Regional background ozone estimation for China through data fusion of observation and simulation

Zhixu Sun a,b, Jiani Tan a,b, Fangting Wang a,b, Rui Li a,b, Xinxin Zhang a,b, Jiaqiang Liao a,b, Yangjun Wang a,b, Ling Huang a,b, Kun Zhang a,b, Joshua S. Fu c, Li Li a,b,*

a School of Environmental and Chemical Engineering, Shanghai University, Shanghai 200444, China
b Key Laboratory of Organic Compound Pollution Control Engineering (MOE), Shanghai University, Shanghai 200444, China
c Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN 37996, USA

HIGHLIGHTS

• The regional background O\textsubscript{3} concentration in China is estimated at 35 ± 4 ppb in 2020.
• O\textsubscript{3,RBG} accounts for 81 % and 55 % of MDA8 O\textsubscript{3} in clean and polluted conditions, respectively.
• O\textsubscript{3,RBG} dominates MDA8 O\textsubscript{3} (>85 %) for all Chinese regions when O\textsubscript{3} is lower than 60 ppb.
• Natural emissions contribute more significantly to O\textsubscript{3,RBG} in China compared to meteorological factors.

GRAPHICAL ABSTRACT

ARTICLE INFO

Editor: Philip K. Hopke

Keywords:
Regional background O\textsubscript{3} Multiple linear regression Chemical transport model Brute force method China

ABSTRACT

Regional background ozone (O\textsubscript{3,RBG}) is an important component of surface ozone (O\textsubscript{3}). However, due to the uncertainties in commonly used Chemical Transport Models (CTMs) and statistical models, accurately assessing O\textsubscript{3,RBG} in China is challenging. In this study, we calculated the O\textsubscript{3,RBG} concentrations with the CTM – Brute Force Method (BFM) and constrained the results with site observations of O\textsubscript{3} with the multiple linear regression (MLR) model. The annual average O\textsubscript{3,RBG} concentration in China region in 2020 is 35 ± 4 ppb, accounting for 81 ± 5 % of the maximum 8-h average O\textsubscript{3} (MDA8 O\textsubscript{3}). We applied the random forest and Shapley additive explanations based on meteorological standardization techniques to separate the contributions of meteorology and natural emissions to O\textsubscript{3,RBG}. Natural emissions contribute more significantly to O\textsubscript{3,RBG} than meteorology in various Chinese regions (30–40 ppb), with higher contributions during the warm season. Meteorological factors show higher contributions in the spring and summer seasons (2–3 ppb) than the other seasons. Temperature and humidity are the primary contributors to O\textsubscript{3,RBG} in regions with severe O\textsubscript{3} pollution in China, with their individual impacts ranging from 30 % to 62 % of the total impacts of all meteorological factors in different seasons.
For policy implications, we tracked the contributions of $O_{3,\text{RBG}}$ and local photochemical reaction contributions ($O_{3,\text{LC}}$) to total $O_3$ concentration at different $O_3$ levels. We found that $O_{3,\text{LC}}$ contribute over 45% to MDA8 $O_3$ on polluted days, supporting the current Chinese policy of reducing $O_3$ peak concentrations by cutting down precursor emissions. However, as the contribution of $O_{3,\text{RBG}}$ is not considered in the policy, additional efforts are needed to achieve the control goal of $O_3$ concentration. As the implementation of stringent $O_3$ control measurements in China, the contribution of $O_{3,\text{RBG}}$ become increasingly significant, suggesting the need for attention to $O_{3,\text{RBG}}$ and regional joint prevention and control.

1. Introduction

Ozone ($O_3$) is a secondary air pollutant formed by the photochemical reactions of nitrogen oxides (NOx) and volatile organic compounds (VOCs) under the presence of sunlight (Zheng et al., 2023). Elevated surface $O_3$ level is detrimental to human health due to its potent oxidizing properties (Gu et al., 2022; Hong and Chen, 2026; Lelieveld et al., 2015; World Health Organization. Regional Office for, E, 2021). It also poses risks to animals, plants, and their habitats, leading to a decline in global food production and disrupting the balance of the ecosystems (Yue et al., 2017). Additionally, $O_3$ plays an important role in the global atmospheric radiation balance (Barnes et al., 2023; Rasmussen et al., 2013) and can consequently influence atmospheric circulations (Rudeva et al., 2023). The annual $O_3$ concentrations in China have been consistently increasing by about 5% per year during 2016–2020 (Guo et al., 2023; Huang et al., 2018; Lu et al., 2020; Silver et al., 2018; Zheng et al., 2017). The concentrations decreased by 1–2 $\mu g/m^3$ in 2020 and 2021 but rebounded by 5.8% in 2022, leading to the exceedance of China’s air quality standard of MDA8 $O_3$ (80ppb) in 91 cities (https://www.mee.gov.cn/). $O_3$ pollution has been a critical challenge facing China and has become the focus of the scientific communities and government agencies.

The surface $O_3$ concentration is the sum of regional background $O_3$ ($O_{3,\text{RBG}}$) and $O_3$ formed via local photochemical reactions ($O_{3,\text{LC}}$). $O_{3,\text{RBG}}$ is defined as the $O_3$ concentrations in the absence of anthropogenic sources (Lu et al., 2019). It is mainly generated from natural sources such as biogenic volatile organic compounds (BVOCs), soil nitrogen oxides (SNOx), lightning NOx (LNOx), wildfires, and methane oxidation (Li et al., 2022; McDonald-Buller et al., 2011), and from stratosphere-troposphere exchange (Knowland et al., 2017) and long-range transport (Mathur et al., 2022). On one hand, the implementation of anthropogenic emission control and management policies by government agencies is promising for the reduction of $O_{3,\text{LC}}$ (Li et al., 2021). On the other hand, $O_{3,\text{RBG}}$ has contributed to the increase of $O_3$ concentrations in the past decade, mainly due to the promoted natural emissions due to changes in meteorological conditions (Chen et al., 2022; Lu et al., 2019). Therefore, investigation of the characteristics of $O_{3,\text{RBG}}$ is of great importance in the mitigation of $O_3$ pollution.

It has been reported that $O_{3,\text{RBG}}$ accounts for about 70–80% of the surface $O_3$ concentrations in China (Chen et al., 2022; Lu et al., 2019), but there leave large uncertainties. Three methods have been commonly used to estimate the $O_{3,\text{RBG}}$ concentrations: (1) Estimation by observed
Fig. 2. (a) Boxplot of the performances of the WRF-CMAQ (O$_{3,\text{SIM}}$) and MLR model (O$_{3,\text{MLR}}$) on simulating O$_3$ concentrations presented by mean bias (MB) of O$_{3,\text{SIM}}$ (in orange) and the MB of O$_{3,\text{MLR}}$ (in blue). The center line of the box represents the median, the square represents the mean, and the upper and lower edges of the box represent the 25th and 75th percentiles, respectively. (b, c) Performances of O$_{3,\text{SIM}}$ and O$_{3,\text{MLR}}$ for China and seven regions presented by normalized mean bias (NMB) and normalized mean error (NME). (d-f) Differences between MLR-adjusted and WRF-CMAQ simulated annual average concentrations of MDA8 O$_3$ (d), O$_{3,\text{RBG}}$ (e) and O$_{3,\text{LC}}$ (f). The values are calculated as (MLR – WRF-CMAQ).

Fig. 3. O$_{3,\text{RBG}}$ and O$_{3,\text{LC}}$ in China and specific regions in 2020. The column chart shows O$_{3,\text{RBG}}$ without a line and O$_{3,\text{LC}}$ with a line. The horizontal axis indicates different regions in China. The left vertical axis represents O$_3$ concentrations levels, with green, red, orange, blue, and pink colors representing spring, summer, autumn, winter, and annual averages, respectively. The purple dots line and the right vertical axis represents the proportion of O$_{3,\text{RBG}}$ to MDA8 O$_3$. 
O$_3$ at remote monitoring sites. However, it is difficult to avoid the influence of extensive human activities (Vingarzan, 2004). (2) Application of statistical methods such as principal component analysis (PCA) (Suciu et al., 2017), and Hidden Markov Models method (HMM) (Rizos et al., 2022). The PCA method determines the principal concentrations of O$_3$ at an observational site using an orthogonal matrix approach and considers it as the O$_3$ background concentrations. The HMM method categorizes the O$_3$ concentrations in a time-series dataset based on features such as anomalies and daily amplitude. This allows for the categorization of O$_3$ concentrations under different scenarios. The O$_3$ concentrations of the most stable and predominant scenario are considered as the background O$_3$. However, the statistical methods are based on mathematical relationships without references to the chemical and physical mechanisms regarding O$_3$ formation. (3) Application of the chemical transport models (CTM) (Atherton et al., 1996). This approach takes advantage of emulating the atmospheric reactions driven by chemical and physical processes and can therefore address the shortcomings of the statistical models. The CTM approach distinguishes the O$_3_{RBG}$ from the total O$_3$ by the Brute Force Method (BFM) (Cheng et al., 2017; Skipper et al., 2021). The BFM approach reduces the emissions of the precursors (for example, anthropogenic NOx emissions) and simulates the changes in the concentrations of air pollutants (for example, O$_3$ concentration) (Zhang

Fig. 4. Spatial distribution of O$_3_{RBG}$ and the proportion of O$_3_{RBG}$ in MDA8 O$_3$. Seasons are represented on the horizontal axis, and O$_3$ categories (MDA8 O$_3$, O$_3_{RBG}$, the proportion of O$_3_{RBG}$ in MDA8 O$_3$) on the vertical axis.
et al., 2014). However, the BFM approach could overlook the nonlinearity between the emission reduction and response to air pollution (Xie et al., 2022; Zheng et al., 2018). Additionally, the CTM results are deviated from site observations of $O_3$ due to the uncertainties in model inputs (for example, emission inventories and meteorological fields), and parameterization (for example, radiation scheme and gas chemistry mechanism) (Huang et al., 2019; Wang et al., 2022b; Xu et al., 2021).

The obstacles above-mentioned quantification methods led to a vague understanding of $O_3$ RBG, which is indicated by the inconsistent estimations from different studies. For example, Wang et al. (2011) used the Goddard Earth Observing System with Chemistry (GEOS-CHEM) model and reported that the annual $O_3$ RBG concentrations ranged between 25–55 ppb in different regions of China. Li et al. (2012) used the Community Multiscale Air Quality Modeling System - Ozone Source Apportionment Technology (CAMx-OSAT) to analyze $O_3$ RBG in the Pearl River Delta region of China and obtained 5–10 ppb higher results than that of Wang et al. (2011). Li et al. (2018) ran the WRF-CHEM model and found that $O_3$ RBG contributed 10–25 ppb in northwestern China, which is 30% lower than the assessments reported by Wang et al. (2011). Chen et al. (2022) quantified the $O_3$ RBG via the relationship between temperature and $O_3$, and found that the $O_3$ RBG values during the warm season in China ranges from 50 to 55 ppb, which are 10 ppb higher than the values estimated by the GEOS-CHEM results by Lu et al. (2019) and Wang et al. (2011).

This study aims to improve our understanding of the $O_3$ RBG concentrations and their role in the mitigation of $O_3$ pollution in China. We simulated the $O_3$ RBG and $O_3$ LC by the CTM-BFM approach for China for the year 2020. The results are then constrained by the observed $O_3$ concentrations from monitoring sites with the multiple linear regression (MLR) model to derive the adjusted $O_3$ RBG and $O_3$ LC. On these bases, we investigated the spatial and seasonal variations of $O_3$ RBG concentrations and discussed the major driving factors. Finally, we discussed the contribution of $O_3$ RBG to total $O_3$ pollution and gave insights into the mitigation of $O_3$ pollution in China.

2. Methodology

2.1. Site observations

Site observations of $O_3$ ($O_3$ OBS) were collected from the National Air Quality Monitoring Stations (AQMS) for the year 2020 (https://quotsoft.net/air/). We collected hourly data from 1462 sites (Fig. 1). The MDA8 $O_3$ concentrations (523,953 data sets in total) were used to evaluate the WRF-CMAQ model performance and build the MLR model. We also obtained the longitude, latitude, and altitude information of the stations for the MLR model. The meteorological data including temperature, relative humidity, wind speed, and wind direction with hourly temporal resolution were collected from the monitoring stations of the National Meteorological Bureau to evaluate the WRF performance (http://data.cma.cn/data). All data were preprocessed to exclude zero and negative values and outliers (see details in Xue et al., 2023).
2.2. Study domain

Fig. 1a presents the distributions of seven regions discussed in this study, including Northeastern China (NEC), North China (NC), East China (EC), Southern China (SC), Central China (CC), Northwestern China (NWC), and Southeast China (SWC). Detailed information about the geographic regions can be found in Table S1. Fig. 1b classifies the regions into three groups according to the topographies, climate zones, and emission sources. The NC, south EC, and southern NEC regions are in the eastern plain and hilly areas in eastern China (EPH region). This region has intensive human activities, and the NOX and VOCs emissions are dominated by anthropogenic sources. The CC, SC, northern NEC, and northern NEC regions are in the high-altitude area in central China (HAA region). This region is composed of mountains and plateaus, and the emissions mainly come from natural sources. The western parts of the NWC and SWC regions are mainly arid areas with low vegetation coverage in western China (ALV region).

2.3. Model configuration

The Weather Research and Forecasting model (WRF, v4.3) (Powers et al., 2017) coupled with the most updated Community Multiscale Air Quality model (CMAQ, v5.3.2) (Appel et al., 2021) system is applied to estimate the O3,RGB and O3,LC concentrations of surface O3 in China. The
The model domain covers China and the surrounding area (Fig. 1), with a spatial resolution of 36 km × 36 km. The WRF model extends 6 grid cells further on each side of the CMAQ model. We simulated the entire year of 2020 and demonstrated the results of the four seasons separately: spring (March, April, May), summer (June, July, August), autumn (September, October, November), and winter (January, February, December). For most regions of China, spring and summer are the warm seasons with the highest temperatures in summer. The warm season in the SC is autumn due to its proximity to the equator (Zhou and Huang, 2014).

The emission inputs of the CMAQ model include both anthropogenic and natural emissions. The anthropogenic emission inventory within China is based on the Multi-resolution Emission Inventory of China (MEIC) for the year 2018 with a spatial resolution of 0.25° × 0.25° (Kang et al., 2016). It provides anthropogenic emissions from industry, power plants, agriculture, mobile, and residence. Emissions from anthropogenic sources outside China are from the Emission Database for Global Atmospheric Research (EDGAR) maintained by the European Commission (Olivier et al., 1994). The SNOx and BVOCs emissions are estimated by the Model of Emissions of Gases and Aerosols from Nature (MEGAN, v3.1) (Guenther, 2007). The emissions from wildfires are derived from the high spatiotemporal resolution fire inventory from NCAR (FINN) (Wiedinmyer et al., 2011). The other emission inputs are obtained from the online calculations of the CMAQ model. LNOx is determined by empirical relationships between lightning and convective precipitation, cloud layers, and others (Kang et al., 2019). The dependence of O$_3$ on sea salt halogens relies on the Sea Spray Emissions Module (Kang et al., 2019). Meanwhile, stratospheric transport is derived based on the correlation between potential vorticity and stratospheric intrusion O$_3$ (Xing et al., 2016).

We obtained the model predicted O$_3$ (O$_3$ SIM), O$_3$ RBG (O$_3$ SRBG), and O$_3$ LC (O$_3$ SLC) concentrations by using the CTM-BFM method. We setup two scenarios, including the Base case and Control case. In the base case, we used the abovementioned emission inventories for anthropogenic and natural emissions. In the control case, we kept the same set-up as the base case but excluded the anthropogenic emissions of China. O$_3$ SRBG is calculated by the control case. O$_3$ SLC is calculated by the difference between the base case and the control case.

Fig. 8. The variations of O$_3$ RBG in different MDA8 O$_3$ ranges. The y-axis represents the proportions of O$_3$ RBG in MDA8 O$_3$. The x-axis represents different MDA8 O$_3$ ranges. The dots line represents the trends of O$_3$ RBG for China and error bars represents the maximum and minimum values of the seven geographical regions. The colors of the dots line indicate the concentrations of O$_3$ RBG. The pink background represents MDA8 O$_3$ ≥ 80 ppb, purple background represents stage1 (67 ppb < MDA8 O$_3$ < 77 ppb), purple background represents stage1 (67 ppb < MDA8 O$_3$ < 77 ppb), green background represents stage2 (50 ppb < MDA8 O$_3$ ≤ 67 ppb), blue background represents stage3 (MDA8 O$_3$ ≤ 50 ppb).
2.4. MLR model

The MLR model has been widely used in the field of atmospheric pollution forecasting and analysis (Han et al., 2020; Hsu et al., 2022; Qian et al., 2022; Skipper et al., 2021). This method establishes a relationship between the dependent and independent variables, deriving corresponding regression coefficients ($\alpha$) acting on the independent variables for a better fit of the dependent variable. This is accomplished through the least squares method, which minimizes the total sum of squared errors to find the optimal function and solve for $\alpha$.

$O_{3, OBS}$ are composed of $O_{3, RGB}$ and $O_{3, LC}$ (Eq. (1)). We employed the CTM-BFM method to simulate the $O_{3, RGB}$ (denoted as $O_{3, RGB}$) and $O_{3, LC}$ (denoted as $O_{3, LC}$) concentrations. These two values are deviated from the true value of $O_{3, RGB}$ and $O_{3, LC}$ due to bias in CTM and the BFM approach. We then establish the MLR model, where $O_{3, RGB}$ and $O_{3, LC}$ are constrained by $O_{3, OBS}$ (Eqs. (2)-(3)). As mentioned earlier, $\alpha$, resolved by MLR, serves as an indicator of the degree to which simulations are constrained by observations. The MLR model is described as follows:

$$O_{3, OBS} = O_{3, RGB} + o_{\alpha} \cdot O_{3, LC}$$

(1)

$$O_{3, RGB} = O_{3, LC} \cdot \alpha (X, Y, Z)$$

(2)

$$O_{3, LC} = O_{3, LC} \cdot \alpha (X, Y, Z)$$

(3)

where $\alpha (X, Y, Z)$ represent $\alpha$ of $O_{3, RGB}$ ($O_{3, LC}$).

The $O_{3, RGB}$ and $\alpha (X, Y, Z)$ values are affected by the geographic information (De Bock et al., 2014; Yan et al., 2020). In this study, we used the standardized longitude ($X$), latitude ($Y$), and altitude ($Z$) to improve the MLR performance and explain the deviation between simulation and observation as shown in Eqs. (4)-(5).

$$\alpha (X, Y, Z) = \alpha_0 (X, Y, Z) + \alpha_1 (X, Y, Z) + \alpha_2 (X, Y, Z) + \alpha_3 (X, Y, Z)$$

(4)

$$\alpha (X, Y, Z) = \alpha_0 (X, Y, Z) + \alpha_1 (X, Y, Z) + \alpha_2 (X, Y, Z) + \alpha_3 (X, Y, Z)$$

(5)

The MLR model solved the $\alpha$ values on the right side of Eqs. (4)-(5). Then, Eqs. (4)-(5) are substituted into Eqs. (2)-(3) to obtain the $O_{3, RGB}$ and $O_{3, LC}$ values. Each CTM grid has different $\alpha$ values based on $X$, $Y$, and $Z$ information. Simultaneously, we obtain $O_{3, RGB}$ ($O_{3, LC}$) constrained by $O_{3, OBS}$ and consider them as our assessments for $O_{3, RGB}$ ($O_{3, LC}$).

3. Results and discussion

3.1. MLR improvements over CMAQ

Fig. 2a, b, c demonstrates the performance of the WRF-CMAQ model in simulating the daily $O_3$ concentrations. The overall performance of the WRF model is acceptable, with slight underestimation in temperature, wind speed, and slightly overestimation in the relative humidity (detailed information in Table S2). The CMAQ model succeeded in capturing the distribution of MDAB $O_3$ in China (Fig. S1) but slightly overestimated the concentrations (Fig. 2a). The Correlation efficient ($R$) was 0.66, which is overall comparable or better than previous studies (Kuo and Fu, 2023). The normalized mean bias (NMB) and normalized mean error (NME) values were 25 % and 37 %, respectively and have both exceed the benchmark by Emery et al. (2017) (Fig. 2b–c).

Evaluations in different regions (Table S3) also indicate general overestimations of the $O_3$ concentrations over the whole of China. The uncertainty of CMAQ $O_3$ modeling in the Chinese region are common. Even within the same study (Kuo and Fu, 2023), the simulation’s R in same region may vary noticeably due to differences in simulation time period (0.4–0.8), Shi et al. (2021) and Xing et al. (2017) conducted simulations with different meteorological scenarios or for different regions nationwide. Their NMB exhibit considerable variations across different regions or months (−50 % to 250 %, averaging 33 % and 32 %). The
estimating the stable O\textsubscript{3} concentrations at a certain temperature range. Our results are 10–20 ppb lower than theirs during December–February and 3–10 ppb lower during May–August. Whether comparing with the O\textsubscript{3} background concentrations obtained solely through the use of BFM in our study, or with findings from other studies, the O\textsubscript{3} background concentrations derived from MLR-BFM in our study are consistently lower. It is worth noting that the differences in research time period and inputs will inevitably affect the results. However, comparing with others can still serve as a valid reference and is necessary, especially when we focus on the final O\textsubscript{3,LMB} obtained.

3.2. Spatial distribution of O\textsubscript{3,RGB} and its contribution to total O\textsubscript{3}

Fig. 3 summarizes the concentrations of O\textsubscript{3,RGB} and O\textsubscript{3,LC} in seven regions of China. Fig. 4 shows the spatial variations of O\textsubscript{3} and O\textsubscript{3,RGB} in different seasons. The O\textsubscript{3,RGB} ranged between 22–45 ppb and averaged at 35 ± 4 ppb in China for the year 2020. Compared to the spatial distribution of O\textsubscript{3} (Fig. 4a), distribution of O\textsubscript{3,RGB} is relatively homogenous (Fig. 4b). The EC region has the highest O\textsubscript{3,RGB} concentrations (37 ppb), followed by the NC (36 ppb), CC (35 ppb), SC (34 ppb) regions and the other regions (32–33 ppb). The contribution of O\textsubscript{3,RGB} to total O\textsubscript{3} ranged between 71–92 % and averaged at 81 ± 5 % over China. The HAA region has a 10 % higher proportion of O\textsubscript{3,RGB} in MDA8 O\textsubscript{3} (85%) than the EPH region (75%) (Fig. 4c).

The HAA region has generally higher O\textsubscript{3,RGB} Concentrations than the EPH region. Research suggests that factors such as the elevated inversion layer and the influence of pollutants’ mesoscale circulation at higher altitudes contribute to elevated O\textsubscript{3} concentrations in mountainous areas (Guo et al., 2013). In addition, stratosphere-troposphere intrusion of O\textsubscript{3} in western and northwestern China at higher altitudes could induce about 5–10 ppb of O\textsubscript{3} during lightning-active months (Lu et al., 2019; Roy et al., 2017; Xu et al., 2016). Due to lower anthropogenic VOCs (AVOCs) emissions in the HAA region, this region has a high proportion of O\textsubscript{3,RGB}.

For the EPH region, the large emissions of SNOx (and soil HONO) are one of the reasons for high O\textsubscript{3,RGB} in the southern of NC and northern of EC regions during summer (Lu et al., 2019; Lyu et al., 2022; Xue et al., 2021; Huang et al., 2023) (Fig. 4f). The concentrations in coastal region (EC and NC) are about 1–4 ppb higher than that in inland region, which may reflect the contribution of marine background O\textsubscript{3} (Lam et al., 2001; Wang et al., 2022a).

3.3. Driving factors of seasonal variations of O\textsubscript{3,RGB}

Meteorology (solar radiation, temperature, precipitation, etc.) and natural emissions are the main drivers of the seasonal distributions of O\textsubscript{3,RGB} in China. To get insights into the contributions of meteorological factors and natural emissions to O\textsubscript{3,RGB}, we adopted meteorological standardization, random forest, and Shapley additive explanations to interpret the driving forces of O\textsubscript{3,RGB}. We derived hourly contributions of meteorology (O\textsubscript{3,RGB} concentration of weather contribution, O\textsubscript{3,wc}) and natural emissions (O\textsubscript{3,RGB} concentration of removing meteorology, O\textsubscript{3,rmw}) to O\textsubscript{3,RGB} (Fig. 5). Additionally, based on previous studies employing this method, we identified the meteorological factors that have a relatively significant impact on O\textsubscript{3} and determined their contributions (Shap value from Shapley additive explanations) to O\textsubscript{3,RGB} (Fig. 6). More details about our machine learning method can be found in Xue et al. (2023), Hou et al. (2022a), and Grange et al. (2018). Model performance evaluation about our machine learning methods is shown in Table S6.

From the results, it is evident that O\textsubscript{3,rmw} contribute more to O\textsubscript{3,RGB} on average than O\textsubscript{3,wc}, a commonality across all regions and seasons. The extreme values of O\textsubscript{3,wc}’s contribution differ more significantly than O\textsubscript{3,rmw}’s, with instances of negative contributions, indicating substantial fluctuations in meteorological contributions. In spring (Fig. 5a), meteorological conditions favor O\textsubscript{3,RGB} production (Feng and Wang, 2020), and O\textsubscript{3,LC}’s overall average contribution is 2 ppb (1.4 ppb to 5 ppb). The relatively high emissions of BVOCs and SNOx (Dai et al., 2018; Mohanty and Panda, 2011; Ruan et al., 2004) result in O\textsubscript{3,rmw} being in a higher concentration range (36–39 ppb). With the arrival of summer (Fig. 5b), O\textsubscript{3,rmw} in most areas (except the NEC region and SC region, discussed later) enters a period of highest concentration (37–40 ppb), associated with further enhanced emissions of BVOCs and SNOx. O\textsubscript{3,wc} in heavily polluted areas like the NC region and EC region also reach peaking concentrations (4.6 ppb, 7.7 ppb). In the NEC region, O\textsubscript{3,rmw} contributes more in spring, aligning with the fact that its O\textsubscript{3,RGB} and MDAB O\textsubscript{3} are highest in spring (Shang et al., 2023). In the SC region, autumn has the highest emissions of BVOCs and SNOx, hence the highest autumn O\textsubscript{3,rmw} (Fig. 5c). Additionally, SC is the only region where O\textsubscript{3,wc} is positive in autumn (1.03 ppb), as the SC region experiences weather conditions in autumn conducive to O\textsubscript{3} production. For other regions, both O\textsubscript{3,rmw} and O\textsubscript{3,wc} contributions are lower in autumn and winter compared to spring and summer, attribute to reduced emissions of BVOCs and SNOx and meteorological conditions less favorable for O\textsubscript{3} production. In winter (Fig. 5d), radiation hits its yearly minimum, temperatures drop, and wind speeds increase. For most regions in China, O\textsubscript{3,wc} contributions reach their yearly minimum.

Fig. 6 illustrates the contributions of various meteorological factors to O\textsubscript{3,RGB} in different regions in different seasons. Regarding the contribution of meteorology to O\textsubscript{3,RGB}, it’s noteworthy that there are apparent negative contributions. Therefore, we additionally take the absolute values of meteorological factors’ contributions to O\textsubscript{3,RGB} in Fig. S2 to illustrate the influencing capacity of meteorological factors on O\textsubscript{3,RGB}. In spring, humidity has the highest contribution in the CC, NC, and SC regions (1.3–1.9 ppb), and its absolute impact is also higher (30 %–40 %). The EC region exhibits a larger contribution from radiation (0.52 ppb), with relative humidity ranking second (0.47 ppb), but the absolute impact of temperature is the highest (29 %). For the NEC region, different from other regions, both radiation and wind speed contribute more (both 1 ppb). In summer, temperature dominates the contributions in two O\textsubscript{3} heavily polluted regions, the EC region and NC region (3.8 ppb and 5.3 ppb), with absolute contributions of 46 % and 52 %, emphasizing the impact of temperature and hot weather on heavily polluted areas. The CC region is also more influenced by humidity (2.9 ppb), but the contribution of temperature shifts from −1.2 ppb in spring to a positive contribution of 1.7 ppb. In autumn, temperature continues to contribute positively in the EC region but decreases to 0.38 ppb. Humidity becomes the more influential meteorological factor in the EC region (45 %). In the northern latitudes, where temperatures are colder in the NC region and CC region, temperature becomes a negative contributor (−1.4 ppb and −1.07 ppb), with the highest absolute impact (34 % and 36 %). In winter, the NC region is still most significantly influenced by temperature (43 %), contributing −4.1 ppb. In the EC region, the contribution from temperature is −4.13 ppb, dominating the meteorological factors, with an absolute impact percentage of 62 %. Overall, for heavily polluted areas like the NC region and EC region, temperature plays a predominant role in influencing O\textsubscript{3}, while humidity has a greater impact in the SC region. For other regions, the CC region is mainly influenced by humidity, and the NEC region is significantly influenced by temperature, radiation, and wind speed.

3.4. Contributions at different O\textsubscript{3} levels

Although China has implemented ambitious emission reduction plans for anthropogenic VOCs and NO\textsubscript{x} over the past decade, the annual variation in O\textsubscript{3} concentration remains fluctuating without a clear downward trend. We analyze the forms of O\textsubscript{3} pollution in China from the perspective of O\textsubscript{3,RGB} and O\textsubscript{3,LC} contributions. Fig. 7 shows the average O\textsubscript{3,RGB} and O\textsubscript{3,LC} on clean days (MDAB O\textsubscript{3} < 80 ppb) and polluted days (MDAB O\textsubscript{3} ≥ 80 ppb) for different regions. On clean days, the average O\textsubscript{3,RGB} concentration is around 34 ppb and contributes 81 % to total O\textsubscript{3}. On polluted days, the average O\textsubscript{3,RGB} value increased to 45 ppb but the
contribution decreased to 55 %, indicating the relatively higher contributions from anthropogenic precursor emissions. There is a significant increase in the proportion of O\textsubscript{3,LC} on pollution days in all regions (an average increase of 250 % compared to clean days). In NC and SC regions, the O\textsubscript{3,LRB} (49 ppb, 47 ppb) on polluted days is higher compared to other areas, corresponding to elevated emissions of biogenic precursors in these two regions. It is worth noting that the NEC region has the highest increase in O\textsubscript{3,LC} contribution on pollution days. This may be due to the fact that, unlike other regions, the contribution of external O\textsubscript{3} in the warm season of NEC exceeds 60 %, the concentration of precursor substances from local anthropogenic sources is lower, and the ability of NO to titrate O\textsubscript{3} is weak (Fang et al., 2021). Detailed assessments for each province can be found in Table S7-S9.

To further analyze the relative contributions of O\textsubscript{3,LRB} and O\textsubscript{3,LC} to environmental O\textsubscript{3} pollution control policies, we sampled MDA8 O\textsubscript{3} every 5 ppb and tracked the changes in O\textsubscript{3,LRB} and O\textsubscript{3,LC} (Fig. 8). Over the past decade, China’s MDA8 O\textsubscript{3} 90th percentile mainly remained within the Stage 1 range (67-77 ppb, Fig. 6, Stage 1). We found that the substantial contribution of O\textsubscript{3,LC} on polluted days leads to extremely high concentrations of MDA8 O\textsubscript{3}, supporting the current Chinese strategy of reducing peak O\textsubscript{3} concentrations. Additionally, O\textsubscript{3,LC} constitutes 35–40 % of MDA8 O\textsubscript{3}, indicating that there is still a considerable amount of O\textsubscript{3,LC} that can be reduced. This supports the Chinese government’s view of strengthening the reduction of anthropogenic VOCs and NO\textsubscript{x} to reduce O\textsubscript{3} concentrations. However, it is worth noting that in Stage 1, the proportion of O\textsubscript{3,LRB} has already reached 60–65 %. This implies that if O\textsubscript{3} concentrations are reduced solely by decreasing anthropogenic VOCs and NO\textsubscript{x}, while neglecting the contribution of O\textsubscript{3,LRB}, the effectiveness may be reduced by more than half. Stage 2 represents the next target for China after escaping the O\textsubscript{3} concentration range from Stage 1. Since the Chinese government has not clearly specified the expected range of future O\textsubscript{3} concentrations, we adopted the World Health Organization’s recommended 50 ppb as the target, considered an acceptable O\textsubscript{3} concentration. We defined the range of 50–67 ppb as Stage 2 (Fig. 8, Stage 2), representing the O\textsubscript{3} concentration reduction that China still needs to achieve. To reach a pollution-free level, China needs to reduce the O\textsubscript{3} concentration by at least 17 ppb. In this process, the proportion of O\textsubscript{3,LRB} will increase from 66 % to 73 %. This emphasizes the importance of O\textsubscript{3,LRB} and underscores the necessity of strengthening inter-regional joint prevention and control measures to transition from Stage 2 to Stage 3 (below 50 ppb, Fig. 8, Stage 3) and achieve the goal.

3.5. Uncertainty analysis

There are two primary factors contributing to the uncertainties in estimating O\textsubscript{3,LRB}. The first is related to uncertainties of MLR. The choice of independent variables affects the way the dependent variable is interpreted. In this study, we standardized the independent variables into consistent concentration units and estimated the impact of O\textsubscript{3} background concentrations, local photochemical reactions, standardized normalized longitude, standardized normalized latitude, and standardized normalized altitude on observed O\textsubscript{3}. However, it is worth noting that O\textsubscript{3} background concentrations and local photochemical reactions predominantly account for the influences, which have been confirmed by others as well (Skipper et al., 2021). Therefore, the inclusion of additional independent variables has minimal overall impact on the relationship between total O\textsubscript{3} and O\textsubscript{3} background concentrations and local photochemical reactions. On the other hand, The training dataset also influences the results of MLR, similar to machine learning, where a large and valid dataset leads to more convincing and satisfactory outcomes (Shang et al., 2023). This is also why we excluded data from the ALV region, as the available training data is too limited compared to the extensive area.

4. Conclusions

In this study, we applied the MLR method to constrain model predicted O\textsubscript{3} and O\textsubscript{3,LRB} from CTM with site observations of O\textsubscript{3}. There has not been a report using this method to assess China’s O\textsubscript{3} background concentrations. Our results demonstrate that this approach has significant improvement on the CTM model performance on predicting O\textsubscript{3} over the whole of China. We estimated the average O\textsubscript{3,LRB} in China to be 35 ± 4 ppb, accounting for 81 ± 5 % of MDA8 O\textsubscript{3}. The O\textsubscript{3,LRB} concentrations are higher in warm seasons (spring 39 ppb, summer 38 ppb) than in cold seasons (winter 31 ppb, autumn 30 ppb). Their contributions to total O\textsubscript{3} are slightly lower in the warm seasons (78 %) than cold seasons (83 %). Natural emissions contribute (31–40 ppb) more significantly than meteorology to O\textsubscript{3,LRB} in various regions and seasons. Meteorological contributions exhibit greater fluctuations, with negative contributions, reaching a seasonal average of 4.6–7.7 ppb during the summer in heavily polluted areas (EC and NC regions). Temperature and relative humidity emerge as the two predominant meteorological factors with the highest impact in urban areas, with their contributions ranging from 30 % to 62 %.

We further investigated the role of O\textsubscript{3,LRB} in controlling O\textsubscript{3} pollution. The average O\textsubscript{3,LRB} concentration is 34 ppb on clean days (daily MDA8 O\textsubscript{3} = 80 ppb) and contributes 81 % to total O\textsubscript{3}. The value increased to 45 ppb on polluted days (daily MDA8 O\textsubscript{3} > 80 ppb) and contributes 55 % to total O\textsubscript{3}. The contribution of O\textsubscript{3,LC} on pollution days over 45 %, supporting the current Chinese policy’s goal of reducing O\textsubscript{3} peak concentrations by minimizing anthropogenic precursor emissions. Within the O\textsubscript{3} concentration 90th percentile range of the past decade, O\textsubscript{3,LC} has contributed 35–40 % to MDA8 O\textsubscript{3}. This supports the Chinese government’s view of strengthening the reduction of anthropogenic precursor emissions to lower O\textsubscript{3} concentrations. However, O\textsubscript{3,LRB}’s contribution to MDA8 O\textsubscript{3} already exceeds half and is expected to rise to 73 % with further reduction in anthropogenic precursor emissions, reaching the WHO-recommended target of 50 ppb. Therefore, there is a critical need to emphasize the importance of including O\textsubscript{3,LRB} in policies, an aspect currently missing, and to prioritize regional joint prevention and control measures.

CRediT authorship contribution statement

Zhiwu Sun: Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft. Jiani Tan: Formal analysis, Investigation, Methodology, Validation, Writing – review & editing. Fangting Wang: Formal analysis. Rui Li: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing. Xinxin Zhang: Formal analysis. Jiaqiang Liao: Formal analysis. Yangjun Wang: Conceptualization, Methodology. Ling Huang: Conceptualization, Investigation. Kun Zhang: Formal analysis, Methodology. Joshua S. Fu: Methodology, Writing – review & editing. Li Li: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Data availability

Data will be made available on request.

Acknowledgements

This study is financially supported by the Shanghai Municipal Bureau of Ecology and Environment Fund Project ((2022)37) and the


