A large role of missing volatile organic compounds reactivity from anthropogenic emissions in ozone pollution regulation

Wenjie Wang1,2*, Bin Yuan1*, Hang Su2, Yafang Cheng2, Jipeng Qi1, Sihang Wang1, Wei Song3, Xinming Wang3, Chaoyang Xue2, Chaoqun Ma2, Fengxia Bao2, Hongli Wang4, Shengrong Lou4, Min Shao1

1 Institute for Environmental and Climate Research, Jinan University, Guangzhou 511443, China
2 Multiphase Chemistry Department, Max Planck Institute for Chemistry, Mainz 55128, Germany
3 State Key Laboratory of Organic Geochemistry, Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, Guangzhou 510640, China
4 State Environmental Protection Key Laboratory of Formation and Prevention of Urban Air Pollution Complex, Shanghai Academy of Environmental Sciences, Shanghai 200233, China

*Correspondence to: Bin Yuan (byuan@jnu.edu.cn); Wenjie Wang (Wenjie.Wang@mpic.de)
Abstract: There are thousands of VOC species in ambient air, while existing techniques can only detect a small part of them (~ several hundred). The large number of unmeasured VOCs prevents us from understanding the photochemistry of ozone and aerosols in the atmosphere. The major sources and photochemical effects of these unmeasured VOCs in urban areas remain unclear. The missing VOC reactivity, which is defined as the total OH reactivity of the unmeasured VOCs, is a good indicator to constrain the photochemical effect of unmeasured VOCs. Here, we identified the dominant role of anthropogenic emission sources in the missing VOC reactivity (accounting for up to 70%) by measuring missing VOC reactivity and tracer-based source analysis in a typical megacity in China. Omitting the missing VOC reactivity from anthropogenic emissions in model simulations will remarkably affect the diagnosis of sensitivity regimes for ozone formation, overestimating the degree of VOC-limited regime by up to 46%. Therefore, a thorough quantification of missing VOC reactivity from various anthropogenic emission sources is urgently needed for constraints of air quality models and the development of effective ozone control strategies.
1 Introduction

Volatile organic compounds (VOCs) are key precursors of major photochemical pollutants, including ozone (O$_3$) and secondary organic aerosols (Atkinson, 2000; Atkinson and Arey, 2003). Severe O$_3$ and particle pollution are frequently related to high emissions of VOCs (Atkinson and Arey, 2003; Monks et al., 2015). There exist thousands of VOC species in ambient air that are emitted from either natural processes or anthropogenic activities (Goldstein and Galbally, 2007). No one instrument can capture all VOCs out there and even when they can be measured there is information missing on identification and properties (Yuan et al., 2017; Wang et al., 2014). Gas chromatograph–mass spectrometer/flame ionization detector (GC–MS/FID) can measure C2-C12 non-methane hydrocarbons (NMHCs) and C2-C6 oxygenated VOCs (OVOCs) while cannot measure NMHCs and OVOCs with larger carbon number (Wang et al., 2014). Proton-transfer-reaction time-of-flight mass spectrometer (PTR-ToF-MS) is able to measure a huge number of OVOCs and aromatics and several alkanes, but cannot measure most alkanes and alkenes, and cannot distinguish isomers (Yuan et al., 2017). The 2,4-dinitrophenylhydrazine (DNPH)/high performance liquid chromatography (HPLC) method can measure several carbonyls but cannot measure non-polar organic species (Wang et al., 2009). The two-dimensional GC is able to measure some intermediate-volatile and semi-volatile non-polar organics (Song et al., 2022). A lack of standard gases prevents these technologies from accurate quantification even if these technologies can identify more VOC species. In general, many branched alkenes, OVOCs with complex functional groups, intermediate-volatile and semi-volatile organics and complex biogenic VOCs cannot currently be well quantified even if they can be identified by instruments. As a result, the total amount of VOCs in ambient air has generally been underestimated. Currently, emission inventories used in air quality models such as the Community Emissions Data System (CEDS) emission inventory and the multi-resolution Emission Inventory for China (MEIC) only include the VOC species that can be measured such as some C1-C9
hydrocarbons and simple-structure OVOCs with small carbon number (<C6). This will lead to an underestimation of the photochemical effect of total VOCs and thus causes uncertainties in predicting secondary pollution. The quantification of the unmeasured VOCs is crucial to assess secondary pollution precisely.

The total OH reactivity \( R_{OH} \), which can be directly measured, is an index for evaluating the amount of reductive pollutants in terms of ambient OH loss. The total OH reactivity is defined as:

\[
R_{OH} = \sum_i k_{OH+Xi} [X_i].
\]

where \( X \) represents a reactive species including carbon monoxide (CO), nitrogen oxides (NOx) and VOCs etc., and \( k_{OH+Xi} \) is the reaction rate constant for the oxidation of species \( X \) by OH. The measured \( R_{OH} \) is higher than that calculated based solely on the measured reactive species, and the difference between them is mostly from unmeasured VOCs (Yang et al., 2017). Missing VOC reactivity (missing VOCR), defined as VOC reactivity (VOCR) of all unmeasured VOCs, can be obtained by subtracting the calculated \( R_{OH} \) from the measured \( R_{OH} \).

\[
\text{missing VOCR} = \text{measured } R_{OH} - \text{calculated } R_{OH}
\]

\[
\text{calculated } R_{OH} = \sum_i k_{OH+\text{reactive species}_i} [\text{reactive species}_i]
\]

where reactive species represents measured VOCs and reactive inorganic species including carbon monoxide (CO), nitric oxide (NO), nitrogen dioxide (NO2), O3, sulfur dioxide (SO2), nitrous acid (HONO), and so on. The missing VOCR provides a constraint for evaluating the photochemical roles of unmeasured VOCs in the atmosphere (Sadanaga et al., 2005; Yang et al., 2016b). The inclusion of the missing VOCR can help to improve the performance of box model and air quality models in simulating photochemistry processes. Relatively high missing VOCR have been found in forests (Di Carlo et al., 2004; Hansen et al., 2014; Nakashima et al., 2014; Nölscher et al., 2016; Praplan et al., 2019), urban areas (Shirley et al., 2006; Yoshino et al., 2006; Dolgorouky et al., 2012; Yang et al., 2017) and suburban areas (Kovacs et al., 2003; Yang et al., 2017; Fuchs et al., 2017; Lou et al., 2010), accounting for 10-75% of total \( R_{OH} \). Given that total VOCR is one part of total \( R_{OH} \), missing VOCR would account
for a larger percentage of total VOC$_R$ (>10%-75%).

The potential sources of missing VOC$_R$ include anthropogenic emissions, biogenic emissions, soil emissions, and photochemical production, etc (Yang et al., 2016b). Previous studies have reported that the missing VOC$_R$ in forest areas was mainly from either direct emissions or photochemical oxidation of biogenic VOCs (Di Carlo et al., 2004; Hansen et al., 2014; Nakashima et al., 2014; Nölscher et al., 2016; Praplan et al., 2019). Nevertheless, the dominant source of the missing VOC$_R$ in urban and suburban areas remains unclear or under debate.

Surface O$_3$ pollution has become a major public health concern in cities worldwide (Paoletti et al., 2014; Lefohn et al., 2018). A critical issue in determining an emission control strategy for ozone pollution is to understand the relative benefits of NOx and VOC emission controls. This is generally framed in terms of ozone precursor sensitivity, i.e., whether ozone production is NOX-limited or VOC-limited (Kleinman, 1994; Sillman et al., 1990). Nevertheless, the effect of missing VOCs on ozone precursor sensitivity has not been well understood yet. Given that the missing VOC$_R$ could potentially account for a large part of total VOC$_R$, clearly clarifying the role of missing VOC$_R$ in determining ozone precursor sensitivity is an urgent need for the diagnosis of ozone sensitivity regimes and formulation of an effective emission reduction roadmap.

China has become a global hot spot of ground-level ozone pollution in recent years (Lu et al., 2018; Wang et al., 2022). Pearl River Delta (PRD) remains one of the most O$_3$-polluted regions in China (Li et al., 2022), although many control measures have been attempted. Here, we measured RO$_H$ in Guangzhou, a megacity in PRD and quantified the missing VOC$_R$. The dominant source of the missing VOC$_R$ and its impact on ozone precursor sensitivity were comprehensively investigated.
Method

2.1 Overview of the measurement

The field campaign was conducted from 25 September to 30 October 2018 continuously at an urban site in downtown Guangzhou (113.2°E, 23°N). The sampling site is located on the ninth floor of a building on the campus of Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, 25 m above the ground level. This site is primarily influenced by industrial and vehicular emissions. ROH, VOCs, NOX, O3, HONO, SO2, CO, photolysis frequencies, and meteorological factors were simultaneously measured during the measurement period.

2.2 ROH measurement

Total ROH was measured by the comparative reactivity method (CRM) (Sinha et al., 2008). The CRM system consists of three major components, namely an inlet and calibration system, a reactor, and a measuring system. Here, pyrrole (C4H5N) was used as the reference substance in CRM and its concentration was quantified by a quadrupole proton-transfer-reaction mass spectrometer (PTR-MS) (Ionicon Analytik GmbH, Innsbruck, Austria). The CRM system was calibrated by propane, propene, toluene standards and 16 VOC mixed standard (acetaldehyde, methanol, ethanol, isoprene, acetone, acetonitrile, methyl vinyl ketone, methyl ethyl ketone, benzene, toluene, o-xylene, α-pinene, 1,2,4-trimethylbenzene, phenol, m-cresol, and naphthalene). Measured and calculated ROH agreed well within 15% for all calibrations. The ROH measurement by the CRM method is interfered from ambient nitric oxide (NO), which produces additional OH radicals via the reaction of HO2 radicals with NO (Sinha et al., 2008). To correct this interference, a series of experiments were conducted by introducing different levels of NO (0–160 ppb) and given amounts of VOC into the CRM reactor. A correction curve was acquired from these NO interference experiments, which can be used to correct the ROH thanks to the simultaneous measurement of
ambient NO concentrations (Supplementary information S1; Fig. S1). The detection limits of the CRM method were around 2.5 s⁻¹, and the total uncertainty was estimated to be about 15%. The CRM method has been successfully applied to measure OH reactivity in urban areas with high NOₓ levels in previous studies (Dolgorouky et al., 2012; Yang et al., 2017; Hansen et al., 2015). The intercomparison between the CRM method and pump–probe technique indicates that the CRM method can be used under high-NOₓ conditions (NOₓ>10 ppb) if a NOₓ-dependent correction is applied (Hansen et al., 2015).

2.3 VOCs measurements

Nonmethane hydrocarbons (NMHCs) were measured using a gas chromatograph–mass spectrometer/flame ionization detector (GC–MS/FID) system coupled with a cryogen-free preconcentration device (Wang et al., 2014). The system contains two-channel sampling and GC column separation, which is able to measure C2–C5 hydrocarbons with the FID in one channel and measure C5–C12 hydrocarbons using MS detector in the other channel. After removal of water vapor, VOCs were trapped at -155 ℃ in a deactivated quartz capillary column (15 cm×0.53 mm ID) and a Porous Layer Open Tubular (PLOT) capillary column (15 cm×0.53 mm ID) for the MS channel and the FID channel, respectively. The system was calibrated weekly by TO-15 (Air Environmental Inc., USA) and PAMS gas standards (Spectra Gases Inc., USA). Detection limits for various compounds were in the range of 0.002–0.070 ppbv. A total of 56 NMHCs species were measured (Table S1). The time resolution of the measurement was 1 h. The uncertainties of VOC measurements by GC–MS/FID are in the range of 15 %–20 %. More details of this method can be found in previous studies (Wang et al., 2014; Yuan et al., 2012).

An online proton-transfer-reaction time-of-flight mass spectrometer (PTR-ToF-MS) (Ionicon Analytic GmbH, Innsbruck, Austria) with H₃O⁺ and NO⁺ ion sources was also used to measure VOCs. During the campaign, the PTR-ToF-MS automatically switched between H₃O⁺ and NO⁺ chemistry every 10–20 min. The H₃O⁺ mode was
used to measure OVOCs and aromatics while the NO⁺ model was used to measure alkanes with more carbons (C8-C20). When running in the H₃O⁺ ionization mode, the drift tube was at a temperature of 50 °C, a pressure of 3.8 mbar, and a voltage of 920 V, leading to an operating E/N (E is the electric field, and N is the number density of the gas in the drift tube) ratio of 120 Td. When running in the NO⁺ ionization mode, the drift tube was at a temperature of 50 °C, a pressure of 3.8 mbar, and a voltage of 470 V, leading to an operating E/N ratio of 60 Td. PTR-ToF-MS technique is capable of measuring oxygenated VOCs (OVOCs) and higher alkanes that GC–MS/FID cannot measure (Wu et al., 2020; Wang et al., 2020a). The time resolution of PTR-ToF-MS measurements was 10 s. A total of 31 VOCs were calibrated using either gas or liquid standards (Table S2). For other measured VOCs, we used the method proposed by Sekimoto et al. (2017) to determine the relationship between VOC sensitivity and kinetic rate constants for proton transfer reactions of H₃O⁺ with VOCs. The fitted line was used to determine the concentrations of those uncalibrated species. The uncertainties of the concentrations for uncalibrated species were about 50 % (Sekimoto et al., 2017). By this method, PTR-ToF-MS can additionally measure 128 VOCs which were included in the analysis of this study. The detailed information for this method can be found in Wu et al. (2020) and all VOC species measured by PTR-ToF-MS were provided in table S4 of that article. The PTR-ToF-MS is capable of measuring additional VOC species that GC–MS/FID cannot measure including alkanes with more carbons (C12–C20) and OVOCs including aldehydes, ketones, carboxylic acids, alcohols, and nitrophenols. Formaldehyde (HCHO) was measured by a custom-built instrument based on the Hantzsch reaction and absorption photometry (Xu et al., 2022).

2.4 Other measurements

Nitrous acid (HONO) was measured by a custom-built LOPAP (Long Path Absorption Photometer) based on wet chemical sampling and photometric detection (Yu et al., 2022). The uncertainty of the measurement was 8 %. NOx, O₃, SO₂, and CO were measured by NOx analyzer (Thermo Scientific, Model 42i), O₃ analyzer (Thermo Scientific, Model 49i), SO₂ analyzer (Thermo Scientific, Model 43i), and CO analyzer
The meteorological data, including temperature (T), relative humidity (RH) and wind speed and direction (WS, WD) were recorded by Vantage Pro2 Weather Station (Davis Instruments Inc., Vantage Pro2) with a time resolution of 1 min. Photolysis frequencies of O$_3$, NO$_2$, HONO, H$_2$O$_2$, HCHO, and NO$_3$ were measured by a spectrometer (Focused Photonics Inc., PFS-100) (Shetter and Müller, 1999; Wang et al., 2019).

2.5 Multiple linear regression

The Multiple Linear Regression (MLR) has been successfully applied to quantify the sources of air pollutants (Li et al., 2019; Yang et al., 2016a). In this study, a tracer-based MLR analysis was used to decouple the individual contributions of anthropogenic emissions, secondary production, biogenic emissions and background level to missing VOC$_R$, as shown in Eq. (4).

$$ Missing \text{VOC}_R = a \Delta \text{CO} + b[O_X] + c[\text{isoprene}_{initial}] + C_{background} $$

(4)

where O$_X$ is defined as O$_3$+NO$_2$. ΔCO, [O$_X$] and [isoprene$_{initial}$] are concentrations of tracers for anthropogenic emissions, secondary production and biogenic emissions, respectively. ΔCO is the relative change between ambient CO and background CO of 150 ppb (Wang et al., 2020a). [isoprene$_{initial}$] represents the initial concentration of isoprene from biogenic emissions that has not undergone any photochemical reactions, which is calculated from observed isoprene and its photochemical products methyl vinyl ketone (MVK) and methacrolein (MACR) (Xie et al., 2008). $C_{background}$ indicates the background level of missing VOC$_R$. $a$, $b$, $c$ and $C_{background}$ are fitted coefficients by the multiple linear regression.

2.6 Observation-based box model

A zero-dimensional box model coupled with the Master Chemical Mechanism (MCM) v3.3.1 chemical mechanism (Jenkin et al., 2003) was used to simulate the photochemical production of RO$_X$ (RO$_X$=OH+HO$_2$+RO$_2$) radicals and O$_3$ during the field campaign. The model was constrained by the observations of meteorological
parameters, photolysis frequencies, VOCs, NO, NO₂, O₃, CO, SO₂, and HONO. The
model runs were performed in a time-dependent mode with a time resolution of 1 hour
and a spin-up of four days. A 24-h lifetime was introduced for all simulated species,
including secondary species and radicals, to approximately simulate dry deposition and
other losses of these species (Lu et al., 2013; Wang et al., 2020b). Sensitivity tests show
that this assumed physical loss lifetime has a relatively small influence on ROₓ radicals
and ozone production rates.

Measured OVOCs such as HCHO, acetaldehyde and acetone were constrained in
the model and unmeasured OVOCs were simulated according to the photochemical
oxidation of NMHCs by OH radicals. RO₂, HO₂ and OH radicals were simulated by the
box model to calculate the net O₃ production rate (P(O₃)) and O₃ loss rate (L(O₃)) as
shown in Equations (5) and (6) as derived by Mihelcic et al. (2003)

\[
P(O_3) = k_{H_2O_2+NO}[H_2O] \cdot [NO] + \sum_i (k_{RO_2+NO}[RO_2] \cdot [NO]) - k_{OH+NO_2}[OH] \cdot [NO_2] - L(O_3)
\]

(5)

\[
L(O_3) = (\theta_j(O^1D) + k_{OH+O_3}[OH] + k_{H_2O_2+O_2}[H_2O] + \sum_j (k_{alkene+O_3}[alkene]) \cdot [O_3]
\]

(6)

where \( \theta \) is the fraction of O^1D from ozone photolysis that reacts with water vapor, and
i and j represent the number of species of RO₂ and alkenes, respectively.

The box model was used to evaluate the impact of missing VOC_R on the O₃
production rate. In the base scenario, the box model was constrained by all measured
inorganic and organic gases but the missing VOC_R was not considered. To consider the
missing VOC_R in the box model, we additionally increased the concentration of
NMHCs to exactly compensate for the missing VOC_R by multiplying a factor, on the
basis of measured NMHC concentrations. We simulated four scenarios by increasing
the concentration of: (1) n-pentane, (2) ethylene, (3) toluene, (4) all measured 56
NMHCs. For the scenario of increasing all 56 NMHCs, concentrations of 56 NMHC
species were increased by multiplying the same factor. Given that the VOC_R of
unconstrained secondary products increases with the increase in the concentration of
NMHCs, several attempts of different values are needed to determine the increasing
3 Results and discussion

3.1 Quantification of missing VOC$_{R}$ during the campaign

Figure 1 shows the time series of measured RO$_{H}$, calculated RO$_{H}$ according to all measured reactive gases, and missing VOC$_{R}$ (the gap between measured and calculated RO$_{H}$) in Guangzhou. By using GC-MS/FID, we measured 56 NMHCs. By using PTR-ToF-MS, we measured 159 VOCs and 128 of them were difficult to be measured before. Besides the alkanes with carbons less than 12, PTR-ToF-MS can also measure alkanes with more carbons (C12–C20). With regard to OVOCs, not only common OVOC species including formaldehyde and C2-C4 carbonyls but also carbonyls with more carbons (C5–C10) and some N-containing OVOC species such as nitrophenol and methyl nitrophenol were measured by PTR-ToF-MS. Thanks to these additional measured VOCs, the measured RO$_{H}$ was close to the calculated RO$_{H}$ within 20% in most periods. In some periods the missing VOC$_{R}$ was negative, which is probably due to the uncertainty in the measurements of RO$_{H}$ and reactive gases. The negative missing VOC$_{R}$ primarily occurred in the afternoon (12:00–17:00) when the photochemistry was most active. Nevertheless, there were still some days exhibiting remarkable missing VOC$_{R}$. The days with missing VOC$_{R}$ of more than 25% of total RO$_{H}$, namely high missing-VOC$_{R}$ days, are indicated by yellow background in Fig. 1a. The largest missing VOC$_{R}$ occurred on October 15$^{th}$, 16$^{th}$, 25$^{th}$ and 26$^{th}$, with average values of 16 s$^{-1}$. During the period of October 24$^{th}$ to 26$^{th}$, the total RO$_{H}$ was highest and the missing VOC$_{R}$ was also relatively high among all days. Figure 1b shows the contribution of different species classifications to total RO$_{H}$ during high missing-VOC$_{R}$ days. Inorganic species, NMHCs and OVOCs account for 34%, 13% and 14% of total RO$_{H}$, respectively, with missing VOC$_{R}$ accounting for 39%. The fraction of missing VOC$_{R}$ (39%) during the high missing–VOC$_{R}$ days is comparable to measurements in Los Angeles 2010 (Griffith et al., 2016) and in Seoul 2016 (Sanchez et al., 2021).
We evaluated the uncertainty of the missing VOC$_R$. The uncertainty of the ROH measurement was 15%. In addition, according to reports of Jet Propulsion Laboratory (Burkholder et al., 2020), reaction rate constants used for the calculation of ROH in Eq (3) have uncertainties of 5%–30%, depending on different species. We took the uncertainties in the reaction rate constants and the measurements of all reactive gases into account when calculating ROH, according to error propagation. As a result, the uncertainties in the missing VOC$_R$ are 3.8 s$^{-1}$ and 5.2 s$^{-1}$ for the whole measurement period and the high missing-VOC$_R$ days, respectively. The average missing VOC$_R$ during the high missing-VOC$_R$ days is 13 s$^{-1}$, which is significantly higher than the uncertainty of 5.2 s$^{-1}$, suggesting that the missing VOC$_R$ really exists during the high missing-VOC$_R$ days.

3.2 The sources of missing VOC$_R$

To explore the sources of missing VOC$_R$ during the whole measurement period, we investigated the correlation between missing VOC$_R$ and tracers characterizing primary emissions (CO, NO$_X$ and NMHCs) and secondary production (O$_X$=O$_3$+NO$_2$ and formic acid). The correlation of missing VOC$_R$ with CO, reactivity of NMHCs (NMHC$_R$) and NO$_X$ is moderate, with correlation coefficient (R) in the range of 0.47–0.56 (Fig. 2a and b, and Fig. S2) while there is no significant correlation of missing VOC$_R$ with O$_X$ and formic acid (Fig. 2c and Fig. S2). Furthermore, there is no significant correlation between missing VOC$_R$ and acetonitrile which is a tracer of biomass burning (de Gouw et al., 2003;Wang et al., 2007) (Fig. S2), indicating that biomass burning was not a major contributor to missing VOC$_R$ during this campaign.

In terms of the diurnal variation, the missing VOC$_R$ was higher in the morning (7:00–10:00) and evening (18:00–22:00) when the anthropogenic emissions, especially vehicle exhaust were intensive, and was lower in the afternoon when the photochemistry was most active (Fig. 2d). The diurnal profile of missing VOC$_R$ was similar to those of CO, NO$_X$ and NMHC$_R$. In contrast, the diurnal profiles of secondary
species including Ox, formic acid and acetic acid, which peaked in the afternoon, evidently differ from the diurnal profile of missing VOC$_R$ (Fig. S3). Further, we investigated the influence of airmass aging on missing VOC$_R$. The ratio of ethylbenzene to m,p-xylene was used to characterize the degree of air mass aging (De Gouw et al., 2005; Yuan et al., 2013). A higher ratio of ethylbenzene to m,p-xylene corresponds to a higher degree of air mass aging as the m,p-xylene has a larger reaction rate constant ($18.9 \times 10^{-12}$ cm$^3$ molecule$^{-1}$s$^{-1}$) than ethylbenzene ($7.0 \times 10^{-12}$ cm$^3$ molecule$^{-1}$s$^{-1}$) when reacting with the major oxidant - OH radicals. As shown in Fig. 2e, missing VOC$_R$ decreases with the ratio of ethylbenzene to m,p-xylene. Given that secondary production generally increased with air mass aging, this result further demonstrates that missing VOC$_R$ was not caused by enhanced secondary production.

During the high missing- VOC$_R$ days, the correlation coefficient for missing VOC$_R$ versus CO is 0.76 (Fig. 3a), which is higher than that in the whole measurement period (0.56) shown in Fig. 2a. We then quantify the sources of missing VOC$_R$ during the high missing- VOC$_R$ days by applying MLR. The fitted coefficients are as follows: $a$ is 0.031 s$^{-1}$ ppb$^{-1}$, $b$ is 0.012 s$^{-1}$ ppb$^{-1}$, $c$ is 1.8 s$^{-1}$ ppb$^{-1}$ and $C_{background}$ is 1.3 s$^{-1}$. The coefficient of determination ($R^2$) for the MLR is 0.68. As shown in Fig. 3b, anthropogenic emissions were the largest contributor to missing VOC$_R$, accounting for 70% of missing VOC$_R$. Secondary production, biogenic emissions and background contribution played a minor role in missing VOC$_R$ (13%, 7%, 10%, respectively). The parametric relationship between missing VOC$_R$ and relevant tracers established by MLR provides a valid approach to estimate the missing VOC$_R$ according to readily available gases including CO, Ox and isoprene.

Although anthropogenic emissions are identified to be the major source of missing VOC$_R$, which species dominantly contribute to the missing VOC$_R$ remains unclear. A potential source is the unmeasured branched alkenes for their high reactivity, previously observed from vehicle exhaust (Nakashima et al., 2010) and gasoline evaporation emissions (Wu et al., 2015). Another possible source is emitted OVOCs with a more complex functional group that cannot be accurately measured. In addition, directly
emitted semi-volatile and intermediate volatility organic compounds are also possible sources of missing VOCR (Stewart et al., 2021).

3.3 The impact of missing VOCR on O₃ sensitivity regimes

The reaction of OH with VOCs is key to the propagation and amplification of OH radicals, thus determining the ozone production rate (Tonnesen and Dennis, 2000). The box model was used to evaluate the impact of missing VOCR on the O₃ production rate during high missing-VOCR days. The setting of model simulations for different scenarios are depicted in Section 2.6. Under the base scenario, on average the measured VOCR of n-pentane, ethylene, toluene and all 56 NMHCs are 0.14 s⁻¹, 0.53 s⁻¹, 0.60 s⁻¹ and 4.6 s⁻¹ respectively. To consider the missing VOCR (on average of 13 s⁻¹) in the model, four scenarios were simulated by additionally increasing n-pentane, ethylene, toluene and 56 NMHCs by a factor of 70, 16, 13.3 and 1.9, respectively. These increasing factors led to an additional increase in VOCR of both NMHCs and unconstrained secondary products, which exactly compensated for the missing VOCR. Figure 4 shows the simulated P(O₃) for the base scenario and the scenarios considering missing VOCR. The daytime average P(O₃) under the scenarios considering missing VOCR is a factor of 1.5–4.5 for the results under the base scenario. The difference in added species has a large effect on P(O₃). Adding toluene causes a larger increase in P(O₃) than adding n-pentane or ethene, as toluene has a stronger ability to amplify the production of radicals.

O₃ precursor sensitivity depends on the dominant loss pathways of ROₓ radicals (ROₓ=OH+HO₂+RO₂). O₃ production is NOₓ-limited if the self-reaction of peroxy radicals (HO₂ and RO₂) dominates the ROₓ sink, and VOC-limited if the reaction of NO₂ with OH dominates (Kleinman et al., 1997; Kleinman et al., 2001). Accordingly, the ratio of ROₓ sink induced by OH+NO₂ reaction to the total rate of the two ROₓ sinks, i.e., $L_N/Q$, is used to identify O₃ sensitivity regimes. O₃ production is NOₓ-limited if $L_N/Q$ is lower than 0.5, otherwise, it is VOC-limited (Kleinman et al., 1997).

$$L_N/Q = \frac{k_{OH+NO_2}[OH][NO_2]}{k_{HO_2+RO_2}[HO_2][RO_2]+k_{HO_2+HO_2}[HO_2]+k_{OH+HO_2}[OH][HO_2]+k_{OH+NO_2}[OH][NO_2]}$$
As shown in Fig. 5a, under the base scenario, LN/Q remained at a stable and high level (>0.9) during the daytime when photochemical production of ozone occurs, indicating O₃ production was VOC-limited. Under the scenarios considering missing VOCr, LN/Q decreased significantly regardless of which VOC species was added, compared to the base scenario. Adding toluene caused the largest decrease in LN/Q, followed by adding all measured NMHC species, adding the alkane and adding the alkene. It is worth noting that adding toluene and all measured NMHC species caused the LN/Q to be close to 0.5 in the afternoon, indicating that the O₃ production shifted to transitional or NOX-limited regimes in these scenarios. Fig. 5b shows the changes in radical sinks before and after considering missing VOCr. All radical sinks including self-reactions of peroxy radicals and OH+NO₂ reaction increased after considering missing VOCr. Nevertheless, the increased proportion of the self-reactions of peroxy radicals was larger than that of OH+NO₂ reaction, leading to a decrease in LN/Q and thus a shift toward NOX-limited regime.

**Figure 5c** shows the dependence of daily peak O₃ concentrations on NOX concentrations, which was calculated by the box model for the base scenario and the scenario considering missing VOCr. The NOX concentration level corresponding to the maximum of O₃ concentrations was determined. This NOX concentration level reflects the threshold to distinguish between VOC-limited and NOX-limited regimes. The larger threshold of NOX represents a higher possibility of ozone production in NOX limited regime. The threshold of NOX for the scenario considering missing VOCr is 46% higher than for the base scenario. Note that the uncertainty in missing VOCr leads to 17% uncertainty in the threshold of NOX for the scenario considering missing VOCr. Overall, Fig. 5 suggests that omitting the missing VOCr will overestimate the degree of the VOC-limited regime and thus overestimate the effect of VOCs abatement in reducing ozone pollution, which in turn may mislead ozone control strategy.
3.4 Atmospheric implications

Although many previous studies have reported that photochemical production processes and biogenic emissions are important sources of missing VOC$_R$ (Lou et al., 2010; Dolgorouky et al., 2012; Yang et al., 2017; Sanchez et al., 2021; Di Carlo et al., 2004), we find that anthropogenic emissions may dominate the missing VOC$_R$ in urban regions. In zero-dimensional box models and three-dimensional chemistry-transport models, the input of VOCs emission information mainly contains well-studied simple-structure alkanes, alkenes and aromatics, while those unmeasured/unknown VOC species have been neglected. This will lead to biases in quantifying ozone production and diagnosing ozone sensitivity regimes. Our study demonstrates that the ambient measurement of RO$_{OH}$ at urban sites can provide quantification of missing VOC$_R$, which can be used in models to account for the missing VOC$_R$ from anthropogenic emissions. In addition, the parametric equation of missing VOC$_R$ derived from MLR method (Eq (4)) here can be used to estimate missing VOC$_R$ according to measurements of CO, O$_X$ and isoprene. Further study should try to parse the specific sources of the missing VOC$_R$, e.g., whether the missing VOC$_R$ is from intermediate-volatility and semivolatile organic compounds emitted from vehicles or whether it is from some other sources. Furthermore, future studies can focus on direct measurements of missing VOC$_R$ for various emission sources to develop a comprehensive emission inventory of missing VOC$_R$, which will help to improve O$_3$ pollution mitigation strategies.

Data availability

The observational data and model code used in this study are available from corresponding authors upon request (byuan@jnu.edu.cn).

Author contributions

BY and WJW designed the research. WJW and BY prepared the manuscript with contributions from other authors. WJW performed data analysis with contributions
from HS, YFC, FXB. CYX, CQM. JPQ, SHW, WS, XMW, HLW, SRL and MS

collected data

Competing interests

At least one of the (co-)authors is a member of the editorial board of Atmospheric Chemistry and Physics.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (grant No., 42121004, 42275103, 42230701, 42175135). This work was also supported by Special Fund Project for Science and Technology Innovation Strategy of Guangdong Province (Grant No.2019B121205004).

References


at a rural site (Wangdu) in the North China Plain: contributions from OH reactants and experimental OH budget, Atmos. Chem. Phys., 17, 645-661, 10.5194/acp-17-645-2017, 2017.


observed and modelled OH and HO2 concentrations in summer 2006, Atmospheric Chemistry and Physics, 13, 1057-1080, 10.5194/acp-13-1057-2013, 2013.


Figure 1. The level of missing VOC\(_R\) during the measurements in Guangzhou. (a) Time series of measured R\(_{OH}\) and calculated R\(_{OH}\) from all measured reactive gases in Guangzhou. Yellow background represents the high missing-VOC\(_R\) days with missing VOC\(_R\) accounting for more than 30% of total R\(_{OH}\). (b) Contributions of different compositions to R\(_{OH}\) in high missing-VOC\(_R\) days. The error bar represents standard deviation of missing VOC\(_R\).
Figure 2. Correlation of missing VOC$_R$ with major tracers during the whole measurement period. (a-c) Correlation of missing VOC$_R$ with CO, OH reactivity of NMHCs (NMHCR) and Ox. Each point represents hourly data. (d) Diurnal variations in missing VOC$_R$, CO, NO$_X$ and NMHCs. (e) The dependence of missing VOC$_R$ on ethylbenzene to m, p-xylene ratio. The red squares indicate the mean values of missing VOC$_R$ in different ranges of ethylbenzene/m,p-xylene with classification width of 0.1, and the error bars represent standard deviation.
Figure 3. The source apportionment of missing VOC$_R$ in high missing-VOC$_R$ days.
(a) Correlation of missing VOC$_R$ with CO. Each point represents hourly data. (b) Contributions of different sources to missing VOC$_R$ according to the MLR.
Figure 4. Simulated daytime mean $P(O_3)$ for the base scenario (without missing $VOC_R$) and the scenario considering missing $VOC_R$, respectively, in high-missing $VOC_R$ days. The missing $VOC_R$ is considered by adding individual species (n-pentane, ethene or toluene) or increasing all measured NMHCs to compensate for the missing $VOC_R$. The error bar represents standard deviation of $P(O_3)$ induced by the uncertainty of missing $VOC_R$. 


Figure 5. The impact of missing \( \text{VOC}_R \) on \( \text{O}_3 \) sensitivity for the high-missing \( \text{VOC}_R \) days. (a) Diurnal variations in \( \ln \frac{\text{Q}}{\text{LN}} \) for the base scenario and the scenarios considering missing \( \text{VOC}_R \). The missing \( \text{VOC}_R \) is considered by adding individual species (n-pentane, ethene or toluene) or increasing all measured NMHCs to fill the missing \( \text{VOC}_R \). The dashed line represents the threshold value of \( \ln \frac{\text{Q}}{\text{LN}} \) that distinguishes VOC-limited and NOx-limited regimes. (b) The averages of radical sinks in the afternoon (12:00-18:00) for the base scenario (red bar) and the scenario considering missing \( \text{VOC}_R \) (blue bar) by increasing all measured NMHCs to fill the missing \( \text{VOC}_R \). (c) Model-simulated dependence of daily peak \( \text{O}_3 \) concentrations on daily mean NOx concentrations for the
base scenario (red curve) and the scenario considering missing VOC$_R$ (blue curve) by increasing all measured NMHCs to fill the missing VOC$_R$. The dashed lines parallel to Y-axis represent the threshold of NO$_X$ levels to distinguish between VOC-limited and NO$_X$-limited regimes. The shaded area represents standard deviation induced by the uncertainty in missing VOC$_R$. 