Geoscientific
Model Development

Model description paper

A Transformer-based agent model of GEOS-Chem v14.2.2 for informative prediction of $PM_{2.5}$ and O_3 levels to future emission scenarios: TGEOS v1.0

Dehao Li[1], Jianbing Jin[1], Guoqiang Wang[2], Mijie Pang[3], Weihong Zhang[4], and Hong Liao[1]

[1]State Key Laboratory of Climate System Prediction and Risk Management, Jiangsu Key Laboratory of Atmospheric Environment Monitoring and Pollution Control, Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, School of Environmental Science and Engineering, Nanjing University of Information Science and Technology, Nanjing, Jiangsu, China
[2]School of Mathematics, Physics and Statistics, Shanghai University of Engineering Science, Shanghai, China
[3]Delft Institute of Applied Mathematics, Delft University of Technology, Delft, the Netherlands
[4]Ecological Environment Monitoring Center of Ningxia Hui Autonomous Region, Yinchuan,
Ningxia Hui Autonomous Region

Correspondence: Jianbing Jin (jianbing.jin@nuist.edu.cn) and Hong Liao (hong.liao@nuist.edu.cn)

Abstract. Efficient and informative air quality modeling in future emission scenarios is vital for effective formulation of emission reduction policies. Traditional chemical transport models (CTMs) struggle with the computational demands required for timely predictions. While advanced emulator techniques greatly accelerate CTM simulating process, they fall short in providing comprehensive estimates of future air quality due to their limited model structure. Additionally, these emulators often have difficulty simultaneously accounting for varying emission variables and the effects of regional transport, which limits their applicability and undermines prediction accuracy. In this study, an informative future air quality prediction model "TGEOS v1.0" based on the Transformer framework is developed as an efficient agent model of GEOS-Chem v14.2.2. TGEOS is able to efficiently estimate key statistical indicators of $PM_{2.5}$ and O_3 concentrations under future emission scenarios and capture potential extreme pollution events, with approximately 2.51 s to execute one-year estimation. The model incorporates sectoral emissions of up to 26 distinct species as well as the impacts of regional emissions and meteorology on pollutant concentrations, enhancing its versatility and predictive accuracy. The spatial and probability distributions predicted by TGEOS are in good agreement with GEOS-Chem, with the correlation coefficients for $PM_{2.5}$ and O_3 exceed 0.98 in high-pollution months. Compared with other machine learning models, TGEOS based on Transformer framework showcases superior performance, underscoring the potential of the Transformer framework in air quality modeling.

1 Introduction

Air pollution constitutes a significant public health emergency due to its detrimental effects on human health, contributing to approximately 6.7 million premature deaths annually (Fuller et al., 2022), particularly in fast developing countries such as China (Lelieveld et al., 2015; WHO, 2023). $PM_{2.5}$ and ozone (O_3), recognized as major air pollutants, demonstrate a strong correlation with cardiovascular diseases and all-cause mortality (Al-Kindi et al., 2020). Elevated exposure to these pollutants considerably exacerbates the public health burden (Bell et al., 2004; Chen et al., 2023a; Wei et al., 2024). $PM_{2.5}$ is a complex pollutant with sources including road dust (Chen et al., 2019), fuel combustion(Bond et al., 2007), and natural sources like wildfires (Burke et al., 2023) and dust storms (Rodríguez and López-Darias, 2024), along with secondary formation through atmospheric reactions (McDuffie et al., 2021). The concentration of ambient

$PM_{2.5}$ is influenced by local and ambient emissions (Qiao et al., 2021). Similarly, as a secondary pollutant resulting from photochemical reactions involving various precursors such as nitrogen oxides (NO_x) and volatile organic compounds (VOC_s) (Wang et al., 2017), O_3 concentrations have been shown to be sensitive to both local and regional precursor emissions (Wei et al., 2019; Wang et al., 2021; Gong et al., 2020). $PM_{2.5}$ and O_3 pollution has emerged as the central environmental topic in China, with multiple incidences of extreme air pollution occurring in typical areas such as the North China Plain (NCP) and the Yangtze River Delta (YRD) (Silver et al., 2018; Lu et al., 2020a). Focusing on severe $PM_{2.5}$ pollution in China, the Chinese government has launched a series of emission control plans (CSC, 2013, 2018). Although the implementation of effective control policies has resulted in reductions in precursor emissions (Zheng et al., 2021) and the consequent decrease in $PM_{2.5}$ concentrations (Xiao et al., 2021), many air pollution issues still persist. For instance, an increase in domestic surface O_3 levels has been observed between 2013 and 2020 (Han et al., 2024), as well as a rebound trend for $PM_{2.5}$ pollution in recent years (Le et al., 2020; Wang et al., 2024). Therefore, it is imperative for China to formulate more holistic and accurate emission control plans to combat air pollution (Wang et al., 2023b; Geng et al., 2024). Concurrently, a model with the ability to rapidly and accurately predict air pollutants concentrations under different emission scenarios is in demand for policymakers.

As comprehensive and reliable tools for simulating atmospheric processes (Seinfeld and Pandis, 2016), Chemical Transport Models (CTMs) are widely used to estimate air pollutant levels under different control measures (Zhang et al., 2023b). CTMs can provide historical, current and future estimates of various air pollutants, including $PM_{2.5}$ and O_3, by solving mass equations with certain input dataset, e.g. emission and meteorology fields (Seinfeld and Pandis, 2016; Cheng et al., 2021; Zeng et al., 2022), thus bridging the connection between the inputs (emission and meteorology data) and the outputs (concentrations of air pollutants) (Shi et al., 2020; Yan et al., 2021). Therefore, CTMs have been used to investigate the sensitivity of air pollutant concentrations to anthropogenic emissions (Thunis et al., 2021) and meteorological conditions (Shi et al., 2020), as well as to guide the formulation of air quality policies by simulating the air quality response to various emission scenarios (Che et al., 2011; Zhang et al., 2022b, 2023b). For example, Zhang et al. (2023b) used the Weather Research and Forecasting and Community Multi-scale Air Quality (WRF-CMAQ) model to simulate O_3 concentrations under different $VOCs$ and NO_x emission reduction scenarios in Northeast China to explore effective control strategies for O_3 pollution; Zhang et al. (2022b) used the Goddard Earth Observing System with Chemistry model (GEOS-Chem), a CTM with the advantage of incorporating advanced gas-aerosol chemistry and consistently evaluating against atmospheric observations (Lu et al., 2020b; Hu et al., 2017), to evaluate the benefits of different NH_3 emission reduction strategies for $PM_{2.5}$ mitigation in China. Although CTMs demonstrate considerable accuracy in air pollution modeling, they also present notable limitations (Salman et al., 2024), including substantial consumption of computational resources and inefficiencies when conducting long-term simulations over extensive areas or high-resolution grids (Thompson and Selin, 2012). Typically, for GEOS-Chem version 14.2.2, on a computational cluster mentioned in Sect. 2.1.2, a standalone 1-year full-chem nested simulation of China at a resolution of $0.5° \times 0.625°$ requires approximately 350 h, and this duration is expected to increase when conducting simulations at finer resolutions or over extended time periods. This limitation makes it impossible for CTMs to meet the needs of policymakers for timely online responses to future air quality under interested scenarios.

To overcome the computational challenge and efficiently retrieve the nonlinear relationship between emissions and concentrations, data-driven statistical emulators have been proposed to accelerate numerical simulations (Castruccio et al., 2014). As a simplified-form of CTM, a reliable emulator can effectively capture the intricate relationships between important CTM inputs and concentration outputs, and rapidly estimate "CTM-aligned" concentrations of pollutants. As an effective framework for constructing CTM-based statistical surrogate models, Response Surface Model (RSM) was originally developed by the US EPA (US EPA, 2006) to establish the relationships between emission rates and the concentration responses of CTM by constructing pollutant response surfaces using polynomial or parametric regression methods. RSM been successfully employed in the response modeling of $PM_{2.5}$ (Wang et al., 2011) and O_3 (Xing et al., 2011) to precursor emissions in China for typical regions. To address the inherent computational burden stemmed from systematically perturbed CTM simulations for model building (Xing et al., 2011), optimized versions of conventional RSM were developed, such as ERSM (Zhao et al., 2015; Xing et al., 2017) and pf-RSM (Xing et al., 2018). Recently, novel machine learning (ML) techniques, for its well performance in simulating complex nonlinear relationships in atmospheric systems (Liu et al., 2021) and dealing with tasks involving multiple variables and objectives (Masmoudi et al., 2020; Huang et al., 2021), have been employed as alternative fitting modules within the RSM framework to further optimize modeling efficiency and estimation accuracy of RSMs (Xing et al., 2020; Li et al., 2022). Based on this advantage, many studies have attempted to build effective emulators using pure ML method (Huang et al., 2021; Zhang et al., 2023a). For example, Zhang et al. (2023a) used ResCNN framewoek to predict annual $PM_{2.5}$ concentration from fossil energy use and reveal the co-benefits of the energy transition, demonstrating the potential of ML method in addressing the emulator modeling task.

Although existing CTM emulators exhibit more efficiency than traditional CTM in estimating the pollutant concentrations to a wide range of emission changes, there are still several issues to be addressed. Firstly, due to the computing limitations (Liu et al., 2022), the temporal resolution for some emulators was constrained with annual scale, which greatly prevent these emulators from providing detailed estimations of air pollutants such as extreme values throughout the year (Guo et al., 2020; Zhao et al., 2022). Secondly, while some emulators have the ability to offer concentration estimations with finer temporal resolution, they still have limitations. On one hand, RSM-based emulators rely on the polynomial assumption, leading to its disadvantage to cope with high-dimension problems. As the number of input variables increases, the complexity of RSM model grows, necessitating a larger number of samples for accurate fitting (Zhao et al., 2015) and potentially leading to multicollinearity issues (Xing et al., 2018). This limitation restricts the applicability of these emulators to more intricate emission scenarios. Therefore, existing RSM-based emulators have primarily concentrated on emissions of a few major pollutants and the add-up emissions (Xing et al., 2020), failing to address air quality response under more detailed scenarios that incorporate sectoral emissions and a broader range of emission species. On the other hand, some studies directly used in-situ observations as targets based on ML method (Du et al., 2023; Zhang et al., 2023a), which is easy to employ and more convenient than those RSM-based emulators. However, these models are constrained by the limited number of observational data stations and are therefore unable to effectively assess air quality in regions where observational infrastructure is lacking (Xu et al., 2022). Furthermore, due to insufficient observational data, these models often do not have enough representative samples to achieve accurate model fitting, which leads to suboptimal predictive performance (Tang et al., 2024). In addition, traditional ML models, such as Multi-Layer Perceptron (MLP) and Random Forest (RF), may not fully capture the nonlinear relationships in complex atmospheric variables (Masmoudi et al., 2020; Natarajan et al., 2024; Abuouelezz et al., 2025), which further undermine their predictions. Thirdly, some current emulators account for each spatial grid or observation site independently while neglect the impact of surrounding emissions (Xing et al., 2018; Li et al., 2022; Zhang et al., 2023a), which have been shown to affect local pollutant concentrations (Cheng et al., 2019). Although certain studies have employed convolutional neural network (CNN) architectures capable of capturing local features to develop models (Xing et al., 2020; Huang et al., 2021; Liu et al., 2022), their applicability to detailed emission response and broader research domains has so far been limited. In summary, given that existing techniques inadequately address the challenges associated with high temporal-resolution prediction, inapplicability of multivariate scenarios, and negligence of emission trans-

port, it still be a significant challenge to develop a comprehensive emulator using more advanced method.

Transformer, as a renowned machine learning architecture characterized by the self-attention mechanism (Vaswani, 2017), has been substantially applied in the natural language process and image classification due to its ability of feature extraction and long-range dependency modeling (Devlin, 2018; Zhou et al., 2024). The self-attention mechanism facilitates the simultaneous evaluation of all positions within the input sequence and allows the model to discern dependencies across various species (Zhou et al., 2024), thus enabling the model to handle sophisticated high-dimensional data. The utilization of Transformer architecture for modeling in atmospheric science has become progressively more prevalent, exemplified by their incorporation into air quality forecasting, e.g. Informer (Zhou et al., 2021) and AirFormer (Liang et al., 2023), along with large-scale meteorological models for numerical weather prediction, such as Pangu (Bi et al., 2022), Fuxi (Chen et al., 2023c), and Fengwu (Chen et al., 2023b). This trend highlights the advantages of Transformer-based models over traditional approaches such as Random Forest and Multilayer Perceptron, especially in terms of their ability to capture complex patterns and relationships in data. However, owing to the stringent requirements in terms of datasets (Narayanan et al., 2021) and hardware resources (Vaswani, 2017), the application of the Transformer architecture in "emission-concentration" predicting research has been limited.

In this study, we proposed an efficient emulator of GEOS-Chem v14.2.2 based on Transformer architecture, with the capability to provide "GC-aligned" air quality predictions under future emission scenarios in China. It is referred to as "TGEOS" throughout this paper. Superior to earlier studies, TGEOS is capable to provide informative predictions about critical statistical indicators of monthly $PM_{2.5}$ and O_3 concentrations (e.g., 75-percentile and max values), and then have a general understanding of probability distribution of future air pollutants. Compared to solely average estimated by previous methods (Liu et al., 2022; Zhang et al., 2023a), probability distributions can provide informative frequency distributions of pollutants (Yang and Wu, 2022). Many studies have used probability distribution curves to represent future states of $PM_{2.5}$ (Li et al., 2024) and O_3 (Zeng et al., 2022) concentrations in diverse emission scenarios, and to explore any extreme pollution events that are typically represented by the high-end tail of the probability distribution curve (Zhang et al., 2018; Lu et al., 2020a), as well as the related health impact (Tian et al., 2022).

Second, TGEOS is suitable for concentration prediction in more comprehensive scenarios that include multiple precursor emissions from multiple sectors. Specifically, in contrast to previous emulators limited by scarce emission variables (Xing et al., 2011, 2020), sectoral emissions for 26 precursor emissions encompassing over 18 VOC species are incorporated into this model, which enhances the model's capacity to

address more flexible demands of policymakers towards interested emission scenarios. Third, given the significant influence of regional transport on local pollutant concentrations (Qiao et al., 2021) and the inability of current technologies to simultaneously consider the impact of regional transport and detailed emission variables, the effects of adjacent grids consist of emission, meteorological conditions, as well as geo-spatial data are taken into account to ensure the accuracy of predictions. In addition, with the use of the Transformer framework, TGEOS demonstrates significantly enhanced predictive accuracy compared to other machine learning models.

This paper is organized as follows: Sect. 2 introduces the dataset and methodology used for this study; Sect. 3 presents an in-depth analysis of TGEOS's performance on the test set, along with a comparative evaluation highlighting its advantages over alternative models; Conclusions are then summarized in Sect. 4.

2 Dataset and methodology

2.1 Dataset

To meet the demands of deep learning model training, we created a multi-scenario dataset based on several meticulously crafted emission scenarios and their corresponding GEOS-Chem simulations. The compilation of this dataset primarily consists of three components: generating multi-scenario emission inventories, conducting GEOS-Chem simulations, and assembling samples. The details of this process are discussed below.

2.1.1 Multi-scenario inventory

As a prerequisite to simulate future air quality, we produced a multi-scenario emission inventory of 36 emission scenarios, including 24 future emission scenarios, 11 fine-tuned scenarios and 1 background scenario. Detailed information on the inventory is shown in Table 1. We first used 24 future emission scenarios based on the DPEC (Dynamic Projection model for Emissions in China) platform (http://meicmodel.org.cn, last access: 13 February 2025) to initially construct the data set. As a dynamic model developed by Tsinghua University (Tong et al., 2020), DPEC can reflect the dynamic changes of China's future emissions under various socioeconomic and policy control scenarios, and provide detailed gridded emission data, including emissions with different control scenarios, emission sectors and spatial coordinate information. The DPEC-provided scenarios are widely recognized as a reliable basis for projecting China's future emission trajectories (Cheng et al., 2021). Firstly, we constructed a scenario set named "DPEC-SSP" to represent emission scenarios under different socio-economic scenarios and different emission control policies. DPEC-SSP was selected from DPECv1.0 (Tong et al., 2020) and consists of

five sub-scenario sets (SSP1 to SSP5) as combinations of different climate scenarios (i.e., SSP1-26, SSP2-45, SSP3-70, SSP4-60, SSP5-85) and pollution control scenarios (i.e., Business-As-Usual, Enhanced-control-policy, Best-Health-Effect). Sectoral emissions of 2030, 2040 and 2050 were selected in each sub-scenario, and each of them was treated as an independent emission scenario for the multi-scenario inventory. Furthermore, to capture policy-driven "clean air" pathways such as carbon peaking and carbon neutrality targets in China, another scenario set, DPEC-CA, was developed based on DPECv1.2 (Cheng et al., 2023) dataset. The DPEC-CA was composed of three sub-scenario sets including "clean air", "on-time peak-clean air", and "early peak-net zero-clean air". Each of these scenario was constructed on SSP1 assumption, without introducing additional climate or air pollution control policies, to reflect different short-term carbon emission reduction policies in China. Compared with DPEC-SSP, DPEC-CA scenarios represent more stringent emission control policies and can provide more "low-value" samples to enrich training set. Similar to DPEC-SSP, three years (2030, 2040, and 2050) were selected from each DPEC-CA sub-scenario, with each year treated as an independent emission scenario.

Since the unit of DPEC-SSP/CA emissions is tons per grid, which is incompatible for GEOS-Chem running, we used MEIC inventory with tons per grid as unit at 2017 as a benchmark (denoted as b-MEIC), and make elementwise divisions between DPEC and b-MEIC to obtain a series of monthly emission coefficient matrices for various species in five sectors, namely power, industry, residential, transportation, and agriculture. Since the majority of grids with emission factor smaller than 2.0 ($> 80\%$) and to prevent abnormal values due to magnitude difference of two inventories, the threshold of emission factors was artificially set to 2.0. Subsequently, we took the Schur product of the coefficient matrices and corresponding part of MEIC inventory used for GEOS-Chem input, with unit of $kg\,m^{-2}\,s^{-1}$, to generate emission inventories projected with DPEC-SSP/CA.

In addition, in order to improve the generalization ability of the model, we designed 11 perturbation scenarios using data assimilation tuning method (denoted as "Tuning scenarios" in Table 1), including emission scenarios with different emission factors ranging from 0 to 2.0 for each emission species and emission sector, thereby expanding coverage of the input space and reducing the risk of extrapolation to unseen values, especially for those predictions under high emission scenarios. These emission factors were generated for representing the spatial variability that widely used in data assimilation (Jin et al., 2023). The detailed process for generating these stochastic emission factors is discussed in the Text S1 in the Supplement.

Focusing on eight key emission variables that predominantly influence $PM_{2.5}$ and O_3 concentrations (Pinder et al., 2007; Wang et al., 2013; Lu et al., 2019; Skyllakou et al., 2021; Lai et al., 2021), we analyzed the distributions across

three emission scenario sets: DPEC-SSP, DPEC-CA, and tuning scenarios. As illustrated in Figs. S2 and S3 in the Supplement, although the magnitudes are generally comparable, noticeable differences among the three curves are evident for each emission variable, thereby confirming the separability of these datasets.

2.1.2 GEOS-Chem configuration

The GEOS-Chem chemical transport model (http://www.geos-chem.org, last access: 15 January 2025, version 14.2.2) (GC) was used to simulate the spatiotemporal distribution of surface $PM_{2.5}$ and O_3 concentrations under different emission scenarios based on year 2017. The nested model was configured with a horizontal resolution 0.5° latitude by 0.625° longitude covering China (from 17.5 to 54° N and 72 to 136° E) and 47 vertical layers. Boundary condition files for model startup were offered by 1-year global GC simulation after spin-up of 6 months with a horizontal resolution of 2° latitude by 2.5° longitude. Assimilated meteorological data from the NASA Global Modeling and Assimilation Office's Modern-Era Retrospective analysis for Research and Applications Version 2 (MERRA-2) (Gelaro et al., 2017) were selected as meteorology fields entering the model through HEMCO interface, with the 3 h temporal resolution and 0.5° × 0.625° spatial resolution. Consistent with previous studies on future air quality assessment (Shi et al., 2021; Liu et al., 2022; Wang et al., 2023a), the meteorological inputs were fixed at 2017 for all simulations in this study to isolate concentration changes attributable solely to emission variations. The Multi-resolution Emission Inventory for China (MEIC, http://meicmodel.org/, last access: 20 January 2025) (Li et al., 2017) and the multi-scenario emission inventories with a horizontal resolution of 0.25° latitude by 0.25° (as detailed in Sect. 2.1.1) were used as the monthly anthropogenic emissions to simulate $PM_{2.5}$ and O_3 concentrations under various emission scenarios. For anthropogenic emissions out of China, we used data from the Community Emissions Data System (CEDS) inventory (Hoesly et al., 2018). All the GC simulations in this research relied on identical configurations: GEOS-Chem version 14.2.2, compiled with OpenMP parallelization (OMP_NUM_THREADS = 32), executed on a Linux cluster node equipped with two Intel(R) Xeon(R) E5-2620 v4 CPUs (8 cores per socket, 32 logical processors total) and 62 GB RAM.

2.1.3 Multi-scenario dataset

Before model training and evaluation, we constructed a multi-scenario dataset by combining emissions with the corresponding GC simulations, as summarized in Table 1. Each sample in this dataset is defined on a grid-cell basis. All emission data were interpolated to a spatial resolution of 0.5° × 0.625° to match the GC simulations. Detailed descriptions of the input features and output targets are provided in Table 2. For each grid cell, 105 sectoral emissions were selected as predictors to represent local emission conditions. At this resolution, a 3 × 3 grid domain already covers a geographic extent for representing influences of neighborhood areas towards local pollutants in a monthly scale, seeing details in Text S2. Therefore, to account for regional transport of precursors, we incorporated the emissions of the eight neighboring cells. In addition, nine key meteorological variables, previously identified as strongly correlated with $PM_{2.5}$ and O_3 concentrations (Shi et al., 2020; Zhang et al., 2022a), were included for each local and neighboring grids, based on the 2017 MERRA-2 reanalysis processed into monthly averages. Spatial information of both the local and adjacent grids was further incorporated to enable the model to capture spatial heterogeneity in emissions and pollutant concentrations. The training targets consisted of twelve statistical indicators on a monthly scale, including the 25th and 75th percentiles, median, mean, maximum, and minimum, derived from the daily averaged concentrations of $PM_{2.5}$ and O_3 in the GC outputs for each scenario. It is important to emphasize that our analysis focuses on air quality responses to anthropogenic emission changes. Consequently, dust-related components were excluded during preprocessing, since dust intrusions can introduce large predictive biases in northern and western China, where they make substantial contributions to $PM_{2.5}$ concentrations (Pang et al., 2023).

2.2 Methodology

2.2.1 Model architecture

Previous emulator studies have often adopted field-based modeling strategies, in which both inputs and outputs are represented as spatially explicit two-dimensional fields (Xing et al., 2020; Huang et al., 2021; Liu et al., 2022). While effective in data-rich settings, such formulations typically require a large number of training samples to robustly learn high-dimensional spatial mappings. In the present study, the number of available scenarios is limited, yielding fewer than 500 samples in total for training. This sample size is insufficient to support stable training of high-capacity field-to-field models, particularly when the input space includes more than 100 variables spanning sectoral emissions and multiple meteorological parameters. Under this data regime, directly modeling full spatial fields would substantially increase the risk of overfitting and unstable generalization. Therefore, rather than adopting a purely field-based representation, we reformulate the problem as a high-dimensional sequential learning task. The dataset is organized as structured multivariate sequences, in which spatial and feature-level information, such as emissions, meteorology, and concentrations over a 3 × 3 neighborhood (nine grid cells), is flattened and treated as a sequence of tokens input to the TGEOS model. This formulation is better aligned with the available sample size and enables more efficient utilization of limited training data,

Table 1. Description of multi-scenario inventory.

Scenario set	Sub-scenario set	Independent scenarios	Description
DPEC scenarios	DPEC-SSP	(1) SSP1 (2030, 2040, 2050)	SSP1-26-BHE control
		(2) SSP2 (2030, 2040, 2050)*	SSP2-45-ECP control
		(3) SSP3 (2030, 2040, 2050)*	SSP3-70-BAU control
		(4) SSP4 (2030, 2040, 2050)	SSP4-60-BAU control
		(5) SSP5 (2030, 2040, 2050)	SSP5-85-BHE control
	DPEC-CA	(1) ep (2030, 2040, 2050)	early_peak-net_zero-clean_air control
		(2) otp (2030, 2040, 2050)	on-time_peak-clean_air control
		(3) ca (2030, 2040, 2050)	clean_air control
Tuning scenarios	–	tun0.05, tun0.2, tun0.4, tun0.6, tun0.8, tun1.0, tun1.2, tun1.4, tun1.6, tun1.8, tun1.95	Random scenarios generated by tuning experiments
Background scenario	–	base2017	Original 2017 emission scenario

* Samples corresponding to six scenarios from SSP2 and SSP3 in the multi-scenario dataset were utilized for model testing, whereas the remaining samples were employed for model training.

Table 2. Targets and features for TGEOS model.

Target number	Training targets	Feature number	Training Features
12	Monthly average, maximum, minimum, median, 25 and 75 percentiles of $PM_{2.5}$ and O_3 concentrations	1045	(1) power, industry, residential, transportation, and agriculture emissions: NH_3, $PM_{2.5}$, OC, PM_{10}, BC, CO, NO, SO_2, RCHO, XYLE, ALK2, CCHO, OLE2, ALK5, HCHO, TOLU, ALK4, ALK3, EOH, ETHE, MOH, ALK1, MEK, OLE1, ACET, MACR, as well as 8 adjacent sectoral emissions. (2) 2 m air temperature (T2M), 10 m northward wind (V10M), 10 m eastward wind (U10M), planetary boundary layer height (PBLH), 2 m specific humidity (QV2M), total precipitation (PRECTOT), relative humidity (RH), evaporation from turbulence (EVAP), and surface pressure (PS), as well as 8 adjacent meteorology. (3) local and adjacent longitude and latitude values, and month in each scenario.

while still preserving key spatial and cross-variable dependencies among neighboring grid points.

The TGEOS model comprises the encoder for feature extraction and the regressor for target mapping, with detailed architecture illustrated in Fig. S1. In order to align with the shape of the dataset, the model was configured with an input feature dimension of 1045 and an output dimension of 12. Six Encoder layers were configured with the model, each of which primarily incorporates a multi-head self-attention mechanism with eight attention heads and a feed-forward network. The multi-head self-attention mechanism was employed to capture the dependency relationships among various positions within the input sequence, while the feed-forward network facilitates additional nonlinear transformations on the features at each position (Vaswani, 2017). By leveraging the multi-head self-attention mechanism, the model can compute the similarity (or attention weights) of each feature in relation to all other features, thus producing a weighted representation for each position and determining the extent to which each position relies on information from others. Moreover, the feed-forward network, consisting of two fully connected layers, enhanced feature representation and improves the model's learning efficacy by incorporating nonlinear activation functions. In this implementation, the ReLU activation function was selected due to its ability to prevent negative values and expedite the model's training process (Nair and Hinton, 2010). Additionally, each sub-module incorporated residual connections and layer normalization to mitigate the risks of gradient disappearance or explosion. The output from the Encoder undergoes global pooling to decrease model complexity. Finally, the output of the encoder is transformed into the specified sequence by the regressor, which is implemented as a linear layer.

As depicted by Fig. 1, the model incorporates local and surrounding sectoral emissions for each grid, along with various meteorological parameters, to predict the probability

distribution of pollutants under different scenarios, which is characterized by a series of concentration indicators including the average, maximum, minimum, median, 25 and 75 percentile of $PM_{2.5}$ and O_3. Previous research has indicated that $PM_{2.5}$ and O_3 concentrations tend to follow characteristic statistical patterns (Zhang et al., 2018; Zeng et al., 2021), with $PM_{2.5}$ generally displaying a right-skewed Gamma-like distribution and O_3 approximating a normal distribution. This distinction is also evident from the comparison of their mean and median values. Based on this insight, we used the TGEOS-predicted statistical indicators to approximate regional probability distribution curves. Specifically, the mean, 25th, and 75th percentiles were applied to capture the overall shape of the distributions, while the minimum and maximum values were incorporated to constrain their ranges.

2.2.2 Model training and evaluation

In this study, four machine learning models were employed independently to evaluate the performance for each kind of model structure. Except for the TGEOS model discussed in this paper, two traditional ML models, e.g. Multilayer Perceptrons (MLP) and Random Forests (RF), as well as two advanced ones, namely Convolutional Neural Network (CNN) and Vision Transformer (ViT), which demonstrated good performance in air quality modeling (Huang et al., 2021; Fang et al., 2023; Pang et al., 2025), were simultaneously employed based on the same training set. For each model, the optimizer and loss function were identical, and hyperparameters were obtained after fine-tuning based on Optuna tool, and specific tuning schedules were listed in Tables S5 to S11 in the Supplement. Training and evaluation of these models were conducted on a GPU-equipped server. Specifically, the benchmark was measured on an NVIDIA GeForce RTX 4080 with 31 GB memory, using PyTorch 2.3.1 with CUDA 12.4 and Python 3.11.5, running under Ubuntu 20.04.6.

The detailed dataset for model construction was derived from the multi-scenario inventory presented in Table 1. Samples from a total of 29 scenarios within the multi-scenario dataset were selected to construct the training set. While samples from six scenarios of SSP2 (SSP2_30, SSP2_40, SSP2_50) and SSP3 (SSP3_30, SSP3_40, SSP3_50) that representing low and high emission scenarios relative to 2017 background scenario were chosen to construct the test set.

In order to optimize the ability of TGEOS model to reproduce GEOS-Chem simulations, the mean squared error (MSE) was adopt as the loss function that measures the squared variance between TGEOS predicted (m_i) and GC simulated (g_i) concentrations to supervise the model training.

$$\mathcal{L}(m) = \frac{1}{N} \sum_{i=1}^{N} (m_i - g_i)^2 \quad (1)$$

The model weights were optimized with respect to the loss function using the Adam optimizer (Kingma, 2014) with an learning rate is 1×10^{-4}. To save the optimal model weights during training, 20 % of the randomly sampled training data were set aside for model validation purposes. The model was trained for 100 epochs with a batch size of 64. To reduce the risk of overfitting, we applied L2 weight regularization on all trainable parameters during training and fine-tuning.

The performance of TGEOS was evaluated using three statistical indices commonly used in evaluating the performance of CTM emulators (Salman et al., 2024), namely, coefficient of determination R-Square (R^2) and Mean Absolute Error (MAE). Their corresponding mathematical formulas are delineated as follows.

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (m_i - g_i)^2}{\sum_{i=1}^{N} (m_i - \overline{m})^2} \quad (2)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |m_i - g_i| \quad (3)$$

Here m_i and g_i denote the TGEOS-predicted and GC-simulated pollutant concentrations, respectively. Indices i means the ith grid cell. $\overline{m}$ is the average of all the model-predicted samples and N refers to the number of samples from the training set.

3 Results and discussions

The overall performance of TGEOS on the test set is shown in Table S1. We found that the model performed well across all target indicators. The R^2 ranges from 0.958 to 0.992, with relatively low RMSE and MAE, averaging 2.808 and 1.588 $\mu\text{g m}^{-3}$, respectively. The following presents detailed analyses: Sect. 3.1 focuses on analyzing the differences between the training set and the test set; Sect. 3.2 and 3.3 involves predicting spatial and probability distributions of $PM_{2.5}$ and O_3 concentrations; Sect. 3.4 is dedicated to comparison of different models.

3.1 Differences between training and test set

Since emission trajectories with different reduction rates may converge at certain time horizons, there exists a potential risk of data leakage arising from similarities in emission and concentration levels across scenarios. To address this concern, we analyzed the Kernel Density Estimation (KDE) curves for six key emission variables, which strongly influence $PM_{2.5}$ and O_3 concentrations (Hu et al., 2023), of the training and test set, as illustrated in Fig. S5. The results indicate that, although the general distribution trends are similar, the densities at different emission levels vary significantly between the two. Furthermore, focusing on the North China Plain (NCP)
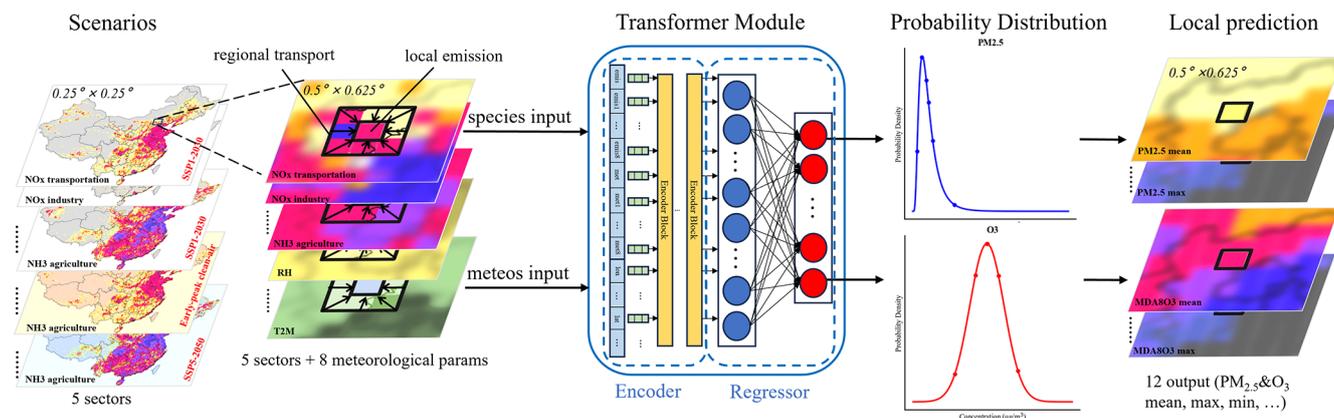
Figure 1. Workflow of TGEOS technique.

where both $PM_{2.5}$ and O_3 pollution are particularly severe, we examined the spatial distribution of mean $PM_{2.5}$ and O_3 concentrations, six critical emissions as well as corresponding absolute differences under the stochastically selected SSP2_2050 test scenario, in comparison with a training scenario (otp2030). The otp2030 scenario was selected by calculating the Euclidean distance between the mean $PM_{2.5}$ and O_3 values of SSP2_2050 and those of each training scenario, and identifying the scenario with the minimum distance. The results are illustrated in Fig. 2 for $PM_{2.5}$ and Fig. S4 for O_3. These pictures indicated that the concentrations of pollutants, as well as emission variables, of the training and test set are exclusive despite some distributional similarities, particularly for samples from highly polluted regions.

In addition, we conducted Kolmogorov-Smirnov (K-S) tests on a total of 12 emission variables, comprising the aforementioned six emissions as well as an additional set of six emissions, with the results summarized in Table S4. Given the large sample size, the p-values for all emission variables are approximately zero (Demir, 2022), making the KS statistic (D-value) a more meaningful indicator. Our analysis shows that the emissions of the two scenarios differ to varying extents, with all D values being greater than zero. It is noteworthy that emission changes are primarily concentrated in major emission regions of eastern China, whereas in many western and southern regions the variations across scenarios are negligible. This spatial heterogeneity implies the presence of redundant samples in the dataset, which could in turn contribute to statistical similarities between scenarios when comparisons are made (D-value < 0.3). Nevertheless, for most emission variables, D values exceed 0.1, suggesting that certain differences still exist between the two scenarios.

3.2 Prediction of spatial distribution of $PM_{2.5}$ and O_3

We first evaluated the spatial distribution performance of TGEOS predictions of $PM_{2.5}$ and O_3 for 6 test scenarios. For the sake of brevity, we presented the results of two test

scenarios, namely SSP2_2040 and SSP3_2040, to represent the low and high emission scenarios. Additionally, we focused on the months with the highest concentrations of $PM_{2.5}$ and O_3 to better visualize the spatial distribution of pollutants. Figures 3 and S6 present the spatial comparison of $PM_{2.5}$ concentration indicators between GC and TGEOS for SSP2_40 and SSP3_40 scenario in January.

The results demonstrate that, across various emission scenarios, the spatial distribution of $PM_{2.5}$ concentrations simulated by TGEOS exhibits a high degree of similarity with that produced by GC. Specifically, TGEOS effectively captures the spatial distribution patterns of $PM_{2.5}$ concentrations, accurately identifying high-pollution zones in central and northern China, alongside low-pollution areas in other regions. Furthermore, the disparities observed in $PM_{2.5}$ concentration levels under distinct emission scenarios indicate that TGEOS has successfully discerned the intricate relationships between precursor emissions and $PM_{2.5}$ concentrations. Beyond its capability to predict monthly mean values, TGEOS also excels in predicting additional statistical indicators associated with $PM_{2.5}$ concentrations, including the maximum concentrations of significant concern to policymakers as well as the 25th and 75th percentile, which reflect the distribution of concentrations. Other statistical indicators, such as the median and minimum values that shown in Fig. S8, are also effectively predicted.

As illustrated in Figs. 5a, 6a, and S10a, there exists a robust statistical correlation between the $PM_{2.5}$ indicators predicted by TGEOS and the actual GC simulations across varying emission scenarios, with R^2 values ranging from 0.976 to 0.995. These results substantiate that $PM_{2.5}$ accurately captures the principal trends and patterns of $PM_{2.5}$ as simulated by GC. The evaluation of model prediction errors, as quantified by the RMSE and MAE, reveals relatively low error levels, with RMSE values ranging from 0.985 to 2.110 and MAE values between 0.685 and 3.243, demonstrating the predictive capabilities of TGEOS with a high degree of accuracy and reliability. The MBE values are ranging from
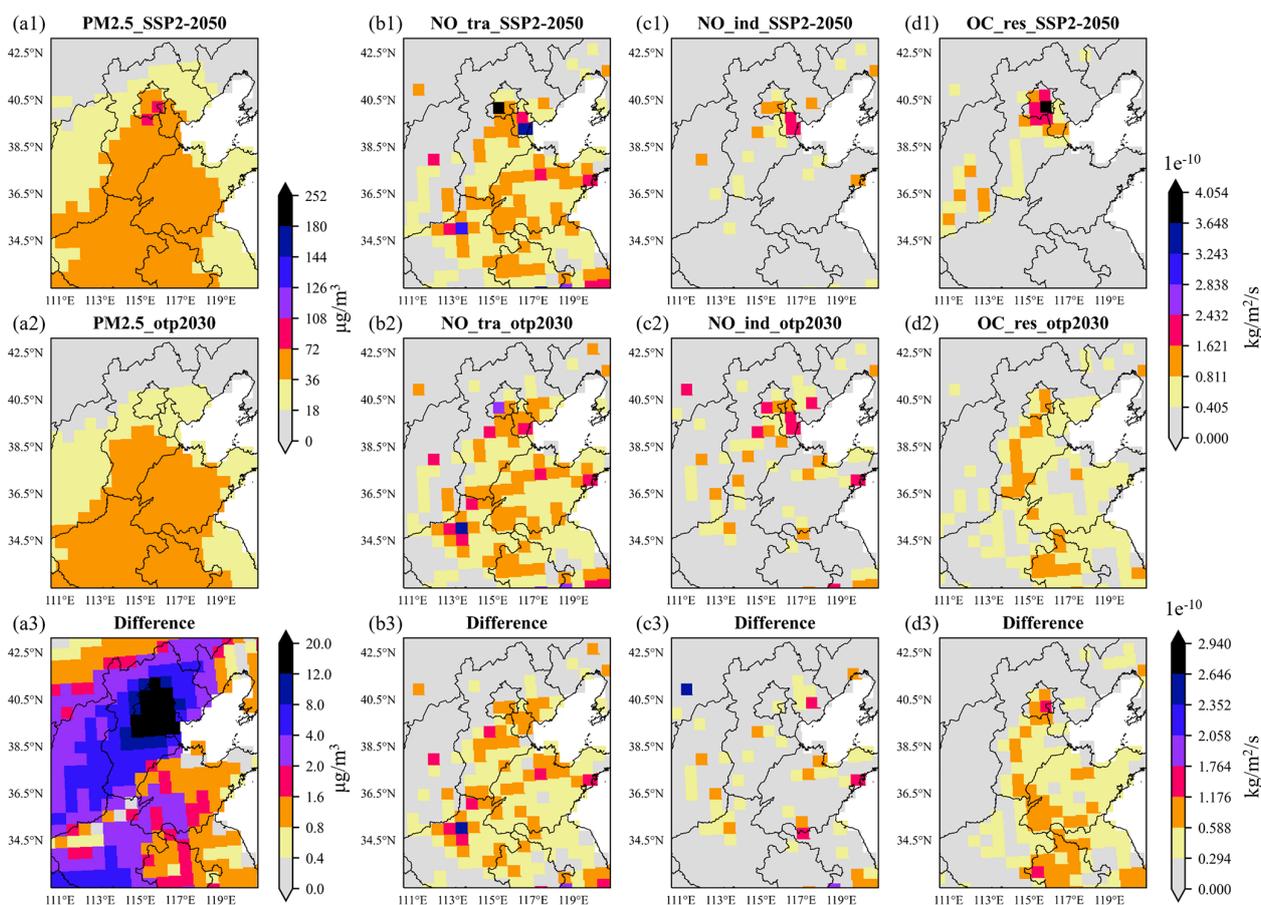
Figure 2. Spatial distributions of mean PM$_{2.5}$ concentration and three emission variables in SSP2_2050 **(a)** and otp2030 **(b)** scenarios in January, along with the quantified absolute differences between two scenarios **(c)**.

−1.453 to 1.420 for PM$_{2.5}$, −0.033 to 1.125 for O$_3$, indicating a slight overall deviations in concentration predictions compared to corresponding GC simulations. Considering that this bias is relatively small compared to the magnitude of the concentrations, the model can be regarded as nearly unbiased. In addition, to evaluate the capability of TGEOS in capturing extreme events, we employed exceedance metrics based on the 90th percentile threshold of the concentration distribution. The results indicate that the model achieves high precision and recall score for both PM$_{2.5}$ and O$_3$ indicators, with all these values larger than 0.85. These values suggest that the majority of the predicted exceedance events correspond to actual exceedances, while nearly all true exceedance events are successfully detected. The high and balanced values of both metrics demonstrate that TGEOS is capable of accurately identifying extreme high-value occurrences with low false alarm and miss rates. Moreover, this performance highlights the robustness of the model in reproducing the upper tail of the distribution, which is particularly important for applications focusing on extreme pollution events.

The spatial comparison of O$_3$ concentration indicators between GC and TGEOS for two scenarios in July are presented in Figs. 4, S7, and S9. Similar to the predictions for PM$_{2.5}$, we observed that TGEOS successfully captures the spatial distribution patterns of O$_3$ as simulated by GC, as well as the concentration differences resulting from various emissions. The scatter density plots presented in Figs. 5b, 6b, and S10b indicate a strong correlation between TGEOS and GC, with R^2 values ranging from 0.966 to 0.996. Additionally, the accuracy of TGEOS predictions is further supported by the relatively low RMSE values, which range from 0.985 to 2.110, and MAE values ranging from 1.593 to 4.933. These results demonstrate that TGEOS is capable of accurately and reliably predicting both PM$_{2.5}$ and O$_3$ concentration distribution across different scenarios, achieving a level of performance comparable to that of GC.

The error graphs of PM$_{2.5}$ indicators for SSP2_2040 and SSP3_2040 are shown in Figs. 3a3 to d3 and S6a3 to d3. We found the model exhibits relatively large errors in predicting the monthly maximum concentrations of PM$_{2.5}$. This is attributed to the inherent randomness of these peak values compared to other indicators, which poses challenges for accurate prediction. Furthermore, our analysis indicates the presence of both overestimation and underestimation within
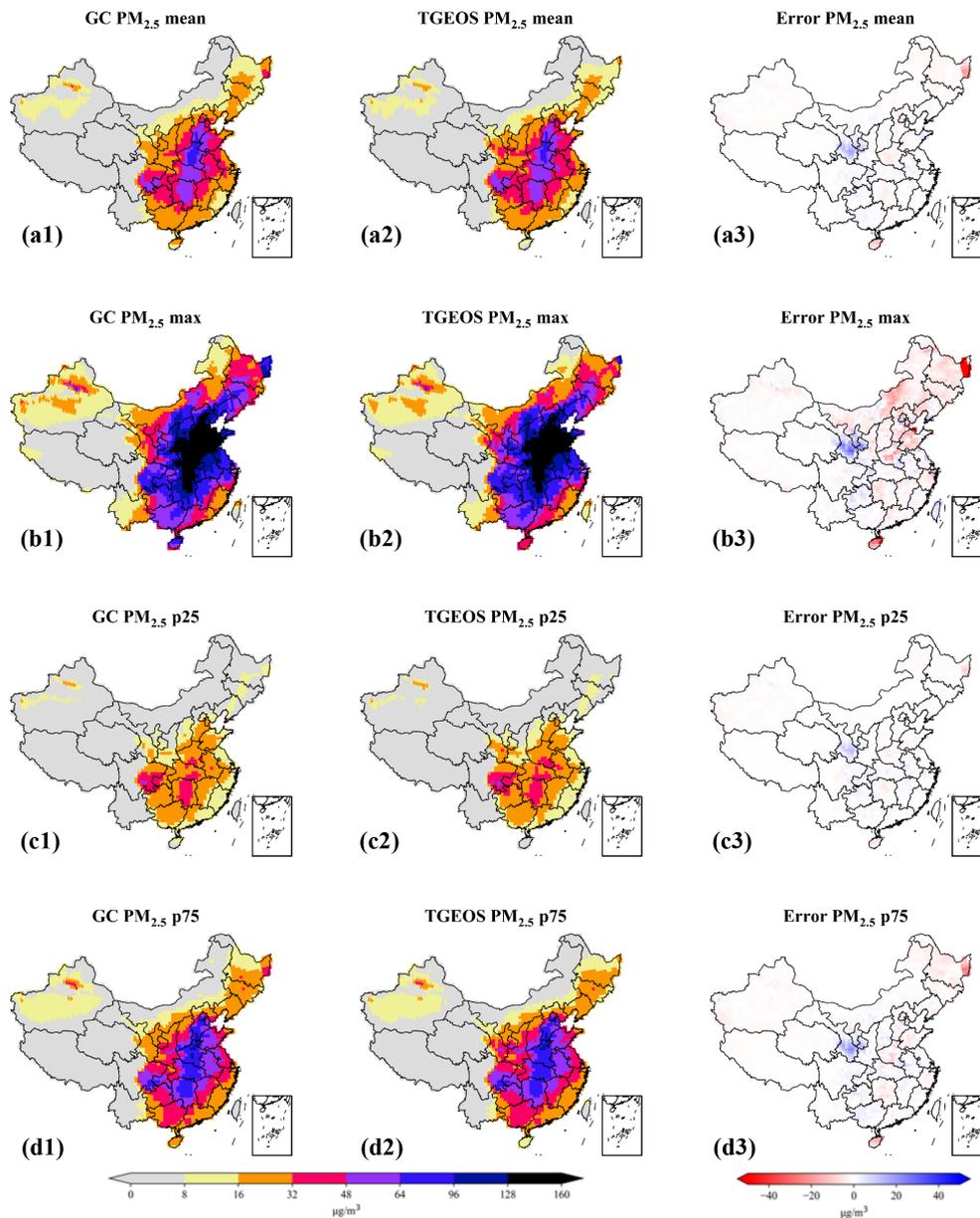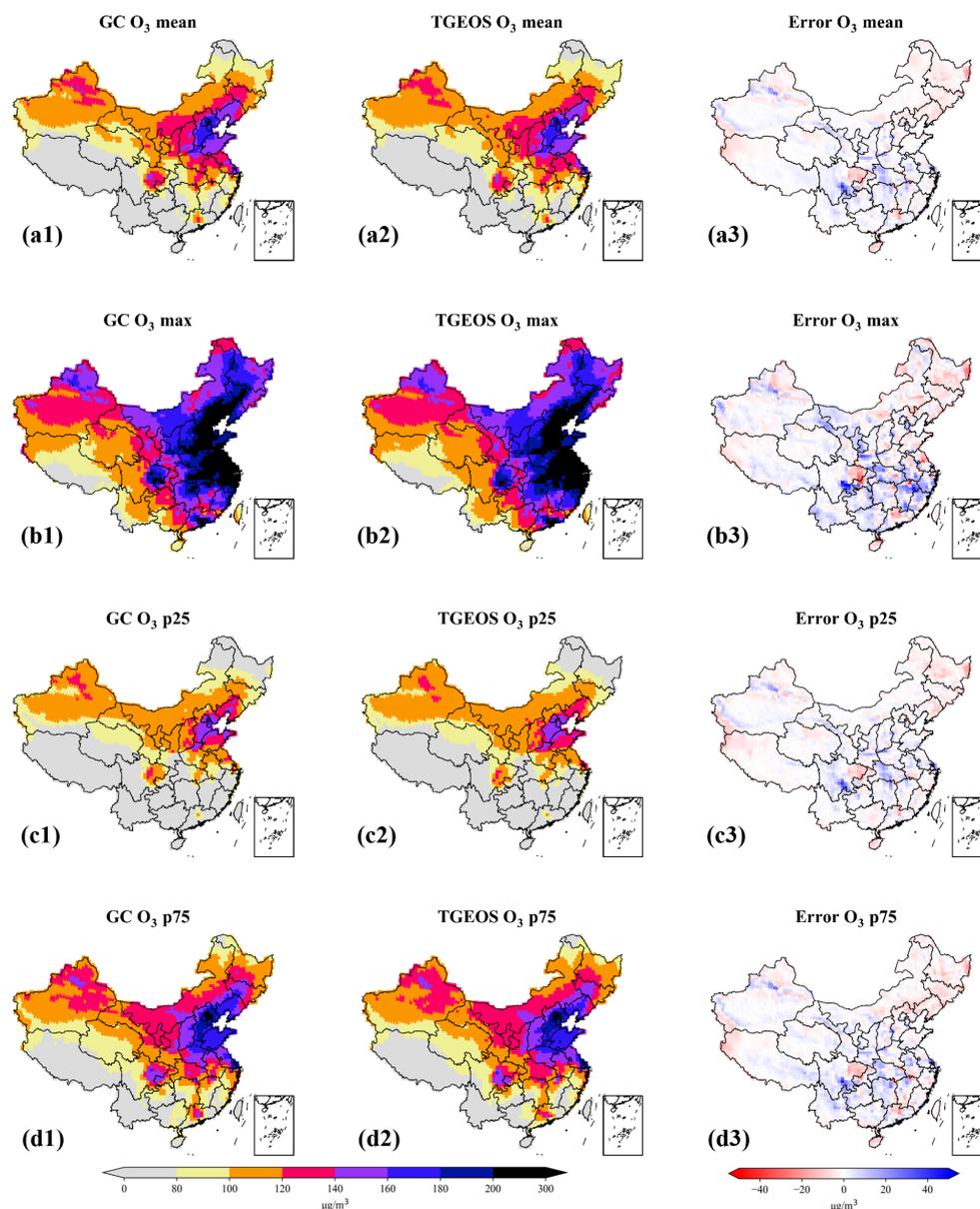
Figure 3. Spatial comparison of GEOS-Chem simulated and TGEOS estimated four statistical indicators of PM$_{2.5}$ concentrations in January under SSP2_2040 scenario (low emissions) and the corresponding error maps. Panels **(a1)** to **(d1)** represent GC simulation while **(a2)** to **(d2)** represent TGEOS prediction, and **(a3)** to **(d3)** represent the error between this two for each indicator, including mean, maximum, 25-percentile, and 75-percentile.

these error graphs. In this study, the GC simulations for each scenario were initialized from a fixed concentration field derived from the 2017 background scenario. As monthly concentrations were treated as independent and did not incorporate the influence of initial fields, discrepancies may arise between model predictions and GC outputs, especially when future concentration levels deviate substantially from the initial state. This effect helps explain the relatively poorer predictive performance under SSP2 scenarios (Fig. 5), as well as the observed patterns of systematic over- and underestima-

tion in the error distributions. Specifically, in the SSP2 scenario (SSP2-45-ECP), stringent environmental policies are projected in the short and medium term (Tong et al., 2020), thereby widening the gap between future and historical emissions and amplifying predictive errors, particularly during the early simulation period. In contrast, under the SSP3 scenario (SSP3-70-BAU), characterized by pessimistic development trajectories and limited investments in environmental protection (Tong et al., 2020), emissions are projected to change slightly, resulting in smaller differences from histor-

Figure 4. Spatial comparison of GEOS-Chem simulated and TGEOS estimated four statistical indicators of O_3 concentrations in July under SSP2_2040 scenario (low emissions) and the corresponding error maps. Panels **(a1)** to **(d1)** represent GC simulation while **(a2)** to **(d2)** represent TGEOS prediction, and **(a3)** to **(d3)** represent the error between this two for each indicator, including mean, maximum, 25-percentile, and 75-percentile.

ical conditions (Fig. 6). Consequently, predictions in SSP3 scenarios are less affected by initialization effects than those in SSP2.

The predictions from TGEOS also demonstrate a clear pattern of seasonal variation. Here, we focus on two statistical indicators that are crucial for fitting the probability distribution curve, namely the 25th percentile and the 75th percentile, and select results from January, April, July, and October to represent the distribution of pollutants during winter, spring, summer, and autumn, respectively. Figures 7 to 8 and

S15 to S16 illustrate the seasonal variation of $PM_{2.5}$ and O_3 indicators predicted by GC and TGEOS under SSP2_2040 and SSP3_2040 scenarios. The R^2 values for the 25th and 75th percentile of $PM_{2.5}$ are 0.964 to 0.994 and 0.973 to 0.996, respectively, while those for O_3 are 0.903 to 0.994 and 0.946 to 0.994, respectively, indicating a strong correlation between predicted and simulated pollutant concentrations across all seasons.

Specifically, TGEOS effectively captures the seasonal trends and patterns of $PM_{2.5}$ and O_3 as simulated by GEOS-
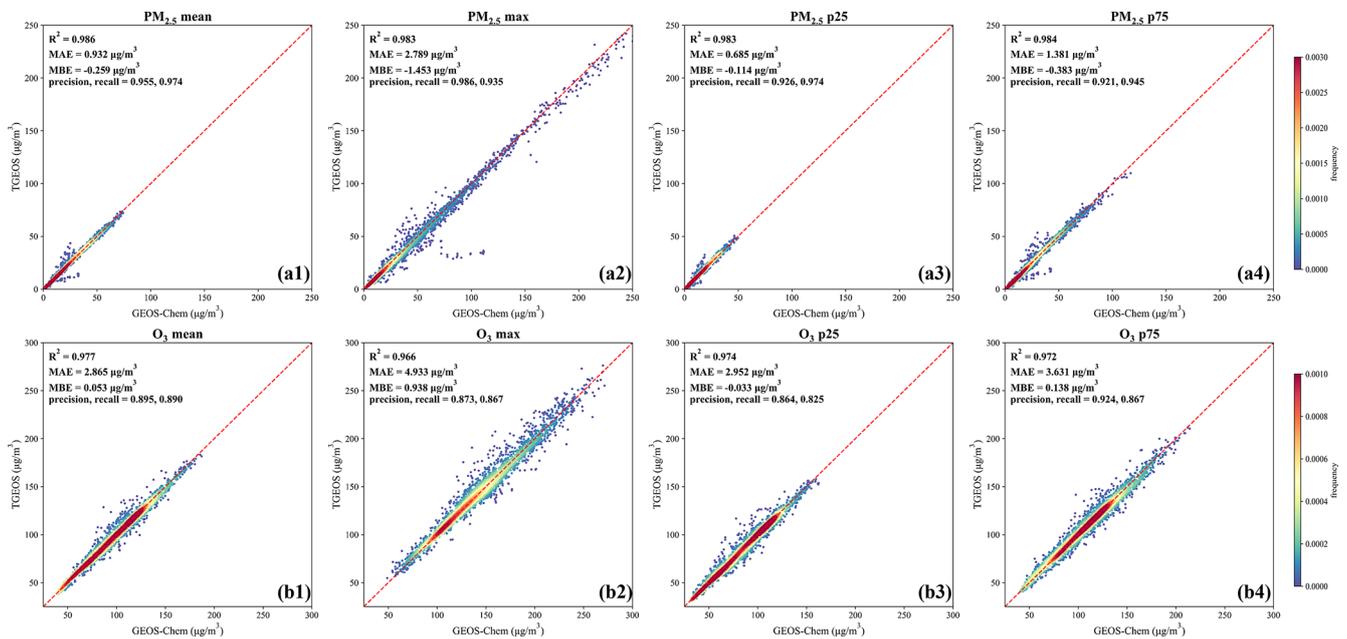
Figure 5. Density scatter plots between GEOS-Chem simulations and TGEOS predictions for eight indicators of $PM_{2.5}$ **(a)** and O_3 **(b)** concentrations in SSP2_2040 scenario. Panels **(a1)** to **(a4)** denote the mean, maximum, 25th percentile, and 75th percentile of January $PM_{2.5}$ concentration; **(b1)** to **(b4)** denote the corresponding statistics for July O_3 concentration.
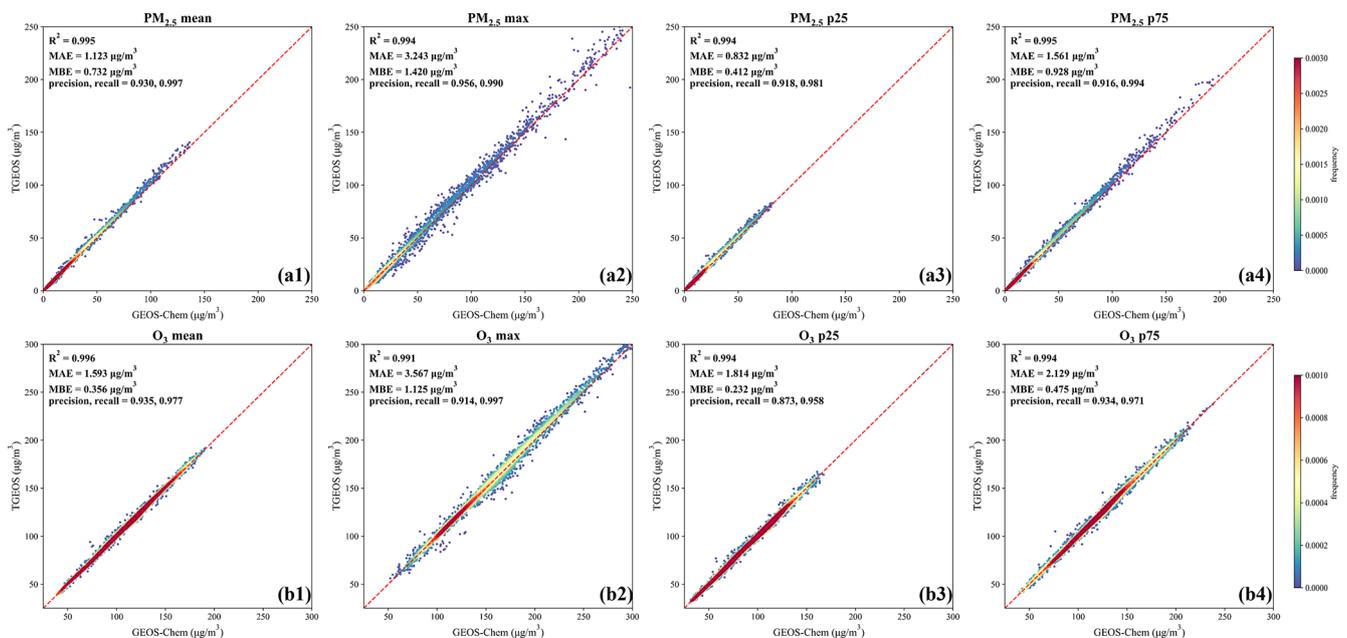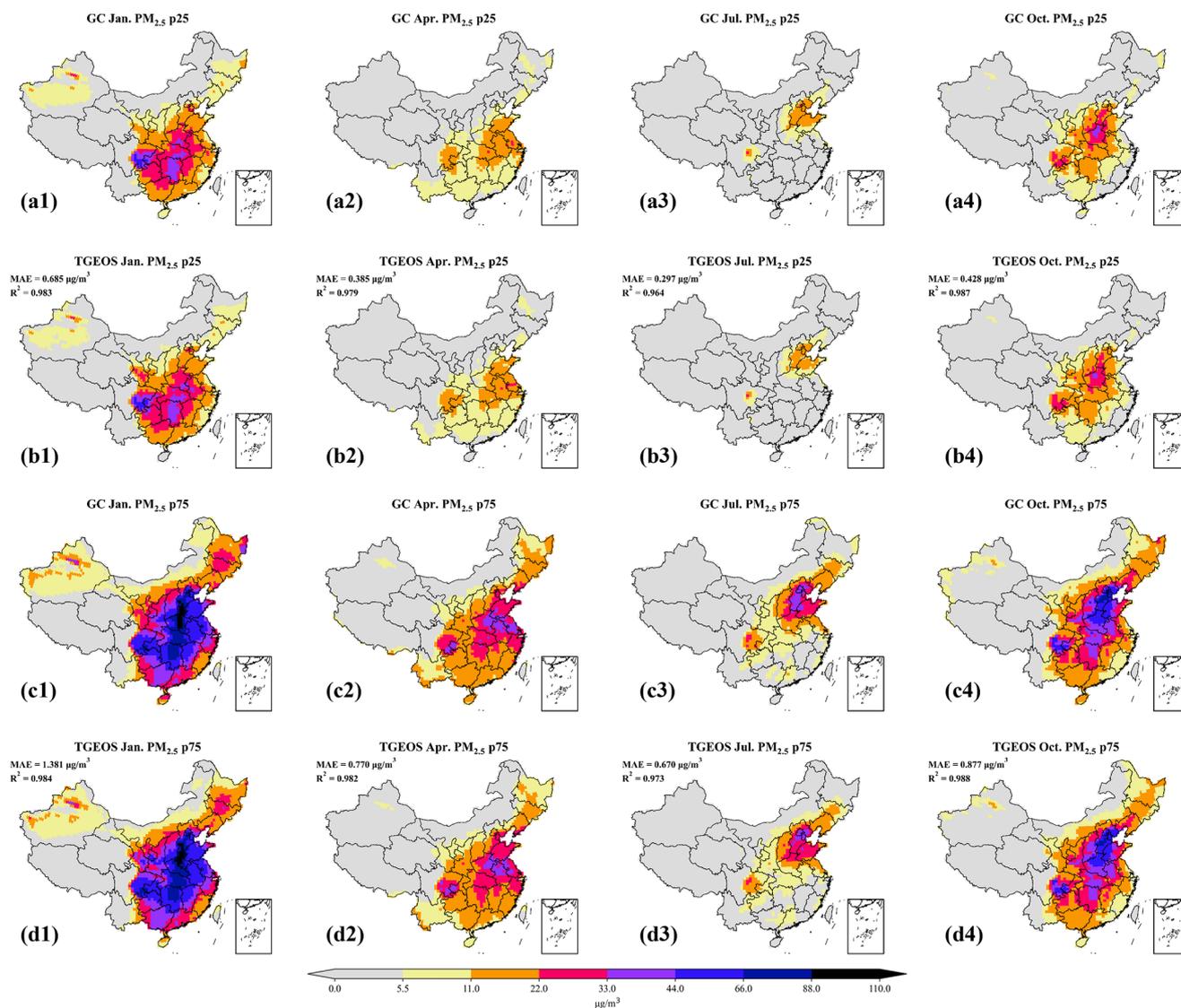

Figure 6. Density scatter plots between GEOS-Chem simulations and TGEOS predictions for eight indicators of $PM_{2.5}$ and O_3 concentrations in SSP3_2040 scenario. Panels **(a1)** to **(a4)** denotes the mean, maximum, 25th percentile, and 75th percentile of January $PM_{2.5}$ concentration; **(b1)** to **(b4)** denotes the corresponding statistics for July O_3 concentration.
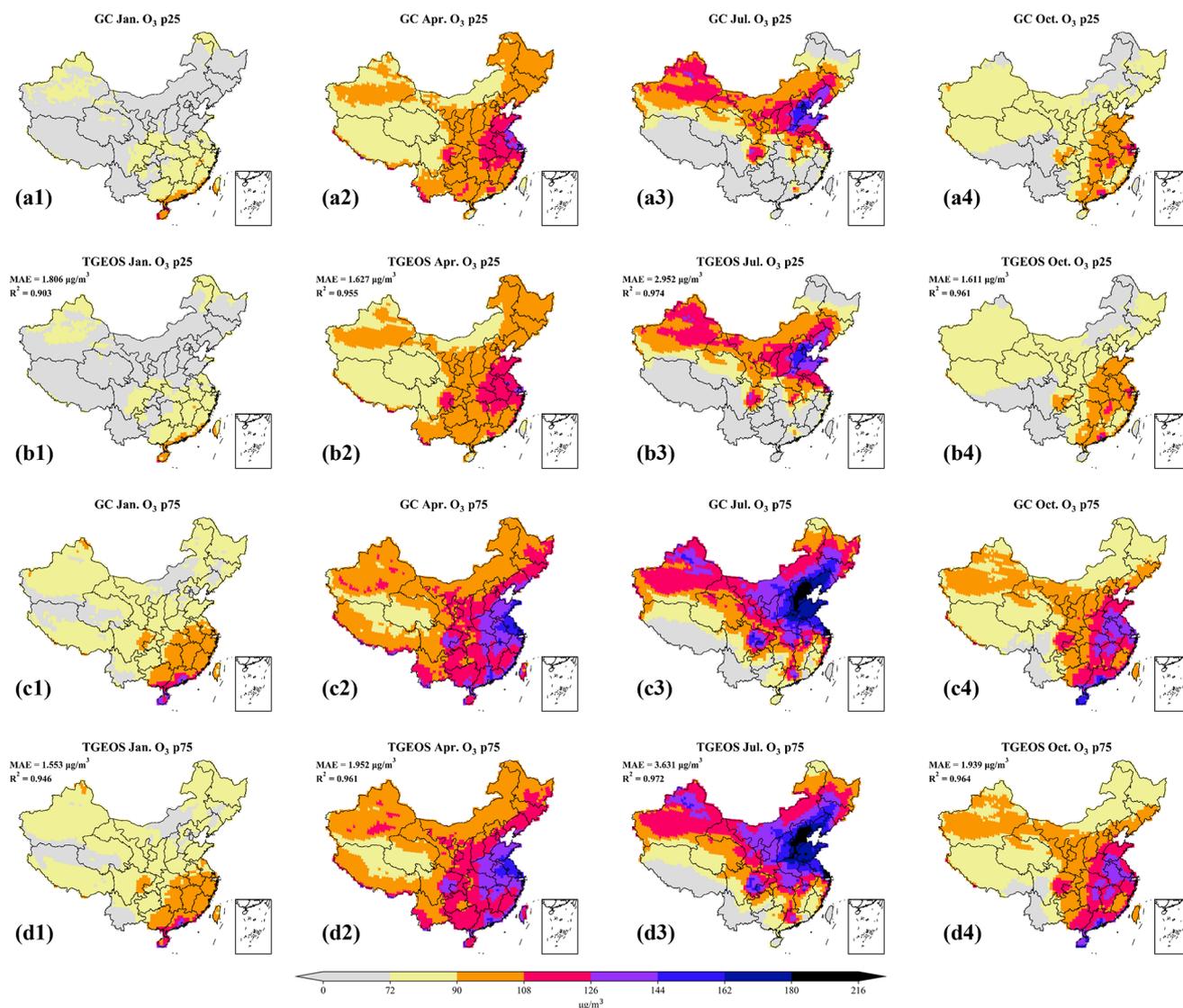
Figure 8. Spatial distribution of 25th and 75th percentile of O_3 concentrations estimated by GEOS-Chem and TGEOS in January, April, July and October under SSP2_2040 scenario. Panels **(a)** and **(c)** illustrate the GEOS-Chem simulations for the 25th and 75th percentile of O_3 from January to October. Panels **(b)** and **(d)** depict the TGEOS estimates for the 25th and 75th percentile of O_3 concentrations during the same months.

3.3 Prediction of probability distribution of $PM_{2.5}$ and O_3

The probability distribution offers a comprehensive representation of pollutant concentrations over a specified time period and effectively captures extreme values, which are typically reflected in the tails of the probability distribution curve. Leveraging this advantage, probability distributions are critical in various air pollution studies, including investigations into future air quality under different emission scenarios (Zeng et al., 2022) or climate changes (Li et al., 2024), and potential mortality in heavily polluted regions (Tian et al., 2022). In this study, we focus on the probability distributions predicted by TGEOS for four key polluted areas: the North

China Plain (NCP, 34–42° N, 113–120° E), Yangtze River Delta (YRD, 26–34° N, 115–123° E), Fenwei Plain (FWP, 33–38° N, 103–114° E), and Sichuan Basin (SCB, 26–34° N, 103–107° E). For each region, the probability density function (PDF) curves were fitted using the TGEOS-predicted monthly indicators averaged over all grid cells in the region. For $PM_{2.5}$, we fitted a right-skewed gamma distribution; for O_3, we fitted a normal distribution. The fitting procedure primarily used the 25th percentile, 75th percentile, and mean as parameters to characterize the distribution shape, with the maximum and minimum values used to constrain the distribution boundaries. These probability distribution curves derived from monthly statistical indicators can be used to pre-

liminarily assess the overall distribution of pollutant concentrations for a given month or quarter under various future emission scenarios.

The $PM_{2.5}$ concentration distributions from GC simulations and TGEOS predictions for the NCP, YRD, FWP, and SCB in SSP2_2040 and SSP3_2040 are illustrated in Fig. 9. Additional results, encompassing four scenarios for the years 2030 and 2050, are presented in Figs. S17 and S19. We found the probability distribution curves of $PM_{2.5}$ concentration that TGEOS predicted in these regions exhibit a strong correlation with corresponding GC curves, indicating that TGEOS model has successfully established the relationship between $PM_{2.5}$ concentration and emissions of precursors in different regions for varying scenarios. The effects of different emission scenarios are clearly reflected. We found that in low-emission scenarios (panels a1 to d1), the $PM_{2.5}$ probability curves for all four regions exhibited significant changes. The reduction in precursor emissions led to a decrease in overall $PM_{2.5}$ concentrations, resulting in an increase in lower values. This caused the peak of the curve to shift to the left relative to the base curve and become sharper. In contrast, in high-emission scenarios (panels a2 to d2), the increase in precursor emissions resulted in higher $PM_{2.5}$ concentrations, shifting the peak to the right and displaying a trend towards flattening.

Figures 10, S18, and S20 illustrate the O_3 concentration distributions predicted by TGEOS for NCP, SCB, YRD, and FWP. Similar to $PM_{2.5}$, the probability distribution curves of O_3 predictions show good agreement with GC curves. TGEOS also established the relationship between O_3 concentrations and emissions in these regions, and successfully predicted the probability distribution of concentrations under various test scenarios. In high emission scenarios, excessive precursor emissions elevate overall O_3 pollution levels, resulting in the occurrence of more high-value concentrations. This flattens the distribution curve of O_3 compared to the base curve, while also lowering its peak value. Conversely, in low emission scenarios, the reduction of precursor pollutants – such as nitrogen oxides (NO_x) and volatile organic compounds (VOC_s), which significantly influence O_3 formation – leads to a decrease in O_3 concentration. This also sharpens the O_3 distribution curve and enhances its peak value. Since the meteorological conditions for all scenarios are fixed in 2017, both concentration variations of $PM_{2.5}$ and O_3 can be attributed to changes in emissions.

Additionally, we observed that the model performs slightly poorly in predicting the probability distribution of pollutants under certain high emission scenarios like Figs. S17 and S20 (panels a2 to d2). As discussed in Sect. 3.3, this discrepancy arises from the limited number of high-emission samples in the dataset, which undermines the model's generalization capabilities. It is also important to emphasize that when predicting O_3 levels under the SSP2_2050 scenario, TGEOS shows a clear underestimation in the YRD region (Fig. S20d1). To investigate this, we compared the emission distributions of SSP2_2050 and the training set. As shown in Fig. S21, several precursor emissions (e.g., CO residential, NO transportation) exhibit much higher densities in the low-emission range under SSP2_2050 (orange line) than in the training set (blue line), where the model has limited training experience. Residual analysis further confirms that the mean residuals of multiple residential emissions (CO, BC, $PM_{2.5}$, PM_{10}, SO_2) are significantly below zero in this regime (Fig. S22), consistent with their density distributions. We therefore attribute the underestimation of O_3 to distributional shifts between training and test data rather than to the direct physical or chemical effects of these species. In particular, the SSP2_2050 scenario contains a substantially larger fraction of samples in the low-emission regime, forcing the model to extrapolate beyond its well-constrained domain and regress toward mean patterns learned from the training set, thereby inducing these prediction biases.

Furthermore, we utilized TGEOS to predict the probability of extreme pollution events under various emission scenarios. According to the China National Ambient Air Quality Standard (GB3095-2012), we set 150 and $160\,\mu g\,m^{-3}$ for $PM_{2.5}$ and O_3 concentrations as thresholds of extreme pollution events, and then calculated the probability of exceeding these thresholds by integrating the fitted probability density functions to the right, represented by the shaded areas shown in the previous images. In the graphs depicting extreme events, the probability of extreme events calculated using the TGEOS curve (represented by the red shaded area) closely matches the probability calculated using the GC curve (represented by the blue shaded area). This concordance demonstrates that TGEOS has effectively learned the distribution patterns of high-concentration pollutants and its capability to predict potential extreme pollution events under future emission scenarios.

Our findings indicate that under low-emission scenarios, the incidence of extreme $PM_{2.5}$ events decreased most significantly in the SCB and YRD regions, as illustrated in Fig. 9b1 and c1. Compared to the background scenario, the incidence in the SCB region decreased by 23.0 %, 25.2 %, and 27.4 %, while in the YRD region, it decreased by 18.1 %, 19.6 %, and 20.6 %, respectively. This indicates that implementing precursor emission reductions in these two regions can effectively control the occurrence of extreme $PM_{2.5}$ events. In contrast, under high-emission scenarios, the increase in incidence for each region was relatively small, approximately 4 % to 5 %. For extreme O_3 events, the reduction effects were most pronounced in the NCP and FWP regions. As presented in panels (a1) and (d1) of Fig. 10, under the three low-emission scenario SSP2 scenarios, the incidence of extreme O_3 events in the NCP region decreased by 17.7 %, 19.9 % and 22.8 %, while in the FWP region, it decreased by 18.6 %, 23.4 % and 27.7 %, respectively. Furthermore, as shown in panel (c1), the YRD region experienced a reduction of approximately 18 %. This demonstrates that precursor emission reductions in areas with high O_3 pollution are
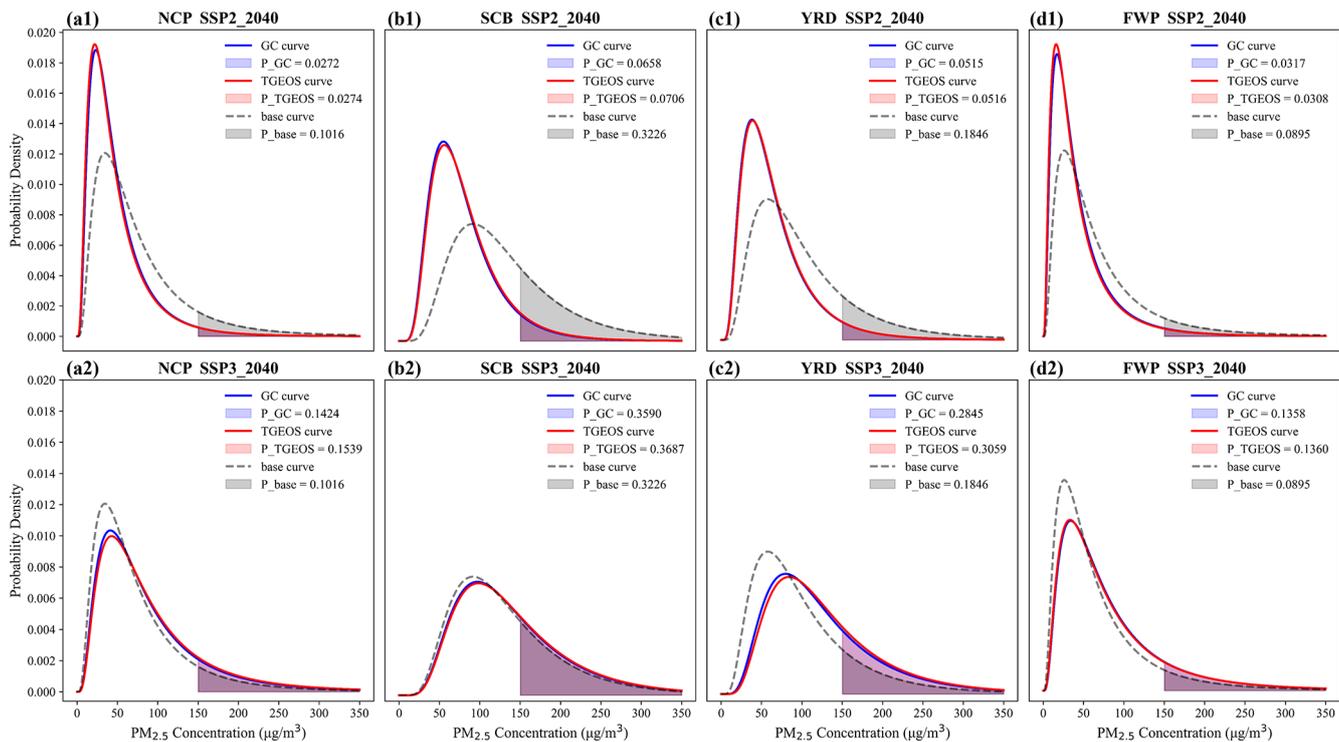
Figure 9. Probability distribution curves of winter $PM_{2.5}$ fitted by GC and TGEOS estimates under two scenarios (SSP2_2040 and SSP3_2040) in four interest regions, including NCP **(a)**, SCB **(b)**, YRD **(c)**, and FWP **(d)**. The blue solid line and the red solid line represent the probability distribution curve of GC and TGEOS results. The black dashed line shows the distribution of pollutants in 2017 (background scenario). P_GC, P_TGEOS and P_base represent the probability of extreme pollution events (calculated from colored areas of each curve) in GC simulation, TGEOS prediction, and 2017 simulation.

highly effective. Conversely, when emission levels increase, the risk of extreme O_3 events in these high-pollution regions rises sharply. Under the three high-emission scenario SSP3 conditions, the risk of extreme O_3 events in the YRD region increased by 12.8 %, 13.3 % and 14.9 %, while in the FWP region, it increased by 8.2 %, 9.7 % and 10.6 %. The changes in the NCP region were less noticeable.

The results above illustrate the impact of different emission scenarios on pollutant concentrations. Under consistent meteorological conditions, significant changes in the concentration distribution of both pollutants can be achieved through straightforward emission reductions, which notably mitigates extreme pollution risks in several heavily polluted areas. Therefore, it is essential to develop strategies to address the current state of air pollution. Furthermore, the TGEOS model shows a high level of similarity to the GC model in predicting pollutant distribution and extreme events, making it a valuable tool for online assessments of related emission reduction policies to enhance decision-making efficiency.

3.4 Comparison of different machine learning models

To validate the performance of the TGEOS model in "emission-concentration" modeling against other machine learning models, four widely used machine learning frameworks, namely Multilayer Perceptrons (MLP), Random Forests (RF), Convolutional Neural Network (CNN), and Vision Transformer (ViT) employed in previous studies (Xing et al., 2020; Huang et al., 2021; Pang et al., 2025), were simultaneously employed based on the multi-scenario dataset mentioned in Sect. 2.1. The MLP model uses 4 hidden layers with 2048, 1024, 512, and 256 neurons, applying ReLU activation and Dropout to prevent overfitting. The RF model uses 300 trees with a maximum depth of 25, a minimum sample split of 4, and a minimum sample per leaf of 2. It uses parallel computation with all CPU cores and performs feature selection by choosing the top 500 important features. The CNN model uses two 3×3 convolutional layers with ReLU activation, followed by adaptive pooling to 29×3. A month embedding is concatenated with the flattened pooled features and passed through three fully connected layers with ReLU applied to the first two. In the ViT model, the 3×3 grid cells are treated as spatial patches; a lightweight CNN is used to generate patch-level embeddings (Wang et al., 2022; Yao et al., 2024); a month token functions as a global CLS token; and the resulting token sequence is then processed by a multi-layer Transformer encoder. It should be noticed that the model inputs for CNN and ViT were reshaped from
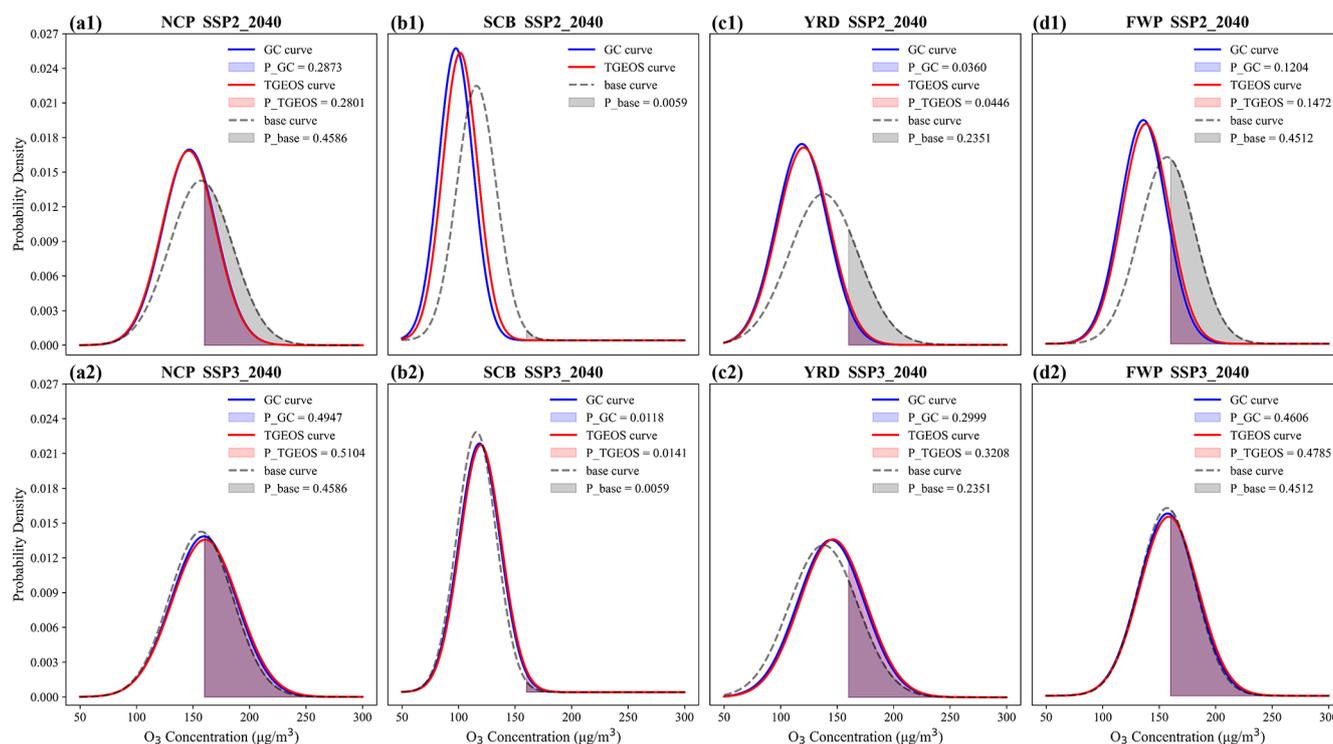
Figure 10. Probability distribution curves of summer O_3 fitted by GC and TGEOS estimates under two scenarios (SSP2_2040 and SSP3_2040) in four interest regions, including NCP **(a)**, SCB **(b)**, YRD **(c)**, and FWP **(d)**. The blue solid line and the red solid line represent the probability distribution curve of GC and TGEOS results. The black dashed line shows the distribution of pollutants in 2017 (background scenario). P_GC, P_TGEOS and P_base represent the probability of extreme pollution events (calculated from colored areas of each curve) in GC simulation, TGEOS prediction, and 2017 simulation.

(1×1045) to $(3 \times 3 \times 116)$ to cater to model architecture, with specific description shown in Fig. S25.

Tables S2 and S3 summarize the performance of the five models on the entire test set. We found that TGEOS outperforms the other four models in both R^2 and MAE metrics. To clearly illustrate the predictive performance of different models, we presented a modified Taylor diagram (Taylor, 2005; Fang et al., 2023) in Fig. 11. This diagram simultaneously displays the Mean Absolute Error (MAE) and correlation coefficient (R) for predictions of $PM_{2.5}$ and O_3 indicators from four models in China domain. Our findings indicate that the RF model performs the poorest. This is primarily due to its reliance on feature importance assessments during feature selection, which overlooks potential underlying features in the data, adversely affecting the model's fitting capability. Additionally, the RF model is sensitive to the distribution of training data, leading to limited extrapolation abilities and poor predictive performance for extreme values. In contrast, the MLP shows a significant improvement in predictive performance relative to the RF model. Leveraging its multi-layer neural network structure, the MLP can more effectively learn complex relationships between multiple features. But this layered structure can struggle when dealing with high-dimensional feature spaces, especially for

highly stochastic indicators such as maximum values, where the MLP still exhibits considerable prediction errors. Compared to MLP and RF models, models based on CNN and ViT frameworks demonstrate better performance, characterized by higher R values as well as lower MAE. However, these models still perform badly for the prediction of indicators reflecting extreme pollutant events such as 75-percentile and maximum, which is mainly because the available spatial information (3×3 grid) is inherently insufficient for these architectures relying heavily on rich spatial structure.

Conversely, the Transformer-based TGEOS model demonstrates superior performance compared to the other models, exhibiting higher R values (exceeding 0.98 and 0.97) and lower MAE values (less than 2.0 and $3 \, \mu g \, m^{-3}$ for the majority indicators of $PM_{2.5}$ and O_3, respectively). These results suggest a higher degree of reliability and accuracy in its predictions. For several indicators where other models perform poorly, such as the maximum, TGEOS demonstrates substantial improvements. The superiority of the Transformer model can be attributed to its greater number of parameters and more complex architecture, which leverage powerful feature extraction capabilities and self-attention mechanisms, allowing it to capture complex relationships in the high-dimensional feature space. It is
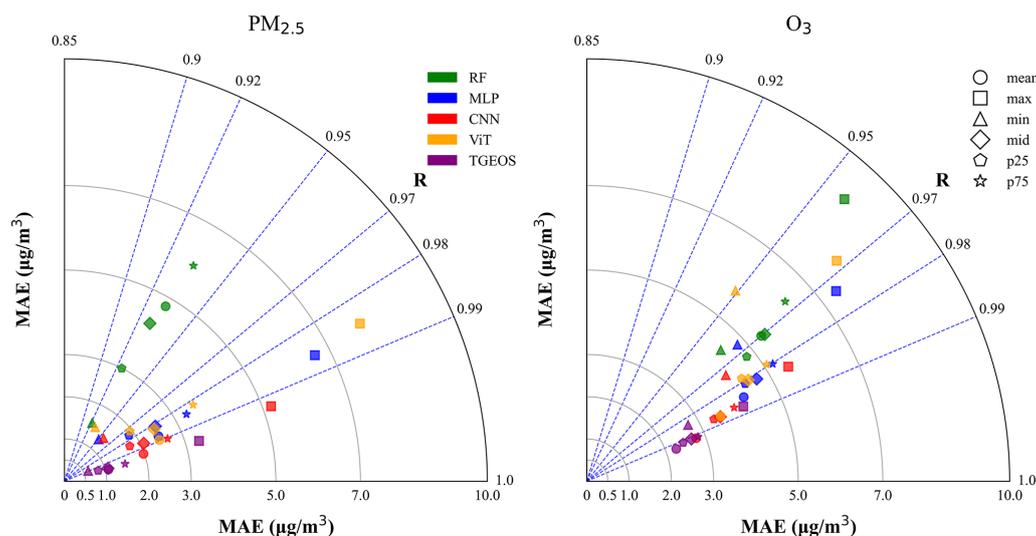
Figure 11. A modified Taylor diagram is presented to jointly illustrate the Mean Absolute Error (MAE) and correlation coefficient (R). The concentration indicators are distinguished using unique markers: circle for mean, square for max, triangle-up for min, diamond for median, pentagon for 25 percentile, and star for 75 percentile. Furthermore, the concentration indicators of winter $PM_{2.5}$ and summer O_3 predicted from different models are visualized using distinct colors, where green represents RF, blue denotes MLP, red indicates CNN, orange means ViT, and purple refers to TGEOS predictions. All indicators are computed based on the six test scenarios.

worth emphasizing that although the capacity of the ViT model in our study was inherently constrained by the limited spatial information available from the compact 3×3 domain, as well as the long-term, monthly timescale that reduces meaningful spatial variability, it still achieved strong predictive performance, with R values exceeding 0.97. This demonstrates the promising representational power of ViT architectures even under suboptimal spatial conditions. Given that ViTs typically rely on richer spatial structures to fully realize their advantages in capturing long-range spatial dependencies, there remains substantial room for further performance gains in settings designed for surface-to-surface, short-term prediction (Pang et al., 2025), where such spatial relationships become more pronounced.

4 Conclusions

In this study, we develop a Transformer-based informative prediction model, TGEOS v1.0, which serves as a GEOS-Chem agent model to represent future air quality under different emission scenarios. Built upon simulations from GEOS-Chem version 14.2.2, TGEOS learns the complex relationships between precursor emissions and the resulting concentrations of $PM_{2.5}$ and O_3. Once trained, the model enables rapid online assessment of the impacts of alternative emission control strategies, producing one-year predictions in approximately 2.51 s. Compared to previous studies that focus solely on average predication, TGEOS can predict the probability distribution of $PM_{2.5}$ and O_3 concentrations in different regions. Leveraging the strengths of

high-dimensional data modeling inherent in the Transformer model, TGEOS is capable to provide more accurate predictions based on more detailed emission scenarios that take into account multiple precursor species, emission sectors, and adjacent emissions.

The air quality prediction of TGEOS for future emission scenarios has good correlation and accuracy with the actual output of GEOS-Chem. The R^2 ranges from 0.958 to 0.992, and the RMSE and MAE are relatively low, with mean values of 2.808 and 1.588 $\mu g\,m^{-3}$, respectively. TGEOS effectively predicted the spatial distribution of $PM_{2.5}$ and O_3 concentration indicators across various emission scenarios, capturing their seasonal variations, and exhibiting an overall spatial distribution pattern that aligns well with the corresponding GC simulations. For another thing, TGEOS accurately predicted the probability distribution of pollutant concentrations in key polluted areas under different emission scenarios, along with potential extreme pollution events, with the probability distribution curves fitted from TGEOS predictions closely matching the corresponding GC curves. In addition, a comparison with traditional machine learning models reveals that TGEOS, built on the Transformer framework, demonstrates superior performance in air quality modeling, with correlation coefficients larger than 0.98 and 0.97 for $PM_{2.5}$ and O_3 predictions, respectively. Consequently, TGEOS offers more accurate and comprehensive predictive capabilities than other machine learning, especially for CNN and ViT architectures prevalent in air-quality modeling.

The TGEOS model still have some limitations to be improved. Firstly, it should be noted that the predictions generated by TGEOS remain incapable of accurately representing

actual air pollutant concentrations, even though TGEOS is highly consistent with GEOS-Chem, since systematic biases have been demonstrated to exist within GEOS-Chem itself (Travis and Jacob, 2019; Miao et al., 2020). Therefore, correcting errors in TEGOS based on near-real observations or reanalysis data is of paramount importance and constitutes a priority for our subsequent research. Additionally, in order to isolate the effects of emission changes on future air quality, the meteorology used in all GC simulations was fixed to the year 2017, following the approach of previous studies (Shi et al., 2021; Wang et al., 2023a). This methodology constrains TGEOS's ability to provide robust predictions under cross-meteorology conditions and prevents it from capturing meteorology–emission interactions and potential "emission–climate" feedbacks. Similar limitations have also been observed in other CTM emulators (Xing et al., 2020; Huang et al., 2021; Liu et al., 2022). Therefore, incorporating diverse climate scenarios that account for meteorological variability will be essential to enhance TGEOS's predictive capability for future air quality under more complex conditions where both emissions and meteorology evolve simultaneously.

Finally, the framework established in this study also reveals promising opportunities for air quality modeling based on ViT models, which may provide substantial performance gains when applied to surface-to-surface and short-term prediction tasks where richer spatial information is available. Although the current study design limited the research domain and temporal scale, future extensions of TGEOS that incorporate fully gridded inputs or higher temporal resolution could make it possible to integrate ViT architectures more effectively. Such developments would enable TGEOS to evolve from a point-based, long-term emulator into a short-term air quality prediction system operating over a larger spatial domain.

References

Abuouelezz, W., Ali, N., Aung, Z., Altunaiji, A., Shah, S. B., and Gliddon, D.: Exploring PM2.5 and PM10 ML forecasting models: a comparative study in the UAE, Scientific Reports, 15, 9797, https://doi.org/10.1038/s41598-025-94013-1, 2025.

Al-Kindi, S. G., Brook, R. D., Biswal, S., and Rajagopalan, S.: Environmental determinants of cardiovascular disease: lessons learned from air pollution, Nature Reviews Cardiology, 17, 656–672, 2020.

Bell, M. L., McDermott, A., Zeger, S. L., Samet, J. M., and Dominici, F.: Ozone and short-term mortality in 95 US urban communities, 1987–2000, Jama, 292, 2372–2378, 2004.

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., and Tian, Q.: Pangu-weather: A 3d high-resolution model for fast and accurate global weather forecast, arXiv [preprint], https://doi.org/10.48550/arXiv.2211.02556, 2022.

Bond, T. C., Bhardwaj, E., Dong, R., Jogani, R., Jung, S., Roden, C., Streets, D. G., and Trautmann, N. M.: Historical emissions of black and organic carbon aerosol from energy-related combustion, 1850–2000, Global Biogeochemical Cycles, 21, https://doi.org/10.1029/2006GB002840, 2007.

Burke, M., Childs, M. L., de la Cuesta, B., Qiu, M., Li, J., Gould, C. F., Heft-Neal, S., and Wara, M.: The contribution of wildfire to PM2.5 trends in the USA, Nature, 622, 761–766, 2023.

Castruccio, S., McInerney, D. J., Stein, M. L., Crouch, F. L., Jacob, R. L., and Moyer, E. J.: Statistical emulation of climate model projections based on precomputed GCM runs, Journal of Climate, 27, 1829–1844, 2014.

Che, W., Zheng, J., Wang, S., Zhong, L., and Lau, A.: Assessment of motor vehicle emission control policies using Model-3/CMAQ model for the Pearl River Delta region, China, Atmospheric Environment, 45, 1740–1751, 2011.

Chen, C., Li, T., Sun, Q., Shi, W., He, M. Z., Wang, J., Liu, J., Zhang, M., Jiang, Q., Wang, M., and Shi, X.: Short-term exposure to ozone and cause-specific mortality risks and thresholds in China: Evidence from nationally representa-

tive data, 2013–2018, Environment International, 171, 107666, https://doi.org/10.1016/j.envint.2022.107666, 2023a.

Chen, K., Han, T., Gong, J., Bai, L., Ling, F., Luo, J.-J., Chen, X., Ma, L., Zhang, T., Su, R., Ci, Y., Li, B., Yang, X., and Ouyang, W.: Fengwu: Pushing the skillful global medium-range weather forecast beyond 10 days lead, arXiv [preprint], https://doi.org/10.48550/arXiv.2304.02948, 2023b.

Chen, L., Zhong, X., Zhang, F., Cheng, Y., Xu, Y., Qi, Y., and Li, H.: FuXi: A cascade machine learning forecasting system for 15-day global weather forecast, npj Climate and Atmospheric Science, 6, 190, https://doi.org/10.1038/s41612-023-00512-1, 2023c.

Chen, S., Zhang, X., Lin, J., Huang, J., Zhao, D., Yuan, T., Huang, K., Luo, Y., Jia, Z., Zang, Z., Qiu, Y., and Xie, L.: Fugitive road dust $PM2.5$ emissions and their potential health impacts, Environmental Science & Technology, 53, 8455–8465, 2019.

Cheng, J., Su, J., Cui, T., Li, X., Dong, X., Sun, F., Yang, Y., Tong, D., Zheng, Y., Li, Y., Li, J., Zhang, Q., and He, K.: Dominant role of emission reduction in $PM_{2.5}$ air quality improvement in Beijing during 2013–2017: a model-based decomposition analysis, Atmospheric Chemistry and Physics, 19, 6125–6146, https://doi.org/10.5194/acp-19-6125-2019, 2019.

Cheng, J., Tong, D., Liu, Y., Yu, S., Yan, L., Zheng, B., Geng, G., He, K., and Zhang, Q.: Comparison of current and future $PM2.5$ air quality in China under CMIP6 and DPEC emission scenarios, Geophysical Research Letters, 48, e2021GL093197, https://doi.org/10.1029/2021GL093197, 2021.

Cheng, J., Tong, D., Liu, Y., Geng, G., Davis, S. J., He, K., and Zhang, Q.: A synergistic approach to air pollution control and carbon neutrality in China can avoid millions of premature deaths annually by 2060, One Earth, 6, 978–989, 2023.

CSC: Air pollution prevention and control action plan, https://www.gov.cn/zwgk/2013-09/12/content_2486773.htm (last access: 20 November 2024), 2013.

CSC: Three-Year Action Plan for Winning the Blue Sky Defense Battle, https://english.mee.gov.cn/News_service/news_release/201807/t20180713_446624.shtml (last access: 17 November 2024), 2018.

Demir, S.: Comparison of normality tests in terms of sample sizes under different skewness and Kurtosis coefficients, International Journal of Assessment Tools in Education, 9, 397–409, 2022.

Devlin, J.: Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv [preprint], https://doi.org/10.48550/arXiv.1810.04805, 2018.

Du, W., Chen, L., Wang, H., Shan, Z., Zhou, Z., Li, W., and Wang, Y.: Deciphering urban traffic impacts on air quality by deep learning and emission inventory, Journal of Environmental Sciences, 124, 745–757, 2023.

Fang, L., Jin, J., Segers, A., Liao, H., Li, K., Xu, B., Han, W., Pang, M., and Lin, H. X.: A gridded air quality forecast through fusing site-available machine learning predictions from RFSML v1.0 and chemical transport model results from GEOS-Chem v13.1.0 using the ensemble Kalman filter, Geoscientific Model Development, 16, 4867–4882, https://doi.org/10.5194/gmd-16-4867-2023, 2023.

Fuller, R., Landrigan, P. J., Balakrishnan, K., Bathan, G., Bose-O'Reilly, S., Brauer, M., Caravanos, J., Chiles, T., Cohen, A., Corra, L., Cropper, M., Ferraro, G., Hanna, J., Hanrahan, D., Hu, H., Hunter, D., Janata, G., Kupka, R., Lanphear, B., Lichtveld, M., Martin, K., Mustapha, A., Sanchez-Triana, E., Sandilya, K.,

Schaefli, L., Shaw, J., Seddon, J., Suk, W., Téllez-Rojo, M. M., and Yan, C.: Pollution and health: a progress update, The Lancet Planetary Health, 6, e535–e547, 2022.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), Journal of Climate, 30, 5419–5454, 2017.

Geng, G., Liu, Y., Liu, Y., Liu, S., Cheng, J., Yan, L., Wu, N., Hu, H., Tong, D., Zheng, B., Yin, Z., He, K., and Zhang, Q.: Efficacy of China's clean air actions to tackle $PM2.5$ pollution between 2013 and 2020, Nature Geoscience, 1–8, 2024.

Gong, C., Liao, H., Zhang, L., Yue, X., Dang, R., and Yang, Y.: Persistent ozone pollution episodes in North China exacerbated by regional transport, Environmental Pollution, 265, 115056, https://doi.org/10.1016/j.envpol.2020.115056, 2020.

Guo, B., Wang, Y., Zhang, X., Che, H., Zhong, J., Chu, Y., and Cheng, L.: Temporal and spatial variations of haze and fog and the characteristics of PM2.5 during heavy pollution episodes in China from 2013 to 2018, Atmospheric Pollution Research, 11, 1847–1856, 2020.

Han, H., Zhang, L., Wang, X., and Lu, X.: Contrasting domestic and global impacts of emission reductions in China on tropospheric ozone, Journal of Geophysical Research: Atmospheres, 129, e2024JD041453, https://doi.org/10.1029/2024JD041453, 2024.

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert, J. J., Vu, L., Andres, R. J., Bolt, R. M., Bond, T. C., Dawidowski, L., Kholod, N., Kurokawa, J.-I., Li, M., Liu, L., Lu, Z., Moura, M. C. P., O'Rourke, P. R., and Zhang, Q.: Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS), Geoscientific Model Development, 11, 369–408, https://doi.org/10.5194/gmd-11-369-2018, 2018.

Hu, L., Jacob, D. J., Liu, X., Zhang, Y., Zhang, L., Kim, P. S., Sulprizio, M. P., and Yantosca, R. M.: Global budget of tropospheric ozone: Evaluating recent model advances with satellite (OMI), aircraft (IAGOS), and ozonesonde observations, Atmospheric Environment, 167, 323–334, 2017.

Hu, W., Zhao, Y., Lu, N., Wang, X., Zheng, B., Henze, D. K., Zhang, L., Fu, T.-M., and Zhai, S.: Changing responses of PM2.5 and ozone to source emissions in the Yangtze River Delta using the adjoint model, Environmental Science & Technology, 58, 628–638, 2023.

Huang, L., Liu, S., Yang, Z., Xing, J., Zhang, J., Bian, J., Li, S., Sahu, S. K., Wang, S., and Liu, T.-Y.: Exploring deep learning for air pollutant emission estimation, Geoscientific Model Development, 14, 4641–4654, https://doi.org/10.5194/gmd-14-4641-2021, 2021.

Jin, J., Fang, L., Li, B., Liao, H., Wang, Y., Han, W., Li, K., Pang, M., Wu, X., and Lin, H. X.: 4DEnVar-based inversion system for ammonia emission estimation in China through assimilating IASI ammonia retrievals, Environmental Research Letters, 18, 034005, https://doi.org/10.1088/1748-9326/acb835, 2023.

Kingma, D. P.: Adam: A method for stochastic optimization, arXiv [preprint], https://doi.org/10.48550/arXiv.1412.6980, 2014.

Lai, A., Lee, M., Carter, E., Chan, Q., Elliott, P., Ezzati, M., Kelly, F., Yan, L., Wu, Y., Yang, X., Zhao, L., Baumgartner, J., and Schauer, J. J.: Chemical investigation of household solid fuel use and outdoor air pollution contributions to personal PM2.5 exposures, Environmental Science & Technology, 55, 15969–15979, 2021.

Le, T., Wang, Y., Liu, L., Yang, J., Yung, Y. L., Li, G., and Seinfeld, J. H.: Unexpected air pollution with marked emission reductions during the COVID-19 outbreak in China, Science, 369, 702–706, 2020.

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., and Pozzer, A.: The contribution of outdoor air pollution sources to premature mortality on a global scale, Nature, 525, 367–371, 2015.

Li, D.: Python source code of TGEOS v1.0, Zenodo [code], https://doi.org/10.5281/zenodo.15422797, 2025a.

Li, D.: The dataset used by TGEOS, Zenodo [data set], https://doi.org/10.5281/zenodo.15717908, 2025b.

Li, D., Wu, Q., Cheng, H., Feng, J., Li, D., Wang, Y., Cao, K., and Wang, L.: Numerical study of the future PM2.5 concentration under climate change and Best-Health-Effect (BHE) scenario, Environmental Pollution, 124391, https://doi.org/10.1016/j.envpol.2024.124391, 2024.

Li, J., Dai, Y., Zhu, Y., Tang, X., Wang, S., Xing, J., Zhao, B., Fan, S., Long, S., and Fang, T.: Improvements of response surface modeling with self-adaptive machine learning method for PM2.5 and O3 predictions, Journal of Environmental Management, 303, 114210, https://doi.org/10.1016/j.jenvman.2021.114210, 2022.

Li, M., Liu, H., Geng, G., Hong, C., Liu, F., Song, Y., Tong, D., Zheng, B., Cui, H., Man, H., Zhang, Q., and He, K.: Anthropogenic emission inventories in China: a review, National Science Review, 4, 834–866, 2017.

Liang, Y., Xia, Y., Ke, S., Wang, Y., Wen, Q., Zhang, J., Zheng, Y., and Zimmermann, R.: Airformer: Predicting nationwide air quality in china with transformers, in: Proceedings of the AAAI Conference on Artificial Intelligence, 37, 14329–14337, 2023.

Liu, C., Zhang, H., Cheng, Z., Shen, J., Zhao, J., Wang, Y., Wang, S., and Cheng, Y.: Emulation of an atmospheric gas-phase chemistry solver through deep learning: Case study of Chinese Mainland, Atmospheric Pollution Research, 12, 101079, https://doi.org/10.1016/j.apr.2021.101079, 2021.

Liu, Z., Dong, M., Xue, W., Ni, X., Qi, Z., Shao, J., Guo, Y., Ma, M., Zhang, Q., and Wang, J.: Interaction patterns between climate action and air cleaning in China: a two-way evaluation based on an ensemble learning approach, Environmental Science & Technology, 56, 9291–9301, 2022.

Lu, X., Zhang, L., Chen, Y., Zhou, M., Zheng, B., Li, K., Liu, Y., Lin, J., Fu, T.-M., and Zhang, Q.: Exploring 2016–2017 surface ozone pollution over China: source contributions and meteorological influences, Atmospheric Chemistry and Physics, 19, 8339–8361, https://doi.org/10.5194/acp-19-8339-2019, 2019.

Lu, X., Zhang, L., Wang, X., Gao, M., Li, K., Zhang, Y., Yue, X., and Zhang, Y.: Rapid increases in warm-season surface ozone and resulting health impact in China since 2013, Environmental Science & Technology Letters, 7, 240–247, 2020a.

Lu, X., Zhang, L., Wu, T., Long, M. S., Wang, J., Jacob, D. J., Zhang, F., Zhang, J., Eastham, S. D., Hu, L., Zhu, L., Liu, X., and Wei, M.: Development of the global atmospheric chemistry general circulation model BCC-GEOS-Chem v1.0: model description and evaluation, Geoscientific Model Development, 13, 3817–3838, https://doi.org/10.5194/gmd-13-3817-2020, 2020b.

Masmoudi, S., Elghazel, H., Taieb, D., Yazar, O., and Kallel, A.: A machine-learning framework for predicting multiple air pollutants' concentrations via multi-target regression and feature selection, Science of the Total Environment, 715, 136991, https://doi.org/10.1016/j.scitotenv.2020.136991, 2020.

McDuffie, E. E., Martin, R. V., Spadaro, J. V., Burnett, R., Smith, S. J., O'Rourke, P., Hammer, M. S., van Donkelaar, A., Bindle, L., Shah, V., Jaeglé, L., Luo, G., Yu, F., Adeniran, J. A., Lin, J., and Brauer, M.: Source sector and fuel contributions to ambient PM2.5 and attributable mortality across multiple spatial scales, Nature Communications, 12, 1–12, 2021.

Miao, R., Chen, Q., Zheng, Y., Cheng, X., Sun, Y., Palmer, P. I., Shrivastava, M., Guo, J., Zhang, Q., Liu, Y., Tan, Z., Ma, X., Chen, S., Zeng, L., Lu, K., and Zhang, Y.: Model bias in simulating major chemical components of $PM_{2.5}$ in China, Atmospheric Chemistry and Physics, 20, 12265–12284, https://doi.org/10.5194/acp-20-12265-2020, 2020.

Nair, V. and Hinton, G. E.: Rectified linear units improve restricted boltzmann machines, in: Proceedings of the 27th international conference on machine learning (ICML-10), 807–814, IBSN 9781605589077, 2010.

Narayanan, D., Shoeybi, M., Casper, J., LeGresley, P., Patwary, M., Korthikanti, V., Vainbrand, D., Kashinkunti, P., Bernauer, J., Catanzaro, B., Phanishayee, A., and Zaharia, M.: Efficient large-scale language model training on gpu clusters using megatron-lm, in: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, 1–15, Association for Computing Machinery, https://doi.org/10.1145/3458817.3476209, 2021.

Natarajan, S. K., Shanmurthy, P., Arockiam, D., Balusamy, B., and Selvarajan, S.: Optimized machine learning model for air quality index prediction in major cities in India, Scientific Reports, 14, 6795, https://doi.org/10.1038/s41598-024-54807-1, 2024.

Pang, M., Jin, J., Segers, A., Jiang, H., Fang, L., Lin, H. X., and Liao, H.: Dust storm forecasting through coupling LOTOS-EUROS with localized ensemble Kalman filter, Atmospheric Environment, 306, 119831, https://doi.org/10.1016/j.atmosenv.2023.119831, 2023.

Pang, M., Jin, J., Segers, A., Lin, H. X., Wang, G., Liao, H., and Han, W.: Zeeman: A Deep Learning Regional Atmospheric Chemistry Transport Model, arXiv [preprint], https://doi.org/10.48550/arXiv.2510.06140, 2025.

Pinder, R. W., Adams, P. J., and Pandis, S. N.: Ammonia emission controls as a cost-effective strategy for reducing atmospheric particulate matter in the eastern United States, Environmental Science & Technology, 41, 380–386, 2007.

Qiao, X., Yuan, Y., Tang, Y., Ying, Q., Guo, H., Zhang, Y., and Zhang, H.: Revealing the origin of fine particulate matter in the Sichuan Basin from a source-oriented modeling perspective, Atmospheric Environment, 244, 117896, https://doi.org/10.1016/j.atmosenv.2020.117896, 2021.

Rodríguez, S. and López-Darias, J.: Extreme Saharan dust events expand northward over the Atlantic and Europe, prompting record-breaking PM_{10} and $PM_{2.5}$ episodes, Atmospheric Chemistry and Physics, 24, 12031–12053, https://doi.org/10.5194/acp-24-12031-2024, 2024.

Salman, A. K., Choi, Y., Park, J., Mousavinezhad, S., Payami, M., Momeni, M., and Ghahremanloo, M.: Deep learning based emulator for simulating CMAQ surface NO2 levels over the CONUS, Atmospheric Environment, 316, 120192, https://doi.org/10.1016/j.atmosenv.2023.120192, 2024.

Seinfeld, J. H. and Pandis, S. N.: Atmospheric chemistry and physics: from air pollution to climate change, John Wiley & Sons, ISBN 978-1-118-94740-1, 2016.

Shi, X., Zheng, Y., Lei, Y., Xue, W., Yan, G., Liu, X., Cai, B., Tong, D., and Wang, J.: Air quality benefits of achieving carbon neutrality in China, Science of the Total Environment, 795, 148784, https://doi.org/10.1016/j.scitotenv.2021.148784, 2021.

Shi, Z., Huang, L., Li, J., Ying, Q., Zhang, H., and Hu, J.: Sensitivity analysis of the surface ozone and fine particulate matter to meteorological parameters in China, Atmospheric Chemistry and Physics, 20, 13455–13466, https://doi.org/10.5194/acp-20-13455-2020, 2020.

Silver, B., Reddington, C., Arnold, S., and Spracklen, D.: Substantial changes in air pollution across China during 2015–2017, Environmental Research Letters, 13, 114012, https://doi.org/10.1088/1748-9326/aae718, 2018.

Skyllakou, K., Rivera, P. G., Dinkelacker, B., Karnezi, E., Kioutsioukis, I., Hernandez, C., Adams, P. J., and Pandis, S. N.: Changes in $PM_{2.5}$ concentrations and their sources in the US from 1990 to 2010, Atmospheric Chemistry and Physics, 21, 17115–17132, https://doi.org/10.5194/acp-21-17115-2021, 2021.

Tang, D., Zhan, Y., and Yang, F.: A review of machine learning for modeling air quality: Overlooked but important issues, Atmospheric Research, 300, 107261, https://doi.org/10.1016/j.atmosres.2024.107261, 2024.

Taylor, K. E.: Taylor diagram primer, Work. Pap., 1–4, Program for Climate Model Diagnosis and Intercomparison (PCMDI), https://pcmdi.llnl.gov/staff/taylor/CV/Taylor_diagram_primer.pdf (last access: 24 January 2025), 2005.

The International GEOS-Chem User Community: geoschem/GCClassic: GCClassic 14.2.2, Zenodo [code], https://doi.org/10.5281/zenodo.10034814, 2023.

Thompson, T. M. and Selin, N. E.: Influence of air quality model resolution on uncertainty associated with health impacts, Atmospheric Chemistry and Physics, 12, 9753–9762, https://doi.org/10.5194/acp-12-9753-2012, 2012.

Thunis, P., Clappier, A., Beekmann, M., Putaud, J. P., Cuvelier, C., Madrazo, J., and de Meij, A.: Non-linear response of $PM_{2.5}$ to changes in NO_x and NH_3 emissions in the Po basin (Italy): consequences for air quality plans, Atmospheric Chemistry and Physics, 21, 9309–9327, https://doi.org/10.5194/acp-21-9309-2021, 2021.

Tian, F., Qi, J., Qian, Z., Li, H., Wang, L., Wang, C., Geiger, S. D., McMillin, S. E., Yin, P., Lin, H., and Zhou, M.: Differentiating the effects of air pollution on daily mortality counts and years of life lost in six Chinese megacities, Science of the Total Environment, 827, 154037, https://doi.org/10.1016/j.scitotenv.2022.154037, 2022.

Tong, D., Cheng, J., Liu, Y., Yu, S., Yan, L., Hong, C., Qin, Y., Zhao, H., Zheng, Y., Geng, G., Li, M., Liu, F., Zhang, Y., Zheng, B., Clarke, L., and Zhang, Q.: Dynamic projection of anthropogenic emissions in China: methodology and 2015–2050 emission pathways under a range of socio-economic, climate policy, and pollu-

tion control scenarios, Atmospheric Chemistry and Physics, 20, 5729–5757, https://doi.org/10.5194/acp-20-5729-2020, 2020.

Travis, K. R. and Jacob, D. J.: Systematic bias in evaluating chemical transport models with maximum daily 8 h average (MDA8) surface ozone for air quality applications: a case study with GEOS-Chem v9.02, Geoscientific Model Development, 12, 3641–3648, https://doi.org/10.5194/gmd-12-3641-2019, 2019.

US EPA: Technical Support Document for the Proposed PM NAAQS Rule: Response Surface Modeling, https://www.epa.gov/sites/default/files/2020-10/documents/pmnaaqs_tsd_rsm_all_021606.pdf (last access: 7 January 2025), 2006.

Vaswani, A.: Attention is all you need, Advances in Neural Information Processing Systems, Curran Associates Inc., https://dl.acm.org/doi/10.5555/3295222.3295349, 2017.

Wang, C., Xu, H., Zhang, X., Wang, L., Zheng, Z., and Liu, H.: Convolutional embedding makes hierarchical vision transformer stronger, in: European conference on computer vision, 739–756, Springer, https://doi.org/10.1007/978-3-031-20044-1_42, 2022.

Wang, F., Han, X., Xie, H., Gao, Y., Guan, X., and Zhang, M.: Investigating trends and causes of simultaneous high pollution from PM2.5 and ozone in China, 2015–2023, Atmospheric Pollution Research, 102351, https://doi.org/10.1016/j.apr.2024.102351, 2024.

Wang, S., Xing, J., Jang, C., Zhu, Y., Fu, J. S., and Hao, J.: Impact assessment of ammonia emissions on inorganic aerosols in East China using response surface modeling technique, Environmental Science & Technology, 45, 9293–9300, 2011.

Wang, S., Wu, D., Wang, X.-M., Fung, J. C.-H., and Yu, J. Z.: Relative contributions of secondary organic aerosol formation from toluene, xylenes, isoprene, and monoterpenes in Hong Kong and Guangzhou in the Pearl River Delta, China: an emission-based box modeling study, Journal of Geophysical Research: Atmospheres, 118, 507–519, 2013.

Wang, T., Xue, L., Brimblecombe, P., Lam, Y. F., Li, L., and Zhang, L.: Ozone pollution in China: A review of concentrations, meteorological influences, chemical precursors, and effects, Science of the Total Environment, 575, 1582–1596, 2017.

Wang, X., Fu, T.-M., Zhang, L., Cao, H., Zhang, Q., Ma, H., Shen, L., Evans, M. J., Ivatt, P. D., Lu, X., Chen, Y., Zhang, L., Feng, X., Yang, X., Zhu, L., and Henze, D. K.: Sensitivities of ozone air pollution in the Beijing–Tianjin–Hebei area to local and upwind precursor emissions using adjoint modeling, Environmental Science & Technology, 55, 5752–5762, 2021.

Wang, Y., Liao, H., Chen, H., and Chen, L.: Future projection of mortality from exposure to PM2.5 and O3 under the carbon neutral pathway: roles of changing emissions and population aging, Geophysical Research Letters, 50, e2023GL104838, https://doi.org/10.1029/2023GL104838, 2023a.

Wang, Y., Zhao, Y., Liu, Y., Jiang, Y., Zheng, B., Xing, J., Liu, Y., Wang, S., and Nielsen, C. P.: Sustained emission reductions have restrained the ozone pollution over China, Nature Geoscience, 16, 967–974, 2023b.

Wei, T., Chen, C., Yang, Y., Li, L., Wang, J., Ye, M., Kan, H., Yang, D., Song, Y., Cai, J., and Hou, D.: Associations between short-term exposure to ambient air pollution and lung function in adults, Journal of Exposure Science & Environmental Epidemiology, 34, 886–894, 2024.

Wei, W., Li, Y., Ren, Y., Cheng, S., and Han, L.: Sensitivity of summer ozone to precursor emission change over Beijing during 2010–2015: A WRF-Chem modeling study, Atmospheric Environment, 218, 116984, https://doi.org/10.1016/j.atmosenv.2019.116984, 2019.

WHO: Air pollution: The invisible health threat, World Health Organization, Geneva, Switzerland, https://www.who.int/news-room/feature-stories/detail/air-pollution--the-invisible-health-threat (last access: 9 November 2024), 2023.

Xiao, Q., Geng, G., Xue, T., Liu, S., Cai, C., He, K., and Zhang, Q.: Tracking PM2.5 and O3 pollution and the related health burden in China 2013–2020, Environmental Science & Technology, 56, 6922–6932, 2021.

Xing, J., Wang, S. X., Jang, C., Zhu, Y., and Hao, J. M.: Nonlinear response of ozone to precursor emission changes in China: a modeling study using response surface methodology, Atmospheric Chemistry and Physics, 11, 5027–5044, https://doi.org/10.5194/acp-11-5027-2011, 2011.

Xing, J., Wang, S., Zhao, B., Wu, W., Ding, D., Jang, C., Zhu, Y., Chang, X., Wang, J., Zhang, F., and Hao, J.: Quantifying nonlinear multiregional contributions to ozone and fine particles using an updated response surface modeling technique, Environmental Science & Technology, 51, 11788–11798, 2017.

Xing, J., Ding, D., Wang, S., Zhao, B., Jang, C., Wu, W., Zhang, F., Zhu, Y., and Hao, J.: Quantification of the enhanced effectiveness of NO_x control from simultaneous reductions of VOC and NH_3 for reducing air pollution in the Beijing–Tianjin–Hebei region, China, Atmospheric Chemistry and Physics, 18, 7799–7814, https://doi.org/10.5194/acp-18-7799-2018, 2018.

Xing, J., Zheng, S., Ding, D., Kelly, J. T., Wang, S., Li, S., Qin, T., Ma, M., Dong, Z., Jang, C., Zhu, Y., Zheng, H., Ren, L., Liu, T.-Y., and Hao, J.: Deep learning for prediction of the air quality response to emission changes, Environmental Science & Technology, 54, 8589–8600, 2020.

Xu, C., Wang, J., Hu, M., and Wang, W.: A new method for interpolation of missing air quality data at monitor stations, Environment International, 169, 107538, https://doi.org/10.1016/j.envint.2022.107538, 2022.

Yan, Y., Zhou, Y., Kong, S., Lin, J., Wu, J., Zheng, H., Zhang, Z., Song, A., Bai, Y., Ling, Z., Liu, D., and Zhao, T.: Effectiveness of emission control in reducing PM2.5 pollution in central China during winter haze episodes under various potential synoptic controls, Atmospheric Chemistry and Physics, 21, 3143–3162, https://doi.org/10.5194/acp-21-3143-2021, 2021.

Yang, S. and Wu, H.: A novel PM2.5 concentrations probability density prediction model combines the least absolute shrinkage and selection operator with quantile regression, Environmental Science and Pollution Research, 29, 78265–78291, 2022.

Yao, T., Li, Y., Pan, Y., and Mei, T.: Hiri-vit: Scaling vision transformer with high resolution inputs, IEEE Transactions on Pattern Analysis and Machine Intelligence, 46, 6431–6442, 2024.

Zeng, L., Yang, Y., Wang, H., Wang, J., Li, J., Ren, L., Li, H., Zhou, Y., Wang, P., and Liao, H.: Intensified modulation of winter aerosol pollution in China by El Niño with short duration, Atmospheric Chemistry and Physics, 21, 10745–10761, https://doi.org/10.5194/acp-21-10745-2021, 2021.

Zeng, X., Gao, Y., Wang, Y., Ma, M., Zhang, J., and Sheng, L.: Characterizing the distinct modulation of future emissions on summer ozone concentrations between urban and rural areas over China, Science of the Total Environment, 820, 153324, https://doi.org/10.1016/j.scitotenv.2022.153324, 2022.

Zhang, D., Wang, Q., Song, S., Chen, S., Li, M., Shen, L., Zheng, S., Cai, B., Wang, S., and Zheng, H.: Machine learning approaches reveal highly heterogeneous air quality co-benefits of the energy transition, Iscience, 26, 2023a.

Zhang, J., Gao, Y., Luo, K., Leung, L. R., Zhang, Y., Wang, K., and Fan, J.: Impacts of compound extreme weather events on ozone in the present and future, Atmospheric Chemistry and Physics, 18, 9861–9877, https://doi.org/10.5194/acp-18-9861-2018, 2018.

Zhang, X., Xiao, X., Wang, F., Brasseur, G., Chen, S., Wang, J., and Gao, M.: Observed sensitivities of PM2.5 and O3 extremes to meteorological conditions in China and implications for the future, Environment International, 168, 107428, https://doi.org/10.1016/j.envint.2022.107428, 2022a.

Zhang, Y., Gao, J., Zhu, Y., Liu, Y., Li, H., Yang, X., Zhong, X., Zhao, M., Wang, W., Che, F., Zhou, D., Wang, S., Zhi, G., Xue, L., and Li, H.: Evolution of ozone formation sensitivity during a persistent regional ozone episode in Northeastern China and its implication for a control strategy, Environmental Science & Technology, 58, 617–627, 2023b.

Zhang, Z., Yan, Y., Kong, S., Deng, Q., Qin, S., Yao, L., Zhao, T., and Qi, S.: Benefits of refined NH_3 emission controls on PM2.5 mitigation in Central China, Science of the Total Environment, 814, 151957, https://doi.org/10.1016/j.scitotenv.2021.151957, 2022b.

Zhao, B., Wang, S. X., Xing, J., Fu, K., Fu, J. S., Jang, C., Zhu, Y., Dong, X. Y., Gao, Y., Wu, W. J., Wang, J. D., and Hao, J. M.: Assessing the nonlinear response of fine particles to precursor emissions: development and application of an extended response surface modeling technique v1.0, Geoscientific Model Development, 8, 115–128, https://doi.org/10.5194/gmd-8-115-2015, 2015.

Zhao, S., Feng, T., Xiao, W., Zhao, S., and Tie, X.: Weather-Climate Anomalies and Regional Transport Contribute to Air Pollution in Northern China During the COVID-19 Lockdown, Journal of Geophysical Research: Atmospheres, 127, e2021JD036345, https://doi.org/10.1029/2021JD036345, 2022.

Zheng, B., Zhang, Q., Geng, G., Chen, C., Shi, Q., Cui, M., Lei, Y., and He, K.: Changes in China's anthropogenic emissions and air quality during the COVID-19 pandemic in 2020, Earth System Science Data, 13, 2895–2907, https://doi.org/10.5194/essd-13-2895-2021, 2021.

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., and Zhang, W.: Informer: Beyond efficient transformer for long sequence time-series forecasting, in: Proceedings of the AAAI conference on artificial intelligence, 35, 11106–11115, https://doi.org/10.48550/arXiv.2012.07436, 2021.

Zhou, S., Wang, W., Zhu, L., Qiao, Q., and Kang, Y.: Deep-learning architecture for PM2.5 concentration prediction: A review, Environmental Science and Ecotechnology, 100400, https://doi.org/10.1016/j.ese.2024.100400, 2024.