

Article

Assessing the Aromatic-Driven Glyoxal Formation and Its Interannual Variability in Summer and Autumn over Eastern China

Xiaoyang Chen ¹, Xi Chen ², Yiming Liu ³ , Chong Shen ⁴ , Shaorou Dong ⁵, Qi Fan ^{3,*}, Shaojia Fan ³, Tao Deng ¹, Xuejiao Deng ¹ and Haibao Huang ⁶

¹ Guangzhou Institute of Tropical and Marine Meteorology of China Meteorological Administration, GBA Academy of Meteorological Research, Guangzhou 510640, China; chenxiaoyang@gd121.cn (X.C.)

² Institute of Atmospheric Environment, Guangdong Provincial Academy of Environmental Science, Guangzhou 510045, China

³ School of Atmospheric Sciences, Sun Yat-sen University, Zhuhai 519082, China

⁴ Guangzhou Ecological and Agricultural Meteorological Center, Guangzhou 511430, China

⁵ Guangdong Provincial Climate Center, Guangzhou 510640, China

⁶ Guangdong Provincial Key Laboratory of Environmental Pollution Control and Remediation Technology (Sun Yat-sen University), Guangzhou 510275, China

* Correspondence: eesfq@mail.sysu.edu.cn

Abstract

Aromatics and their key oxidation intermediate such as formaldehyde and dicarbonyl compounds (glyoxal and methyglyoxal) are crucial precursors for ozone (O₃) and secondary organic aerosols (SOA). However, the spatial–temporal variation in aromatics' contribution to these intermediate species and O₃/SOA over Eastern China during the past decades remains insufficiently quantified. This study combines satellite observations of formaldehyde and glyoxal column densities (2008–2014) with an innovative tracer method implemented in the Community Multiscale Air Quality (CMAQ) modeling system to quantify aromatic-driven dicarbonyl chemistry. Simulations of summer and autumn in 2010, 2012, 2014, and 2016 are conducted to demonstrate the change in aromatics and its impact through the years. Estimated primary and intermediate VOCs show good consistency with measurements at a supersite; and the simulated vertical column density of formaldehyde and glyoxal agree with satellite observations in spatial distributions. The contribution of aromatic hydrocarbons to the columnar concentration of glyoxal has seen a significant increase since 2010, which can, to some extent, explain the interannual trend of glyoxal column concentrations in key regions of Beijing–Tianjin–Hebei (BTH), Yangtze River Delta (YRD), and Pearl River Delta (PRD). A cross-comparison reveals a good consistency between the observed glyoxal columnar concentrations to formaldehyde columnar concentration ratio (R_{GF}) from satellite measurements and the high contribution areas of aromatics to glyoxal: pronounced values are observed in the above three key regions in Eastern China. Additionally, the applicability of R_{GF} and its indicative nature in Eastern China was discussed, revealing notable seasonal and regional variations in R_{GF} . Revised R_{GF} thresholds ([0.015–0.03] for models vs. [0.04–0.06] for satellites) improve summer precursor classification, while a threshold of >0.04 could distinguish the areas with high anthropogenic impacts during autumn. These findings advance understanding of VOC oxidation pathways in polluted regions, providing critical insights for ozone and secondary organic aerosol mitigation strategies. The integrated satellite model approach demonstrates the growing atmospheric influence of aromatics amid changing emission patterns in Eastern China.



Academic Editor: Pavel Kishcha

Received: 3 July 2025

Revised: 1 September 2025

Accepted: 9 September 2025

Published: 12 September 2025

Citation: Chen, X.; Chen, X.; Liu, Y.; Shen, C.; Dong, S.; Fan, Q.; Fan, S.; Deng, T.; Deng, X.; Huang, H. Assessing the Aromatic-Driven Glyoxal Formation and Its Interannual Variability in Summer and Autumn over Eastern China. *Remote Sens.* **2025**, *17*, 3174. <https://doi.org/10.3390/rs17183174>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: aromatics; glyoxal; formaldehyde; column density; satellite; CMAQ

1. Introduction

Aromatics serve as crucial precursors to both ozone (O_3) and secondary organic aerosols (SOA). Upon oxygenation, aromatics can generate RO_2 radicals. RO_2 radicals participate in the NO_x cycles, facilitating the production of increased O_3 via the photolysis of NO_2 [1]. Furthermore, the RO_2 radicals can undergo additional oxygenation to form semi-volatile or low-volatility organic compounds, subsequently leading to the formation of SOA through gas particle partitioning. Notably, certain dicarbonyl compounds, including glyoxal (GLY) and methylglyoxal (MGLY), which are vital intermediates derived from aromatics, can also contribute to SOA formation through the aqueous phase processes occurring on cloud droplets and aerosol surfaces.

Since 2013, China's ground level national air quality monitoring network has shown a rapid rise in near-surface ozone (O_3) concentrations [2,3]. Between 2013 and 2017, under the country's strong and effective emission control measures in the Air Pollution Prevention and Control Action Plan, the emissions and pollution problems related to sulfur dioxide (SO_2), nitrogen oxides (NO_x), and particulate matter saw significant improvements. However, the decline in particulate matter concentrations, which led to changes in photochemical and heterogeneous processes, combined with a lack of corresponding emission control measures for volatile organic compounds (VOCs), exacerbated O_3 pollution during this period [4,5]. Eastern China, being the region with the highest anthropogenic activity, experiences extensive pollutant emissions. There are multiple key economic regions in Eastern China, within which the Beijing–Tianjin–Hebei region (BTH), Yangtze River Delta (YRD), and Pearl River Delta (PRD) are focused on in this study. In these regions, the precursors' sensitivity of ground level O_3 in highly developed urban areas generally leans towards VOC-limited or a mixed sensitivity [6,7]. In areas with such ozone precursor sensitivities, a decrease in NO_x concentrations or an increase in VOC concentrations may lead to a rise in ozone levels [8]. In the period from 2018 to 2020, due to the decelerated decrease in particulate matter emissions and concentrations, combined with the introduction and implementation of the “13th Five-Year Plan for the Prevention and Control of Volatile Organic Compound Pollution”, the intensification of near-surface O_3 pollution has slowed but continues to fluctuate upwards [9]. There is an urgent need to further implement precise VOC emission reductions, achieve coordinated reductions in VOCs and NO_x , and address the O_3 pollution issue.

VOCs play an instrumental role in the formation process of secondary pollutants, acting as vital precursors for both ozone and SOA. However, the emission patterns of VOCs are intricate, with considerable challenges in accurately measuring their components. Furthermore, many species remain elusive to precise quantification. Unlike the emission reduction measures and outcomes of the aforementioned inorganic pollutants, the uncertainty surrounding VOCs emission sources remains substantial. The measures and efforts towards emission reduction are still in their nascent stages, without a discernible decline in overall emissions. VOCs are poised to be pivotal in controlling the concentration of secondary pollutants both in the present and future, marking them as a focal point in air pollution management.

Among the various components of VOCs, aromatics possess a notably high potential for ozone formation. Although aromatics do not rank at the forefront in terms of concentration or volume fraction metrics, their significant reactivity plays a crucial role in influencing and contributing to ozone levels. Furthermore, aromatics serve as essential precursors to secondary organic aerosols (SOA). Some studies suggest that aromatics are the

primary precursors of anthropogenic secondary organic aerosols (ASOA) [10]. RO₂ radicals formed from the oxidation of aromatics can be oxidized by peroxy radicals to produce low-volatility organic compounds that directly enter the particulate phase, or they can be oxidized to generate semi-volatile intermediate species that produce SOA through gas particle partitioning processes. Aromatics can also be oxidized to produce oxygenated volatile organic compounds (OVOCs), such as glyoxal and methylglyoxal, which then generate SOA via heterogeneous processes (liquid-phase uptake and catalytic oxidation) on aerosol surfaces. Therefore, aromatics represent a critical component of VOCs that significantly impact both ozone (O₃) and SOA. However, there remains considerable uncertainty in the anthropogenic emission volumes of aromatics. The spatiotemporal distribution patterns, mechanisms related to aerosols, and the historical evolution of aromatics' impacts on ozone and SOA are still inadequately researched, necessitating further comprehensive studies.

Glyoxal serves as a primary oxidation product of aromatic hydrocarbons, while also being generated through the oxidation of various VOCs including isoprene, acetylene, terpenes, and other unsaturated compounds. Globally, the major sources contributing to glyoxal formation are distributed as follows: isoprene (46.7%), acetylene (19.8%), primary glyoxal emissions (17.1%), ethene (5.6%), terpenes (4.0%), and aromatics (4.4%) [11]. However, in highly urbanized areas, aromatics demonstrate substantially greater contributions to glyoxal production. For instance, emission inventory-based calculations for the Pearl River Delta in 2006 revealed that aromatics accounted for approximately 75% of total glyoxal formation [12]. This pronounced spatial variability establishes glyoxal as an effective indicator of aromatic hydrocarbon levels. Consequently, satellite-observed glyoxal column densities currently represent one of the most valuable indirect datasets for analyzing the spatiotemporal patterns and temporal evolution of atmospheric aromatics.

Both formaldehyde (HCHO) and glyoxal column concentrations serve as excellent indicators, effectively reflecting regional atmospheric photochemical activity and VOC abundance while highlighting spatial variations across different areas. Recent studies have introduced the ratio of glyoxal to formaldehyde column concentrations (R_{GF}) as a novel diagnostic tool. This metric has demonstrated utility in distinguishing whether biogenic or anthropogenic volatile organic compounds (AVOCs) dominate regional atmospheric chemistry. Vrekoussis et al. (2010) [13], utilizing GOME-2 satellite data, established that global regions with R_{GF} values between 0.04 and 0.06 are predominantly influenced by biogenic VOCs, whereas areas with $R_{GF} < 0.04$ exhibit stronger anthropogenic signatures. Subsequent research has confirmed R_{GF} 's diagnostic potential, though its applicability and interpretation require season- and environment-specific refinement. For instance, Kaiser et al. (2015) [14] found R_{GF} to be less effective as an indicator during U.S. summers when isoprene (ISOP) dominates and masks anthropogenic signals. Conversely, ground-based observations by DiGangi et al. (2012) [15] yielded partially contradictory conclusions to Vrekoussis et al.'s findings [13]. These discrepancies underscore the need for further investigation into R_{GF} 's behavior.

In this study, satellite-based formaldehyde and glyoxal column density during years 2008–2014 are utilized. A tracer method is developed and implemented in a chemical transport model system for quantifying the contribution of aromatics to dicarbonyls. An innovative method integrating a satellite model approach provides a robust VOC-source apportionment tool for identifying anthropogenic vs. biogenic precursors' contribution to the formation of intermediate VOCs. The revised R_{GF} diagnostic framework with quantified aromatics' contributions to formaldehyde and glyoxal columns significantly enhances precursor identification, offering scientific support for precise secondary pollution mitigation targeting anthropogenic aromatics.

2. Materials and Methods

2.1. WRF-CMAQ Modeling System

An offline model system coupling the Weather Research and Forecasting model (WRF) with the Community Multi-scale Air Quality (CMAQ) model system is utilized to conduct the numerical experiments in this study. This WRF–CMAQ modeling system was described and the model configuration including physical and chemical parameterizations can be found in a previous study [16]. It is further revised for the development of a diagnostic function in this study, which tracks and distinguishes the contribution of aromatics to dicarbonyls (glyoxal and methyglyoxal) and aerosol concentrations. The WRF–CMAQ system is based on WRF model version 3.7.1 and CMAQ model version 5.1 [17]. The configured domain covers Eastern China, with a grid spacing of 27 km (Figure 1). We primarily focus on the three key regions in Eastern China: the BTH which indicates Beijing city, Tianjin city, and Hebei Province; the YRD which includes Shanghai city, Jiangsu Province, and Zhejiang Province; and the PRD which includes nine cities: Guangzhou, Shenzhen, Foshan, Zhaoqing, Zhongshan, Jiangmen, Zhuhai, Dongguan, and Huizhou. The WRF model utilizes historical reanalysis data for initial and boundary conditions, allowing for a more refined numerical simulation of the meteorological field within the simulation domain area. The results of the WRF meteorological field simulation are further processed for certain parameters through the Meteorology-Chemistry Interface Processor (MCIP), providing the necessary meteorological parameters for conducting the CMAQ simulations. Anthropogenic emission inputs are obtained from the Multi-resolution Emission Inventory for China (MEIC) developed by Tsinghua University [18], which dynamically couples changes in emission characteristics, socio-economic development, and industrial process technology updates. This study utilizes version 1.3 of the MEIC emission inventory with SAPRC07 mechanism classification, which provides detailed pollutant emission data for mainland China at $0.25^\circ \times 0.25^\circ$ grid resolution for four years of 2010, 2012, 2014, and 2016. Biogenic emissions are calculated using the MEGAN model v2.1 [19,20] with meteorological field inputs from WRF model simulations. The biogenic emissions vary according to meteorological simulation results for July and October from 2010 to 2016.

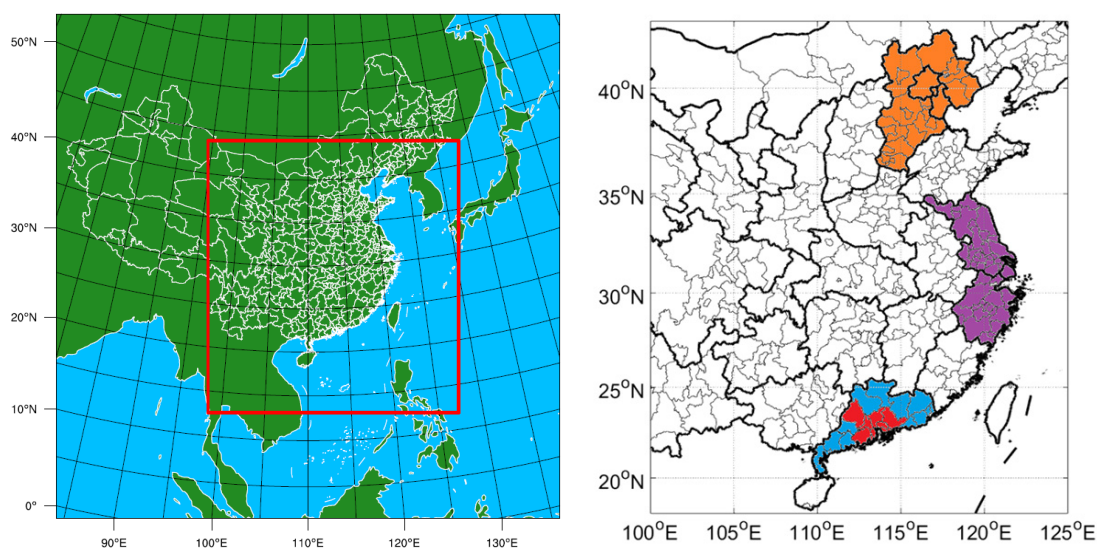


Figure 1. (Left) Domain configuration of the WRF–CMAQ modeling system. (Right) Key study regions in Eastern China (orange: BTH, purple: YRD, red: PRD, blue: Guangdong province).

The original version of the SAPRC07tic mechanism implemented in CMAQ v5.1 features 179 simulated species and 789 reaction equations, providing detailed descriptions of aromatics and their oxidation products. Based on SAPRC07tic, this study developed

a tracer method to identify contributions of aromatics to glyoxal (GLY), methylglyoxal (MGLY), and dicarbonyl SOA which is produced by the aqueous uptake onto aerosol surface (aqSOA) [21]. Utilizing this method, our WRF–CMAQ model system can output the concentrations of dicarbonyl compounds, dicarbonyl aqueous aerosols (AGLYJ, AMGLYJ), dicarbonyl compounds originating from aromatics (GLYaro, MGLYaro), dicarbonyl aqueous aerosols originating from aromatics (AGLYJaro, AMGLYJaro), and respective concentrations from primary emissions of dicarbonyls (GLYpri, MGLYpri, AGLYpri, AMGLYpri). The detailed construction process of this method is described follow.

To quantitatively assess the contribution of aromatic hydrocarbons to the spatiotemporal variations in glyoxal and formaldehyde column concentrations, and to guide the inverse application of satellite-derived and simulated glyoxal/formaldehyde column results for validating the reliability of aromatic emission estimates and simulated concentration patterns, thereby establishing a foundation for investigating the impact of aromatics on ozone and secondary organic aerosol (SOA) formation—this study developed a tracer method within the CMAQ model to track aromatic-derived dicarbonyl compounds and aqueous-phase SOA.

The specific approach involves rewriting the relevant aromatic species and reactions into independent equations for each component. Taking the reaction between glyoxal (GLY) and OH radicals as an example (Equation (1)):



In the model, glyoxal (GLY) reacts with OH radicals to form peroxy radical (HO₂), carbon monoxide (CO), and peroxyacyl radical (HCOCO₃), with the reaction rate $K_{\text{GLY,OH}}$ being a temperature-dependent function. The change in glyoxal concentration over a simulation timestep can be expressed by Equation (2):

$$-\frac{\partial \text{GLY}}{\partial t} = K_{\text{GLY,OH}} \times C_{\text{GLY}} \times C_{\text{OH}} \quad (2)$$

where C_{GLY} and C_{OH} represent the concentrations of glyoxal and OH radicals, respectively. To trace aromatic-derived contributions, we introduced tagged species for aromatics and primary emissions into the SAPRC07tic mechanism. Thus, the glyoxal concentration can be decomposed into $\text{GLY} = \text{GLY}_{\text{aro}} + \text{GLY}_{\text{pri}} + \text{GLY}_{\text{other}}$. Accordingly, Equation (2) can be rewritten as Equation (3):

$$-\frac{\partial (\text{GLY}_{\text{aro}} + \text{GLY}_{\text{pri}} + \text{GLY}_{\text{other}})}{\partial t} = K_{\text{GLY,OH}} \times (C_{\text{GLY}_{\text{aro}}} + C_{\text{GLY}_{\text{pri}}} + C_{\text{GLY}_{\text{other}}}) \times C_{\text{OH}} \quad (3)$$

If the tagged species are treated as independent chemical entities, their reactions can be expressed as Equations (4)–(9):



$$-\frac{\partial \text{GLY}_{\text{aro}}}{\partial t} = K_{\text{GLY,OH}} \times C_{\text{GLY}_{\text{aro}}} \times C_{\text{OH}} \quad (5)$$



$$-\frac{\partial \text{GLY}_{\text{pri}}}{\partial t} = K_{\text{GLY,OH}} \times C_{\text{GLY}_{\text{pri}}} \times C_{\text{OH}} \quad (7)$$



$$-\frac{\partial \text{GLY}_{\text{other}}}{\partial t} = K_{\text{GLY,OH}} \times C_{\text{GLY}_{\text{other}}} \times C_{\text{OH}} \quad (9)$$

If the sum of the split reactions equals the original unsplit reaction (i.e., Equation (3)), the equations can be decomposed. In atmospheric chemistry models, species concentrations remain constant within a single timestep, and gas-phase reactions occur simultaneously. The model completes all calculations for the current timestep before synchronously updating all species concentrations for the next step. Therefore, C_{OH} in Equations (5), (7), and (9) remains identical. Thus, the sum of Equations (4), (6), and (8) can be expressed as Equation (10):

$$\begin{aligned} -\frac{\partial \text{GLY}_{\text{aro}} + \partial \text{GLY}_{\text{pri}} + \partial \text{GLY}_{\text{other}}}{\partial t} &= K_{\text{GLY,OH}} \times (C_{\text{GLY}_{\text{aro}}} + C_{\text{GLY}_{\text{pri}}} + C_{\text{GLY}_{\text{other}}}) \times C_{\text{OH}} \\ &= -\frac{\partial (\text{GLY}_{\text{aro}} + \text{GLY}_{\text{pri}} + \text{GLY}_{\text{other}})}{\partial t} \end{aligned} \quad (10)$$

This confirms that GLY can be decomposed into contributions from aromatics GLY_{aro} , primary emissions GLY_{pri} , and other sources $\text{GLY}_{\text{other}}$. In the gas-phase mechanism, the tagged species replace the original species in calculations. We extended this tagging and splitting approach to all species involved in the formation of dicarbonyl compounds (glyoxal and methylglyoxal) from aromatics (Figure 2), modifying a total of 110 chemical reactions in the original SAPRC07tic mechanism (Table S2). This led to the development of the SAPRC07tic-AROttoGLY mechanism, enabling the model to track the contributions of aromatic-derived glyoxal and methylglyoxal to the final concentrations of dicarbonyl compounds and aqueous-phase SOA during simulations.

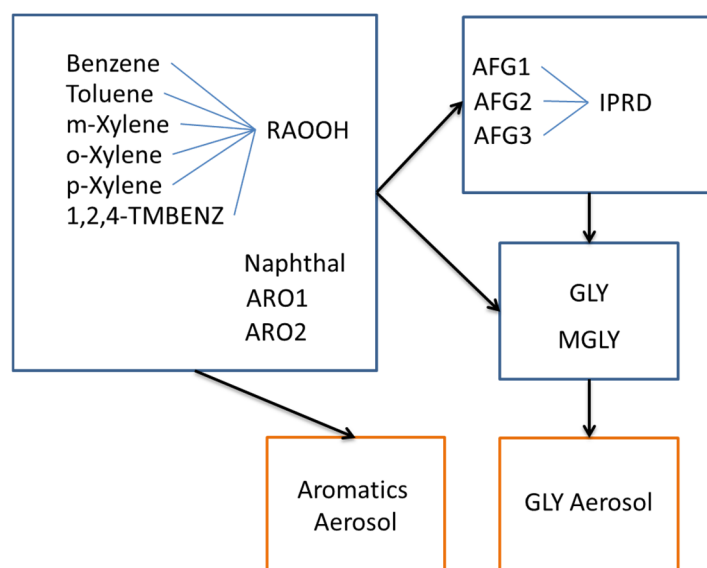


Figure 2. Schematic diagram of tracer-tagged species and processes for aromatic-derived dicarbonyl compounds.

2.2. Observation Data

2.2.1. In Situ Observation

An observational campaign was conducted at the Heshan Supersite in autumn 2014, during which 79 VOCs were sampled and analyzed to obtain hourly concentration data. Valid measurements spanned from 00:00 on 22 October to 08:00 on 20 November 2014. Located southwest to Guangzhou and Foshan at [112.93°E, 22.72°N] in the Pearl River Delta core region, the Heshan Supersite lies downwind of the Guangzhou–Foshan area during autumn and winter. Within the PRD, Guangzhou and Foshan urban centers exhibit high population density with intensive residential and industrial activities, resulting in substantial pollutant emissions [22]. Autumn and winter seasons also witness the most frequent and severe pollution episodes in the PRD [23,24]. Situated in a suburban downwind

location with no significant local anthropogenic emission sources, the Heshan Supersite serves as an ideal platform for investigating regional pollutant transport and chemical processes in the PRD.

The observed VOC species during the campaign are listed in Table S1. Since air quality models typically classify VOCs by functional groups rather than individual species, direct comparison between modeled and observed species requires regrouping the 79 measured VOCs according to the model's chemical mechanism. Among these, 15 species are explicitly simulated in the model: ethene, propene, isoprene, acetylene, benzene, toluene, m/p-xylene, o-xylene, 1,2,4-trimethylbenzene, formaldehyde, acetaldehyde (CCHO), acrolein, 2-methylacrolein (MACR), acetone, and methyl vinyl ketone (MVK). The lumped species and their corresponding measured compounds are mapped in Table S1.

2.2.2. Satellite Data

In July 2004, NASA launched the AURA satellite, the third primary spacecraft in the Earth Observing System (EOS) following the well-known TERRA and AQUA satellites. The AURA satellite carries a sensor of Ozone Monitoring Instrument (OMI), which was developed by Dutch Space (Noordwijk, The Netherlands) and TNO Science & Industry (The Hague, The Netherlands). The OMI instrument was dedicated to atmospheric chemistry research.

The European Space Agency (ESA) and EUMETSAT (European Organisation for the Exploitation of Meteorological Satellites) launched the MetOp satellite series to detect atmospheric physical parameters and vertical distributions of ozone and other trace gases. Three satellites, MetOp-A, MetOp-B, and MetOp-C, were successively launched in October 2006, September 2012, and November 2018, respectively. All MetOp satellites are equipped with the Global Ozone Monitoring Experiment-2 (GOME-2) instrument, which was developed by Selex ES in Florence, Italy, for atmospheric chemistry observations. Originally, each successor satellite was intended to replace its predecessor to maintain continuous observations over more than a decade. The GOME-2 observations from MetOp-A and MetOp-B since 2007 serve as crucial historical references for this study.

OMI covers a spectral range of 270–500 nm, while GOME-2 observes from 240 to 790 nm (details available at <https://science.nasa.gov/mission/aura/omi/> and http://www.esa.int/Our_Activities/Observing_the_Earth/Meteorological_missions/MetOp/About_GOME-2, accessed on 8 September 2025). The tropospheric formaldehyde column data employed in this study were retrieved and produced by the Royal Belgian Institute for Space Aeronomy (BIRA-IASB) using OMI observations [25,26]. Similarly, the glyoxal (GLY) column dataset was retrieved and generated by the Institute of Environmental Physics at the University of Bremen [27]. Both satellite-derived products (HCHO and GLY) are Level 3 data with a spatial resolution of $0.25^\circ \times 0.25^\circ$. To facilitate comparative analysis with model simulations, we performed bilinear interpolation to regrid the satellite observations to match the model's grid resolution.

3. Results

3.1. Model Validation

An intensive VOC measurement campaign was conducted at the Heshan supersite in the Pearl River Delta from 22 October to 20 November 2014. Located downwind of the Guangzhou–Foshan metropolitan area in the PRD core region, this site provided ideal conditions for studying urban pollution transport. We matched model-simulated species with observed species according to the correspondence shown in Table S1 to validate model performance.

Figure 3 presents the diurnal variation characteristics of observed versus simulated concentrations for aromatic hydrocarbons and key species. The model shows varying performance across different species: while toluene simulations agree well with observations (Figure 3d) and the formaldehyde diurnal trend is reasonably captured (Figure 3b), benzene and 1,2,4-trimethylbenzene are notably overestimated (Figure 3c,g). It should be noted that the observational dataset covers fewer species than the model represents—for instance, model species ARO1 and ARO2 encompass broader chemical groups, resulting in higher simulated concentrations compared to the observations (Figure 3h,i).

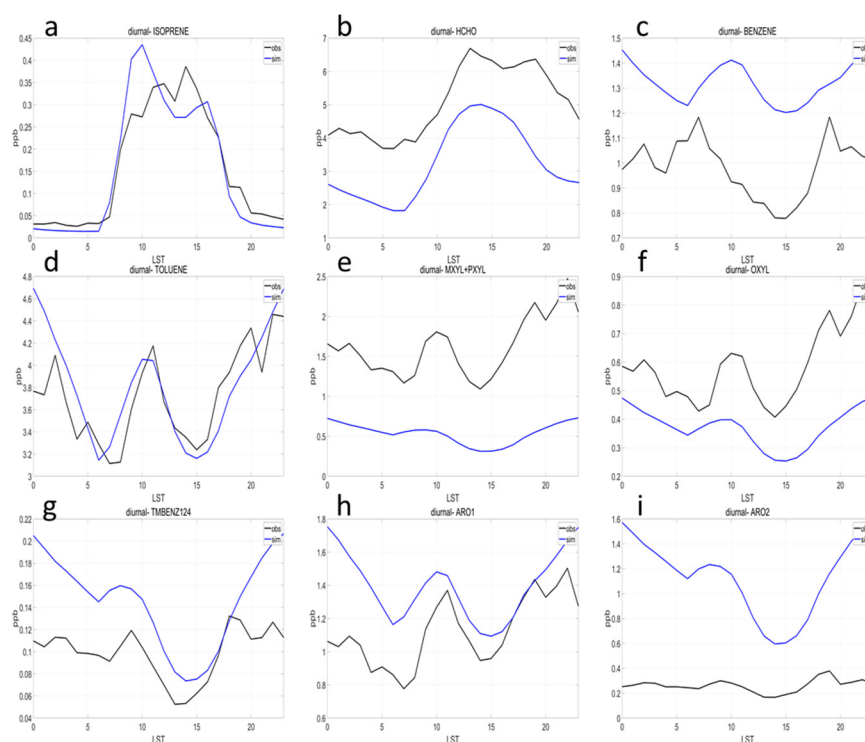


Figure 3. Comparison of observed (black lines) and simulated (blue lines) diurnal variations in VOC species at Heshan supersite during the autumn 2014 campaign: (a) isoprene, (b) formaldehyde, (c) benzene, (d) toluene, (e) m/p-xylene, (f) o-xylene, (g) 1,2,4-trimethylbenzene, (h) ARO1, (i) ARO2.

As shown in Figure 3a and Figure S1a–d, the model demonstrates good agreement with observations for isoprene (ISOP), methacrolein (MACR), methyl vinyl ketone (MVK), and ethene (ETHE). Both diurnal patterns and hourly time series indicate the model’s capability in reproducing these species’ characteristics. Isoprene, as one of the typical biogenic volatile organic compounds (BVOCs) representative, serves as an important indicator. Its accurate simulation suggests reasonable representation of both BVOC emissions and meteorological fields in the model. MACR and MVK, primarily originating from the oxidation of isoprene and terpenes, provide crucial validation for the model’s gas-phase chemical mechanism. Ethene, being an important ozone precursor, possesses high ozone formation potential (OFP).

The successful simulation of these non-aromatic VOCs (isoprene, MACR, MVK, and ethene) helps minimize errors from non-aromatic sources, thereby establishing a more reliable foundation for the subsequent analysis of aromatic hydrocarbons’ impact on ozone formation.

Table S3 shows the comparison between model simulations (SIM) against air quality monitoring network observations (OBS), revealing varying levels of agreement across different pollutants and regions, as assessed by the Index of Agreement (IOA) and Mean Bias (MB). For ozone, the model demonstrates reasonable performance, with IOA values

ranging from 0.78 to 0.89, indicating moderate to strong agreement, particularly in the BTH region (IOA: 0.80–0.88) and the PRD (IOA: 0.79–0.89). However, the model systematically overestimates O_3 concentrations, as evidenced by positive MB values (e.g., -0.5 to $+28.5 \mu\text{g m}^{-3}$ in BTH and $+4.1$ to $+24.6 \mu\text{g m}^{-3}$ in PRD), except for the YRD where a significant underestimation ($-21.7 \mu\text{g m}^{-3}$) occurs. For nitrogen dioxide (NO_2), agreement is weaker (IOA: 0.59–0.85), with notable overestimation in urban areas like YRD (MB: $+8.0$ to $+10.2 \mu\text{g m}^{-3}$) and BTH (MB: $+0.6$ to $+6.5 \mu\text{g m}^{-3}$), while the regional average of Eastern China shows slight underestimation (MB: -9.1 to $-2.4 \mu\text{g m}^{-3}$). The most pronounced discrepancies occur for $PM_{2.5}$, where IOA values are lower (0.48–0.76), reflecting moderate agreement. The model consistently underestimates concentrations across all regions, particularly in BTH (MB: -43.9 to $-23.5 \mu\text{g m}^{-3}$) and PRD (MB: -25.4 to $-11.4 \mu\text{g m}^{-3}$). These results suggest that while the model captures general trends for O_3 and NO_2 , its performance for $PM_{2.5}$ is less reliable, likely due to challenges in representing emission sources or aerosol processes. For example, the fugitive dust emission, which has a large impact in BTH, is not fully incorporated in the model system yet. Further refinement is needed to reduce biases, especially for particulate matter simulations in high-pollution regions.

3.2. Formaldehyde and Glyoxal Column Concentrations

To facilitate comparison with satellite data, we calculated the vertical column density (VCD) of simulated formaldehyde and glyoxal using Equation (11) [28]:

$$\text{VCD} = \sum_{l=1}^{24} \frac{P_l \times Z_l \times A \times C_{ppb,l}}{R \times T_l \times 10^{13}} \quad (11)$$

where $C_{ppb,l}$ represents the volume mixing ratio (in ppbv) of formaldehyde or glyoxal at the L -th model layer. P_l , Z_l , and T_l denote the pressure (Pa), thickness (m), and temperature (K) of the L -th layer, respectively. A is Avogadro's constant (6.02×10^{23}), and R represents the gas constant ($8.314 \text{ J} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$). The integration is performed from the first layer to the topmost layer, with the WRF–CMAQ model employing 24 vertical layers in this study. To align with satellite overpass times (OMI at 13:45 local time; GOME-2 at 09:30 local time), model outputs at 14:00 and 10:00 local time were used for comparative analysis, respectively. Unless otherwise specified, all comparisons in this chapter are based on the 14:00 model results versus OMI observations, with GOME-2 comparisons being explicitly noted when used.

Formaldehyde serves as a crucial intermediate in atmospheric photochemical processes. In anthropogenically influenced regions, HCHO primarily originates from (1) oxidation of biogenic volatile organic compounds such as isoprene and terpenes, and (2) oxidation of anthropogenic VOCs including alkenes and aromatics. This makes HCHO an excellent indicator for both VOC concentration levels and photochemical activity. Research indicates that aromatics contribute significantly to annual HCHO production in highly urbanized regions. For instance, in the Pearl River Delta region, aromatics accounted for approximately 19% of total chemical HCHO production in 2006 [12]. However, uncertainties remain in aromatic gas-phase mechanisms [29], and their contribution to HCHO exhibits seasonal variations. Typically, in the summer, reduced sensitivity of HCHO to aromatics is observed due to dominant BVOC emissions (mainly isoprene); in the winter, aromatics account for about 8% of total VOC sensitivity for HCHO [30].

Figure 4 presents the spatial distribution of formaldehyde column concentrations over China from 2010 to 2016, as retrieved from OMI satellite observations and estimated by WRF–CMAQ simulations. The summer (July) average column concentrations exhibit a pronounced increasing trend across North China, Central China, East China, and Southwest China during 2008–2014. After 2014, regional divergence emerged with continued modest

increases in the BTH and YRD regions, while Central China showed no significant trend. Autumn (October) HCHO columns are systematically lower than summer values over Eastern China except in PRD, with a distinct spatial pattern featuring high-value zones extending from southern BTH through the Henan–Shandong border region to Central China and further southward to South China. Similarly to summer trends, autumn HCHO columns demonstrate an overall increasing trend since 2008.

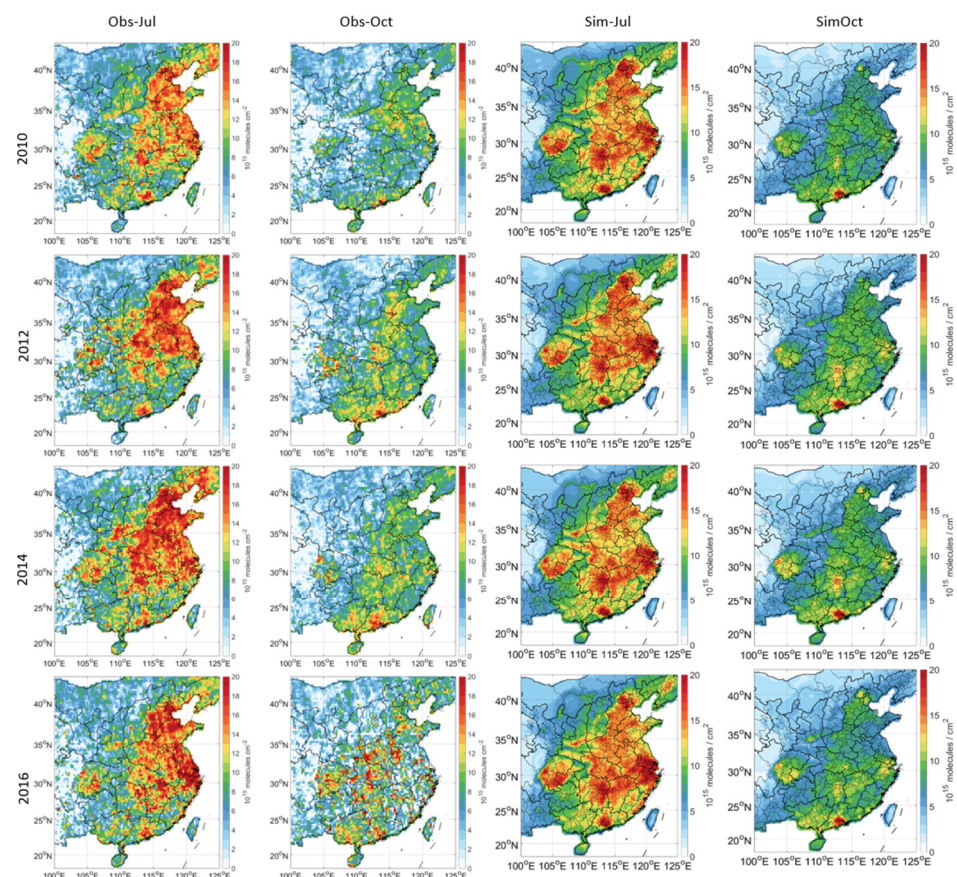


Figure 4. Spatial distributions of OMI satellite-retrieved and modeled formaldehyde column density during summer (July) and autumn (October) of year 2010, 2012, 2014, and 2016.

Table 1 illustrates the interannual variations in HCHO columns in three key Eastern China regions and major cities, with quality control measures excluding values below the detection limit of 2.5×10^{15} molecules cm^{-2} [13]. The 2010 Asian Games in Guangzhou prompted substantial emission control measures, likely contributing to the record-low HCHO columns observed in the Pearl River Delta (PRD) during summer and autumn of that year. Summer HCHO levels show minimal differences among the three key regions before 2014, though post-2014 data reveals higher columns in YRD and PRD compared to BTH's. Autumn exhibits elevated concentrations in PRD relative to YRD and BTH, consistent with PRD's severe autumn photochemical pollution season when peak ozone levels typically occur. The synchronous behavior of HCHO columns and ozone pollution are particularly evident during autumn.

Model simulations for 2010–2016 successfully capture the spatial distribution of summer HCHO columns, accurately reproducing high-concentration zones in BTH, YRD, PRD, Central China, and the Chengdu–Chongqing region. Summer averages across Eastern China show modest interannual variability (13.0 – 13.6×10^{15} molecules cm^{-2}) with an overall trend of $+0.05 \times 10^{15}$ molecules $\text{cm}^{-2} \text{yr}^{-1}$. Regional July averages are 12.8 , 15.2 , and 13.5×10^{15} molecules cm^{-2} for BTH, YRD, and PRD, respectively, with trends of -0.40 ,

+0.42, and $+0.21 \times 10^{15}$ molecules $\text{cm}^{-2} \text{yr}^{-1}$, generally consistent with OMI observations. Autumn simulations reproduce the broad spatial patterns but underestimated Southern BTH and Central China while overestimating PRD concentrations (12.75 vs. observed 11.93×10^{15} molecules cm^{-2}). The Eastern China autumn average is 8.3×10^{15} molecules cm^{-2} ($+0.14 \times 10^{15}$ molecules $\text{cm}^{-2} \text{yr}^{-1}$), with regional values of 6.6, 8.7, and 12.7×10^{15} molecules cm^{-2} for BTH, YRD, and PRD. While the model captures BTH and YRD autumn trends reasonably well, it produced an opposite trend in PRD ($+0.14$ vs. observed -0.06×10^{15} molecules $\text{cm}^{-2} \text{yr}^{-1}$).

Table 1. Observed and modeled VCDs of formaldehyde and glyoxal.

		OMI-Obs HCHO VCD (10^{15} Molecules cm^{-2})	Modeled HCHO VCD (10^{15} Molecules cm^{-2})	Difference	OMI-Obs GLY VCD (10^{14} Molecules cm^{-2})	Modeled GLY VCD (10^{14} Molecules cm^{-2})	Difference
2010-Jul	BTH	14.49	13.74	−0.75	4.03	2.75	−1.28
	YRD	12.72	14.76	2.04	4.24	2.49	−1.75
	PRD	9.51	13.46	3.95	3.50	2.93	−0.57
2012-Jul	BTH	11.73	12.46	0.73	3.20	2.48	−0.72
	YRD	12.93	15.13	2.20	3.89	3.04	−0.85
	PRD	12.38	12.83	0.45	3.81	2.73	−1.08
2014-Jul	BTH	13.14	12.65	−0.49	3.84	2.82	−1.02
	YRD	14.45	14.79	0.34	4.03	3.19	−0.84
	PRD	11.79	13.82	2.03	3.51	3.44	−0.07
2016-Jul	BTH	12.31	12.36	0.05	-	2.62	-
	YRD	16.03	16.28	0.25	-	3.58	-
	PRD	12.54	13.85	1.31	-	3.49	-
Avg-Juls	BTH	12.92	12.80	−0.12	3.69	2.68	−1.01
	YRD	14.03	15.24	1.21	4.05	2.91	−1.15
	PRD	11.56	13.49	1.94	3.61	3.03	−0.57
2010-Oct	BTH	7.16	5.94	−1.22	3.61	1.12	−2.49
	YRD	8.54	8.20	−0.34	4.11	1.85	−2.26
	PRD	10.34	12.11	1.77	5.63	2.91	−2.72
2012-Oct	BTH	7.56	6.08	−1.48	3.56	1.25	−2.31
	YRD	8.16	9.33	1.17	3.91	2.33	−1.58
	PRD	12.67	13.08	0.41	5.44	3.36	−2.08
2014-Oct	BTH	7.28	7.27	−0.01	3.42	1.62	−1.80
	YRD	8.64	8.88	0.24	3.71	2.49	−1.22
	PRD	13.30	13.30	0.00	5.21	4.05	−1.16
2016-Oct	BTH	7.85	6.97	−0.88	-	1.74	-
	YRD	8.70	8.44	−0.26	-	1.75	-
	PRD	11.42	12.50	1.08	-	3.43	-
Avg-Octs	BTH	7.46	6.57	−0.90	3.53	1.33	−2.20
	YRD	8.51	8.71	0.20	3.91	2.22	−1.69
	PRD	11.93	12.75	0.81	5.43	3.44	−1.99

Linear regression analysis of 2008–2016 data reveals distinct regional trends. Summer (July) increases averaged 0.06 , 0.17 , and 0.25×10^{15} molecules $\text{cm}^{-2} \text{yr}^{-1}$ for BTH, YRD, and PRD, respectively. Overall BTH concentrations rise annually due to significant increases in surrounding areas like Tianjin and Tangshan. Autumn trends are more muted, with only BTH showing substantial increases (0.10×10^{15} molecules $\text{cm}^{-2} \text{yr}^{-1}$), potentially influenced by the low 2008 baseline following emission controls for Beijing Olympic Games. The YRD exhibited dual autumn peaks in 2009 and 2013.

The seasonal variation in glyoxal column concentrations exhibits smaller amplitude compared to that of formaldehyde columns, as shown in Figure 5. The spatial distribution of high GLY concentrations remains relatively consistent across seasons, with both sum-

mer (July) and autumn (October) satellite observations showing pronounced high-value areas extending from the BTH region to YRD. Continuous hotspots can be found in PRD, particularly during autumn. Multi-year average OMI GLY columns for summer reached 3.46 , 3.97 , and 3.72×10^{14} molecules cm^{-2} in BTH, YRD, and PRD, respectively, while autumn averages are 3.39 , 3.91 , and 5.27×10^{14} molecules cm^{-2} in these regions. The muted seasonal contrast in GLY columns (except in PRD) likely reflects compensating contributions from biogenic and anthropogenic precursors. While BVOCs dominate summer production, AVOCs become more important in autumn. The exceptional autumn enhancement in PRD (5.27 vs. 3.72×10^{14} molecules cm^{-2}) underscores the region's intense autumn photochemical activity and associated pollution severity.

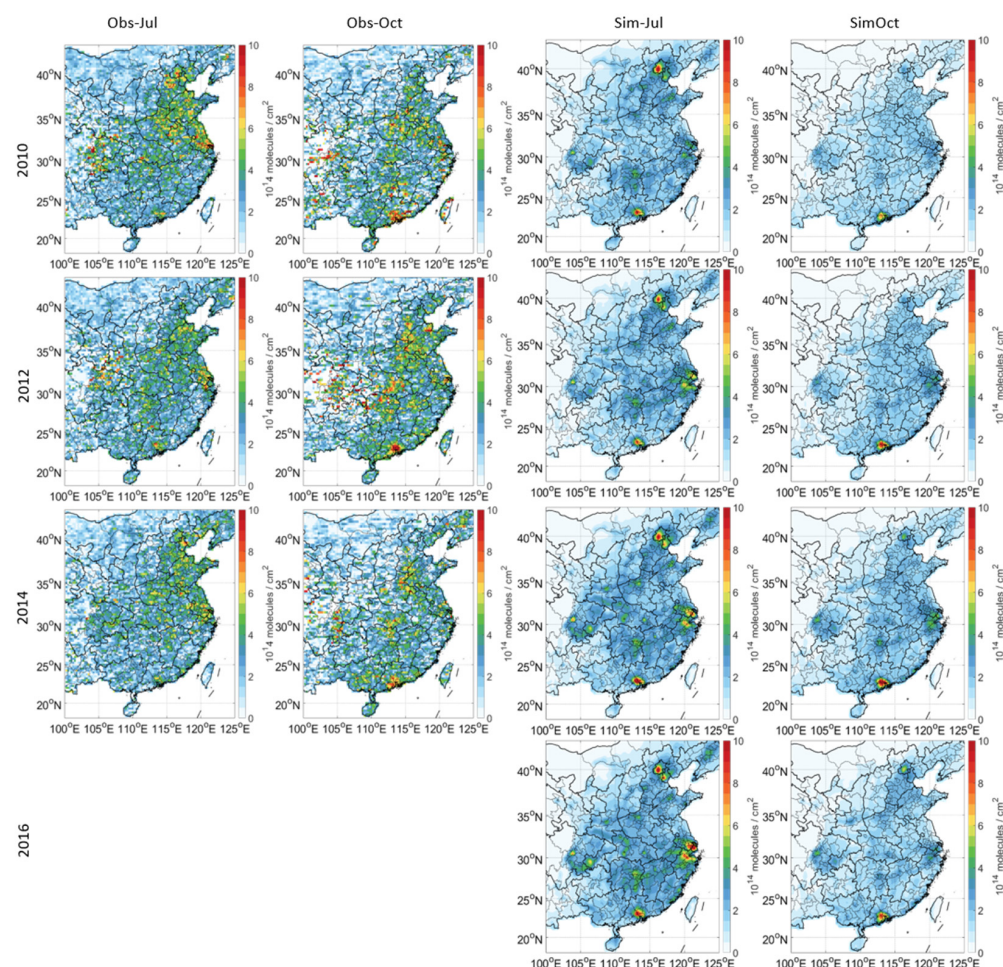


Figure 5. Spatial distributions of OMI satellite-retrieved and modeled glyoxal column density during summer (Jul) and autumn (Oct) of year 2010, 2012, 2014, and 2016.

Quality-controlled column results (excluding values below 2×10^{14} molecules cm^{-2} detection limit) reveal complex interannual trends. Summer OMI GLY columns show modest increases in BTH (0.05×10^{13} molecules $\text{cm}^{-2} \text{yr}^{-1}$) but decreases in YRD (-1.04) and PRD (-1.47×10^{13} molecules $\text{cm}^{-2} \text{yr}^{-1}$), with patterns broadly corroborated by GOME-2A observations (1.03 , -0.84 , -1.32×10^{13} molecules $\text{cm}^{-2} \text{yr}^{-1}$, respectively). Autumn trends diverge more substantially between sensors—OMI recorded 0.37 (BTH), -1.20 (YRD), and -1.41×10^{13} molecules $\text{cm}^{-2} \text{yr}^{-1}$ (PRD)—whereas GOME-2A suggested stronger increases (2.66 , 1.00 , 0.88×10^{13} molecules $\text{cm}^{-2} \text{yr}^{-1}$). These discrepancies, along with the unexpected similarity between morning (09:30) GOME-2A and afternoon (13:45) OMI measurements (despite stronger photochemistry expected later in the day), highlight substantial satellite retrieval uncertainties requiring cautious interpretation.

WRF–CMAQ simulations systematically underestimate GLY columns, particularly in southern BTH, YRD, and Central China during both seasons. Summer modeled averages reach only 2.67 (BTH), 3.08 (YRD), and 3.15×10^{14} molecules cm^{-2} (PRD), with trends of -0.06 , 3.4 , and 2.4×10^{13} molecules $\text{cm}^{-2} \text{yr}^{-1}$, respectively. Autumn underestimations are more severe (1.43 , 2.10 , 3.44×10^{14} molecules cm^{-2}) despite positive trends in BTH (2.25) and PRD (2.24×10^{13} molecules $\text{cm}^{-2} \text{yr}^{-1}$). This seasonal contrast suggests potential BVOC emission uncertainties, particularly given known biases in MODIS-based MEGAN modeling that overestimates coniferous while underestimating broadleaf emissions in Southern China [31]. However, ground observations discussed in Section 3.1 show good model–observation agreement for isoprene and other BVOC oxidation products in PRD during autumn 2014, implying regional heterogeneity in the dominant factors driving GLY underestimations, potentially including aromatic emission uncertainties [32] alongside the noted BVOC issues. The overall model performance presents an intriguing contrast between HCHO overestimations and GLY underestimations, reflecting substantial uncertainties in both gas-phase mechanisms and precursor emissions for these key oxygenated VOCs. The discrepancy between satellite-derived and modeled glyoxal columns aligns with the satellite instruments and the uncertainty in emission inventories, especially the aromatics emission. Glyoxal columns observed by SCIAMACHY are $\sim 60\%$ higher than those by OMI in PRD [12]. Modeled GLY VCD was largely underestimated (even 10–20 folds) than SCIAMACHY observed ones [32]. Chan also intercompared modeled GLY VCD using different emission inventories, finding it underestimated GLY VCD by a factor of 2 in PRD when using the same inventory as Liu’s study, which was INTEx-B for year 2006. Our modeled GLY VCD is $\sim 63\%$ to 67% of the OMI-observed GLY VCD in PRD, which is consistent with those studies. Li et al. [33] found their RAMS–CMAQ system underestimated GLY VCD by a factor of 4–10 in July, especially in BTH, YRD, and Sichuan–Chongqing regions. The emission inventories used in Li et al. [33] was similar with the ones used in our studies, based on MEIC in China mainland but with different version and time periods. We have better performance in July, underestimating $\sim 27\%$, 28% , and 16% in BTH, YRD, and PRD, respectively. But the underestimations in October are much significant. Liu et al. [32] and Li et al. [33] indicated that increasing aromatic emission could partly improve the underestimation of GLY VCD. However, the scale factors of aromatic emissions needed to be very large (e.g., >10) to fit the gap in GLY VCD, meanwhile resulting in the large overestimations in HCHO VCD. In general, the discrepancy aligned in satellite-observed and modeled GLY VCD is largely attributed to the uncertainties in satellite instruments, retrieval algorithms, and emission inventories.

3.3. Contribution of Aromatics to Glyoxal

The systematic underestimation of glyoxal columns in the model simulations consequently lead to reduced simulated ratios of glyoxal to formaldehyde column densities (R_{GF}), as illustrated in Table 2. A pronounced declining trend in R_{GF} is observed across China from 2010 to 2014 during both summer and autumn seasons. However, between 2014 and 2016, the three key regions exhibit varying degrees of declining R_{GF} trends, a pattern that synchronized with both the interannual variations in simulated GLY columns and aromatic emission changes.

Over Eastern China, the multi-year average R_{GF} values for summer and autumn are 0.018 and 0.021, respectively, consistently lower than satellite observations. While the model captures the seasonal pattern of higher autumn R_{GF} values, the magnitude of this seasonal difference is less pronounced than observed. Notably, summer R_{GF} values in Henan and Shandong provinces are lower than surrounding regions, aligning with satellite observations that identify these areas as low- R_{GF} zones. This consistency suggests that

the model's R_{GF} outputs remain useful for comparing dominant precursor types across regions during summer, though requiring adjusted threshold criteria. Our results indicate that applying a modified threshold range of [0.02–0.03] yields conclusions comparable to satellite findings (which used 0.04–0.06): regions like Henan, Shandong, and Northern Anhui with R_{GF} consistently below 0.02 are clearly dominated by anthropogenic emissions.

Table 2. Source of glyoxal VCD and comparison of observed R_{GF} and simulated R_{GF} .

	BTH				YRD			
	GLYaro	GLYpri	obs R_{GF}	sim R_{GF}	GLYaro	GLYpri	obs R_{GF}	sim R_{GF}
2010JUL	0.413	0.025	0.027	0.019	0.513	0.031	0.034	0.015
2012JUL	0.505	0.023	0.027	0.018	0.474	0.026	0.029	0.019
2014JUL	0.461	0.020	0.026	0.021	0.526	0.019	0.028	0.020
2016JUL	0.489	0.016	-	0.020	0.439	0.017	-	0.020
2010OCT	0.725	0.065	0.047	0.018	0.706	0.043	0.049	0.021
2012OCT	0.730	0.059	0.044	0.019	0.691	0.036	0.049	0.024
2014OCT	0.767	0.035	0.042	0.021	0.719	0.029	0.044	0.026
2016OCT	0.804	0.030	-	0.023	0.820	0.035	-	0.019

	PRD				Eastern China			
	GLYaro	GLYpri	obs R_{GF}	sim R_{GF}	GLYaro	GLYpri	obs R_{GF}	sim R_{GF}
2010JUL	0.670	0.014	0.036	0.017	0.453	0.034	0.032	0.017
2012JUL	0.715	0.013	0.032	0.017	0.495	0.029	0.030	0.017
2014JUL	0.673	0.008	0.029	0.020	0.477	0.024	0.028	0.019
2016JUL	0.545	0.008	-	0.021	0.430	0.021	-	0.019
2010OCT	0.665	0.016	0.056	0.022	0.645	0.051	0.047	0.019
2012OCT	0.673	0.011	0.042	0.022	0.636	0.044	0.045	0.021
2014OCT	0.704	0.007	0.040	0.026	0.665	0.034	0.041	0.023
2016OCT	0.703	0.007	-	0.023	0.730	0.033	-	0.020

The central urban areas of BTH, YRD, and PRD display significantly elevated R_{GF} compared to their peripheries, mirroring satellite observations. This spatial pattern cautions against simplistic threshold-based interpretations within metropolitan regions, where intensive AVOC emissions drive both high GLY columns and R_{GF} values. In these hotspots, elevated R_{GF} strongly indicates anthropogenic dominance. To further elucidate the specific VOC components driving GLY columns and R_{GF} patterns, particularly whether aromatics dominate historical GLY variations and contribute to regional R_{GF} differences, more detailed source apportionment analysis is required.

Satellite-derived R_{GF} suffers from amplified uncertainties due to compounding errors from cloud cover, water vapor, aerosol interference, and surface reflectance effects (Figure 6, top). Nevertheless, clear patterns emerge; autumn R_{GF} consistently exceeds summer values across Eastern China, with 2008–2014 multi-year averages of 0.031 (summer) and 0.046 (autumn), both showing a -0.001 yr^{-1} trend. These multi-year averages effectively filter noise, revealing robust spatial patterns (Figure 6, bottom). Summer R_{GF} maxima occurs in YRD and PRD (0.032 and 0.035, respectively), while BTH shows minima (BTH: 0.026). All regional values remain below 0.04, consistent with Vrekoussis et al. (2010) [13]. The spatial contrast between high-NO_x regions (BTH/YRD with lower R_{GF}) and PRD (higher R_{GF}) is particularly striking in summer. Autumn spatial patterns differed markedly, with high- R_{GF} zones covering Beijing/Southern BTH (0.049), Northern/Central YRD (0.054), and PRD (0.050). This autumn homogenization of R_{GF} values across regions, along with their spatial correspondence with aromatic emission and concentration patterns, suggests stronger aromatic influences during autumn.

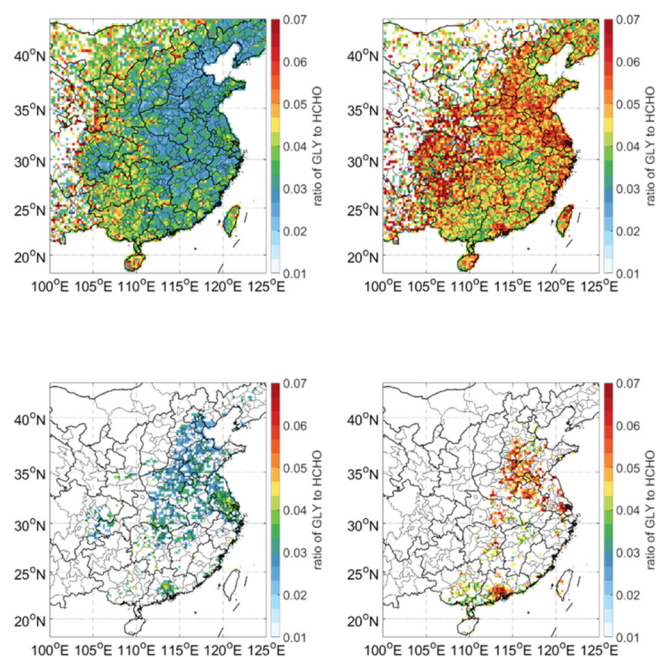


Figure 6. Ratio of glyoxal to formaldehyde column densities (R_{GF}) from OMI satellite observations (2008–2014) in July (**left** panel) and October (**right** panel): (**Top**) 7-year average R_{GF} ; (**bottom**) grid cells retaining valid values across all 7 years.

The ozone formation potential (OFP) is utilized for distinguishing if a region is dominated by anthropogenic or biogenic precursors. OFP of individual VOC species in the model system is derived by their emission and the maximum incremental reactivity (MIR) listed in Table S4. The OFP of aromatics and BVOC, as well as the total OFP, are shown in Figure S2. The results indicate that regions with relatively high proportions of anthropogenic OFP in Eastern China include BTH, Shandong, Henan, Central and Southern YRD, Fujian, Jiangxi, PRD, and Chengdu–Chongqing, among others. This spatial distribution of anthropogenic/biogenic sources aligns well with the satellite-observed spatial characteristics of R_{GF} (Figure 6). Therefore, the localized R_{GF} thresholds are identified by distinguishing R_{GF} between highly anthropogenic influenced regions and biogenic-dominated regions.

The spatial distribution of regions with high primary glyoxal emission contribute slightly to column concentrations (Figure 7), notably in Jiangsu and Anhui provinces. This attenuation stems from the characteristic emission profile of primary GLY sources, concentrated near the surface where emissions undergo rapid depletion through gas-phase chemical reactions and deposition processes, creating a net consumption effect at higher altitudes. In contrast, GLY produced from BVOCs and AVOCs oxidation provides positive contributions to concentrations throughout the atmospheric column. Consequently, vertical integration diminishes primary emissions' relative contribution to column densities compared to surface concentrations. The correlation between GLY columns and emissions of aromatics and primary GLY is presented in Figure S2. The OMI-observed GLY columns shown in Figure S2 do not exclude values smaller than the detection limit of 2×10^{14} molecules cm^{-2} . Thus, the regional averages are necessarily smaller than those presented in Table 1. It can be seen that the modeled GLY columns in the three key regions are more consistent with the non-QCed OMI columns, with smaller underestimation. The primary GLY emissions have been consistently declining since 2008 according to the MEIC inventory. However, the aromatics emissions slightly increased during 2008–2014. Thus, the simulated GLY columns increased accordingly from 2010 to 2014. Although the simulated GLY columns slightly decreased from 2014 to 2016, attributing to the decline of aromatics and

primary glyoxal emissions, the ratio of aromatic-GLY to total GLY column kept increasing from 2014 to 2016, especially in autumn month.

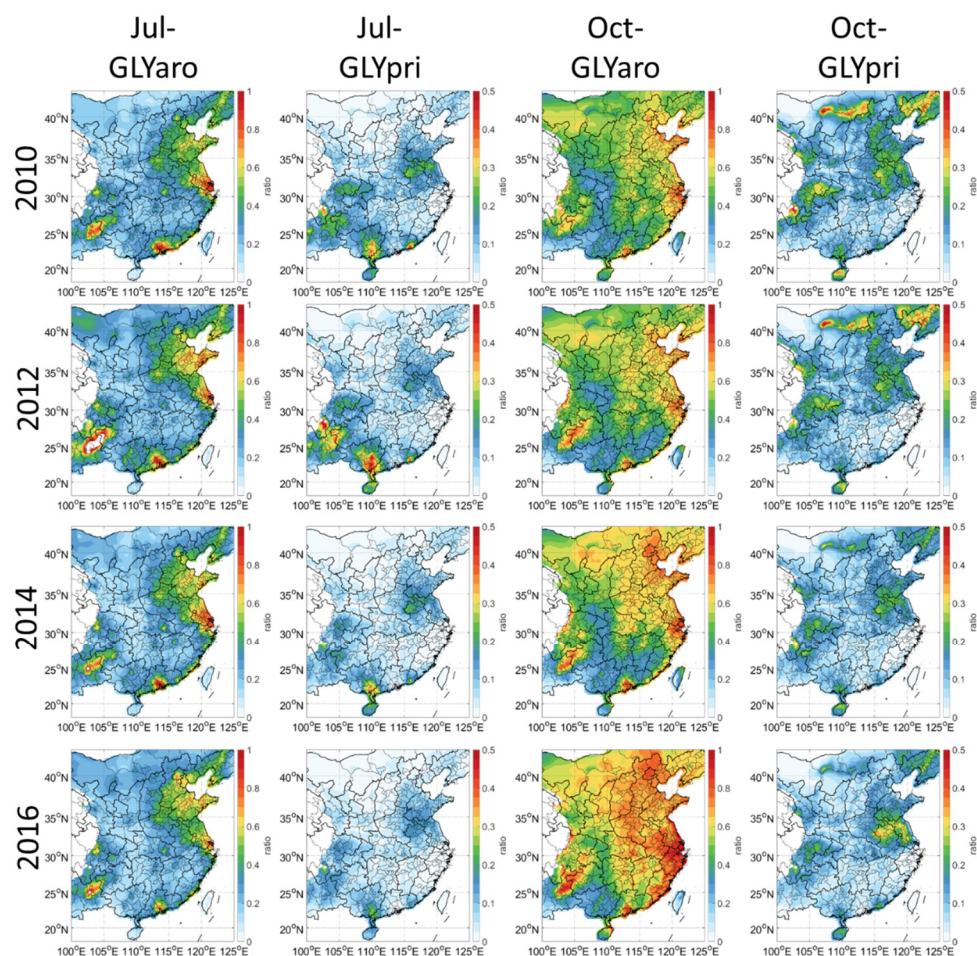


Figure 7. Simulated source contributions to glyoxal vertical column densities in July (summer) and October (autumn): the ratio (ranging 0–1) of aromatic-glyoxal to bulk glyoxal (GLYaro); the ratio (ranging 0–1) of glyoxal from primary glyoxal emissions to bulk glyoxal (GLYpri).

This vertical distribution effect also amplifies the spatial extent and proportional contribution of aromatics to GLY columns relative to surface impacts. During summer, elevated BVOC emissions further reduce aromatics' relative contribution compared to autumn conditions. Notably, the high-contribution areas of aromatics to GLY columns align remarkably with satellite-observed R_{GF} hotspots: concentrated in YRD and PRD during summer, while expanding across BTH, YRD, and PRD in autumn.

The convergence between model results and satellite data strongly suggests aromatics as a major driver of GLY column concentrations, particularly in urban clusters and during autumn when R_{GF} maxima coincide with aromatic contribution peaks. As previously discussed, the declining R_{GF} trends observed across regions may reflect growing anthropogenic influences over time. This interpretation is reinforced by the concurrent rise in aromatics' proportional contribution to GLY columns, a particularly telling correlation given that aromatics derive almost exclusively from anthropogenic emission sources. The combined evidence positions aromatic hydrocarbons as both a dominant influence on spatiotemporal GLY patterns and a sensitive indicator of anthropogenic impacts on atmospheric composition.

4. Conclusions

This study developed a tracer method within the CMAQ air quality modeling system to quantitatively assess the contribution of aromatic hydrocarbons to oxygenated volatile organic compound (OVOC) intermediates, specifically glyoxal (GLY) and methylglyoxal (MGLY). Using the WRF–CMAQ modeling framework, simulations were conducted for typical summer and autumn months in 2010, 2012, 2014, and 2016 over Eastern China. By integrating field observations, satellite data, and numerical modeling results, the historical trends of aromatic hydrocarbons and their impacts in Eastern China were systematically analyzed:

- (1) The aromatic tagging method was implemented based on the SAPRC07tic chemical mechanism and AERO6i aerosol module in CMAQ, enabling quantitative diagnosis of aromatic contributions to dicarbonyl concentrations. Multi-year emission inventories from 2010 to 2016 were incorporated into the simulations. Model validation against ground-based VOC measurements in autumn 2014 demonstrated robust performance of the chemical mechanism, while simulated formaldehyde and glyoxal column concentrations showed good spatiotemporal consistency with satellite observations.
- (2) The results revealed a significant increase in aromatic emissions and ambient concentrations since 2010, with aromatics playing a dominant role in shaping glyoxal column density hotspots across three key regions: the Beijing–Tianjin–Hebei (BTH) metropolitan area, Yangtze River Delta (YRD), and Pearl River Delta (PRD). Analysis of the satellite-derived R_{GF} provided quantitative insights into regional and seasonal variations in aromatic influences. The study also evaluated the applicability of traditional R_{GF} threshold methods for precursor classification in Eastern China. During summer, high- R_{GF} areas outside urban cores aligned well with BVOC-dominated regions, while low- R_{GF} zones (e.g., Southern BTH, Shandong, Northern YRD) reflected anthropogenic dominance. In contrast, autumn exhibited uniformly elevated R_{GF} values across Eastern China, corresponding to anthropogenic source regions and areas with substantial aromatic contributions. By combining satellite retrievals and numerical modeling, this work refined the appropriate R_{GF} threshold ranges for summer and autumn in Eastern China.
- (3) The WRF–CMAQ system effectively reproduced the observed spatiotemporal patterns of HCHO and GLY columns, particularly capturing interannual trends and regional HCHO levels with high fidelity. However, simulated GLY columns were systematically underestimated, with greater discrepancies in autumn compared to summer. While the model successfully replicated GLY spatial characteristics in PRD, significant differences emerged between simulated and observed R_{GF} values due to the pervasive GLY underestimation. Adjusted threshold ranges ([0.015–0.03] for model simulations versus [0.04–0.06] for satellite observations) improved summer precursor classification, though autumn R_{GF} thresholds remained unreliable for model applications.
- (4) Tracer-based analysis quantified the contributions of aromatics and primary emissions to both surface and column GLY concentrations. The results showed increasing aromatic contributions to GLY levels over time, highlighting their growing importance in atmospheric chemical processes and potential impacts on photochemical pollution and aerosol formation. In contrast, primary GLY emissions exhibited declining contributions due to sustained emission controls. Aromatics demonstrated stronger influences on GLY columns in urban areas than those in rural areas, spatially coincident with R_{GF} hotspots. These findings establish aromatics as a key driver of R_{GF} anomalies and a critical factor in secondary pollution formation.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs17183174/s1>, Figure S1: Time series of observed (black dots) and simulated (blue lines) VOC concentrations at Heshan supersite in 2014: (a) isoprene, (b) methacrolein, (c) methyl vinyl ketone, (d) ethene, (e) benzene, (f) toluene, (g) o-xylene, (h) ARO1; Figure S2. Spatial Distribution of Average OFP (10^N tons month⁻¹) from 2010 to 2016: (a) Total OFP in July, (b) Contribution ratio of aromatics to OFP in July, (c) Contribution ratio of biogenic BVOCs to OFP in July, (d) Total OFP in October, (e) Contribution ratio of aromatics to OFP in October, (f) Contribution ratio of biogenic BVOCs to OFP in October; Figure S3. October average VCD of OMI-observed (purple bars), simulated GLY (green bars), simulated aromatic-GLY (cyan bars), and annual emission of aromatics (blue curves) and glyoxal (red curves) in BTH (top), YRD (middle), and PRD (bottom); Table S1: Measured VOCs at Heshan supersite and their mapping to model species; Table S2. Chemical mechanism of revised SAPRC07tic with tagged aromatics; Table S3. Model performance evaluation against air quality monitoring network; Table S4. The MIR of SAPRC07 mechanism species.

Author Contributions: Conceptualization, X.C. (Xiaoyang Chen) and Q.F.; methodology, X.C. (Xiaoyang Chen), X.C. (Xi Chen) and Y.L.; software, X.C. (Xiaoyang Chen) and C.S.; validation, X.C. (Xiaoyang Chen), C.S., and Y.L.; formal analysis, X.C. (Xiaoyang Chen), X.C. (Xi Chen), Y.L., C.S., and S.D.; investigation, X.C. (Xiaoyang Chen), X.C. (Xi Chen), and C.S.; resources, T.D., X.D., S.F., and H.H.; data curation, X.C. (Xiaoyang Chen), X.C. (Xi Chen), Q.F., and S.F.; writing—original draft preparation, X.C. (Xiaoyang Chen), and X.C. (Xi Chen); writing—review and editing, X.C. (Xiaoyang Chen), X.C. (Xi Chen), Y.L., C.S., and S.D.; visualization, X.C. (Xiaoyang Chen); supervision, Q.F.; project administration, Q.F.; funding acquisition, T.D. and X.C. (Xiaoyang Chen) All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Technologies Research and Development Program of China (2023YFC3709201), Guangdong Basic and Applied Basic Research Foundation (2024A1515510025, 2023A1515110103), National Natural Science Foundation of China (42405194, 42375182), Science and Technology Planning Project of Guangzhou (2025A04J4711), Science and Technology Project of Guangdong Meteorological Service (GRMC2023Q03), and Research Fund Program of Guangdong Provincial Key Laboratory of Environmental Pollution Control and Remediation Technology (2023B1212060016).

Data Availability Statement: The data presented in this study are available on request from Xiaoyang Chen (chenxiaoyang@gd121.cn).

Acknowledgments: We give thanks for the support from the high-performance computing platform of Guangdong Meteorological Service, and the high-performance grid-computing platform of School of Atmospheric Sciences, Sun Yat-sen University.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wang, T.; Xue, L.; Brimblecombe, P.; Lam, Y.F.; Li, L.; Zhang, L. Ozone pollution in China: A review of concentrations, meteorological influences, chemical precursors, and effects. *Sci. Total Environ.* **2016**, *575*, 1582–1596. [\[CrossRef\]](#)
2. Liu, Y.; Wang, T. Worsening urban ozone pollution in China from 2013 to 2017—Part 1: The complex and varying roles of meteorology. *Atmos. Chem. Phys.* **2020**, *20*, 6305–6321. [\[CrossRef\]](#)
3. Lu, X.; Zhang, L.; Wang, X.; Gao, M.; Li, K.; Zhang, Y.; Yue, X.; Zhang, Y. Rapid Increases in Warm-Season Surface Ozone and Resulting Health Impact in China since 2013. *Environ. Sci. Technol. Lett.* **2020**, *7*, 240–247. [\[CrossRef\]](#)
4. Li, K.; Jacob, D.J.; Liao, H.; Shen, L.; Zhang, Q.; Bates, K.H. Anthropogenic drivers of 2013–2017 trends in summer surface ozone in China. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 422–427. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Liu, Y.; Wang, T. Worsening urban ozone pollution in China from 2013 to 2017—Part 2: The effects of emission changes and implications for multi-pollutant control. *Atmos. Chem. Phys.* **2020**, *20*, 6323–6337. [\[CrossRef\]](#)
6. Wang, S.; Zhao, M.; Xing, J.; Wu, Y.; Zhou, Y.; Lei, Y.; He, K.; Fu, L.; Hao, J. Quantifying the air pollutants emission reduction during the 2008 olympic games in Beijing. *Environ. Sci. Technol.* **2010**, *44*, 2490–2496. [\[CrossRef\]](#)
7. Jin, X.; Holloway, T. Spatial and temporal variability of ozone sensitivity over China observed from the Ozone Monitoring Instrument. *J. Geophys. Res. Atmos.* **2015**, *120*, 7229–7246. [\[CrossRef\]](#)

8. Wang, Y.; Zhao, Y.; Liu, Y.; Jiang, Y.; Zheng, B.; Xing, J.; Liu, Y.; Wang, S.; Nielsen, C.P. Sustained emission reductions have restrained the ozone pollution over China. *Nat. Geosci.* **2023**, *16*, 967–974. [\[CrossRef\]](#)
9. Liu, Y.; Geng, G.; Cheng, J.; Liu, Y.; Xiao, Q.; Liu, L.; Shi, Q.; Tong, D.; He, K.; Zhang, Q. Drivers of Increasing Ozone during the Two Phases of Clean Air Actions in China 2013–2020. *Environ. Sci. Technol.* **2023**, *57*, 8954–8964. [\[CrossRef\]](#)
10. Wu, R.; Xie, S. Spatial Distribution of Secondary Organic Aerosol Formation Potential in China Derived from Speciated Anthropogenic Volatile Organic Compound Emissions. *Environ. Sci. Technol.* **2018**, *52*, 8146–8156. [\[CrossRef\]](#)
11. Fu, T.-M.; Jacob, D.J.; Wittrock, F.; Burrows, J.P.; Vrekoussis, M.; Henze, D.K. Global budgets of atmospheric glyoxal and methylglyoxal, and implications for formation of secondary organic aerosols. *J. Geophys. Res. Atmos.* **2008**, *113*, D15303. [\[CrossRef\]](#)
12. Chan Miller, C.; Jacob, D.J.; González Abad, G.; Chance, K. Hotspot of glyoxal over the Pearl River delta seen from the OMI satellite instrument: Implications for emissions of aromatic hydrocarbons. *Atmos. Chem. Phys.* **2016**, *16*, 4631–4639. [\[CrossRef\]](#)
13. Vrekoussis, M.; Wittrock, F.; Richter, A.; Burrows, J.P. GOME-2 observations of oxygenated VOCs: What can we learn from the ratio glyoxal to formaldehyde on a global scale? *Atmos. Chem. Phys.* **2010**, *10*, 10145–10160. [\[CrossRef\]](#)
14. Kaiser, J.; Wolfe, G.M.; Min, K.E.; Brown, S.S.; Miller, C.C.; Jacob, D.J.; deGouw, J.A.; Graus, M.; Hanisco, T.F.; Holloway, J.; et al. Reassessing the ratio of glyoxal to formaldehyde as an indicator of hydrocarbon precursor speciation. *Atmos. Chem. Phys.* **2015**, *15*, 7571–7583. [\[CrossRef\]](#)
15. DiGangi, J.P.; Henry, S.B.; Kammrath, A.; Boyle, E.S.; Kaser, L.; Schnitzhofer, R.; Graus, M.; Turnipseed, A.; Park, J.H.; Weber, R.J.; et al. Observations of glyoxal and formaldehyde as metrics for the anthropogenic impact on rural photochemistry. *Atmos. Chem. Phys.* **2012**, *12*, 9529–9543. [\[CrossRef\]](#)
16. Chen, X.; Zhang, Y.; Zhao, J.; Liu, Y.; Shen, C.; Wu, L.; Wang, X.; Fan, Q.; Zhou, S.; Hang, J. Regional modeling of secondary organic aerosol formation over eastern China: The impact of uptake coefficients of dicarbonyls and semivolatile process of primary organic aerosol. *Sci. Total Environ.* **2021**, *793*, 148176. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Byun, D.; Schere, K.L. Review of the governing equations, computational algorithms, and other components of the models-3 community multiscale air quality (CMAQ) modeling system. *Appl. Mech. Rev.* **2006**, *59*, 51–77. [\[CrossRef\]](#)
18. Zheng, B.; Tong, D.; Li, M.; Liu, F.; Hong, C.; Geng, G.; Li, H.; Li, X.; Peng, L.; Qi, J.; et al. Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions. *Atmos. Chem. Phys.* **2018**, *18*, 14095–14111. [\[CrossRef\]](#)
19. Guenther, A.; Karl, T.; Harley, P.; Wiedinmyer, C.; Palmer, P.I.; Geron, C. Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature). *Atmos. Chem. Phys.* **2006**, *6*, 3181–3210. [\[CrossRef\]](#)
20. Guenther, A.B.; Jiang, X.; Heald, C.L.; Sakulyanontvittaya, T.; Duhl, T.; Emmons, L.K.; Wang, X. The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): An extended and updated framework for modeling biogenic emissions. *Geosci. Model. Dev.* **2012**, *5*, 1471–1492. [\[CrossRef\]](#)
21. Fahey, K.M.; Carlton, A.G.; Pye, H.O.T.; Baek, J.; Hutzell, W.T.; Stanier, C.O.; Baker, K.R.; Wyatt Appel, K.; Jaoui, M.; Offenberg, J.H. A framework for expanding aqueous chemistry in the Community Multiscale Air Quality (CMAQ) model version 5.1. *Geosci. Model. Dev.* **2017**, *10*, 1587–1605. [\[CrossRef\]](#)
22. Zheng, J.; Zhang, L.; Che, W.; Zheng, Z.; Yin, S. A highly resolved temporal and spatial air pollutant emission inventory for the Pearl River Delta region, China and its uncertainty assessment. *Atmos. Environ.* **2009**, *43*, 5112–5122. [\[CrossRef\]](#)
23. Griffith, S.M.; Huang, X.H.H.; Louie, P.K.K.; Yu, J.Z. Characterizing the thermodynamic and chemical composition factors controlling PM_{2.5} nitrate: Insights gained from two years of online measurements in Hong Kong. *Atmos. Environ.* **2015**, *122*, 864–875. [\[CrossRef\]](#)
24. Li, J.; Lu, K.; Lv, W.; Li, J.; Zhong, L.; Ou, Y.; Chen, D.; Huang, X.; Zhang, Y. Fast increasing of surface ozone concentrations in Pearl River Delta characterized by a regional air quality monitoring network during 2006–2011. *J. Environ. Sci.* **2014**, *26*, 23–26. [\[CrossRef\]](#) [\[PubMed\]](#)
25. De Smedt, I.; Müller, J.F.; Stavrou, T.; van der A, R.; Eskes, H.; Van Roozendael, M. Twelve years of global observations of formaldehyde in the troposphere using GOME and SCIAMACHY sensors. *Atmos. Chem. Phys.* **2008**, *8*, 4947–4963. [\[CrossRef\]](#)
26. De Smedt, I.; Stavrou, T.; Hendrick, F.; Danckaert, T.; Vlemmix, T.; Pinardi, G.; Theys, N.; Lerot, C.; Gielen, C.; Vigouroux, C.; et al. Diurnal, seasonal and long-term variations of global formaldehyde columns inferred from combined OMI and GOME-2 observations. *Atmos. Chem. Phys.* **2015**, *15*, 12519–12545. [\[CrossRef\]](#)
27. Alvarado, L.M.A.; Richter, A.; Vrekoussis, M.; Wittrock, F.; Hilboll, A.; Schreier, S.F.; Burrows, J.P. An improved glyoxal retrieval from OMI measurements. *Atmos. Meas. Tech.* **2014**, *7*, 4133–4150. [\[CrossRef\]](#)
28. Ge, B.Z.; Wang, Z.F.; Xu, X.B.; Wu, J.B.; Yu, X.L.; Li, J. Wet deposition of acidifying substances in different regions of China and the rest of East Asia: Modeling with updated NAQPMS. *Environ. Pollut.* **2014**, *187*, 10–21. [\[CrossRef\]](#)
29. Luecken, D.J.; Hutzell, W.T.; Strum, M.L.; Pouliot, G.A. Regional sources of atmospheric formaldehyde and acetaldehyde, and implications for atmospheric modeling. *Atmos. Environ.* **2012**, *47*, 477–490. [\[CrossRef\]](#)
30. Luecken, D.J.; Napelenok, S.L.; Strum, M.; Scheffe, R.; Phillips, S. Sensitivity of Ambient Atmospheric Formaldehyde and Ozone to Precursor Species and Source Types Across the United States. *Environ. Sci. Technol.* **2018**, *52*, 4668–4675. [\[CrossRef\]](#)

31. Chen, W.H.; Guenther, A.B.; Wang, X.M.; Chen, Y.H.; Gu, D.S.; Chang, M.; Zhou, S.Z.; Wu, L.L.; Zhang, Y.Q. Regional to Global Biogenic Isoprene Emission Responses to Changes in Vegetation From 2000 to 2015. *J. Geophys. Res. Atmos.* **2018**, *123*, 3757–3771. [\[CrossRef\]](#)
32. Liu, Z.; Wang, Y.; Vrekoussis, M.; Richter, A.; Wittrock, F.; Burrows, J.P.; Shao, M.; Chang, C.C.; Liu, S.C.; Wang, H. Exploring the missing source of glyoxal (CHOCHO) over China. *Geophys. Res. Lett.* **2012**, *39*, L10812. [\[CrossRef\]](#)
33. Li, J.; Zhang, M.; Tang, G.; Wu, F.; Alvarado, L.M.A.; Vrekoussis, M.; Richter, A.; Burrows, J.P. Investigating missing sources of glyoxal over China using a regional air quality model (RAMS-CMAQ). *J. Environ. Sci.* **2018**, *71*, 108–118. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.