Dustformer: A PM10 Concentration Prediction Model Integrating Spatiotemporal Modeling and Mixture-of-Experts Mechanism

Mingchao Zhang[1], Yongsheng Wang[1,*], Delong Zhang[2], Guolin Zhang[1]

[1] College of Intelligent Science and Technology (College of Cyberspace Security), Inner Mongolia University of Technology, Hohhot 010080, China

[2] Meteorological Data Center of Inner Mongolia Autonomous Region, Hohhot 010051, China

[*] Corresponding author: Yongsheng Wang (wangys@imut.edu.cn).

***Abstract*—Elevated PM10 concentrations, driven by both anthropogenic emissions and frequent sand and dust events, pose serious threats to public health and infrastructure; accurate PM10 forecasting is therefore essential for pollution control and resilient energy operations. This study proposes Dustformer, a spatiotemporal forecasting framework designed to fuse ground observations, meteorological data, satellite remote sensing data, and topography data. The proposed architecture integrates a Mixture-of-Experts (MoE) Transformer with dynamic routing to capture complex multimodal temporal patterns, coupled with an attention-weighted Dynamic Graph Neural Network (DGNN) to explicitly model evolving inter-station pollutant diffusion paths. Additionally, a robust sample screening mechanism is incorporated to mitigate the impact of data noise. Extensive experiments on real-world datasets demonstrate that Dustformer outperforms several representative baseline methods in terms of predictive accuracy and robustness. Furthermore, systematic ablation studies confirm the indispensability of each module, revealing that adaptive temporal expert routing and dynamic spatial dependency learning are critical for minimizing forecasting errors.**

***Index Terms*—PM10 Prediction, Spatiotemporal Graph Neural Network, Attention Mechanism, Mixture-of-Experts Model**

I. Introduction

Frequent sand and dust events, together with anthropogenic emissions, contribute to elevated concentrations of inhalable particulate matter, particularly PM10 (aerodynamic diameter 10m), which degrade visibility, attenuate surface solar radiation, and pose long-term cardiovascular and respiratory risks to exposed populations [1]–[4]. Dust deposition also reduces photovoltaic conversion efficiency and accelerates abrasive erosion of wind-turbine components, threatening the reliable operation of large-scale renewable energy systems and transmission infrastructure [5]–[7]. These multifaceted impacts motivate accurate short-to-medium-term PM10 forecasting to support environmental management and resilient energy operations.

This work was supported by the National Natural Science Foundation of China (Project No. 62366039), Inner Mongolia Autonomous Region Science and Technology Projects (Project Nos. 2023YFSH0066 and 2025YFHH0045), and the Scientific Research Project of Inner Mongolia Higher Education Institutions (Project No. JMZD202301).

Accurate PM10 prediction is difficult because emissions, transport, and geography interact across multiple spatial and temporal scales, producing nonlinear, nonstationary, and heterogeneous patterns with both long-range dependencies and abrupt events. Classical sequential learners such as LSTM and GRU can model temporal dynamics but often struggle with very long-range dependencies, are sensitive to nonstationarity, and may require careful decomposition or long training to avoid overfitting [8], [9]. Tree-based and ensemble methods (RF, XGBoost) handle nonlinear feature interactions and missing data robustly, yet they lack native mechanisms for capturing temporal continuity and spatiotemporal structure and therefore depend heavily on handcrafted features or lagged inputs [10]. CNNs extract local patterns effectively but assume Euclidean grid structures, limiting their ability to exploit irregular, non-Euclidean station networks without additional graph or interpolation-based preprocessing [11], [12]. Hybrid architectures (e.g., GC-LSTM, 1D-CNN + BiLSTM, IDW-BiLSTM, Bi-AGRU with attention) combine complementary strengths but increase model complexity and computational cost, and they can still fail to generalize under heavy non-stationarity or across heterogeneous regions [13]–[16]. Signal-decomposition approaches (EMD, EEMD) reduce non-stationarity and improve some metrics but introduce extra preprocessing steps, computational overhead, and challenges for multivariate, long-sequence forecasting [17], [18]. Transformer variants (Reformer, Informer, Flashformer, Flowformer) address long-range temporal dependence and parallelism via attention mechanisms, yet many neglect irregular spatial topology and can be brittle when temporal regimes change rapidly or when multimodal temporal patterns are present [19]–[25].

In this work, we propose Dustformer, a unified spatiotemporal forecasting framework designed to address these challenges. Dustformer integrates multi-source heterogeneous data—including ground observations, meteorology, satellite remote sensing, and topography—through a rigorous spatial-temporal alignment scheme. Methodologically, it employs a Mixture-of-Experts(MoE) Transformer to adaptively model multimodal and nonstationary temporal patterns, coupled with an attention-based Dynamic Graph Neural Network(DGNN) to capture evolving inter-station pollutant diffusion. Experimental results on real-world datasets demonstrate that Dustformer significantly outperforms representative spatiotemporal baselines in 8-hour PM10 single-step forecasting tasks. The main contributions are summarized as follows:

- We introduce Dustformer, a novel framework that jointly models adaptive temporal dynamics via MoE and dynamic spatial dependencies via DGNN. This unified architecture effectively captures the complex, non-linear transport and evolution mechanisms of PM10.
- We propose a robust multi-source data fusion framework that effectively integrates static topographic constraints with dynamic meteorological and remote sensing features. This strategy resolves spatiotemporal resolution mismatches and significantly enriches the feature representation for accurate forecasting.
- Extensive experiments conducted on multiple real-world datasets demonstrate the superiority of Dustformer over state-of-the-art methods. The proposed model achieves superior prediction accuracy, highlighting its practical effectiveness in air quality prediction.

II. Data Processing and Feature Construction

Meteorological data used in this study were obtained from the Meteorological Data Center of the Inner Mongolia Autonomous Region (http://nm.cma.gov.cn/zfxxgk/), including seven key variables: inhalable particulate matter (PM10), relative humidity (RHU), wind speed (WS), wind direction (WD), atmospheric pressure (PRS), air temperature (TEM), and horizontal visibility (VIS). Observations were collected at Monitoring stations across 10 regions, corresponding to the graph nodes shown in Fig. 1. In addition to ground-based observations, FY-4 AGRI satellite dust products (e.g., DST and IDDI-BK) are incorporated to supply spatially continuous dust information, together with a high-resolution DEM from the National Geomatics Center of China. FY-4 gridded observations are temporally aggregated and spatially interpolated to station locations, while the DEM is used to derive static topographic covariates, including elevation and local slope, for each monitoring station.

A. Spatiotemporal Alignment and Multi-Source Data Mapping

PM10 concentrations and meteorological variables (RHU, WS, WD, PRS, TEM, and VIS) were collected from multiple monitoring stations. To ensure spatiotemporal consistency among heterogeneous data sources, each PM10 station located at (ϕ_p, λ_p) is associated with its nearest meteorological station $s^* \in S$ by minimizing the Haversine spherical distance:

$$s^* = \underset{s \in S}{\operatorname{argmin}}\, d_H(s,p), \tag{1}$$

where matches are accepted only when $d_H < D_{\text{max}}$ (with $D_{\text{max}} = 200\text{km}$). To account for slight discrepancies in observation times, each PM10 record is paired with the closest meteorological observation within a maximum time difference $|t_s - t_p| < T_{\text{max}}$ (with T_{max} set to 1 hour); unmatched records are discarded.

Subsequently, to incorporate spatially continuous dust information, FY-4 AGRI dust storm detection (DSD) products, including DST and IDDI-BK, are processed. Given the 4-km gridded resolution of FY-4 data, hourly aggregation is first applied to maintain temporal consistency with ground observations. Inverse distance weighting (IDW) is then employed to interpolate gridded remote sensing values to station locations [26]:

$$v_p = \sum_{i=1}^{n} w_i v_i, \quad w_i = \frac{\frac{1}{d_i^\alpha}}{\sum_{j=1}^{n} \frac{1}{d_j^\alpha}}, \tag{2}$$

where v_i denotes the value of the i-th pixel, d_i is the distance between pixel i and station p, and α is the interpolation power. This procedure ensures consistent spatial mapping of remote sensing indicators to ground-based monitoring stations.

Furthermore, to capture terrain-induced modulation of airflow and pollutant dispersion, a high-resolution digital elevation model (DEM) is obtained from the National Geomatics Center of China. We spatially align the DEM grid with the coordinates of each monitoring station. Specifically, elevation values are derived via bilinear interpolation, while the local slope θ_p for station p is calculated based on the surface gradient of the elevation field E:

$$\theta_p = \arctan\left(\sqrt{\left(\frac{\partial E}{\partial x}\right)^2 + \left(\frac{\partial E}{\partial y}\right)^2}\right). \tag{3}$$

B. Feature Construction and Time-Series Sample Generation

After spatial mapping and temporal alignment, a unified multi-source feature set is constructed for each station and time step. The integrated feature vector comprises ground-based PM10 measurements, matched meteorological variables, interpolated FY-4 remote sensing indicators (DST and IDDI-BK), and static topographic attributes (elevation and slope).

To improve data quality, missing values are filled using feature-wise mean imputation. Furthermore, to mitigate the impact of sensor malfunctions, a physical range check is applied following standard quality assurance procedures [27]. Specifically, observations falling outside the valid instrument range—strictly defined as negative values or readings exceeding the sensor's measurement limit—are identified as non-physical artifacts and filtered out. This strategy effectively removes sensor noise while strictly preserving genuine high-concentration dust episodes. All numerical features are standardized using Z-score normalization prior to model training. Finally, time-series samples are generated using a sliding window approach: for station i at prediction time t, the model input consists of the concatenation of multi-source feature vectors from the preceding 24 time steps, while the output is the PM10 concentration at time t:

$$X^{(i)} = \left[x_{t-24}^{(i)}, x_{t-24+1}^{(i)}, \ldots, x_{t-1}^{(i)}\right], \quad y^{(i)} = PM10_t^{(i)}, \tag{4}$$

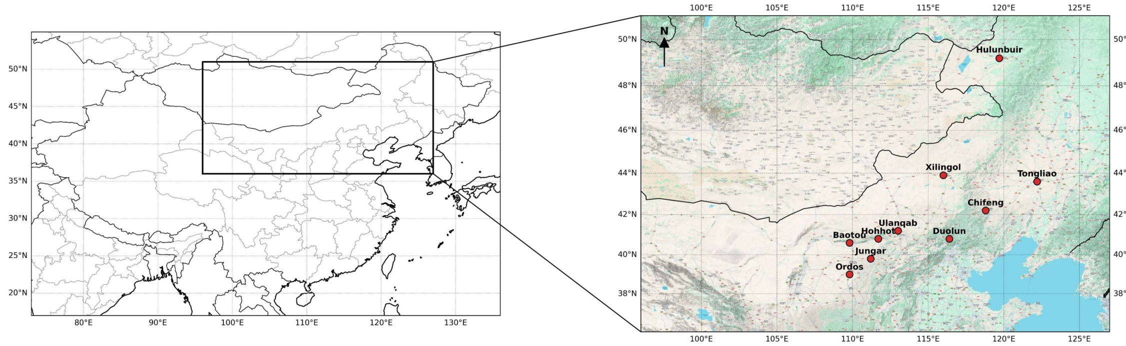

Fig. 1. Location of each major monitoring station.

here, $x_t^{(i)}$ denotes the fused feature vector for station i at time t.

III. METHODOLOGY AND TECHNICAL

A. Model Structure Design: Dustformer

To comprehensively capture the spatiotemporal evolution of air pollution, this paper proposes a multi-source information-driven spatiotemporal joint prediction model, termed Dustformer. As illustrated in Fig. 2, the model explicitly integrates a graph-based spatial representation with temporal sequence modeling, while fusing multi-source meteorological data to enable accurate PM10 concentration prediction.

B. Temporal Modeling Module

Unlike standard Transformer blocks that utilize a shared Feed-Forward Network (FFN) for all inputs, our module consists of parallel sub-experts. Each sub-expert maintains independent parameters and implicitly specializes in modeling distinct pollution regimes—such as steady background fluctuations, rapid accumulation, or dissipation. For any given input token representation, a learnable gating network dynamically calculates the routing weights, determining which experts are best suited to process the current temporal state. The gating score is computed as:

$$g = \text{Softmax}(W_g \cdot \text{MeanPool}(x) + b_g) \in \mathbb{R}^{N_e}, \quad (5)$$

then, the top-2 experts with the highest weights are selected for forward propagation, and the final output is:

$$y = \sum_{i \in k} g_i \cdot r_i(x), \quad (6)$$

where k denotes the set of the selected top-k experts, g_i represents the normalized gating coefficient for the i-th expert, and r_i is the output of the corresponding expert network.

This adaptive activation mechanism is visualized in Fig. 3 As observed, the gating network effectively decouples the pollution process: Expert 1 dominates during the stable periods (low PM10 concentration), while the model automatically switches dominance to Expert 3 during the rapid rising phase of the dust storm. This demonstrates that the MoE module successfully alleviates the challenges of nonlinearity by decomposing complex time-series data into manageable sub-patterns via input-adaptive routing.

C. Spatial Modeling Module

The spatial distribution of PM10 concentration during dust events is not only affected by local pollution sources but also by the combined effects of large-scale diffusion of dust air masses, wind field transportation, and topographic guidance. Traditional temporal modeling methods often only focus on the historical information of a single point and cannot describe the spatial influence relations between different monitoring points, thus limiting the global prediction ability of the model.

To effectively characterize the transmission relations of pollutants between multiple stations, Dustformer designs a Dynamic Graph Neural Network with adaptive attention weights in the spatial dimension, which constructs a transmission structure graph between stations to simulate the regional diffusion pattern of dust pollution and fully integrates the spatial distribution features of remote sensing information.

Given N monitoring stations, the spherical distance d_{ij} between any two stations is first calculated based on their geographical longitude and latitude (ϕ_i, λ_i). If $d_{ij} < d_{th}$ (the set threshold is 200 km), $A_{ij} = 1$; otherwise, $A_{ij} = 0$. An initial adjacency matrix $A \in \mathbb{R}^{N \times N}$ is constructed, which is used to characterize the possible influence paths of dust in geographical space, such as the transportation chain from the source area to the downwind area. However, the pollution diffusion path in the actual dust process is non-uniform and dynamic, and static distance information alone cannot fully model it. To this end, this paper introduces a learnable attention weight matrix $W_{attn} \in \mathbb{R}^{N \times N}$ to dynamically adjust the weight of each edge, and normalization is achieved through softmax to form the final spatial transmission graph:

$$\tilde{A}_{ij} = \frac{\exp(W_{attn}^{ij}) \cdot A_{ij}}{\sum_{k \in N(i)} \exp(W_{attn}^{ik}) \cdot A_{ik}}, \quad (7)$$

where $\tilde{A}_{ij}$ is the normalized edge weight. W_{attn}^{ik} denotes the learnable interaction parameter, and A_{ij} is the binary adjacency matrix acting as a structural mask. The softmax function is applied only to the valid neighbors $k \in N(i)$ to

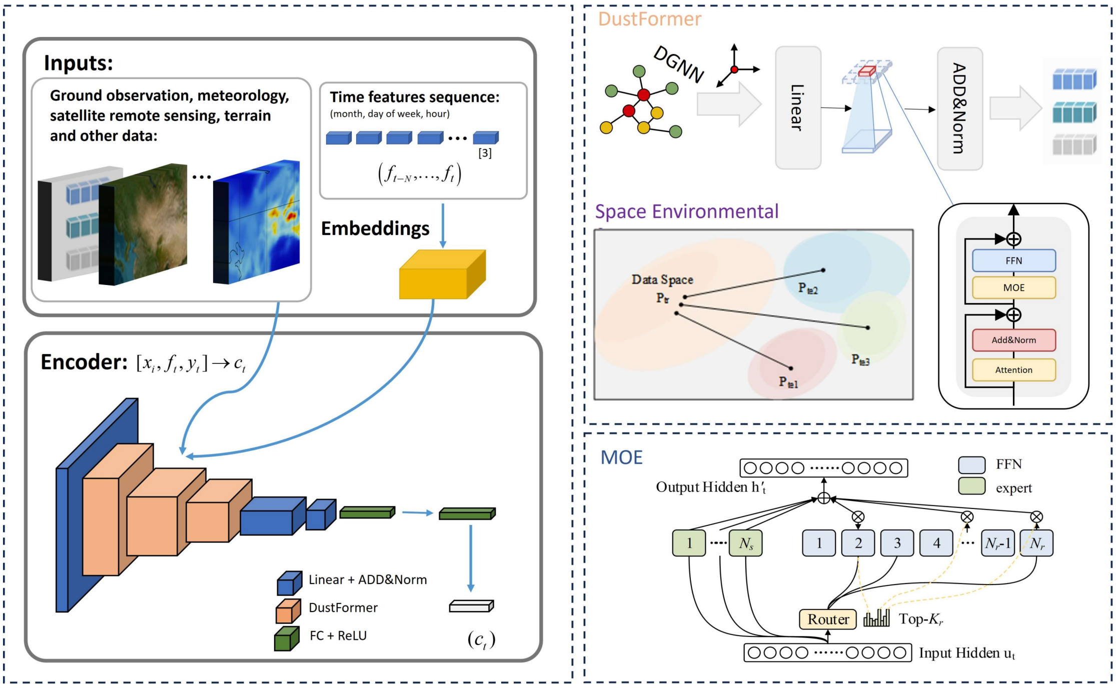

Fig. 2. Dustformer model diagram.

ensure that the sum of attention weights equals 1, guaranteeing numerical stability during feature aggregation.

This mechanism allows the model to automatically learn key transmission paths during training, such as the main dust channels detected by remote sensing or wind direction-dominated paths.

Leveraging the constructed spatial attention graph $\tilde{A}_{ij}$, Dustformer uses a multi-layer graph neural network to iteratively update the spatial features of each station. Initially, a learnable static embedding e_i is assigned to each station. At each layer l, the representation $h_i^{(l)}$ of station i is updated by weighting the embeddings of its neighbor nodes:

$$h_i^{(l+1)} = \sigma\left(\sum_{j\in N(i)} \tilde{A}_{ij}\cdot W^{(l)}\cdot h_j^{(l)} + b^{(l)}\right), \tag{8}$$

where $W^{(l)}$ is the weight matrix and σ is the activation function. Through this process, each station "perceives" the pollution trends in its surrounding area. Finally, to preserve the unique identity of each station, we add a residual connection to the output:

$$h_i^{(l+1)} := h_i^{(l+1)} + e_i, \tag{9}$$

where e_i is the initial learnable embedding of station i.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

To comprehensively verify the effectiveness and superiority of the proposed Dustformer model, systematic experiments were conducted on the constructed multi-source fusion dataset. The evaluation is organized into four parts: experimental setup, comparative prediction accuracy, parameter sensitivity analysis, and module ablation studies.

A. Experimental Setup and Dataset Partitioning

To ensure temporal consistency and rigorous evaluation, the dataset is chronologically partitioned into three subsets: a training set(January 2022 to June 2023), a validation set (July 2023 to September 2023), and a test set (October 2023 to December 2023). During preprocessing, ground observations, meteorological data, remote sensing imagery, and topographic data are temporally aligned and normalized. Input sequences are generated using a sliding window mechanism. The validation set is employed for hyperparameter tuning and early stopping, while the test set serves to evaluate the model's generalization ability on unseen samples.

B. Main Prediction Performance Comparison

We compared Dustformer against a comprehensive set of baselines, including spatiotemporal graph models (STGNN, GCN), ensemble tree models (LightGBM, Random Forest), and deep temporal models (Flashformer, Transformer). The quantitative results are summarized in Table I.

As presented in Table I, Dustformer demonstrates significant superiority over all baseline models, achieving state-of-the-art performance across all evaluation metrics. This performance advantage validates the effectiveness of the proposed architecture, which jointly models spatiotemporal dependencies through the integration of attention mechanism, DGNN, and the Mixture-of-Experts mechanism.

To provide a more intuitive assessment of prediction alignment, Fig. 4 visualizes the comparative fitting curves of

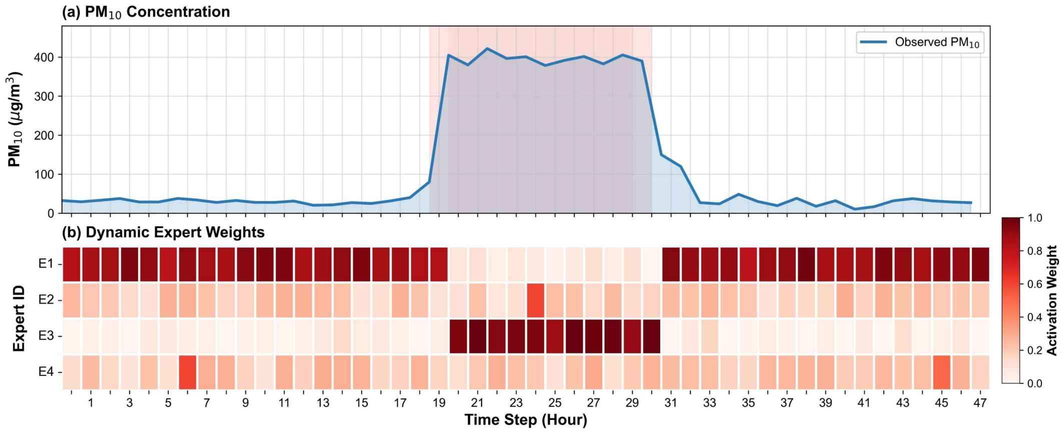

Fig. 3. Dynamic Evolution of Expert Activation Weights during Stable Periods and Dust Storms.

TABLE I
COMPARISON OF PREDICTION RESULTS FROM DIFFERENT MODELS

Model	MAE	MSE	RMSE	R^2
Dustformer	0.785	1.943	1.394	0.967
STGNN	0.970	3.746	1.935	0.923
Flashformer	1.387	9.250	3.041	0.884
LightGBM	1.274	4.18	2.048	0.910
Random Forest	1.802	9.876	3.142	0.873
SVR	2.002	6.367	2.523	0.850
Transformer	1.179	7.395	2.719	0.908
MLP	2.029	6.248	2.499	0.822
ARIMA	1.911	7.768	2.787	0.883
GCN	1.058	5.733	2.394	0.916

TABLE II
PARAMETER SENSITIVITY ANALYSIS

Parameter	Setting	MAE	MSE	Time (ms)
MoE Experts	4	0.852	2.017	82.3
	8	0.785	1.943	125.6
	16	0.791	1.951	201.2
Attn. Heads	2	1.013	2.345	68.7
	4	0.785	1.943	92.5
	8	0.772	1.921	156.3
Graph Weight	7:3	0.927	2.153	91.6
	5:5	0.785	1.943	88.2
	3:7	0.894	2.098	75.4

Dustformer against representative baseline models over a continuous 300-hour test period. The ground truth PM10 series (black solid line) exhibits high volatility with sharp peaks and deep valleys, presenting a significant challenge for forecasting. As observed across subplots Fig. 4(a) through (i), the red dashed line corresponding to Dustformer tracks the true value trajectory with remarkable fidelity. Crucially, Dustformer demonstrates superior capability in capturing abrupt changes and extreme values where baseline models often falter. For instance, compared to spatiotemporal graph models like STGNN (Fig. 4(a)) or pure temporal models like Flashformer (Fig. 4(b)), Dustformer shows significantly less lag during rapid fluctuations and tighter alignment with peak concentrations. This visual evidence strongly corroborates the quantitative metrics in Table I affirming the model's robustness in handling non-stationary air quality data.

Further analysis highlights that explicitly modeling spatial topology is critical. Spatiotemporal graph-based approaches (e.g., STGNN, GCN) rank second in Table I and show reasonable trend-following in Fig. 4(a) and (i), significantly outperforming pure temporal models such as Flashformer and Transformer. This indicates that capturing cross-regional transmission dynamics via graph structures is essential for accurate diffusion modeling.

It is also noteworthy that LightGBM exhibits competitive performance (MSE=4.18), surpassing complex temporal models like Flashformer. This observation suggests that while attention mechanisms excel at capturing temporal dynamics, they struggle with spatial heterogeneity in the absence of explicit structural guidance. In contrast, ensemble tree models can effectively exploit structured meteorological and geographical features. However, traditional machine learning models (Random Forest, SVR) perform the worst, underscoring the necessity of deep spatiotemporal fusion architectures for high-precision forecasting.

C. Hyperparameter Sensitivity Analysis

To identify the optimal configuration between accuracy and computational efficiency, we analyzed three key hyperparameters: the number of MoE experts, attention heads, and dynamic

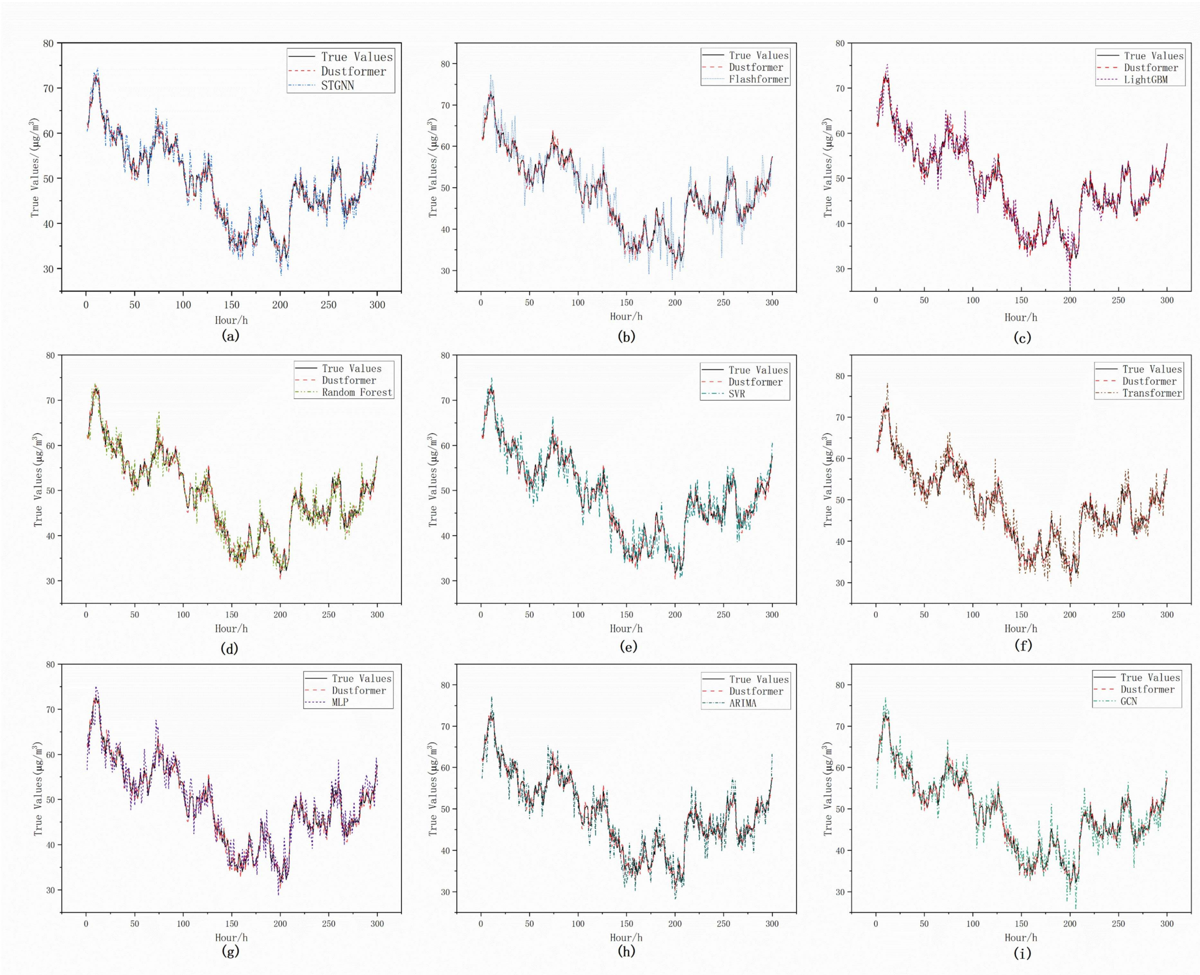

Fig. 4. Renderings of different model.

graph weights. The experiments were conducted under single-sample inference settings, with results detailed in Table II.

Regarding the Mixture-of-Experts, increasing the number of experts from 4 to 8 yields significant accuracy gains. However, expanding to 16 experts provides negligible improvement while distinctly increasing inference time. Consequently, 8 experts represent the optimal trade-off point, balancing model capacity with computational cost. Similarly, for attention heads, insufficient heads (e.g., 2) fail to capture multi-scale temporal dependencies, while 8 heads double the computational cost with only marginal gains. Thus, 4 heads are selected as the default configuration.

For the dynamic graph weights (geography vs. meteorology), the analysis in Table II and Fig. 5(b) presents a similar trade-off pattern. A meteorology-biased ratio (3:7) yields the fastest execution but compromises accuracy due to weak spatial constraints, whereas a geography-dominant ratio (7:3) incurs the highest time cost without providing performance benefits. Consequently, the balanced 5:5 ratio is selected as the optimal setting, achieving the lowest error (MAE=0.785) by effectively harmonizing static topology with dynamic weather evolution.

D. Module Ablation Experiment

To verify the contribution of individual components, we conducted ablation experiments focusing on neural modules and the inclusion of DEM data.

The ablation results in Table III demonstrate a clear synergistic effect among the proposed modules. The full Dustformer configuration consistently outperforms all partial combinations. Specifically, removing the Dynamic Graph module or the attention mechanism module leads to noticeable performance degradation. This indicates that temporal dynamics and spatial diffusion must be modeled simultaneously to achieve optimal accuracy.

Furthermore, the introduction of DEM data plays a pivotal role in enhancing model robustness. As shown in the ablation study results (Fig. 5(c)), incorporating static topographic information significantly improves the overall prediction accuracy. This validates that static topographic covariates provide essential physical constraints for modeling airflow and pollutant dispersion, enabling the model to capture terrain-induced

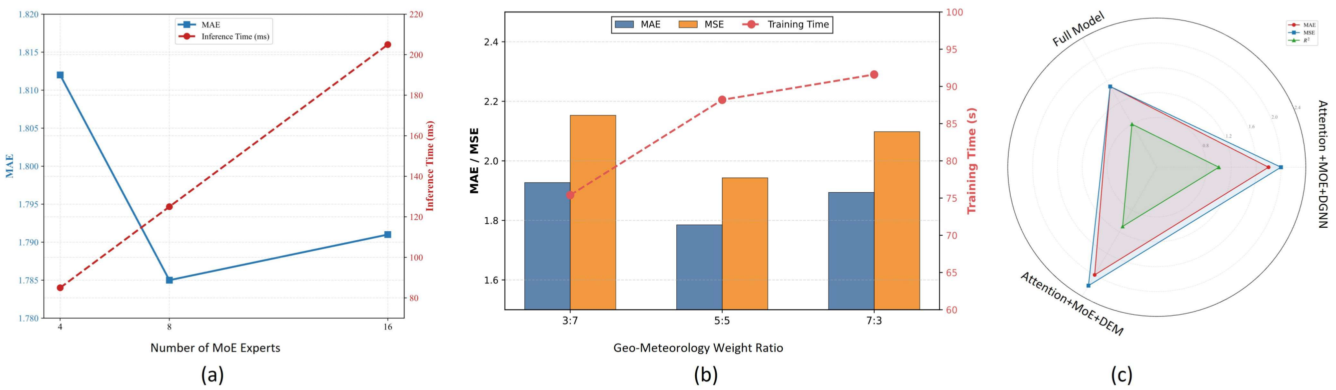

Fig. 5. Model Architecture Optimization Experiments Visualization. (a) Trade-off Between MoE Expert Number and Precision-Efficiency. (b) Sensitivity Analysis of Model Performance to Weight Ratios (c) Multi-module Combination Synergy Verification by Radar Chart.

TABLE III
ABLATION STUDY OF DIFFERENT COMPONENTS

Attention	**DGNN**	**MOE**	**DEM**	**MAE**	**MSE**	**R^2**
✓		✓		0.882	2.150	0.932
	✓	✓		0.853	2.001	0.953
✓	✓			0.905	2.214	0.915
✓		✓	✓	0.837	1.985	0.951
	✓	✓	✓	0.858	2.036	0.930
✓	✓	✓	✓	0.785	1.943	0.967

diffusion patterns that are often overlooked by purely data-driven approaches.

V. CONCLUSION

In this study, we proposed Dustformer, a novel spatiotemporal forecasting framework designed to enhance PM10 concentration prediction by integrating ground observations, meteorological data, FY-4 satellite products, and DEM data. By coupling a Mixture-of-Experts Transformer with a learnable Dynamic Graph Neural Network, the proposed model effectively captures long-range temporal dependencies and dynamic spatial diffusion patterns associated with dust events.

Our extensive experiments reveal a critical trade-off between prediction accuracy and computational efficiency. Specifically, the hyperparameter analysis demonstrates that a balanced configuration—comprising 8 MoE experts and a 5:5 dynamic weight ratio between geographical and meteorological constraints—achieves the optimal performance sweet spot. This setting significantly reduces prediction error while maintaining a reasonable inference cost, outperforming static graph baselines and single-modal approaches.

In future work, we plan to extend the Dustformer framework to multi-pollutant prediction tasks (e.g., PM2.5 and O3) and explore the integration of physical diffusion equations to further enhance interpretability and robustness in extreme weather scenarios.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Meteorological Data Center of Inner Mongolia Autonomous Region for providing the high-quality meteorological and PM10 monitoring datasets, which are fundamental to the development and validation of the Dustformer model proposed in this study. The authors also gratefully acknowledge the anonymous reviewers and the editorial board members for their insightful feedback and constructive suggestions, which have significantly enhanced the clarity, rigor, and overall quality of the manuscript. Additionally, we thank our research colleagues for their continuous support, technical assistance, and valuable discussions throughout the entire research process.

REFERENCES

[1] C. Xu, "Spatiotemporal evolution characteristics and source contribution of sandstorms in northern China," Ph.D. dissertation, Lanzhou Univ., Lanzhou, China, 2020.

[2] Y. O. Khaniabadi, G. Goudarzi, S. M. Daryanoosh, *et al.*, "Exposure to PM10, NO2, and O3 and impacts on human health," *Environ. Sci. Pollut. Res.*, vol. 24, pp. 2781–2789, 2017.

[3] L. Huang, D. Liu, and D. J. Cai, "Health risk assessment of exposure to multiple pollutants in Guangzhou," *China Environ. Sci.*, vol. 42, no. 11, pp. 5418–5426, 2022.

[4] G. Z. Ju, J. R. Wang, and C. Cui, "Study on the impact of extreme weather events on renewable energy generation and grid operation," *Smart Power*, vol. 50, no. 11, pp. 77–83, 2022.

[5] M. Z. Zhao, L. L. Wu, S. Wang, *et al.*, "Experimental study on the effect of desert dust on photovoltaic module output performance," *Thermal Sci. Technol.*, vol. 20, no. 4, pp. 340–347, 2021.

[6] E. P. Liang, G. S. Ma, Y. Li, *et al.*, "Review of the mechanical effects of sandstorm environment on key components of wind turbines," *Sci. Sin. Phys. Mech. Astron.*, vol. 53, no. 3, pp. 133–145, 2023.

[7] J. Y. Wang, K. Ding, Y. L. Sun, *et al.*, "Resilience assessment and enhancement of new energy bases under severe sandstorms," *Power Syst. Technol.*, pp. 1–16, 2024.

[8] P. D. Sampson and C. Barbour, "Kalman filtering for air quality management," *Environ. Monit. Assess.*, vol. 97, no. 1–3, pp. 55–62, Jan. 2004.

[9] C. Wen, S. Liu, X. Yao, *et al.*, "A novel spatiotemporal convolutional long short-term neural network for air pollution prediction," *Sci. Total Environ.*, vol. 654, pp. 1091–1099, 2019.

[10] J. Becerra-Rico, M. A. Aceves-Fernández, K. Esquivel-Escalante, *et al.*, "Airborne particle pollution predictive model using Gated Recurrent Unit (GRU) deep neural networks," *Earth Sci. Inf.*, vol. 13, pp. 821–834, 2020.

[11] J. Bi, K. E. Knowland, C. A. Keller, and Y. Liu, "Combining machine learning and numerical simulation for high-resolution PM2.5 concentration forecast," *Environ. Sci. Technol.*, vol. 56, no. 3, pp. 1544–1556, 2022.

[12] J. Peng, H. Han, Y. Yi, H. Huang, and L. Xie, "Machine learning and deep learning modeling and simulation for predicting PM2.5 concentrations," *Chemosphere*, vol. 308, Art. no. 136353, 2022.

[13] K. Zhang, X. Yang, H. Cao, *et al.*, "Multi-step forecast of PM2.5 and PM10 concentrations using convolutional neural network integrated with spatial–temporal attention and residual learning," *Environ. Int.*, vol. 171, Art. no. 107691, 2023.

[14] S. Chae, J. Shin, S. Kwon, *et al.*, "PM10 and PM2.5 real-time prediction models using an interpolated convolutional neural network," *Sci. Rep.*, vol. 11, no. 1, Art. no. 11952, 2021.

[15] Y. Qi, Q. Li, H. Karimian, and D. Liu, "A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory," *Sci. Total Environ.*, vol. 664, pp. 1–10, 2019.

[16] S. Du, T. Li, Y. Yang, and S. J. Horng, "Deep air quality forecasting using hybrid deep learning framework," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 6, pp. 2412–2424, 2019.

[17] K. Zhang, J. Thé, G. Xie, and H. Yu, "Multi-step ahead forecasting of regional air quality using spatial-temporal deep neural networks: a case study of Huaihai Economic Zone," *J. Clean. Prod.*, vol. 277, Art. no. 123231, 2020.

[18] Z. Ebrahimi-Khusfi, R. Taghizadeh-Mehrjardi, and A. R. Nafarzadegan, "Accuracy, uncertainty, and interpretability assessments of ANFIS models to predict dust concentration in semi-arid regions," *Environ. Sci. Pollut. Res.*, vol. 28, no. 6, pp. 6796–6810, 2021.

[19] E. H. Liu, Y. J. Fu, Z. Zhang, *et al.*, "PM2.5 concentration prediction network based on Transformer attention mechanism," *J. Saf. Environ.*, vol. 23, no. 10, pp. 3760–3768, 2023.

[20] M. Y. Liu, B. W. Cui, Y. K. Wang, *et al.*, "PM2.5 concentration prediction using Transformer model with attention mechanism," *Environ. Sci.*, pp. 1–15, 2024.

[21] N. Kitaev, L. Kaiser, and A. Levskaya, "Reformer: The efficient transformer," *arXiv preprint arXiv:2001.04451*, 2020.

[22] H. Zhou, S. Zhang, J. Peng, *et al.*, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 12, pp. 11106–11115, 2021.

[23] W. Hua, Z. Dai, H. Liu, *et al.*, "Transformer quality in linear time," in *Proc. Int. Conf. Mach. Learn.*, PMLR, 2022, pp. 9099–9117.

[24] Z. Huang, X. Shi, C. Zhang, *et al.*, "Flowformer: A transformer architecture for optical flow," in *Eur. Conf. Comput. Vis.*, Cham: Springer, 2022, pp. 668–685.

[25] M. A. A. Al-qaness, A. Dahou, A. A. Ewees, *et al.*, "ResInformer: residual transformer-based artificial time-series forecasting model for PM2.5 concentration in three major Chinese cities," *Mathematics*, vol. 11, no. 2, Art. no. 476, 2023.

[26] D. Shepard, "A two-dimensional interpolation function for irregularly-spaced data," in Proc. 23rd ACM National Conference, 1968, pp. 517–524.

[27] M. A. Shafer, C. A. Fiebrich, D. S. Arndt, et al., "Quality assurance procedures in the Oklahoma Mesonetwork," in *Journal of Atmospheric and Oceanic Technology*, vol. 17, no. 4, pp. 474–494, 2000.