformer: decomposition transformers with auto-correlation for long-term series forecasting. *Adv. neural Inf. Process. Syst.* 34, 22419–22430. doi:10.48550/arxiv.2106.13008

Yu, S., Peng, J., Ge, Y., Yu, X., Ding, F., Li, S., et al. (2024). A traffic state prediction method based on spatial–temporal data mining of floating car data by using autoformer architecture. *Computer-Aided Civ. Infrastructure Eng.* 39, 2774–2787. doi:10.1111/mice.13179

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., et al. (2020). Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sens. Environ.* 241, 111716. doi:10.1016/j.rse.2020.111716

Zaini, N., Ean, L. W., Ahmed, A. N., and Malek, M. A. (2022). A systematic literature review of deep learning neural network for time series air quality forecasting. *Environ. Sci. Pollut. Res.* 29, 4958–4990. doi:10.1007/s11356-021-17442-1

Zeng, Q., Wang, L., Zhu, S., Gao, Y., Qiu, X., and Chen, L. (2023). Long-term pm2. 5 concentrations forecasting using ceemdan and deep transformer neural network. *Atmos. Pollut. Res.* 14, 101839. doi:10.1016/j.apr.2023.101839

Zhang, Z., and Zhang, S. (2023). Modeling air quality pm2. 5 forecasting using deep sparse attention-based transformer networks. *Int. J. Environ. Sci. Technol.* 20, 13535–13550. doi:10.1007/s13762-023-04900-1

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., et al. (2021). Informer: beyond efficient transformer for long sequence time-series forecasting. *Proc. AAAI Conf. Artif. Intell.* 35, 11106–11115. doi:10.1609/aaai.v35i12.17325

Zhou, T., Ma, Z., Wen, Q., Sun, L., Yao, T., Yin, W., et al. (2022a). Film: frequency improved legendre memory model for long-term time series forecasting. *Adv. neural Inf. Process. Syst.* 35, 12677–12690. doi:10.48550/arxiv.2205.08897

Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., and Jin, R. (2022b). Fedformer: frequency enhanced decomposed transformer for long-term series forecasting, in International conference on machine learning *(PMLR)*, 27268–27286.

Zhou, Y., De, S., Ewa, G., Perera, C., and Moessner, K. (2018). Data-driven air quality characterization for urban environments: a case study. *IEEE Access* 6, 77996–78006. doi:10.1109/access.2018.2884647