Geophysical Research Letters

RESEARCH LETTER
10.1029/2026GL123448

iTransformer-Informed Hourly Surface Ozone Forecasting Using Geostationary Satellite Observations

Fan Cheng[1,2,3], Jing Wei[1], Fuzhong Weng[4], Shan Zhou[3], Zeyu Yang[5], Yang Cheng[6], Jian Gao[7], Xinghua Qiu[1], and Tong Zhu[1]

[1]MEEKL-AERM, College of Environmental Sciences and Engineering, Institute of Tibetan Plateau, and Center for Environment and Health, Peking University, Beijing, China, [2]School of Environmental Science and Engineering, Shanghai Jiao Tong University, Shanghai, China, [3]School of the Environment and Sustainable Engineering, Eastern Institute of Technology, Ningbo, China, [4]CMA Earth System Modeling and Prediction Centre, China Meteorological Administration, Beijing, China, [5]School of Environment, Tsinghua University, Beijing, China, [6]School of Astronautics, Harbin Institute of Technology, Harbin, China, [7]Chinese Research Academy of Environmental Sciences, Beijing, China

Abstract Previous machine learning approaches for surface ozone (O_3) forecasting are often limited to site-specific applications and struggle to capture regional pollution dynamics. Here, we introduce an inverted Transformer (iTransformer), a time-series model that integrates hourly satellite-derived surface O_3 with key meteorological and environmental drivers. The model is applied to eastern China and demonstrates strong skill in 72-hr surface O_3 forecasting, achieving an overall correlation of 0.86 (mean bias = 0.29 μg/m^3). It effectively captures spatial patterns, diurnal variability, and high-ozone episodes, maintaining stable performance over time. During pollution events, the model reproduces both localized and large-scale ozone enhancements with high fidelity. Forecast biases remain low, particularly at shorter lead times, indicating robust predictive skill. Overall, the results highlight the ability of advanced deep learning architectures to bridge satellite observations and air quality forecasting, offering a scalable framework for near-real-time ozone prediction and supporting timely environmental and public health decision-making.

Plain Language Summary Surface ozone (O_3) is a harmful air pollutant that threatens human health and ecosystems, making reliable short-term forecasts essential for early warnings. However, many existing machine learning models rely mainly on ground monitoring sites and often struggle to represent regional pollution patterns. In this study, we develop a deep learning model, the inverted Transformer (iTransformer), that combines satellite-derived hourly surface O_3 estimates with meteorological and environmental information to forecast O_3 concentrations up to 72 hr in advance across eastern China. The model successfully captures regional pollution patterns and daily O_3 variations and performs well during severe pollution episodes. Our results show that integrating satellite observations with deep learning can improve regional air-quality forecasting and provide timely information to support public health protection and air pollution management.

Key Points:

- A novel time-series iTransformer framework enables hourly surface O_3 forecasting
- The model captures regional O_3 patterns and diurnal variability up to 72 hr ahead
- Satellite-driven AI forecasting supports regional air-quality forecasts and early pollution warnings

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

J. Wei,
jingwei@pku.edu.cn

Citation:

Cheng, F., Wei, J., Weng, F., Zhou, S., Yang, Z., Cheng, Y., et al. (2026). iTransformer-informed hourly surface ozone forecasting using geostationary satellite observations. *Geophysical Research Letters*, *53*, e2026GL123448. https://doi.org/10.1029/2026GL123448

Received 5 APR 2026
Accepted 11 MAY 2026

Author Contributions:

Data curation: Fan Cheng
Formal analysis: Fan Cheng, Zeyu Yang, Yang Cheng
Funding acquisition: Jing Wei
Methodology: Fan Cheng, Jing Wei
Project administration: Jing Wei
Software: Fan Cheng
Supervision: Jing Wei
Validation: Fan Cheng, Zeyu Yang, Yang Cheng
Writing – original draft: Fan Cheng
Writing – review & editing: Jing Wei, Shan Zhou, Jian Gao, Xinghua Qiu, Tong Zhu

1. Introduction

Surface ozone (O_3) is a major atmospheric pollutant primarily through photochemical processes driven by nitrogen oxides (NO_x) and volatile organic compounds (VOCs) (Wang et al., 2017, 2022). Its amount is also influenced by horizontal air transport and the vertical transport of tropospheric nitrogen dioxide (NO_2) (He et al., 2022; Zhang et al., 2022). High O_3 concentration has adverse effects on ecosystems and plants (Ainsworth et al., 2012; Lin et al., 2018; Yue & Unger, 2014) and imposes notable health burdens. During 2013–2020, O_3 pollution across China has been a serious environmental concern, with the maximum daily 8-hr average (MDA8) O_3 concentration rising by an average increase of 2.49 μg/m^3/yr ($p < 0.001$) (Wei et al., 2022). Critically, over 99% of the Chinese population has been exposed to O_3 levels exceeding the WHO-recommended long-term air quality guideline (AQG) of 60 μg/m^3 (Yang et al., 2025). Given the health risks of ozone pollution, accurate forecasting of ambient O_3 concentrations is essential for protecting public health, enabling individuals to avoid periods of high pollution and reduce exposure by adjusting their outdoor activities (Shams et al., 2025; Yuval et al., 2012).

Chemical transport models (CTMs) have been widely used to forecast surface O_3 concentrations by simulating atmospheric dynamics, physical and chemical processes, and integrating emissions, transport, deposition, and other relevant mechanisms (Sayeed et al., 2022; Wang et al., 2023). For example, NASA GEOS-CF (Chin

et al., 2003; Keller et al., 2021) produces global 5-day forecasts of hourly surface O_3 at a resolution of 0.25°. The CMAQ model serves as the core of the U.S. National air quality forecast system, generating regional forecasts of hourly surface O_3 with a fine spatial resolution of ∼12 km across the United States (Appel et al., 2007). The CAMS delivers daily global forecasts of key pollutants, including O_3, at a spatial resolution of 0.75° using an ensemble of chemistry-transport models within the ECMWF Integrated Forecast System (IFS) framework (Inness et al., 2019). However, these models still suffer from substantial uncertainties and high computational costs, particularly under high-pollution conditions (Gao & Zhou, 2024; Kitayama et al., 2019; Rao et al., 2020).

Recently, machine learning (ML), especially deep learning (DL), has demonstrated strong potential for forecasting O_3 concentrations due to its powerful feature extraction abilities and capability to capture nonlinear relationships (Shinde & Shah, 2018). Sayeed et al. employed a deep convolutional neural network (CNN) model to forecast hourly surface O_3 concentrations 24 hr ahead using meteorological and air pollution variables from the previous day collected at 21 continuous ambient monitoring stations (Sayeed et al., 2020). Yafouz et al. (2022) adopted various ML models to forecast hourly surface O_3 concentrations at the site scale in 2019, using data from three ground-based monitoring stations that included O_3 concentrations and pertinent meteorological variables (Yafouz et al., 2022). Liu et al. employed multi-source time series data from ground-based monitoring stations, including O_3 and other pollutants, as inputs to forecast hourly O_3 concentrations by integrating fuzzy cognitive maps (FCMs) with evidential reasoning theory (X. Liu et al., 2022). Mu et al. used ground-based O_3 time series, together with other atmospheric and pollution variables, as inputs to a seasonal-trend decomposition (STL)-Transformer model for hourly surface O_3 forecasting (Mu et al., 2024). Guo et al. (2025) constructed a hybrid deep learning framework integrating CNN and long short-term memory (LSTM) networks for daily O_3 concentrations forecasting from ground-based monitoring stations in Liaocheng City (Guo et al., 2025). However, most previous research has focused on site-level forecasts, leaving a significant gap in raster-scale surface O_3 forecasting, largely due to a lack of spatially continuous hourly surface O_3 data.

This study introduces a time-series-enhanced deep learning model, the Inverted Transformer (iTransformer), for raster-scale surface O_3 forecasting across eastern China, with lead times of up to 72 hr (3 days). iTransformer offers unique advantages over other deep learning models by simultaneously capturing long-range global dependencies and modeling complex nonlinear temporal and inter-variable relationships. Its inverted-token design further enhances multivariate dependency representation and improves suitability for time-series data, thereby enabling superior performance and interpretability in forecasting applications (Y. Liu et al., 2024). The iTransformer model is trained using our satellite-derived hourly surface O_3 data (Cheng et al., 2025), along with meteorological variables and other auxiliary factors relevant to ozone pollution. To enhance performance under high-pollution conditions, the training data are further augmented by involving additional high O_3 concentration samples. Last, we evaluate the model's forecasting accuracy and uncertainty from multiple perspectives and demonstrate its ability to identify high O_3 pollution events. Additionally, we investigate the spatiotemporal variability of these events under different high-pollution O_3 scenarios, providing insights into the dynamic behavior of O_3 concentrations across different regions.

2. Data and Methods

2.1. Input Data

In this study, eastern China was selected as the study area because it concentrates the majority of the country's population and the most frequent O_3 pollution episodes, making it the region of greatest public concern, while its dense ground-based monitoring network enables robust and comprehensive model validation, unlike the sparsely instrumented western regions. We employed our satellite-derived 24-hr surface O_3 data set at 1-km resolution as the training data set for surface O_3 forecasting (Cheng et al., 2025). It is generated using an advanced Four-Dimensional Space-Time Extra-Trees (4D-STET) model, which integrates geostationary satellite observations, surface O_3 ground-based measurements, major precursors, anthropogenic emissions, meteorological variables, and modeled surface mass concentrations of O_3. The hourly surface O_3 estimates are reliable, yielding an average out-of-sample 10-fold cross-validated coefficient of determination (CV-R^2) of 0.92 for daytime and 0.82 for nighttime, and an overall CV-R^2 of 0.89 for the whole 24-hr period (Cheng et al., 2025).

Several key meteorological variables known to affect surface O_3 concentrations were collected, including downward surface radiation (DSR), ultraviolet radiation (UV), 2 m temperature (TEM), relative humidity (RH), total precipitation (TP), wind speed (WS) and direction (WD), and surface pressure (SP), derived from the ERA5

10.1029/2026GL123448

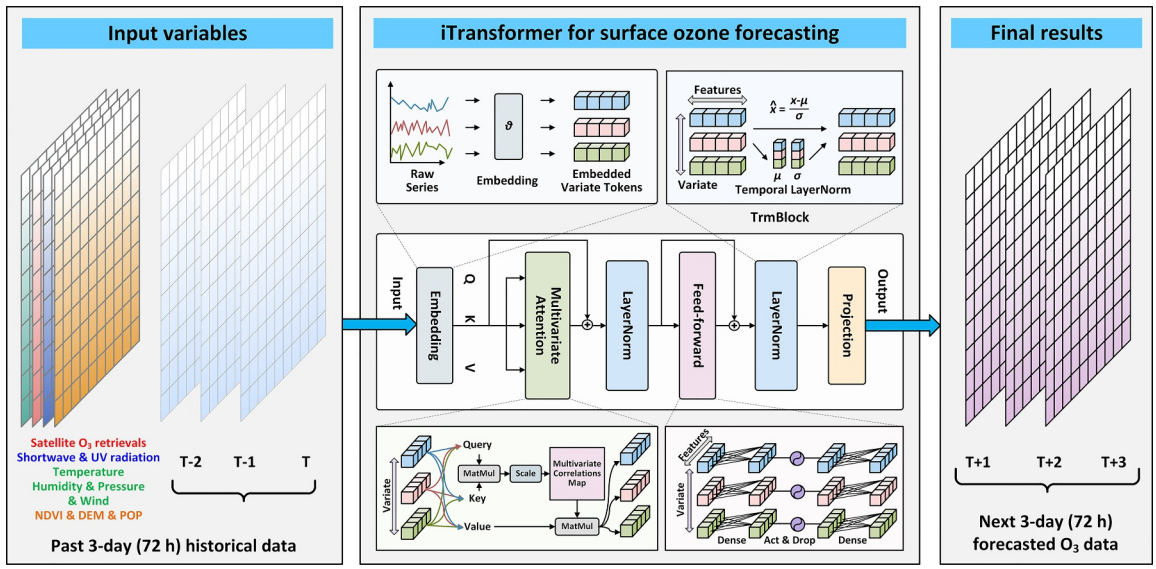

Figure 1. iTransformer-based framework for hourly surface O_3 concentration forecasting.

reanalysis (Hersbach et al., 2020). Additionally, auxiliary factors were introduced to improve model performance, including the Normalized Difference Vegetation Index (NDVI) from MODIS, the Digital Elevation Model (DEM) from SRTM, and population distribution (POP) data from LandScan™. All data were collected during the warm O_3 season (May–September) in 2019, comprising a total of 3,672 hr. To objectively assess the contribution of each predictor, we conducted SHAP analysis, which highlights the dominant roles of prior-hour O_3 and meteorological variables in driving model performance (Text S1 in Supporting Information S1).

2.2. Surface O_3 Forecasting Framework

Time-series deep learning models offer greater potential for capturing and forecasting the long-term dynamic changes in atmospheric pollutants. The Transformer model, introduced by Google (Vaswani et al., 2017), was initially designed for natural language processing (NLP) tasks, particularly machine translation. Its self-attention mechanism allows modeling of long-range dependencies among input tokens, while its encoder-decoder architecture supports efficient sequence-to-sequence processing (K. Han et al., 2023; Wei et al., 2024). In the standard Transformer architecture, variables measured at the same time step are usually treated as individual tokens (Vaswani et al., 2017). However, this convention imposes a narrow receptive field and fails to account for time-unaligned events across simultaneous time points, thereby limiting the model's capacity for representing multivariate correlations. To mitigate these issues, this research employs the advanced inverted Transformer (iTransformer) model to forecast hourly surface O_3 concentrations up to 72 hr (3 days) in advance over eastern China. The iTransformer takes an inverted perspective on time series by embedding the entire temporal sequence of each variable into a single token, leading to better characterization of long-term dependencies and cross-variable interactions. The architecture comprises four main components and is designed to effectively extract O_3-related signals from multivariate inputs, thereby improving the accuracy of time-series forecasts.

iTransformer-O_3 forecasting model development. The model execution primarily follows the workflow outlined in Figure 1. First, unlike standard approaches that treat individual time steps as tokens, iTransformer represents the entire temporal sequence of each variable as a single token, which is then embedded into high-dimensional vector representations suitable for model processing. Spatial information is implicitly preserved through the grid-based input structure, with each grid cell modeled as an independent time series and shared model parameters across all cells. However, explicit spatial dependencies (e.g., local adjacency) are not directly encoded, which remains a limitation to be addressed in future work using spatial attention mechanisms or diffusion-based models. Next, an eight-head multi-head attention mechanism captures the dependencies among the input variables. Through dynamic weighting, the model learns to identify the specific contributions of different variables to surface O_3 concentrations. The attention outputs are then normalized using a LayerNorm layer to standardize the distribution of activations (zero mean and unit variance), thereby accelerating convergence and improving

generalization by mitigating internal covariate shifts. Subsequently, a feedforward network (FFN) applies nonlinear transformations to refine these feature representations further. Following another LayerNorm operation, one encoder block is completed. This output is iteratively processed if multiple encoders are stacked. Finally, a projection layer maps the high-dimensional encoder outputs to a lower-dimensional space, generating the final 72–hr forecasts of surface O_3 concentrations and effectively translating the abstract features into actionable outputs.

Enhancing O_3 forecasts under high-polluted conditions. To better capture extreme O_3 events in the forecasts, we implemented a two-phase forecasting framework with a targeted enhancement for high-concentration scenarios. First, a baseline model was trained on the complete data set using standard temporal sequences. Next, we developed an optimized process for high-pollution scenarios through three key steps: Critical time steps were identified as those exceeding the 99th percentile of hourly O_3 concentrations within each 72–hr spatiotemporal batch. A constrained temporal permutation was then applied to emphasize these extreme values during training while preserving their absolute positions, and non-critical time segments (<99th percentile) were rearranged to produce a single augmented sample. Finally, the enhanced forecasts from this optimized phase were selectively fused into the baseline forecasts at corresponding grid cells, ensuring that the original model output was retained for all non-extreme periods. The 99th percentile threshold was determined exclusively from the training data set and used only to construct augmented training samples. No enhancement was applied during testing or operational forecasting, as the model learned to predict extreme events from the augmented training data without requiring prior identification of high-pollution scenarios.

2.3. Model Training and Evaluation

This study partitioned the input data set into training (60%), validation (10%), and testing (30%) subsets. Specifically, the training set comprised data from May to July 2019 (2,208 hr), the validation set contained data from 1 to 15 August 2019 (360 hr), and the test set comprised data from August 16 to 30 September 2019 (1,104 hr, 30%). Model parameters were optimized using the training set, while the validation set was used to assess performance, adjust hyperparameters (e.g., learning rate), and implement early stopping to prevent overfitting; the best-performing model on this set was saved. Finally, the test set was served for final, independent evaluation, generating 72–hr surface O_3 forecasts.

The pre-trained iTransformer model was retrained on sequences from the past 72 hr (from t-71 to t) across 13 variables, including satellite-derived O_3 concentrations (the target variable), shortwave and UV radiation, temperature, pressure, winds, NDVI, DEM, and POP. The input data were batched into a shape of (72, 13, 169, 253), representing the hourly time steps, variables, and spatial grid dimensions, respectively. The model output, which maintained the same structure, was used to extract the forecasting O_3 concentrations for the subsequent 72 hr (from t+1 to t+72):

$$O_3^{t+1:t+72}(x,y) = \text{iTransformer}\left(X^{t-71:t}(x,y), O_3^{t-71,t}(x,y)\right) \quad (1)$$

$$X \sim [\text{DSR}, \text{UV}, \text{TEM}, \text{BLH}, \text{RH}, \text{PRE}, \text{SP}, \text{WD}, \text{WS}, \text{DEM}, \text{POP}, \text{NDVI}] \quad (2)$$

The output tensor $O_3 \in R^{72 \times 169 \times 253}$ denotes the forecasted surface O_3 concentrations over the next 72 hr, where the input tensor $X \in R^{72 \times 13 \times 169 \times 253}$ comprises a 72-hr time series of 13 meteorological and geographical variables. Model architecture, training configuration, and evaluation metrics are described in detail in Text S2 of Supporting Information S1.

3. Results

3.1. Overall Forecasting Accuracy

On the first forecasting day (Day 1: 0–24 hr; Figure S2 in Supporting Information S1), the model demonstrated high and consistent accuracy. The forecasting performance peaked at the initial forecast hour (00 hr), achieving an average R of 0.94, an RMSE (MAE) of 12.64 (7.53) μg/m^3, and a near-ideal regression slope of 0.96. This high accuracy persisted throughout the day: The R values showed a gradual decline, remaining above 0.90 until the 06-hr lead time and around 0.89 thereafter, while the slopes consistently exceeded 0.85 across all 24 hr. Overall, Day 1 forecasts exhibited a strong linear relationship with the true values, yielding an average R of 0.90, a slope of

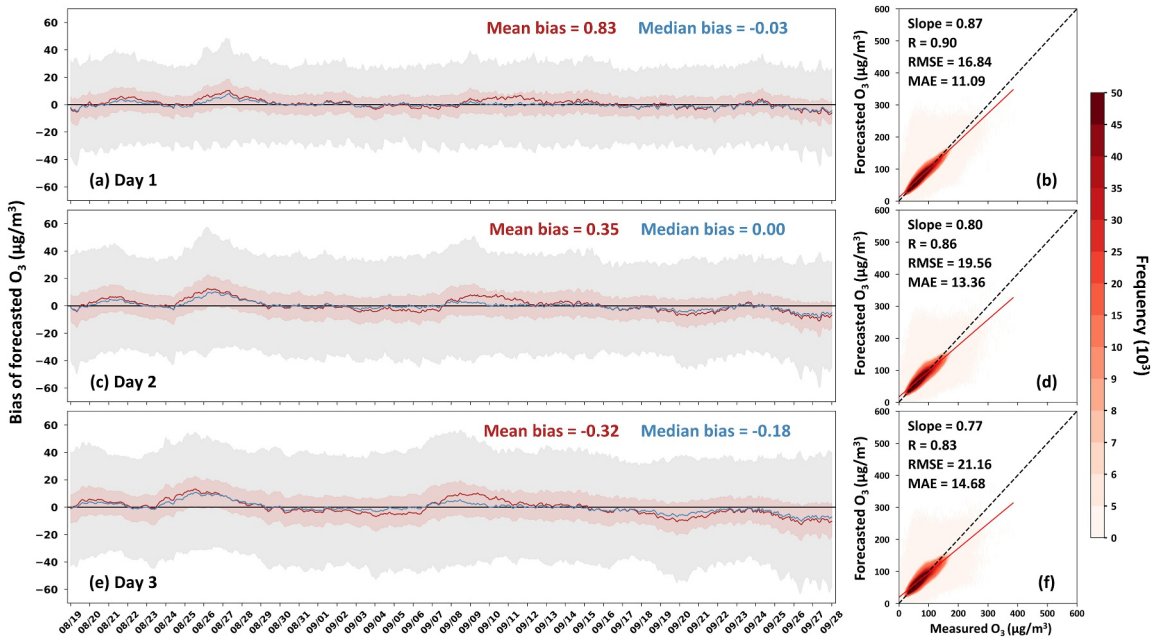

Figure 2. Forecasting performance of the iTransformer-O_3 model. Time series of hourly forecast biases (left panels) and density scatter plots comparing forecasted and observed O_3 concentrations (right panels) for the first (a, b), second (c, d), and third (e, f) forecasting days. Red and blue text indicate mean and median bias, shaded areas represent the 25th–75th percentile range of hourly bias, and dashed lines denote the 1:1 reference.

0.87, and an RMSE (MAE) of 16.84 (11.09) μg/m^3, respectively (Figure S3 in Supporting Information S1). The bias analysis further confirms the model's reliability (Figure 2a), where the differences between forecasted and true O_3 concentrations on Day 1 mostly fell between −20 and 10 μg/m^3. Throughout the evaluation period (19 August to 28 September 2019), the model exhibited relatively stable bias patterns, with only minor temporal shifts. It began with a slight underestimation (mean and median biases = −0.01 to −5.5 μg/m^3) from 21 to 28 August, transitioned to mild overestimation (mean bias up to 10.4 μg/m^3) until mid-September, and reverted to underestimation on a few specific days (e.g., 22, 26–28 September). Despite these temporal variations, the overall mean and median biases were 0.83 and −0.03 μg/m^3, respectively, underscoring the robustness of the forecasting.

On the second forecasting day (Day 2: 24–48 hr; Figure S4 in Supporting Information S1), the model maintained strong performance, albeit with a slight deterioration from Day 1. High accuracy was observed during hours 24–31 (R: 0.86–0.88; slope: 0.80–0.84), followed by stable performance from hours 32–47 (R: ∼0.85; slope: 0.79–0.80). Overall, Day 2 forecasts yielded an average R of 0.86 and an RMSE (MAE) of 19.56 (13.36) μg/m^3, respectively (Figure S3 in Supporting Information S1). The bias distribution (Figure 2c) resembled that of Day 1 but exhibited more pronounced deviations at certain times. Nevertheless, the overall bias remained low, with a mean of 0.35 μg/m^3 and a median of 0 μg/m^3, confirming the model's stability. The model continued to demonstrate reliable performance on the third forecasting day (Day 3: 48–72 hr; Figure S5 in Supporting Information S1), reaching an average R of 0.83 and an RMSE (MAE) of 21.16 (14.68) μg/m^3. The forecasting accuracy was highest at 48 hr (R = 0.85, slope = 0.79), remained consistent at 0.84 during hours 49–51, and stabilized thereafter (hours 52–71: R = 0.83, slope = 0.76–0.77). The bias distribution (Figure 2e) followed a pattern like previous days, with a mean of −0.32 μg/m^3 and a median of −0.18 μg/m^3. These results illustrate the model's capability to generate robust O_3 forecasts even at a 72-hr lead time.

At the hourly scale, forecasting accuracy reaches its peak during the first 48 hr, with R values ranging from 0.85 to 0.96, and both RMSE and MAE between 12.64–17.79 and 12.07–13.74 μg/m^3, respectively (Figures S3 and S5 in Supporting Information S1). Across the entire 0–72 hr period, the model's performance was consistently high, indicated by R values above 0.83, and RMSE and MSE below 22 and 15 μg/m^3, respectively. Overall mean R, RMSE, and MAE for the 0–72 hr period were 0.86, 19.07, and 13.11 μg/m^3, respectively (Figure S3 in Supporting Information S1). Importantly, the inclusion of extended high-value samples significantly improved model performance under highly polluted (150–200 μg/m^3) and extremely polluted (>200 μg/m^3) O_3 conditions. Compared

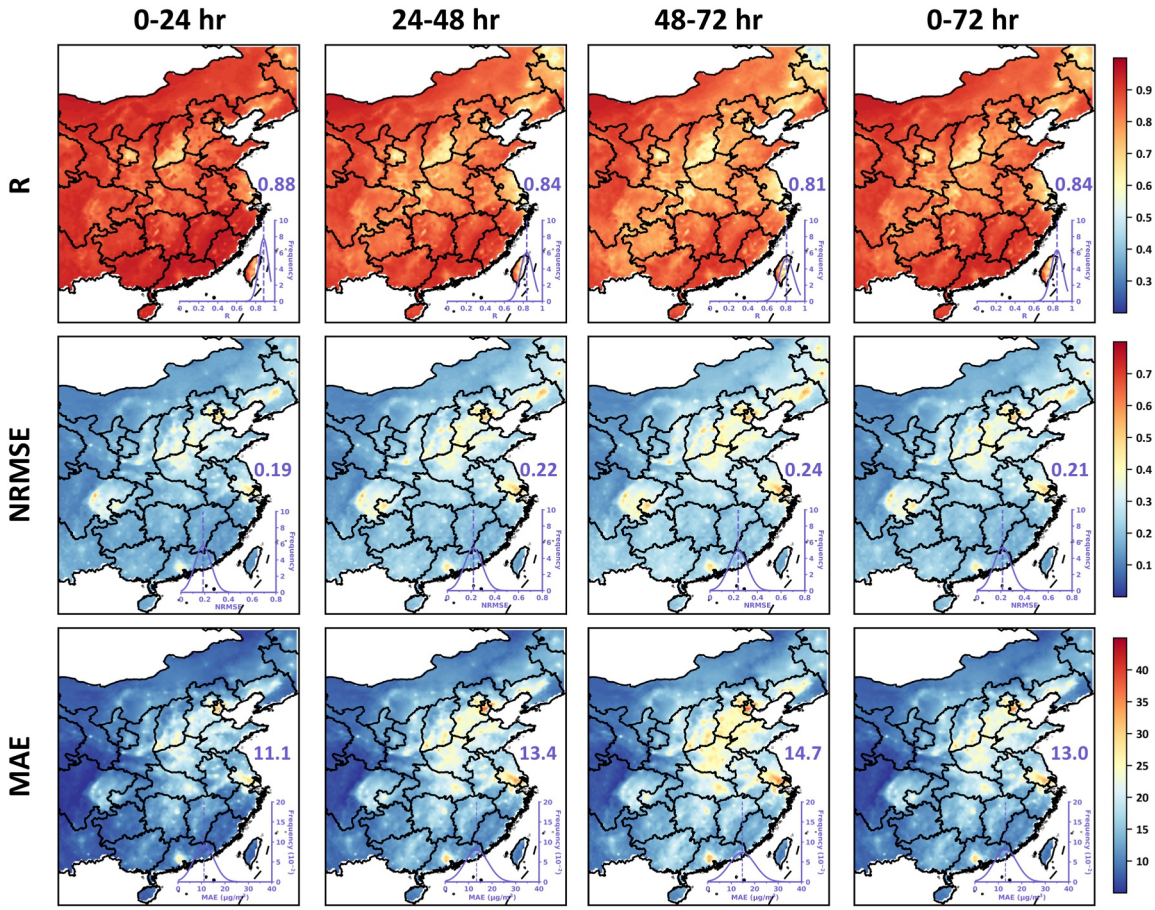

Figure 3. Spatial performance of hourly O_3 forecasts across eastern China. Maps show the Pearson correlation coefficient (R), normalized root mean square error (NRMSE), and mean absolute error (MAE) for the first forecasting day (0–24 hr), second forecasting day (24–48 hr), and third forecasting day (48–72 hr), and the entire 72-hr period (0–72 hr). Inset plots show frequency histograms of R, NRMSE, and MAE.

to the scenario without such sample extension, the proportion of forecasts within 20% underestimation increased by 2% and 13% across the 0–72 hr period, respectively. For O_3 concentrations exceeding 200 μg/m^3, the proportion rose by 8%, 14%, and 15% for Day 1, Day 2, and Day 3, respectively (Figure S6 in Supporting Information S1), highlighting the model's enhanced accuracy under extreme pollution conditions.

Furthermore, Figure 3 shows the model's spatial performance in forecasting surface O_3 across different forecasting periods. On Day 1 (0–24 hr), high correlations were widespread across about 93% of the region, particularly in Jiangxi, Fujian, Guangdong, and Inner Mongolia provinces, with R values exceeding 0.8. Most areas show low forecast uncertainty, with 92% of the region exhibiting NRMSE values below 0.3% and 83% of areas showing small biases, with MAE below 15 μg/m^3. Larger errors were observed in specific areas, including central Sichuan, southern Hebei, Shanxi, and Shanghai. Overall, Day 1 showed superior accuracy, with an average R of 0.88, NRMSE of 0.19, and MAE of 11.1 μg/m^3. The model's robust performance persisted on Day 2 (24–48 hr) and Day 3 (48–72 hr), maintaining strong spatial consistency. The average R values decreased slightly to 0.84 and 0.81, respectively, while the corresponding NRMSE (MAE) values increased modestly to 0.22 (13.4 μg/m^3) and 0.24 (14.7 μg/m^3). Regions with high correlation ($R > 0.8$) remained extensive, covering 76% and 60% of the domain, primarily across southeastern and northeastern China. Similarly, 86% and 80% of the region exhibited NRMSE values below 0.3. The spatial pattern of bias remained similar to Day 1, but with a gradual increase in magnitude. Nonetheless, the proportion of regions with small biases (MAE < 15 μg/m^3) remained substantial, at 70% and 59% on Day 2 and Day 3, respectively. Correspondingly, our hourly surface O_3 forecasts and ground-based measurements show high agreement against in situ measurements at monitoring stations broadly across Eastern China (Figure S7 in Supporting Information S1), achieving the highest accuracy

during 0–24 hr (mean $R = 0.73$, NRMSE = 0.40) and maintaining stable performance over the full 0–72 hr period (mean $R = 0.69$, NRMSE = 0.43). We further examined the relationships between city-level population and evaluation metrics (Figure S8 in Supporting Information S1). Both bias and NRMSE increase with population, whereas R shows the opposite trend, indicating generally larger forecast errors in more densely populated regions.

Overall, the 0–72 hr forecasts demonstrated robust spatial performance, with 80%, 60%, and 44% of regions simultaneously attaining strong correlations with low uncertainties (e.g., $R > 0.8$, NRMSE < 0.3, and MAE < 15 μg/m^3), yielding an average R of 0.84, NRMSE of 0.21, and MAE of 13.0 μg/m^3. At the regional scale, forecast performance was strong across major urban clusters. In BTH and YRD, the model reached average R values of 0.81 and 0.78, respectively, with slight positive biases (averaging 2.71 and 2.70 μg/m^3). In Northeast China, the model exhibited an R value of 0.77 and a slight negative bias of approximately −0.47 μg/m^3. PRD exhibited the highest accuracy ($R = 0.90$), but with an average underestimation of −2.19 μg/m^3 (Table S3 in Supporting Information S1). These findings highlight the model's effectiveness in delivering stable, accurate multi-day surface O_3 forecasts.

3.2. Model Performance During Severe Pollution Episodes

The model effectively captured the evolution of severe surface O_3 episodes by reproducing the characteristic warm-season diurnal cycle, for example, a midday increase peaking between 15:00 and 16:00 CST (China Standard Time), a steady nocturnal decline to a minimum around 07:00, and a subsequent morning rebound (Figure S9 in Supporting Information S1). This behavior is consistent with the well-established pattern that high-O_3 pollution events primarily occur during daytime, whereas nighttime episodes are relatively rare (Lyu et al., 2019; Xia et al., 2021; Xue et al., 2024). During a large-scale high-polluted O_3 episode (Case 1, 25–28 September 2019; Figure 4), pronounced spatial heterogeneity was evident at 12:00, with peak O_3 concentrations (140–180 μg/m^3) observed in the PRD, Jiangxi, and Hunan provinces. Moderate levels (∼100 μg/m^3) were recorded in the northwest, while the lowest values (<80 μg/m^3) were found in the southwest. By 16:00 CST, O_3 levels intensified, with concentrations exceeding 160 μg/m^3 across most of eastern China, and extreme values (>220 μg/m^3) were observed in the North China Plain, particularly in Beijing and northern Henan, and in the PRD region. The subsequent nocturnal decay phase, reaching minima (20–60 μg/m^3) around 07:00 CST due to diminished photochemical activity. Throughout the 3-day episode, the temporal cycle remained consistent, though the spatial distribution of high-pollution zones shifted. New hotspots emerged in Shandong and Shanghai, replacing prior maxima, a pattern indicative of meteorologically driven pollutant transport. Notably, the PRD experienced persistent severe pollution throughout this period. Regarding the differences in forecasted O_3 concentrations, Case 1 exhibited a generally consistent spatial pattern from Day 1 to Day 3. Despite the large-scale pollution during the daytime, forecast biases across most of eastern and southern China remained relatively small, generally within ±10%, indicating that the model captured the high-pollution event well. However, as time progressed, forecast biases increased. By night, a pronounced underestimation occurred in the northern regions, while overestimation emerged in the southern areas.

Overall, the model demonstrated consistently low mean (median) biases, ranging from −0.84 (−0.86) μg/m^3 to −4.88 (−3.29) μg/m^3 (Figure S10a in Supporting Information S1). It exhibited high initial accuracy on the first day, with most deviations within ±10 μg/m^3. Although daytime underestimation gradually increased over the following days, the bias remained predominantly within a modest range, indicating a stable, forecasted error pattern. The model achieved excellent overall forecast performance, with a strong correlation of 0.92 and low associated errors (i.e., RMSE = 14.33 μg/m^3, MAE = 10.71 μg/m^3).

Additional high-ozone pollution episodes (Cases 2–4; September 2019) further demonstrate the model's robustness in capturing the spatiotemporal evolution and diurnal variability of surface O_3 (Figures S11–S15 in Supporting Information S1). Forecast–observation differences show coherent spatial patterns and clear diurnal dependence, with relatively small biases during daytime peak O_3 conditions and larger relative differences at night due to low background concentrations. Overall, forecast errors remain well constrained with high correlation and stable performance across 72-hr forecasts, indicating strong generalization capability (Text S3 in Supporting Information S1).

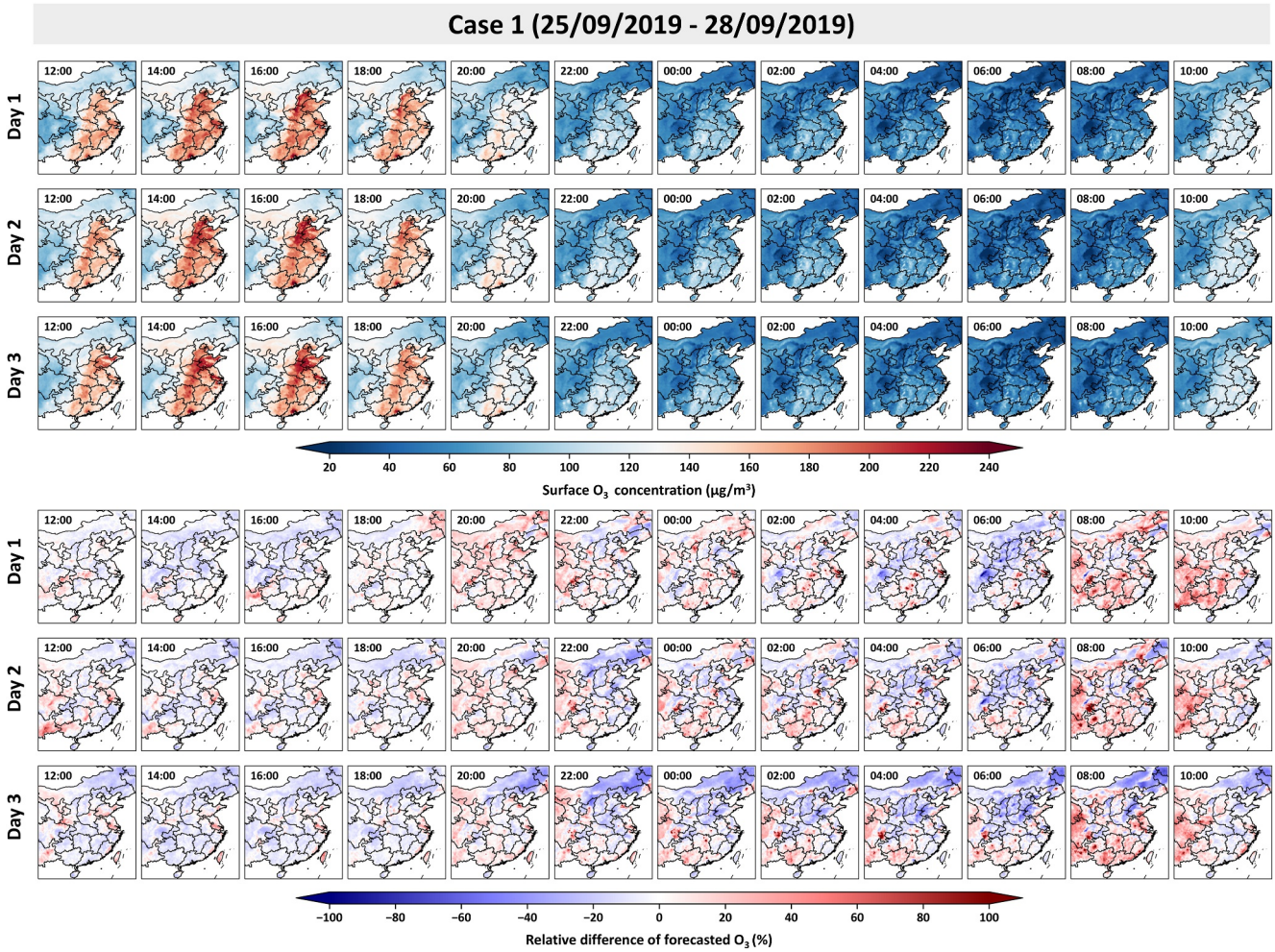

Figure 4. Spatial distribution of forecasted hourly surface O_3 concentration (upper panel) and their relative differences from observations (lower panel) at two-hourly intervals for the 72-hr period during the large-scale high-polluted O_3 episode (25–28 September 2019).

4. Discussion and Conclusions

We position our forecasting framework relative to previous studies to highlight its methodological advances and performance (Text S4 and Table S4 in Supporting Information S1). Most existing approaches focus on site- or city-scale O_3 forecasting using ground-based data, limiting their ability to capture regional transport and provide early warnings (Balashov et al., 2017; Biancofiore et al., 2015; Booth et al., 2025; Durão et al., 2016; Kumar et al., 2024; Pernak et al., 2019; Zhang et al., 2023; Zhang & Awang, 2025). In contrast, our framework enables regional-scale 72-hr forecasting across eastern China using high-resolution, satellite-derived hourly O_3 data, and maintains robust performance across extended lead times and achieves higher accuracy than related studies (Table S4 in Supporting Information S1). Moreover, our model outperforms traditional machine learning and deep learning models when applied to the same data set (Table S5 in Supporting Information S1).

Traditional machine learning and hybrid models often suffer from error accumulation and limited representation of spatial heterogeneity, leading to reduced performance in long-term forecasts (Biancofiore et al., 2015; W. Han et al., 2023; Li et al., 2024, 2025; Zhang et al., 2023). By integrating satellite-derived high-resolution hourly surface O_3 data within a time-series iTransformer framework, our approach enables accurate regional 72-hr forecasts while maintaining robust performance under extended lead times. The model effectively captures diurnal variability and spatial heterogeneity, particularly during severe pollution episodes, and reliably identifies evolving pollution hotspots. These capabilities demonstrate strong potential for air-quality forecasting at spatial and temporal scales relevant to population exposure and public health. More broadly, this work highlights the

promise of satellite-driven AI frameworks to translate continuous Earth observations into actionable regional O_3 forecasts and support timely public health warnings. Nevertheless, performance may be constrained under abrupt meteorological changes or extreme pollution conditions.

Future work will focus on integrating near-real-time satellite O_3 retrievals with advanced deep learning architectures and chemical transport model (CTM) forecasts, including data from geostationary platforms such as GEMS. Such integration will enable more timely and operational air-quality forecasting, support early warning systems, and help reduce population exposure and associated health risks during severe O_3 pollution events.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

The hourly surface O_3 retrievals from ChinaHighO_3 are available from Cheng et al. (2025). The ERA5 reanalysis data can be accessed from Copernicus Climate Change Service (2019). The SRTM DEM data is available from NASA JPL (2013). The LandScan^{TM} population data can be found in Rose et al. (2019). The Normalized Difference Vegetation Index (NDVI) is available from Didan (2021).

Acknowledgments
This work was supported by the Jing-Jin-Ji Regional Integrated Environmental Improvement-National Science and Technology Major Project (2026ZD1212900), the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China (JYB2025XDXM906), the National Key Technology and Development Program of Corps (2025AA001), and the Fundamental Research Funds for the Central Universities, Peking University.

References

Ainsworth, E. A., Yendrek, C. R., Sitch, S., Collins, W. J., & Emberson, L. D. (2012). The effects of tropospheric ozone on net primary productivity and implications for climate change. *Annual Review of Plant Biology*, *63*(1), 637–661. https://doi.org/10.1146/annurev-arplant-042110-103829

Appel, K. W., Gilliland, A. B., Sarwar, G., & Gilliam, R. C. (2007). Evaluation of the community multiscale air quality (CMAQ) model version 4.5: Sensitivities impacting model performance. *Atmospheric Environment*, *41*(40), 9603–9615. https://doi.org/10.1016/j.atmosenv.2007.08.044

Balashov, N. V., Thompson, A. M., & Young, G. S. (2017). Probabilistic forecasting of surface ozone with a novel statistical approach. *Journal of Applied Meteorology and Climatology*, *56*(2), 297–316. https://doi.org/10.1175/JAMC-D-16-0110.1

Biancofiore, F., Verdecchia, M., Di Carlo, P., Tomassetti, B., Aruffo, E., Busilacchio, M., et al. (2015). Analysis of surface ozone using a recurrent neural network. *Science of the Total Environment*, *514*, 379–387. https://doi.org/10.1016/j.scitotenv.2015.01.106

Booth, A., James, P., McGough, S., & Solaiman, E. (2025). Cross-regional deep learning for air quality forecasting: A comparative study of CO, NO_2, O_3, $PM_{2.5}$, and PM_{10}. *Forecasting*, *7*(4), 66. https://doi.org/10.3390/forecast7040066

Cheng, F., Li, Z., Yang, Z., Li, R., Wang, D., Jia, A., et al. (2025). First retrieval of 24-hourly 1-km-resolution gapless surface ozone (O_3) from space in China using artificial intelligence: Diurnal variations and implications for air quality and phytotoxicity. *Remote Sensing of Environment*, *316*, 114482. https://doi.org/10.1016/j.rse.2024.114482

Chin, M., Ginoux, P., Lucchesi, R., Huebert, B., Weber, R., Anderson, T., et al. (2003). A global aerosol model forecast for the ACE-Asia field experiment. *Journal of Geophysical Research*, *108*(D23). https://doi.org/10.1029/2003JD003642

Copernicus Climate Change Service. (2019). ERA5-Land hourly data from 1950 to present [Dataset]. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. https://doi.org/10.24381/CDS.E2161BAC

Didan, K. (2021). MODIS/Terra vegetation indices 16-Day L3 global 250m SIN grid V061 [Dataset]. *NASA Land Processes Distributed Active Archive Center*. https://doi.org/10.5067/MODIS/MOD13Q1.061

Durão, R. M., Mendes, M. T., & João Pereira, M. (2016). Forecasting O_3 levels in industrial area surroundings up to 24 h in advance, combining classification trees and MLP models. *Atmospheric Pollution Research*, *7*(6), 961–970. https://doi.org/10.1016/j.apr.2016.05.008

Gao, Z., & Zhou, X. (2024). A review of the CAMx, CMAQ, WRF-Chem and NAQPMS models: Application, evaluation and uncertainty factors. *Environmental Pollution*, *343*, 123183. https://doi.org/10.1016/j.envpol.2023.123183

Guo, Q., He, Z., & Wang, Z. (2025). Assessing the effectiveness of long short-term memory and artificial neural network in predicting daily ozone concentrations in Liaocheng City. *Scientific Reports*, *15*(1), 6798. https://doi.org/10.1038/s41598-025-91329-w

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., et al. (2023). A survey on vision transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(1), 87–110. https://doi.org/10.1109/TPAMI.2022.3152247

Han, W., He, T., Jiang, Z., Zhu, R., Jones, D., Miyazaki, K., & Shen, Y. (2023). The capability of deep learning model to predict ozone across continents in China, the United States and Europe. *Geophysical Research Letters*, *50*(24), e2023GL104928. https://doi.org/10.1029/2023GL104928

He, C., Lu, X., Wang, H., Wang, H., Li, Y., He, G., et al. (2022). The unexpected high frequency of nocturnal surface ozone enhancement events over China: Characteristics and mechanisms. *Atmospheric Chemistry and Physics*, *22*(23), 15243–15261. https://doi.org/10.5194/acp-22-15243-2022

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., et al. (2019). The CAMS reanalysis of atmospheric composition. *Atmospheric Chemistry and Physics*, *19*(6), 3515–3556. https://doi.org/10.5194/acp-19-3515-2019

Keller, C. A., Knowland, K. E., Duncan, B. N., Liu, J., Anderson, D. C., Das, S., et al. (2021). Description of the NASA GEOS composition forecast modeling system GEOS-CF v1.0. *Journal of Advances in Modeling Earth Systems*, *13*(4), e2020MS002413. https://doi.org/10.1029/2020MS002413

Kitayama, K., Morino, Y., Yamaji, K., & Chatani, S. (2019). Uncertainties in O_3 concentrations simulated by CMAQ over Japan using four chemical mechanisms. *Atmospheric Environment*, *198*, 448–462. https://doi.org/10.1016/j.atmosenv.2018.11.003

Kumar, S., Kour, V., Jain, N., & Misra, R. (2024). A novel hybrid CNN-transformer framework for spatiotemporal prediction of O_3. In *2024 IEEE International Conference on Big Data (BigData)* (pp. 4327–4333). IEEE. https://doi.org/10.1109/BigData62323.2024.10825690

Li, Q., Li, J., Wang, Z., Liu, B., Wang, W., & Wang, Z. (2024). Development of a city-level surface ozone forecasting system using deep learning techniques and air quality model: Application in eastern China. *Atmospheric Environment*, *339*, 120865. https://doi.org/10.1016/j.atmosenv.2024.120865

Li, Z., Bi, J., Liu, Y., & Hu, X. (2025). Forecasting O_3 and NO_2 concentrations with spatiotemporally continuous coverage in southeastern China using a Machine learning approach. *Environment International*, *195*, 109249. https://doi.org/10.1016/j.envint.2024.109249

Lin, Y., Jiang, F., Zhao, J., Zhu, G., He, X., Ma, X., et al. (2018). Impacts of O_3 on premature mortality and crop yield loss across China. *Atmospheric Environment*, *194*, 41–47. https://doi.org/10.1016/j.atmosenv.2018.09.024

Liu, X., Zhang, Y., Wang, J., Huang, H., & Yin, H. (2022). Multi-source and multivariate ozone prediction based on fuzzy cognitive maps and evidential reasoning theory. *Applied Soft Computing*, *119*, 108600. https://doi.org/10.1016/j.asoc.2022.108600

Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., & Long, M. (2024). iTransformer: Inverted transformers are effective for time series forecasting. *In International Conference on Learning Representations (ICLR)* (pp. 11116–11140). Retrieved from https://openreview.net/forum?id=JePfAI8fah

Lyu, X., Wang, N., Guo, H., Xue, L., Jiang, F., Zeren, Y., et al. (2019). Causes of a continuous summertime O_3 pollution event in Jinan, a central city in the North China plain. *Atmospheric Chemistry and Physics*, *19*(5), 3025–3042. https://doi.org/10.5194/acp-19-3025-2019

Mu, L., Bi, S., Ding, X., & Xu, Y. (2024). Transformer-based ozone multivariate prediction considering interpretable and priori knowledge: A case study of Beijing, China. *Journal of Environmental Management*, *366*, 121883. https://doi.org/10.1016/j.jenvman.2024.121883

NASA JPL. (2013). NASA shuttle radar topography mission global 3 arc second [Dataset]. *NASA Land Processes Distributed Active Archive Center*. https://doi.org/10.5067/MEASURES/SRTM/SRTMGL3.003

Pernak, R., Alvarado, M., Lonsdale, C., Mountain, M., Hegarty, J., & Nehrkorn, T. (2019). Forecasting surface O_3 in Texas urban areas using random forest and generalized additive models. *Aerosol and Air Quality Research*, *19*(12), 2815–2826. https://doi.org/10.4209/aaqr.2018.12.0464

Rao, S. T., Luo, H., Astitha, M., Hogrefe, C., Garcia, V., & Mathur, R. (2020). On the limit to the accuracy of regional-scale air quality models. *Atmospheric Chemistry and Physics*, *20*(3), 1627–1639. https://doi.org/10.5194/acp-20-1627-2020

Rose, A., McKee, J., Sims, K., Bright, E., Reith, A., & Urban, M. (2019). LandScan global 2019 [Dataset]. *Oak Ridge National Laboratory*. https://doi.org/10.48690/1524214

Sayeed, A., Choi, Y., Eslami, E., Lops, Y., Roy, A., & Jung, J. (2020). Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance. *Neural Networks*, *121*, 396–408. https://doi.org/10.1016/j.neunet.2019.09.033

Sayeed, A., Eslami, E., Lops, Y., & Choi, Y. (2022). CMAQ-CNN: A new-generation of post-processing techniques for chemical transport models using deep neural networks. *Atmospheric Environment*, *273*, 118961. https://doi.org/10.1016/j.atmosenv.2022.118961

Shams, S. R., Choi, Y., Singh, D., Kayastha, S., & Park, J. (2025). Deep learning-based forecasting of daily maximum ozone levels and assessment of socioeconomic and health impacts in South Korea. *Science of the Total Environment*, *983*, 179684. https://doi.org/10.1016/j.scitotenv.2025.179684

Shinde, P. P., & Shah, S. (2018). A review of machine learning and deep learning applications. In *2018 fourth International Conference on Computing Communication Control and Automation (ICCUBEA)* (pp. 1–6). https://doi.org/10.1109/ICCUBEA.2018.8697857

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*. https://doi.org/10.48550/arXiv.1706.03762

Wang, T., Xue, L., Brimblecombe, P., Lam, Y. F., Li, L., & Zhang, L. (2017). Ozone pollution in China: A review of concentrations, meteorological influences, chemical precursors, and effects. *The Science of the Total Environment*, *575*, 1582–1596. https://doi.org/10.1016/j.scitotenv.2016.10.081

Wang, T., Xue, L., Feng, Z., Dai, J., Zhang, Y., & Tan, Y. (2022). Ground-level ozone pollution in China: A synthesis of recent findings on influencing factors and impacts. *Environmental Research Letters*, *17*(6), 063003. https://doi.org/10.1088/1748-9326/ac69fe

Wang, Y., Jiang, S., Huang, L., Lu, G., Kasemsan, M., Yaluk, E. A., et al. (2023). Differences between VOCs and NO_x transport contributions, their impacts on O_3, and implications for O_3 pollution mitigation based on CMAQ simulation over the Yangtze River Delta, China. *Science of the Total Environment*, *872*, 162118. https://doi.org/10.1016/j.scitotenv.2023.162118

Wei, J., Li, Z., Li, K., Dickerson, R. R., Pinker, R. T., Wang, J., et al. (2022). Full-coverage mapping and spatiotemporal variations of ground-level ozone (O_3) pollution from 2013 to 2020 across China. *Remote Sensing of Environment*, *270*, 112775. https://doi.org/10.1016/j.rse.2021.112775

Wei, J., Wang, Z., Li, Z., Li, Z., Pang, S., Xi, X., et al. (2024). Global aerosol retrieval over land from Landsat imagery integrating transformer and Google Earth Engine. *Remote Sensing of Environment*, *315*, 114404. https://doi.org/10.1016/j.rse.2024.114404

Xia, N., Du, E., Guo, Z., & De Vries, W. (2021). The diurnal cycle of summer tropospheric ozone concentrations across Chinese cities: Spatial patterns and main drivers. *Environmental Pollution*, *286*, 117547. https://doi.org/10.1016/j.envpol.2021.117547

Xue, J., Liao, H., & Li, J. (2024). Nighttime ozone excess days and the associated meteorological conditions in the North China plain from 2015 to 2022. *Atmospheric Environment*, *330*, 120561. https://doi.org/10.1016/j.atmosenv.2024.120561

Yafouz, A., AlDahoul, N., Birima, A. H., Ahmed, A. N., Sherif, M., Sefelnasr, A., et al. (2022). Comprehensive comparison of various machine learning algorithms for short-term ozone concentration prediction. *Alexandria Engineering Journal*, *61*(6), 4607–4622. https://doi.org/10.1016/j.aej.2021.10.021

Yang, Z., Li, Z., Cheng, F., Lv, Q., Li, K., Zhang, T., et al. (2025). Two-decade surface ozone (O_3) pollution in China: Enhanced fine-scale estimations and environmental health implications. *Remote Sensing of Environment*, *317*, 114459. https://doi.org/10.1016/j.rse.2024.114459

Yue, X., & Unger, N. (2014). Ozone vegetation damage effects on gross primary productivity in the United States. *Atmospheric Chemistry and Physics*, *14*(17), 9137–9153. https://doi.org/10.5194/acp-14-9137-2014

Yuval, Broday, D. M., & Alpert, P. (2012). Exploring the applicability of future air quality predictions based on synoptic system forecasts. *Environmental Pollution*, *166*, 65–74. https://doi.org/10.1016/j.envpol.2012.03.010

Zhang, A., Fu, T., Feng, X., Guo, J., Liu, C., Chen, J., et al. (2023). Deep learning-based ensemble forecasts and predictability assessments for surface ozone pollution. *Geophysical Research Letters*, *50*(8), e2022GL102611. https://doi.org/10.1029/2022GL102611

Zhang, R., & Awang, N. (2025). A novel ST-iTransformer model for spatio-temporal ambient air pollution forecasting. *Journal of Big Data*, *12*(1), 101. https://doi.org/10.1186/s40537-025-01150-5

Zhang, Y., Yu, S., Chen, X., Li, Z., Li, M., Song, Z., et al. (2022). Local production, downward and regional transport aggravated surface ozone pollution during the historical orange-alert large-scale ozone episode in eastern China. *Environmental Chemistry Letters*, *20*(3), 1577–1588. https://doi.org/10.1007/s10311-022-01421-0

References From the Supporting Information

Ahmad, M., Rappenglück, B., Osibanjo, O., & Retama, A. (2025). Machine-learning models for surface ozone forecast in Mexico City. *Atmosphere*, *16*(8), 931. https://doi.org/10.3390/atmos16080931

Bais, A. F., Bernhard, G., McKenzie, R. L., Aucamp, P. J., Young, P. J., Ilyas, M., et al. (2019). Ozone–climate interactions and effects on solar ultraviolet radiation. *Photochemical and Photobiological Sciences*, *18*(3), 602–640. https://doi.org/10.1039/C8PP90059K

Hu, X.-M., Doughty, D. C., Sanchez, K. J., Joseph, E., & Fuentes, J. D. (2012). Ozone variability in the atmospheric boundary layer in Maryland and its implications for vertical transport model. *Atmospheric Environment*, *46*, 354–364. https://doi.org/10.1016/j.atmosenv.2011.09.054

Kavassalis, S. C., & Murphy, J. G. (2017). Understanding ozone-meteorology correlations: A role for dry deposition. *Geophysical Research Letters*, *44*(6), 2922–2931. https://doi.org/10.1002/2016GL071791

Li, X., Chee, S., Hao, J., Abbatt, J. P. D., Jiang, J., & Smith, J. N. (2019). Relative humidity effect on the formation of highly oxidized molecules and new particles during monoterpene oxidation. *Atmospheric Chemistry and Physics*, *19*(3), 1555–1570. https://doi.org/10.5194/acp-19-1555-2019

Matsunaga, S. N., Guenther, A. B., Izawa, Y., Wiedinmyer, C., Greenberg, J. P., & Kawamura, K. (2007). Importance of wet precipitation as a removal and transport process for atmospheric water soluble carbonyls. https://doi.org/10.1016/j.atmosenv.2006.08.054

Porter, W. C., & Heald, C. L. (2019). The mechanisms and meteorological drivers of the summertime ozone–temperature relationship. *Atmospheric Chemistry and Physics*, *19*(21), 13367–13381. https://doi.org/10.5194/acp-19-13367-2019

Sims, K., Reith, A., Bright, E., Kaufman, J., Pyle, J., Epting, J., et al. (2023). LandScan global 2022 (version 2022) [Dataset]. *Oak Ridge National Laboratory*. https://doi.org/10.48690/1529167

Sun, Y., Wang, Y.-S., Li, X., An, J.-L., Xin, J.-Y., & Hu, B. (2006). An analysis for vertical distribution of O_3, NOx and CO in the atmosphere during a serious air pollution in Beijing. *Chinese Journal of Geophysics*, *49*(6), 1475–1482. https://doi.org/10.1002/cjg2.974

Vukovich, F. M., Bach, W. D., Crissman, B. W., & King, W. J. (1977). On the relationship between high ozone in the rural surface layer and high pressure systems. *Atmospheric Environment*, *11*(10), 967–983. https://doi.org/10.1016/0004-6981(77)90026-9

Yoo, J.-M., Lee, Y.-R., Kim, D., Jeong, M.-J., Stockwell, W. R., Kundu, P. K., et al. (2014). New indices for wet scavenging of air pollutants (O_3, CO, NO_2, SO_2, and PM_{10}) by summertime rain. *Atmospheric Environment*, *82*, 226–237. https://doi.org/10.1016/j.atmosenv.2013.10.022

Zhang, Z., Johansson, C., Engardt, M., Stafoggia, M., & Ma, X. (2024). Improving 3-day deterministic air pollution forecasts using machine learning algorithms. *Atmospheric Chemistry and Physics*, *24*(2), 807–851. https://doi.org/10.5194/acp-24-807-2024