MODELING STUDY OF PRECURSOR-SPECIFIC SECONDARY ORGANIC AEROSOL AND ORGANIC TRACERS

A Dissertation

by

JIE ZHANG

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Qi Ying
Committee Members,	Shankararaman Chellam
	Garrett McKay
	Xiaohong Liu
Head of Department,	Zachary Grasley

December 2021

Major Subject: Civil Engineering

ABSTRACT

In this study, the community multiscale air quality (CMAQ), with modifications to track precursor-specific SOA, was applied to model SOA and organic tracer formation from aromatic compounds, isoprene, monoterpenes, and sesquiterpenes. The model predicted aromatic and monoterpene SOA showed strong correlations with the measured daily corresponding organic tracers, which indicates that the tracer-method is a good approach to evaluate model predictions in precursor-specific SOA. However, the tracer-to-SOA ratios (f_{SOA}) derived from the modeling results show large variation based on different SOA components considered, and the f_{SOA} values showed significant difference from those determined in chamber experiments due to the difference between chamber conditions and ambient atmosphere. The f_{SOA} in the ambient air can be assessed by the modified CMAQ model with abilities to simulate organic tracers and SOA simultaneously.

The modeled aromatic SOA tracer, 2,3-dihydroxy-4-oxopentanoic acid (DHOPA), agree well with the field measurements (MFB = 0.15; R = 0.8), and approximately two-thirds of it is from the oxidation of toluene. The modeled f_{SOA} shows a strong dependence on the OA loading when only semivolatile aromatic SOA components are included, while this dependence becomes weaker when non-volatile oligomers and dicarbonyl SOA products are considered. To predict total aromatic SOA, a constant f_{SOA} of 0.002 is determined, and the common-used chamber-determined f_{SOA} value of 0.004 could lead to an underestimation of SOA by a factor of 2.

The isoprene-SOA scheme in the CMAQ model is expanded to simulate the unique isoprene tracers 2-methyltetrols (2-MT) and 2-methylglyceric acid (2-MG) by treating them as semivolatile species and including a non-heterogeneous formation pathway. The modeled f_{SOA} of sum of 2-MT and 2-MG in the total isoprene-SOA varies gently, between 0.01-0.02 in polluted

regions, suggesting that the chamber-determined f_{SOA} of 0.063 may lead to large underestimations of overall isoprene SOA.

The monoterpene (MT) and sesquiterpene (SQT) SOA was simulated by the CMAQ model with five explicit and one lumped MT species and SQT, and the contribution from each oxidation pathway was tracked in the MT SOA formation. Three MT tracers (pinic acid, PA; pinonic acid, PNA; and 3-methyl-1,2,3-butanetricarboxylic acid, MBTCA) and one SQT tracer (β-caryophyllinic acid, BCARYA) were modeled to assess the f_{SOA} values to estimate MT and SQT SOA. The f_{SOA} shows significant OA dependence, suggesting that using a constant f_{SOA} could lead to large errors in estimating terpene SOA. Instead, power-law equations directly link the tracer concentrations to the corresponding SOA concentrations were proposed and lead to good SOA estimations.

ACKNOWLEDGEMENTS

I would like to express my special appreciation and gratitude to my advisor, Dr. Qi Ying, for his guidance and continuous support on my research and Ph.D. learning. He is a good professor who always show his patience and encourages to his students. I am also deeply grateful to my committee members, Dr. Shankararaman Chellam, Dr. Garrett McKay, and Dr. Xiaohong Liu, for their insightful comments and suggestions.

Additionally, I express my appreciation to Dr. Jian Zhen Yu, Professor in HKUST, for providing significant guidance and comments on my research. And I would like to thank my group fellows and visiting scholars, Minsu Choi, Dr. Mingjie Kang (Fudan University), and Kezheng Liao (HKUST) for their contributions and help.

Finally, I would like to thank my wife, Dr. Ruyue Song for her encourages and accompanies. And I am grateful for the love from my parents and my big family.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professors Dr. Qi Ying, Dr. Shankararaman Chellam, Dr. Garrett McKay of the Department of Civil and Environmental Engineering and Dr. Xiaohong Liu of Department of Atmospheric Science.

The chemical mechanism of toluene-OH initial oxidation in Section 2 was provided by Professor Renyi Zhang of the Department of Atmospheric Science and were published in 2017. The observation data used in Section 3-6 was provided by Professor Jian Zhen Yu of the Department of Chemistry in the Hong Kong University of Science and Technology.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a grant from National Institutes of Health (R01 ES029509).

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

1 INTRODUCTION

Organic aerosols (OA) are generated from direct emissions (i.e., primary organic aerosol, or POA) and gas-to-particle partitioning of the semi-volatile oxidation products from parent volatile organic compounds (VOCs). Secondary organic aerosol (SOA) contributions to total OA loading vary from 20% to 80%, with significant spatial and seasonal variation [1-4]. Chemical transport models (CTMs) have been widely used to quantitatively study the regional and global impacts of carbonaceous aerosols [5-8]. Correctly predicting SOA in these models is challenging, especially in polluted urban areas, as many precursors contribute to SOA formation, and the ability of each precursor to form SOA is different. Major precursors of SOA include aromatic compounds [9-11], isoprene [2, 12-14], monoterpenes [15, 16], and sesquiterpenes [17, 18]. Also, dicarbonyls such as glyoxal (GLY) and methylglyoxal (MGLY) are found to contribute significantly to SOA formation through aqueous and heterogeneous processes [19, 20]. Predicted SOA concentrations are affected by the model representation of the emission, photochemical oxidation, gas-to-particle partitioning, and the multiphase reaction processes [21-23]. Many of these physical and chemical processes remain uncertain due to an incomplete understanding of the SOA formation mechanisms and large differences between the atmospheric conditions in the ambient environment and the chamber conditions under which the SOA formation experiments were conducted to determine parameters used in the models [24, 25].

Chemical mechanism of VOC oxidation is one of the significant driven factors to represent SOA formation, however, it remains uncertain for a lot of reactions. For example, toluene is one of the most abundant VOCs emitted in the urban atmosphere. Emitted from a variety of sources, including vehicular exhaust and evaporation, biomass burning, and solvent usage [26-28], it is a major precursor to O_3 and SOA formation [29-31]. The OH addition products, whose pathways account for approximately 90% in the reaction of toluene with OH [32, 33], subsequently react via three different pathways to produce o-cresol, bicyclic peroxy radicals, or aromatic oxides. However, the relative importance of the cresol and bicyclic peroxy radical pathways is still under debate. So far, the impact of different representations of the toluene oxidation chemistry on model predictions of SOA has not been studied.

Even though large uncertainties exist in modeled SOA concentrations, predicted SOA concentrations are not well constrained by observations. Techniques to apportion the observed total organic aerosol (OA) concentrations to POA and SOA include the minimum OC/EC ratio method [34] and the positive matrix factorization (PMF) analysis of the aerosol mass spectrums [35, 36]. The apportioned SOA concentrations, however, are not directly related to specific precursors. Most of the regional model evaluations of SOA predictions were carried out by comparing model predictions with these bulk observation-based SOA estimations [21, 37]. Few studies are reported to evaluate the precursor-specific SOA predictions with the corresponding observations.

Kleindienst, et al. [38] developed a widely-used precursor-specific tracer method to estimate the contributions of different precursors to ambient SOA concentrations. The mass fraction of the quantified precursor-specific tracers to total SOA formed from the precursor was determined in smog chamber experiments [38, 39]. Based on these source-specific ratios and the measured ambient concentrations of the SOA tracers, contributions of the precursor to ambient SOA can be estimated. However, since the ambient conditions are different from those in the chamber experiments, it remains unclear how accurate these tracer-based estimations of SOA are and whether the modeled precursor contributions to SOA can match these tracer-based estimations.

Among the tracers identified, 2,3-dihydroxy-4-oxopentanoic acid (DHOPA; C5H8O5) is a widely used tracer for SOA formed from monoaromatic hydrocarbons. Numerous studies have measured concentrations of DHOPA in various atmospheric environments and used it to estimate aromatic SOA concentrations.[11, 38-42] However, since the ambient conditions are different from those in the chamber experiments,[24, 25] the f_{SOA} determined in chamber experiments might not be directly applicable to estimate SOA in the ambient environment.

The chemical composition of isoprene SOA has been widely investigated in modeling studies and chamber experiments.[12, 43-45] Among the identified isoprene SOA species, 2-methyltetrols (2-MT, including 2-methylthreitol and 2-methylerythritol) and 2-methylglyceric acid (2-MG) are considered as unique tracer compounds,[38, 46] and are used to estimate the overall ambient isoprene-derived SOA,[47, 48] with the assumption that the mass fractions of the tracers in the ambient isoprene SOA (f_{SOA}, or f_{SOC} for secondary organic carbons, SOC) measured in the chamber experiments are similar to those in the ambient air. The current model was developed to simulate 2-MG and 2-MT formation in the aerosol liquid water, which are considered to be non-volatile [49, 50]. However, the scheme cannot explain the detection of these tracers in the dry condition chamber experiments [38, 46, 51-53] and ambient measurements in the gas phase [11].

Among of many compounds identified from the monoterpene SOA,[54] pinic acid (PA) and pinonic acid (PNA) are often considered as organic tracers for major bicyclic monoterpenes (α-pinene (APIN), β-pinene (BPIN), Δ^3-carene (CARN3), and sabinene (SABI)).[16, 38, 54, 55] 3-methyl-1,2,3-butanetricarboxylic acid (MBTCA) is also considered a unique tracer for α-pinene and β-pinene [56]. MBTCA is mainly generated from gas-phase oxidation of cis-PNA[57] and aqueous oxidation of PA may also contribute to it.[58] For sesquiterpene (SQT), the most commonly used organic tracer is β-caryophyllinic acid (BCARYA) [59], generated from the photochemical oxidation of β-caryophyllene, one of the major species in biogenic SQTs [60]. Although this source apportionment method is easy to apply, there are concerns about the validity of the assumption that the chamber-derived f_{SOA} could be applied under ambient conditions.[11, 61] One of the approaches to test this assumption is use CTMs to simultaneously predict the formation and gas-particle partitioning of the precursor-specific tracers and the total amount of SOA produced from a specific precursor.

The first objective of this research is to assess the potential increase of toluene SOA yield with the mechanism proposed by Ji et al using a photochemical box model. And the impact of this new mechanism on regional O_3 and SOA predictions is assessed using the Community Multiscale Air Quality (CMAQ) model, a regional chemical transport model developed by the United States Environmental Protection Agency (US EPA).

The second objective of this research is to compare the model predicted SOA derived from aromatic compounds and monoterpenes with the hourly and daily average concentrations of the specific SOA tracers and estimated the tracer-to-SOA ratio based on two different sets of model results using two different emission inventories. This study is the first evaluation of model-predicted precursor-resolved SOA with source-specific tracers.

The third objective of this research is to simulate DHOPA directly and compare the modeled concentrations with ambient measurements. The modeled DHOPA and aromatic SOA concentrations were used to calculate f_{SOA} and the results revealed that the organic aerosol loading significantly influences f_{SOA}. This is the first study to directly model DHOPA and assessed the f_{SOA} values used to estimate ambient aromatic SOA.

The fourth objective of this research is to expand the isoprene SOA model of Pye et al.[49] to simulate the formation of 2-MT and 2-MG by treating these two species as semivolatile and including a non-aqueous formation pathway. This is the first regional chemical transport modeling study to evaluate the potential volatility and the non-aqueous formation of the two isoprene markers in relatively polluted atmospheres.

The fifth objective of this research is to simulate the concentrations of the monoterpene and sesquiterpene tracers along with monoterpene and SQT SOA. This is first study to evaluate the CTM capability in reproducing the observed organic tracer concentrations from monoterpenes and SQTs against ambient measurements. The modeled spatial and temporal variability of the mass fraction f_{SOA} are evaluated against ambient measurement data over a large geographical area impacted by different levels of anthropogenic emissions.

The summary of SOA tracers simulated in this study is listed in Table 1-1.

Table 1-1 Organic tracers simulated in the research.

Tracer compounds	Abbrev.	Parent VOCs	f_{SOA}
2,3-dihydroxy-4-oxopentanoic acid	DHOPA	Aromatics	0.002
2-methyltetrols 2-methylglyceric acid	2-MT 2-MG	Isoprene	0.01 – 0.02
Pinic acid Pinonic acid	PA PNA	α-pinene + β-pinene	0.033
3-methyl-1,2,3-butanetricarboxylic acid	MBTCA	Monoterpenes	0.015
β-caryophyllinic acid	BCARYA	Sesquiterpenes	0.018

In conclusion, this study will aid in understanding roles of SOA formation in the ambient environment. The f_{SOA} value for each precursor-specific SOA will be assessed using CTMs and provide useful information on the adjustment of f_{SOA} based on the difference of conditions in the chambers and ambient air. This study will improve the estimation of precursor-specific SOA using organic tracer field measurements.

2 ASSESSING THE UNCERTAINTIES IN OZONE AND SOA PREDICTIONS DUE TO DIFFERENT BRANCHING RATIOS OF THE CRESOL PATHWAY IN THE TOLUENE-OH OXIDATION MECHANISM *

Oxidation of toluene by OH radicals plays a significant role in forming ozone (O_3) and secondary organic aerosol (SOA) in polluted urban atmospheres. However, the branching ratio of the cresol formation pathway after OH addition to the aromatic ring remains uncertain, affecting model predictions of O_3 and SOA. In this study, SOA formation under low (18%) and high (48%) cresol branching ratio conditions are determined by modeling chamber experiments on toluene SOA formation, using a photochemical box model with the semi-explicit Master Chemical Mechanism (MCM) v3.2 and an SOA module for the equilibrium gas-to-particle partitioning of semi-volatile products. The modeled SOA concentrations are fitted to determine the SOA yields and saturation concentrations using the classical two-product representation. These parameters are then applied in the Community Multiscale Air Quality (CMAQ) model to assess the impact of the cresol branching ratio on SOA formation. The reaction products of ARO1 (the lump species that includes mostly toluene) with OH are also modified to reflect the higher cresol branching ratio. Two sets of CMAQ simulations for China (C0, with low SOA yields and unmodified ARO1 + OH reaction, and C1, with high SOA yields and modified ARO1 + OH reaction) are conducted for January and July 2013. Predicted monoaromatic compound concentrations in major urban areas are ~4-7 ppb in January and ~1.5-3 ppb in July, which generally agree with measurements. The

---

* Reprinted with permission from “Assessing the Uncertainties in Ozone and SOA Predictions due to Different Branching Ratios of the Cresol Pathway in the Toluene-OH Oxidation Mechanism” by Zhang, J., Choi, M., Ji, Y., Zhang, R., Zhang, R., and Ying, Q., 2021. ACS Earth and Space Chemistry. 5, 8, 1958-1970, Copyright [2021] by American Chemical Society.

higher cresol branching ratio simulations leads to slightly lower OH radicals and O_3 predictions. Less than 1 ppb decrease of monthly average daily maximum 8-hour (DM8H) and peak hour O_3 is found in the urban areas in July and in broader spatial coverage in January. Increase in January ARO1 SOA is approximately 1.2 μg m^{-3}, which corresponds to a relative increase of 40-70% to ARO1 SOA or a ~10% increase of total SOA. This change reflects the combined effect of increasing ARO1 SOA due to higher yields and reduced formation of semi-volatile organic products and glyoxal and methylglyoxal due to lower OH radicals.

2.1 Introduction

Photochemical oxidation of volatile organic compounds (VOCs) by hydroxyl (OH) radical results in tropospheric ozone (O_3) and secondary organic aerosol (SOA) formation.[62-64] Toluene is one of the most abundant anthropogenic VOCs in the urban atmosphere. Emitted from a variety of sources, including vehicular exhaust, fuel evaporation, biomass burning, and solvent usage,[26-28] it is a major precursor to O_3 and SOA formation.[29-31] It is generally accepted that the reaction of toluene with OH mainly proceeds with the addition pathway (~90%), while the H-abstraction reaction (from the methyl group) accounts for approximately 10%.[32, 33] The OH addition products subsequently react via three different pathways to produce cresol, bicyclic peroxy radicals, and aromatic oxides. The o-cresol and bicyclic peroxy radical pathways are considered more important than the aromatic oxide pathway (~10%). However, the relative importance of the cresol and bicyclic peroxy radical pathways is still under debate. Several chamber experiments reported that the cresol pathway is relatively minor. For example, Klotz, et al. [65] measured a cresol yield of 18%, which is currently used as the cresol pathway branching ratio in the MCM mechanism, together with a branching ratio of 65% for the bicyclic peroxy radical pathway. While similar cresol yields between 15% - 18% were found in other experimental studies,[66, 67] higher cresol yields between

25% - 28% were also reported previously.[68, 69] Recently, Ji et al.[70] reassessed the toluene-OH oxidation mechanism using both chamber experiments and quantum mechanical calculations and proposed a lower limit of the cresol branching ratio of approximately 48%. In a separate study, Qi et al. [71] recommended an overall cresol branching ratio of 42-53%, based on the real-time spectrum data from a single photon ionization time-of-flight mass spectrometer (SPIMS) and a single particle aerosol mass spectrometry (SPAMS) instruments.

Differences in how models represent the toluene OH oxidation pathways can affect their predictions of O_3. Sarwar et al.[72] used a regional atmospheric modeling system to evaluate the difference in O_3 prediction due to a new condensed toluene mechanism in Carbon Bond 05 (CB05) mechanism with a lower cresol yield (18%) than the original mechanism (36%). An increase of 1-2 ppb of monthly mean daily maximum 8-h (DM8H) O_3 in populated US cities was estimated, and the higher O_3 reduced the under-prediction biases in the urban areas. However, they only tested the changes under summer conditions. It is still unclear how sensitive the modeled results are under wintertime NO_x saturated conditions in polluted atmospheres. Differences in the representation of toluene OH oxidation products are also expected to affect the formation of SOA. It has been shown that cresol can form highly oxygenated low-volatility products, and account for 20-40% of the SOA produced from toluene.[29] It is expected that, for the models that use detailed oxidation products to predict SOA formation,[73] higher cresol yields could lead to higher model predictions of toluene SOA.[29] For the models that use lumped semi-volatile products whose yields are derived from chamber experiments, the difference in cresol yields can still affect the predicted SOA concentrations indirectly due to changes in the concentrations of OH and other oxidants. So far, the impact of different representations of the toluene oxidation chemistry on model predictions of SOA has not been studied.

In this study, we first used a photochemical box model to assess the potential change of toluene SOA yield with the mechanism proposed by Ji et al.[70] We incorporated the reported branching ratio of the cresol pathway in the box model to evaluate how well the model can reproduce the observed SOA yields from toluene oxidation under high and low NO_x conditions in chamber experiments. We then derived parameters for modeling SOA in regional models based on the predicted SOA yields with a higher cresol branching ratio in the box model simulations and assessed the impact of this new mechanism on regional O_3 and SOA predictions using the Community Multiscale Air Quality (CMAQ) model, a regional chemical transport model developed by the United States Environmental Protection Agency (US EPA). The sensitivity of the model predictions of toluene SOA to important parameters such as the saturation vapor pressure is explored.

2.2 Methods

2.2.1 The photochemical box model simulation of SOA formation from toluene

The photochemical box model used in this study to simulate SOA formed from photooxidation of toluene by OH is based on the Master Chemical Mechanism (MCM) v3.2[33, 74, 75] for the gas phase photochemical reactions and is equipped with an equilibrium gas-to-particle partitioning SOA module. The MCM is a semi-explicit photochemical mechanism representing the oxidation of 142 non-methane hydrocarbons using detailed reactions and reaction products. It has a total of 16892 reactions (excluding inorganic reactions) and 5710 organic species. The toluene mechanism includes 774 reactions and 256 organic species. The initial oxidation of toluene by OH includes four branches. The branching ratio for the methyl group H-abstraction pathway is 7%, as shown in reaction (R1),

$$C_6H_5CH_3 + OH \overset{O_2}{\rightarrow} C_6H_5CH_2O_2,\ k=1.8\times10^{-12}\exp(340/T)*0.07 \qquad \text{(R1)}$$

T is the temperature in K and k is the reaction rate coefficient in units of cm^3 $molecule^{-1}$ s^{-1}. Reaction (R2) is the cresol formation pathway and has a branching ratio of 0.18,

$$C_6H_5CH_3 + OH \overset{O_2}{\rightarrow} CRESOL + HO_2,\ k=1.8\times10^{-12}\exp(340/T)*0.18 \qquad (R2)$$

The SOA module calculates the equilibrium partitioning of individual semi-volatile products generated during the oxidation of parent VOCs using the saturation vapor pressures estimated by the EPI Suite program from the US EPA.[76] It also considers the activity coefficients of SOA constituents using the UNIFAC model.[77] Details of the MCM-SOA module have been described by Li et al.[73] Particle and vapor wall losses are not considered in the current model because the yield data used in the analysis (Section 2.2.1) have already been corrected to account for wall losses.

In this study, the branching ratios of four toluene primary oxidation pathways in the MCM model are changed to match those reported by Ji et al.[70] As shown in Table 2-1, the branching ratio for the cresol pathway is increased from 18% to 47.9% and the methyl group H-abstraction pathway is increased to 11.3%, based on the reported benzaldehyde yield in Ji et al.[70] The other two pathways are reduced accordingly so that the total branching ratios add to 100%.

Table 2-1 Branching ratios of the original and modified toluene reaction with hydroxyl radical in MCM v3.2.

Toluene primary oxidation pathway		Branching ratio	
		Original	New
TOLUENE + OH =	C6H5CH2O2	7%	11.3%
	CRESOL + HO2	18%	47.9%
	TLBIPERO2	65%	35.4%
	TLEPOXMUC + HO2	10%	5.4%

Chamber experiments of SOA formation from photooxidation of toluene reported by Hildebrandt et al.[31] under high-NO_x (experiments 2-6) and low-NO_x conditions (experiments 7-

11) are simulated using the original and the modified MCM mechanism with SOA formation. The initial concentrations of toluene and NO, H_2O_2 (as OH source), and chamber conditions (temperature and relative humidity) are strictly based on the reported values and are summarized in Table S2-1 in Appendix A.

The photolysis rate of NO_2 (J_{NO2}) is not described in the chamber experiments by Hildebrandt et al., but a value of 0.06 min^{-1} is reported in a previous study using the same chamber.[78] However, the conditions of the black lights used in prior experiments might not be the same as those used by Hildebrandt et al.[31] Thus, the J_{NO2} in the MCM box model (with the original toluene + OH reaction) to simulate the high-NO_x chamber experiments is adjusted to minimize the sum of squared errors (SSE) between the predicted and reported OH concentrations in Hildebrandt et al.[31] The optimum J_{NO2} is found to be 0.0767 min^{-1} (Figure S2-1 in Appendix A), and the predicted OH in each experiment generally agrees with the experimental values (Figure S2-2 in Appendix A). More detailed discussions regarding the photolysis rates used to simulate the chamber experiments are included in Section S2.1 in Appendix A.

The low-NO_x experiments are simulated using the same optimized J_{NO2}. The comparison between predicted and estimated OH concentrations is shown in Figure S2-2 in Appendix A. The OH concentrations in the chamber experiments are in the range of $0.5\text{-}3.5\times10^6$ molecules cm^{-3}. All the simulated OH in the low-NO_x cases are within a factor of 0.5 to 2 of the reported OH.

The SOA mass yields (Y, defined as $Y = C_{OM}/\Delta VOC$, where C_{OM} is the SOA concentration) at C_{OM} of 10 and 20 μg m^{-3} are determined for each high-NO_x experiment and compared with vapor and particle wall-loss corrected chamber SOA data, as shown in Figure 2-1. With the original toluene-OH oxidation branch ratio, the box model simulations underpredict Y with a mean bias (MB) of -0.105. The agreement between observed and predicted Y is improved significantly (MB = 0.015) when the updated branching ratios are used. The modified toluene mechanism also leads to higher Y values for the low-NO_x experiments, although both the original and the modified mechanisms underpredict the observed yields, as shown in Figure S2-3 in Appendix A. SOA from glyoxal (GLY) and methylglyoxal (MGLY) surface uptake[73] is included in the MCM simulation. However, they only account for less than 2% of the SOA formed because the chamber relative humidity (RH) values are all less than 20%, far below the deliquescence RH of the seed particles.

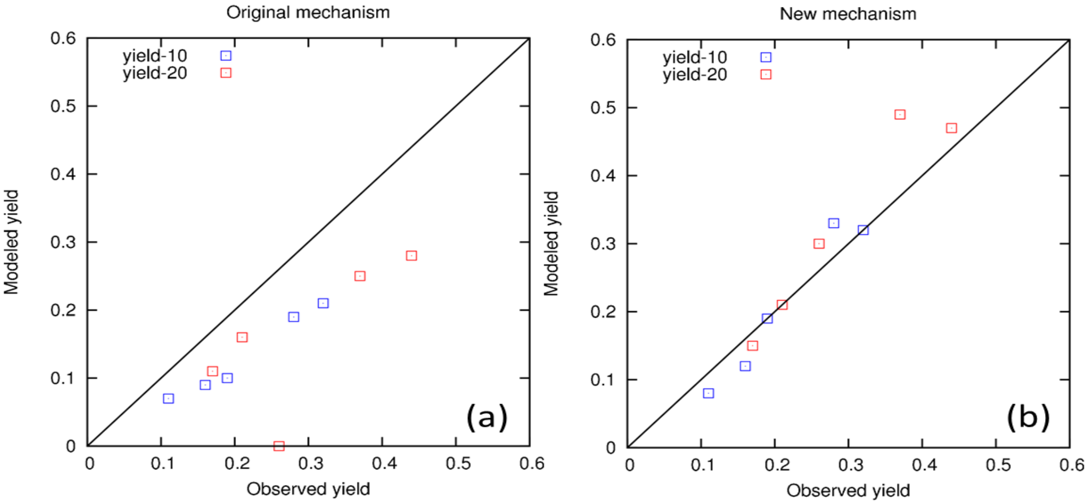


Figure 2-1 Observed and modeled toluene SOA yields under high-NO_x conditions using (a) the original toluene + OH mechanism and (b) the modified mechanism with a higher branching ratio for the cresol pathway. Yield-10 and yield-20 represent SOA yield when the SOA concentration in the experiments or the model simulations are 10 and 20 μg m^{-3}, respectively. The outlier on (a) is because the modeled SOA is less than 20 μg m^{-3}.

To assess the change in the branching ratios on SOA predictions in regional scales, we fit the calculated SOA yields from the simulations of the chamber experiments at 298 K to the two-product (2p) model that describes the equilibrium partitioning of semi-volatile organic vapor between gas and particle organic phases,[79] as shown in Equation (2.1),

$$Y = \frac{\alpha_1}{1 + C_1^*/c_{OM}} + \frac{\alpha_2}{1 + C_2^*/c_{OM}} \qquad (2.1)$$

where α_1 and α_2 are the mass-based stoichiometric yields of semi-volatile products, and C_1^* and C_2^* are the saturation mass concentration (μg m^{-3}) of two lumped SVOC products to be determined. These parameters are used to replace the existing parameters in a regional chemical transport model, as described in Section 2.2.2. The detailed results of the photochemical box model simulation of the chamber experiments are described in the Results section.

2.2.2 Regional air quality model

The Community Multiscale Air Quality (CMAQ) model v5.0.1[80] with the SAPRC-11 photochemical mechanism[81] and AERO6 aerosol module is modified to evaluate the impact of the higher branching ratio for cresol formation in the toluene oxidation on regional O_3 and SOA.

In this version of the SAPRC-11, toluene is included in the lumped species ARO1 (Aromatics with $k_{OH} < 1.36 \times 10^{-11}$ cm^3 $molec^{-1}$ s^{-1}). The rate constants and product yields of the ARO1 + OH reaction in the original mechanism are based on an assumed mixture of species lumped into ARO1.[81] The yields of the major products are listed in Table 2-2 as Case 0. To assess the impact of a higher cresol branching ratio on O_3, Dr. William Carter of the University of California, Riverside (personal communication), the developer of the SAPRC-11 mechanism, provided a modified ARO1 + OH reaction by combining the OH ring addition and non-ring pathways and assuming that the formation of cresol (from toluene) and other phenolic compounds (from other monoalkylbenzenes) is the only fate of the OH ring addition pathway. The molar yields of major products of this theoretical upper-limit case are also listed in Table 2-2.

To make the reaction more consistent with the modified MCM box model toluene yield of 47.9%, we update the ARO1 + OH yield of cresol and other related products using a weighted average of the yields for the original mechanism in Case C0 (with a weighting factor w_1) and the theoretical upper-limit case (weighting factor w_2). The weighting factors $w_1 = 0.604$ and $w_2 = 0.396$ are calculated based on the original cresol yield of 0.18 and the upper-limit cresol yield of 0.935 for the toluene reaction with OH, which was also provided by Dr. Carter, so that $0.180w_1 + 0.935w_2 = 0.479$, and $w_1 + w_2 = 1$. The molar yields of this modified ARO1 + OH reaction are included in Table 2-2 as Case C1.

Table 2-2 Molar yields of major ARO1 + OH reaction products used in three different CMAQ simulation cases

Products	Product yields		
	Original (Case C0)	New[^] (Cases C1, C2)	Upper-limit[$]
RO2XC	0.089	0.059	0.014
RO2C	0.622	0.415	0.099
HO2	0.209	0.477	0.887
CRES	0.135	0.365	0.717
XYNL	0.032	0.087	0.170
xGLY	0.268	0.162	0.000
xMGLY	0.231	0.140	0.000
xHO2	0.612	0.396	0.065[%]
xAFG1[#]	0.255	0.154	0.000
xAFG2	0.244	0.147	0.000
yRAOOH[§]	0.567	0.342	0.000
OH	0.084	0.051	0.000
AFG3[&]	0.084	0.051	0.000
AFG5[*]	0.042	0.025	0.000

[^] The new mechanism was interpolated between the original case and the upper-limit case.

[$] The upper-limit estimation assumed all OH + ring adducts produce cresol (toluene) or other phenolic compounds (for other monoalkylbenzenes).

[%] Includes contributions from non-ring reactions; Non-ring xHO2 yield is 6.53% based on SAPRC mechanism generator for toluene (available at http://mechgen.cert.ucr.edu/).

[#] Monounsaturated dialdehydes or aldehyde-ketones formed from peroxy radical reactions with NO and NO_3 and RO_2 (xAFG1 for most photoreactive, and xAFG2 for least photoreactive).

[§] Organic hydroperoxides formed following RO_2+HO_2 reactions, or formation of H-shift disproportionation products.

[&] Di-unsaturated dicarbonyl aromatic fragmentation products that are assumed not to photolyze rapidly.

[*] 3-hexene-2,5-dione and other monounsaturated diketone aromatic products.

In CMAQ model v5.0.1, SOA formation from ARO1 is treated using the 2p model approach, as described in Section 2.2.1. To assess the impact of the higher cresol yield on SOA in models with an explicit representation of the toluene oxidation products using the lumped model approach, the SOA yield parameters from the original and modified MCM mechanisms for the photooxidation of toluene are applied in model simulations. A detailed description of the determination of the parameters is included in Section 2.3.1. The yield parameters in the default CMAQ mechanism based on Ng et al.[30] are not used in the study, as the objective of this study is to evaluate the potential changes in SOA due to changes in the cresol yield.

In addition, the CMAQ model used in this study also includes the formation of non-volatile oligomers from the lumped semi-volatile products[82] and extensions to model SOA formation from reactive surface uptake of GLY and MGLY.[83] The model uses a precursor tracking scheme[84] to determine the complete contribution to SOA from individual precursors using precursor-specific reactive species in the expanded gas and aerosol mechanisms. An illustration of the scheme for ARO1 is shown in Figure 2-2.

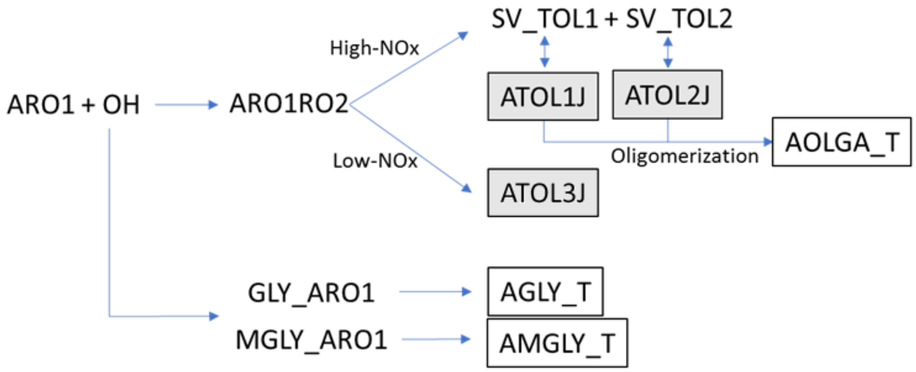


Figure 2-2 The precursor-tracking scheme to model SOA formation from ARO1 in the CMAQ model. The shaded boxes represent products from the traditional 2-product model. AGLY_T and AMGLY_T are SOA products from irreversible surface uptake of glyoxal and methylglyoxal, respectively. SOA formation from other major precursors is also tracked similarly.

2.2.3 Model application

The CMAQ model is applied to simulate air quality in China in January and July 2013 to study the impact of the toluene chemistry on O_3 and SOA in winter and summer. Two sets of simulations are performed for each month. The first simulation (base case, Case C0) uses the unmodified ARO1 + OH reaction, and the ARO1 SOA yields are derived from the original MCM mechanism. In the second simulation, ARO1 + OH is based on Case C1 described in Section 2.2.2, and the SOA yields are from the modified MCM mechanism with the higher cresol branching ratio.

Previous studies showed the significance of the cresol pathway contributing to toluene SOA.[29] As cresol does not form SOA in this version of the CMAQ model directly, the changes in the product yields listed in Table 2-2 do not directly affect SOA predictions, although they indirectly affect SOA formation from changes in OH and O_3 concentrations. Another sensitivity case (Case C2), where the gas phase mechanism is based on Case C1 but the ARO1 SOA yields are the same as those in Case C0 is also included in the study. This is designed to isolate the impacts on increased SOA yields and the reduced gas phase reactivity on predicted SOA concentrations.

The 197×127 36×36 km^2 spatial resolution domain covers China and the surrounding regions. Eighteen stretching vertical layers with a first layer height of approximately 35 m reach a model top nearly 20 km above the surface. Meteorological and emission inputs to the model have been described by Hu et al.,[85] and a summary is provided in the following. Emissions in China are generated using the Multiscale Emission Inventory of China (MEIC) v1.0, which is based on activities in the base year of 2012. Emissions from other countries in the model domain are generated using the Regional Emission inventory in Asia version 2 (REAS2). The MEIC and

REAS2 inventories are in 0.25x0.25^{o} grids and are re-projected to the Lambert conformal coordinates in 36x36 km^2 resolution. Windblown dust emissions in the entire domain are generated by CMAQ inline module.[86] Biogenic emissions are generated using the Model for Emissions of Gaseous and Aerosols from Nature (MEGAN) v2.10.[87] The Weather Research and Forecasting model v3.6 are used to generate the meteorological inputs to the CMAQ model. Initial and boundary conditions are generated using the vertical profiles distributed with the CMAQ model. The first 5 days of the simulation results are treated as spin-up and are excluded in the final analysis.

The emission inventory and meteorological fields have been used in several previous studies using the same host CMAQ model for O_3, NO_x, CO, $PM_{2.5}$, secondary inorganic aerosol, and SOA in China during 2013.[85, 88, 89] The performance of the model has been extensively evaluated against observations in these studies. In general, the model can reproduce the observed concentrations of O_3 and $PM_{2.5}$ in most urban areas. Model performance evaluation for the primary and secondary inorganic components is done mostly for Beijing due to limited observation data. Concentrations of total SOA in January 2013 are still underestimated.[90]

To further evaluate the model capability in reproducing the observed high $PM_{2.5}$ concentrations in January and July 2013, we obtained the hourly $PM_{2.5}$ concentrations in five major Chinese cities measured at the US Consulates (downloaded from https://www.airnow.gov/international/us-embassies-and-consulates/) and compared with predicted concentrations from Case C1, as shown in Figure S2-4 and Figure S2-5, for January and July, respectively. Predicted concentrations show reasonable agreement with the measurements, as indicated by the model performance statistics shown in Table S2-2 in Appendix A with Mean Fractional Biases (MFB) between -0.55 and 0.06 and Mean Fractional Errors (MFE) less than 0.7, and can capture high concentrations up to 400-500 μg m^{-3}. This agreement provides further evidence that the emissions and meteorological conditions and the formation of secondary $PM_{2.5}$ are generally well represented. The relatively larger underprediction of $PM_{2.5}$ in Beijing than those in the other cities, as indicated by the MFB, is mostly caused by the failure of the model in predicting the extremely high concentrations of $PM_{2.5}$ ~900 μg m^{-3} in Beijing on January 12-13. To the best of the authors' knowledge, none of the published studies are yet able to reproduce this high $PM_{2.5}$ pollution event.

2.3 Results

2.3.1 Determination of two-product SOA model parameters

In this study, the 2p model parameters from the original and modified MCM are used in the regional simulations to assess how the increased cresol yield would affect regional SOA predictions without the necessity of using the CMAQ-MCM-SOA model of Li et al.[73] The five high-NO_x and six low-NO_x chamber experiments are simulated using a standard temperature of 298 K instead of the reported chamber temperatures to determine the SOA yield parameters.

Figure 2-3 shows the simulated SOA yield data from all five high-NO_x simulations, as well as the 2p model parameters determined using the MATLAB curve fitting app (cftool). For the SOA yields simulated by the original MCM toluene mechanism, α_1 and α_2 are 0.239 and 0.738, respectively. The two products have drastically different C* values of 10.15 and 2147 μg m^{-3}, making the second product essentially always in the gas phase under realistic atmospheric loadings of organics. The data from the modified MCM, when fit to the 2p model, as shown in equation (1), lead to two products with very similar C*. Thus, they are finally determined using a one-product representation with α_1 = 0.770 and C_1^* = 21.1 μg m^{-3}. In the range of C_{OM} from 5 to 50 μg m^{-3}, the high-NO_x toluene SOA yields predicted by the modified MCM mechanism are approximately 1.8-

2.5 times those of the original mechanism. Both the original and modified MCM mechanisms predict higher SOA yields than the default parameters used in the CMAQ model, based on those experiments of Ng et al. [30] The modified MCM mechanism predicts SOA yields 2.4-4.0 times of those based on the default CMAQ parameters in the C_{OM} range of 5-50 µg m^{-3}.

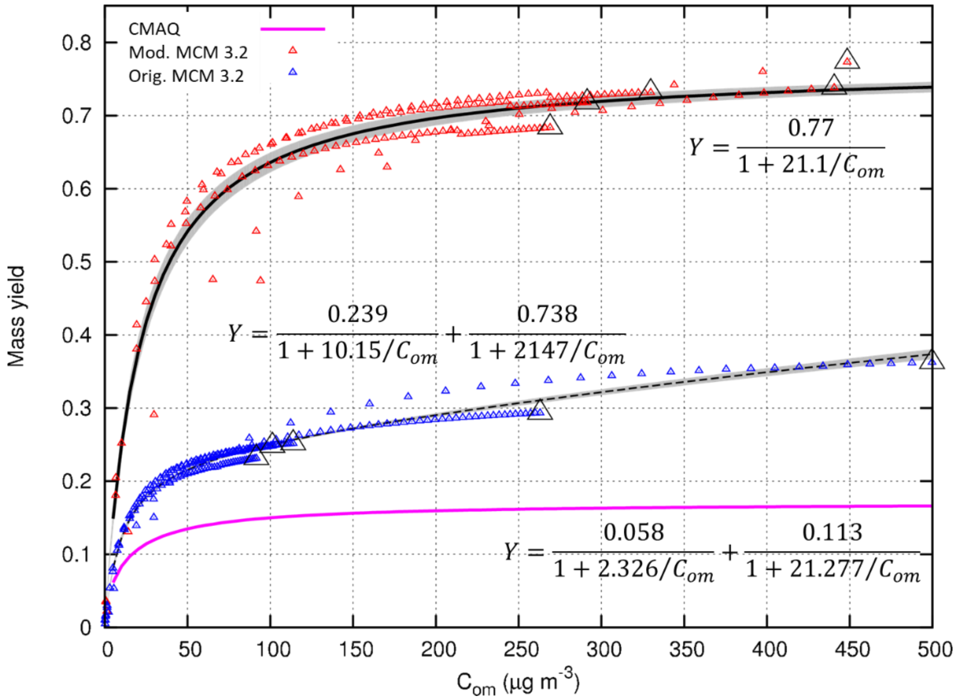


Figure 2-3 Parameters for ARO1 SOA yield under high-NO$_x$ condition used in the original CMAQ model (pink line) and those derived based on the photochemical box model simulations using the original MCM and the modified MCM mechanism with a higher branching ratio for the cresol pathway. Small triangles are 10-min resolution yield data from the simulations, and the large triangles show the yields at the end of each simulation.

SOA formation under low-NO$_x$ conditions is much slower than that under high-NO$_x$ conditions, and neither the original nor modified MCM mechanism could correctly predict the observed SOA yields in the chamber experiments. As shown in Figure S2-3 in Appendix A, the predicted SOA yields from the original MCM mechanism can be fitted with a one-product model with α_1=0.1 and C_1^*=9.592 μg m^{-3}, while the data from the modified mechanism can be described using a non-volatile product with a fixed mass yield of 0.179. However, it is still 50% lower than the fixed mass yield of 0.360 used in the CMAQ model. Table S2-3 in Appendix A shows a summary of the ARO1 SOA yields used in this study.

An additional set of box model simulations is conducted using data from Ng et al.[30] to illustrate that the MCM-SOA mechanism can simulate SOA formation under different experimental conditions. A more detailed discussion of this evaluation is included in Section S2.3 of Appendix A.

2.3.2 Compare with observations

Before estimating the impacts of the changes in ARO1 SOA yields on regional O_3 and SOA, it is necessary to assess if the current model gives reasonable estimations of the concentrations of monoaromatic precursors. However, direct observations of monoaromatic compounds are very limited for the year 2013. Therefore, to provide an understanding of the general range of the concentrations of these species, we summarize the observed concentrations of toluene and ethylbenzene (major species in ARO1) and xylenes (o+m/p, major species in ARO2) in six Chinese cities (see Figure 2-4a for the locations of the cities) in recent years, as shown in Table S2-4 in Appendix A. In most cities, the toluene concentrations are between 1-5 ppb, and ethylbenzene concentrations are approximately 20% to 40% of toluene. The concentrations of xylenes are usually lower than toluene, in the range of 0.5-2.0 ppb. For the same city, concentrations in urban locations are usually higher than those in suburban locations.

The predicted monthly ARO1 and ARO2 concentrations in seven major cities from Case C1 are listed in Table S2-5 in Appendix A. In January, ARO1 and ARO2 concentrations are approximately 3-4 ppb and 1.5-2.5 ppb, respectively. Concentrations in the summer are lower, with ARO1 and ARO2 in the ranges of 1.5-2.0 ppb and 0.7-1.4 ppb, respectively. These concentrations are in general agreement with the observations. However, as intensive air pollution controls have been implemented after 2013 throughout the country, concentrations in 2013 are expected to be higher than in later years. Thus, concentrations of the aromatics are likely underestimated, which can lead to the underestimation of toluene SOA.

Table 2-3 Observed toluene (monoaromatics) SOA concentrations in China in recent years based on the organic tracer DHOPA*

Location	Season	Conc. (μg m^{-3})	Reference
Beijing (urban)	Summer, 2008	3.3	Guo et al.[91]
Beijing (rural)	Summer, 2008	3.1	
Shanghai	Summer, 2014	4.6	Gao et al.[92]
Shanghai	Winter, 2018	3-23	He et al.[93]
Shanghai	Jan. 2010/2011	0.06-0.08	Feng et al.[94]
	Apr.-May, 2010	0.14-0.28	
	Jul. 2010	0.48-0.52	
	Oct.-Nov., 2010	0.48-0.66	
Guangzhou	Summer, 2008	4.5	Ding et al.[95]
Guangzhou	Fall-Winter, 2008	3.3	
Beijing (urban)	Summer, 2016	3.3	Tang et al.[96]
Beijing (rural)	Summer, 2016	2.2	
Tianjin (urban)	Summer, 2018	0.22	Wang et al.[97]
Tianjin (rural)	Summer, 2018	0.36	
Tianjin (urban)	Autumn, 2018	0.44	
Tianjin (rural)	Autumn, 2018	0.52	

* In all studies, the organic tracer 2,3-Dihydroxy-4-oxopentanoic acid (DHOPA) was used to estimate SOA concentrations. Some studies considered this as toluene SOA while others reported this as SOA from aromatic compounds. The concentrations in the original references are secondary organic carbon (SOC) and are converted to SOA using an OM/OC ratio of 2.0.[38]

2,3-Dihydroxy-4-oxopentanoic acid (DHOPA) is considered as a specific tracer for SOA formed from the oxidation of toluene[38] and later studies show that it can also form from other monoaromatic compounds.[39] DHOPA has been measured in field studies to estimate aromatic SOA in several Chinese cities, as summarized in Table 2-3. Among the seven references, five report toluene (or aromatics) SOA concentrations of 2-5 µg m^{-3}, and in Shanghai, the concentration can be as high as 23 µg m^{-3} on polluted days. These reported concentrations are in closer agreement with the predictions from Case 1, as shown in Table 2-4. It should be noted that the moderate difference between the two mechanisms and the large uncertainty of the toluene SOA estimation from DHOPA, the comparison of the predicted toluene SOA against these tracer-based estimations are not sufficient to draw a conclusion which mechanism is better in predicting toluene SOA.

Table 2-4 Predicted monthly average SOA concentrations from ARO1 (Case C1 and C0) and ARO2 (Case C1 only) in 7 cities in China. Units are µg m^{-3}.

	January			July		
Site	ARO1 C1	ARO1 C0	ARO2 C1	ARO1 C1	ARO1 C0	ARO2 C1
Beijing	0.50	0.34	1.22	1.12	0.79	1.21
	(0.23,0.71)*	(0.16,0.48)	(0.47,1.92)	(0.90,1.50)	(0.64,1.06)	(0.95,1.75)
Shenyang	0.80	0.47	1.64	0.79	0.56	0.92
	(0.51,0.96)	(0.31,0.57)	(1.03,2.08)	(0.73,0.84)	(0.51,0.59)	(0.80,1.05)
Nanjing	1.90	1.28	2.88	0.53	0.37	0.86
	(1.72,2.09)	(1.17,1.40)	(2.58,3.18)	(0.43,0.72)	(0.30,0.50)	(0.76,1.11)
Chengdu	2.57	1.69	3.65	1.15	0.83	1.50
	(1.44,3.14)	(0.96,2.09)	(1.90,4.29)	(0.53,1.15)	(0.39,0.83)	(0.67,1.50)
Guangzhou	1.98	1.38	2.35	0.25	0.21	0.37
	(1.85,2.06)	(1.28,1.43)	(2.15,2.40)	(0.07,0.34)	(0.05,0.28)	(0.13,0.47)
Shanghai	1.04	0.74	1.53	0.78	0.55	1.18
	(0.93,1.23)	(0.67,0.87)	(1.38,1.84)	(0.66,0.97)	(0.46,0.68)	(1.06,1.50)
Zhengzhou	1.65	1.08	3.14	1.02	0.74	1.19
	(1.37,1.98)	(0.91,1.30)	(2.54,3.61)	(0.80,1.02)	(0.58,0.74)	(0.98,1.19)

*Numbers in the parenthesis represent the range of concentrations within the 3×3 grid cells with the urban center in the middle.

In July, average concentrations in Shanghai are between 0.7-1 µg m^{-3} for SOA from ARO1, and 1.7-2.5 µg m^{-3} if SOA from ARO2 is included. This is slightly lower than the monoaromatic

SOA concentrations measured in Shanghai (~4.6 μg m^{-3}) using the tracer-based method on August 4-7, 2014[92]. The predicted monthly average SOA concentration from ARO1 in Beijing also range from approximately 1-1.5 μg m^{-3} (1.6-3.2 μg m^{-3} if ARO2 SOA is included), which is also close to the SOA-tracer based estimation of approximately 3 μg m^{-3} in Beijing in early summer 2016 [96]. In Guangzhou, Ding et al. [95] estimate that the SOA due to aromatics is about 4.5 μg m^{-3} in summer 2008. The CMAQ-predicted monthly-average SOA from ARO1 and ARO2 in Guangzhou are both ~0.1-0.5 μg m^{-3}, lower than the estimation by Ding et al. Two of the studies [94, 97] report summer aromatics SOA concentrations in Shanghai and Tianjin in the range of 0.2-0.5 μg m^{-3}, which is almost a factor of 10 lower than other reported values. For wintertime monoaromatic SOA, Ding et al. [95] estimate that the average aromatics SOA concentration in fall-winter Guangzhou is approximately 3 μg m^{-3}, which is similar to the predicted ARO1 SOA of ~2 μg m^{-3}, but is lower than the prediction if SOA from ARO2 (~2.5 μg m^{-3}) was also included. He et al. [93] find that aromatics SOA concentrations in Shanghai vary from ~3 μg m^{-3} on relatively clean days to 20 μg m^{-3} on very polluted days. In comparison, the predicted monthly ARO1+ARO2 SOA in winter Shanghai is approximately 2.5-3 μg m^{-3}.

2.3.3 Impacts on regional ozone

Figure 2-4a and 4b illustrate the severe O_3 pollution in the North China Plain (NCP) in July 2013. The highest monthly average of daily maximum 8-hour (DM8H) and peak hour O_3 concentration in Case 1 reaches 83 and 90 ppb, respectively. The modified ARO1 reaction with OH in SAPRC-11 leads to a small decrease in DM8H and peak hour O_3 in large urban areas with large emissions of toluene and their immediate downwind regions where high O_3 concentrations are predicted, mostly located along the east coast of China from the Yangtze River Delta (YRD) region to the Bohai Bay region north of the Shandong Peninsula. The decreases of DM8H and peak hour O_3 in the polluted regions are higher than 0.5 ppb (~1%), and could be large as 0.7 ppb and 0.9 ppb, respectively. The decreases of O_3 transition to no obvious changes as one moves away from the urban center. The change of O_3 due to the modified ARO1 reaction correlates with changes in the OH radical concentrations. Figure S2-6 in Appendix A shows a relative OH decrease of ~1% and a maximum of 2.1% in O_3 in large urban areas. The small decrease of OH in large urban areas is expected because of the slightly lower yields of OH radicals and reactive ring opening products such as the di-carbonyl aromatic ring opening products (AFG3) in the modified reaction.

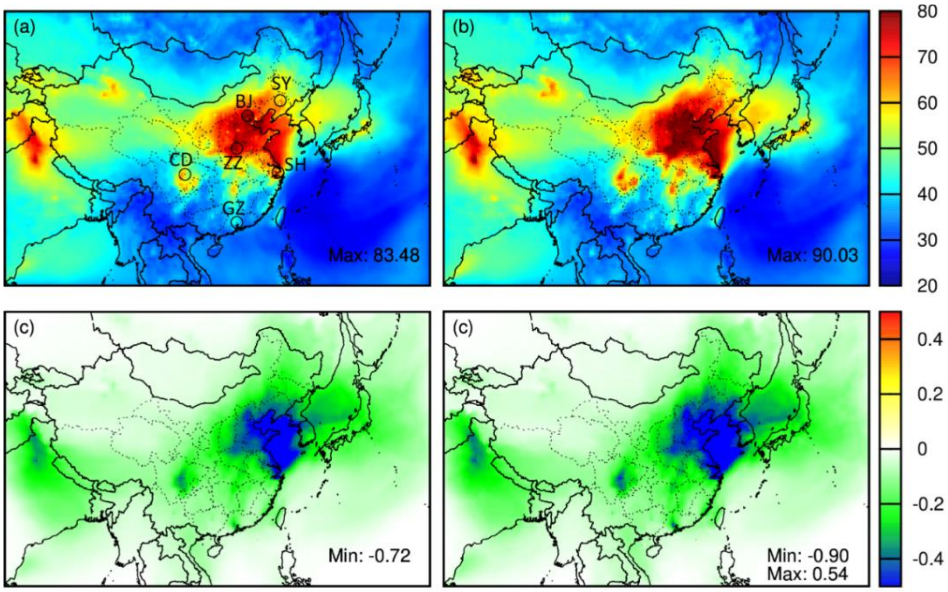


Figure 2-4 Monthly average daily maximum 8-h O_3 (DM8H O_3) concentrations (a) and daily peak hour O_3 (b) predicted with the modified SARPC-11 mechanism (Case 1) for July 2013; Difference of the monthly average DM8H O_3 (c) and peak hour O_3 (d) between the modified and original SAPRC-11 mechanisms (Case 1 - Case 0) for July 2013. Units are ppb. Relative difference of monthly average DM8H O_3 (e) and peak hour O_3 (f) between the modified and original SAPRC-11 mechanisms ((Case 1 – Case 0)/Case 0).

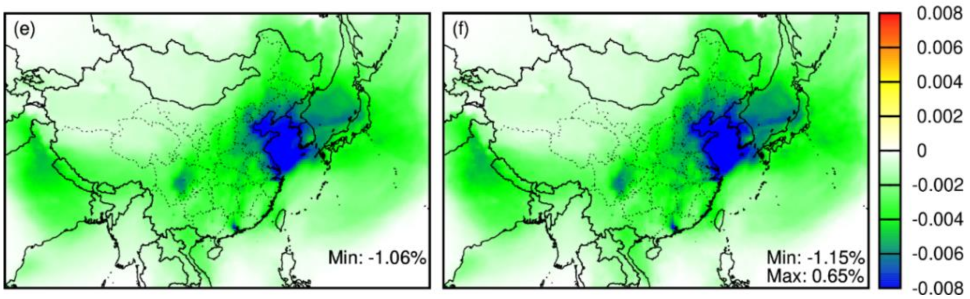


Figure 2-4 Continued.

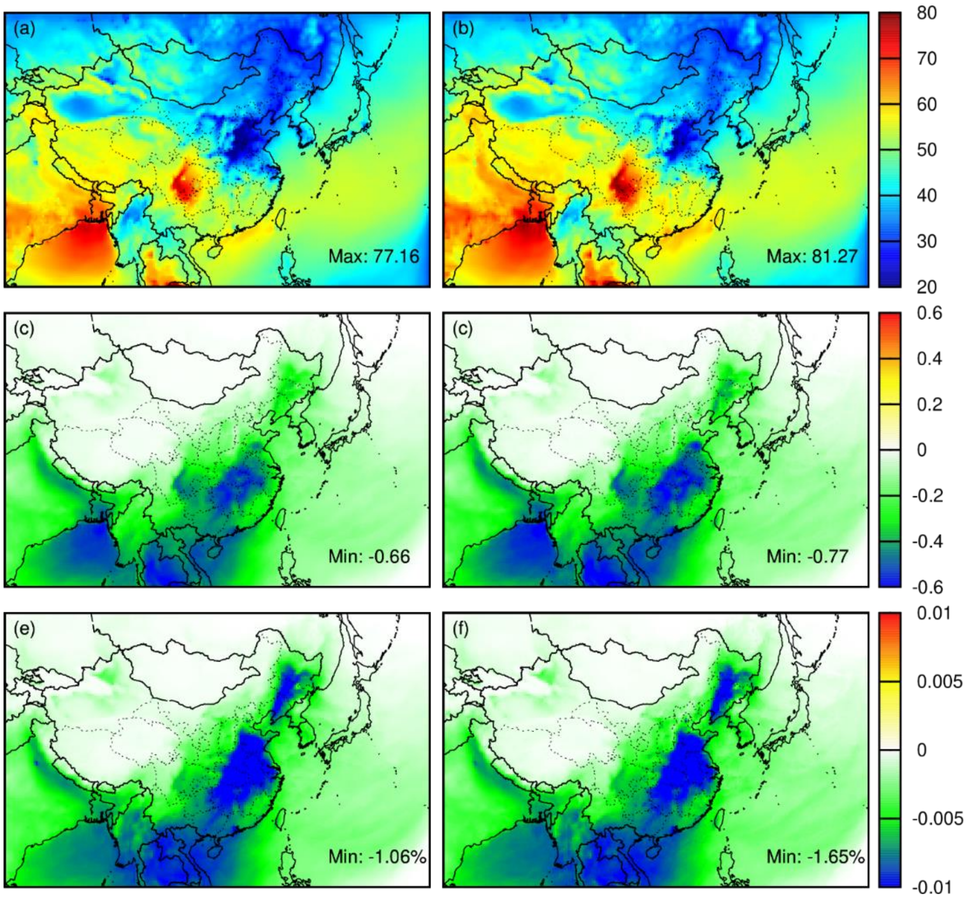


Figure 2-5 Monthly average daily maximum 8-h O_3 (DM8H O_3) concentrations (a) and daily peak hour O_3 (b) predicted with the modified SARPC-11 mechanism (Case 1) for January 2013; Difference of the monthly average DM8H O_3 (c) and peak hour O_3 (d) between the modified and original SAPRC-11 mechanisms (Case 1 – Case 0) for January 2013. Units are ppb. Relative difference of monthly average DM8H O_3 (e) and peak hour O_3 (f) between the modified and original SAPRC-11 mechanisms ((Case 1 – Case 0)/Case 0).

Figure 2-5a and 5b show the regional distribution of DM8H and peak hour O_3 for January 2013 in Case 1 and the relative increases in O_3 (Case 1 – Case 0). O_3 concentrations in northern China are lower, but high concentrations still occurred in southern China and the Sichuan Basin. The decrease in O_3 is not limited to urban areas. Large areas in central and south China show decreases in DM8H by approximately 0.5 ppb and up to 0.7 ppb, and peak hour O_3 decreases could be as high as 0.8 ppb. The relative changes in the O_3 decreased areas are more than 1%, and the maximum decrease of peak hour O_3 is 1.7 %. The difference in the regional distribution of the impact of ARO1 + OH reaction modifications on O_3 between summer and winter months is likely due to the contributions of ARO1 oxidation to the overall VOC oxidation. In the summer month, biogenic VOCs such as isoprene and terpenes are abundant, and they react with OH fast and have large maximum incremental reactivity. Based on the analysis of Wang et al.,[98] biogenic emissions could account for more than 50% of the O_3 formation attributed to VOC in China in the summer. Based on the IRR analysis of the OH consumption by VOCs (Figure S2-7 in Appendix A), ARO1 only accounted for less than 2% of the OH consumed by VOCs in most areas and about 4-8% in urban areas in the summer month. In the winter month, however, biogenic emissions in northern and central China were much lower, and the relative contribution of ARO1 became much more important, reaching 6-8% in most areas.

2.3.4 *Impacts on regional SOA*

Figure 2-6(a) shows that in January, higher concentrations of the ARO1 SOA are distributed in the mid-south and northeast areas of China, with a maximum monthly average concentration of approximately 3.6 $\mu g\ m^{-3}$ in Case C1. The increase of ARO1 SOA (Case C1 – C0) is widely distributed spatially, as shown in Figure 2-6(c), with a maximum regional increase of 1.1 μg m^{-3}. Due to the higher cresol yield, the relative increase of ARO1 SOA is approximately 40-70% in regions where the predicted toluene SOA is greater than 0.05 μg m^{-3}. The impact on higher yields on ARO1 SOA in July is relatively small because higher temperature tends to drive the semi-volatile products into the gas phase. Relatively higher ARO1 SOA concentrations are in east China, with a maximum monthly average concentration of approximately 1.6 μg m^{-3}, based on Case C1 (Figure 2-6(b)). As shown in Figure 2-6(d), the increase in ARO1 SOA is also small, about 0.2-0.3 μg $m^{-3,}$ with a maximum increase of 0.4 μg m^{-3}. The general increase in the North China Plain is approximate 30-40% of the ARO1 SOA concentration in Case C0.

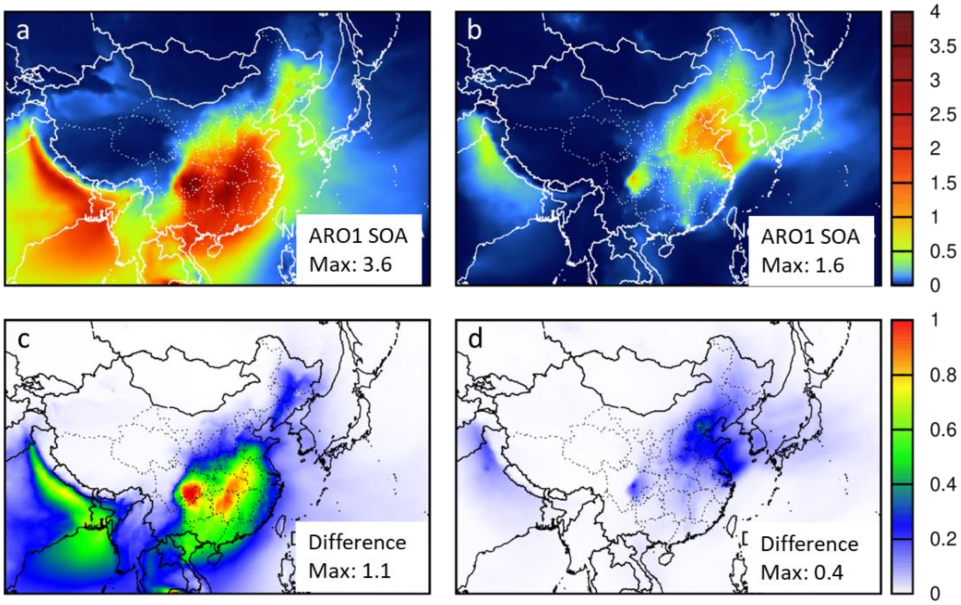


Figure 2-6 Predicted monthly average SOA from ARO1 for January 2013 (a) and July 2013 (b) from Case C1, and the increase in ARO1 toluene SOA (Case C1- Case C0) for January (c) and July 2013 (d). Units are μg m^{-3}. The relative increase in ARO1 toluene SOA ((Case C1 – Case C0)/Case C0) for January and July 2013 are shown in (e) and (f), respectively. The ARO1 SOA includes the two-product model components (ATOL1J, ATOL2J) under high NO_x conditions and a non-volatile component (ATOL3J) under low-NO_x conditions, as well as non-volatile components from surface uptake of glyoxal and methylglyoxal formed from oxidation of ARO1 and oligomers formed from the semi-volatile components. White spaces in panels (e) and (f) are grid cells with ARO1 SOA concentrations less than 0.05 μg m^{-3} in Case C0.

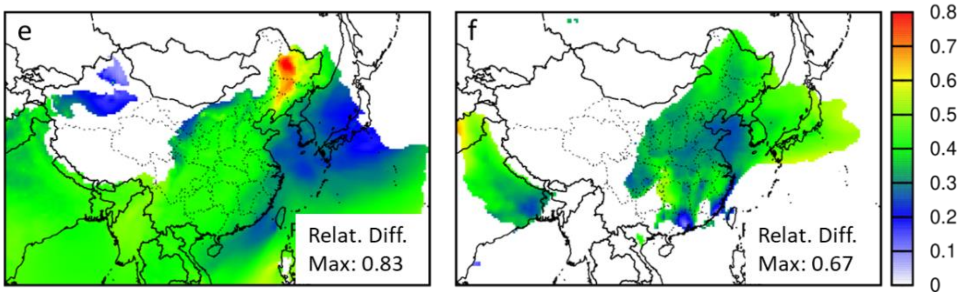


Figure 2-6 Continued.

The increase of total SOA, as shown in Figure S2-8 in Appendix A, is slightly higher than the increase due to ARO1 SOA alone. The relative increase in total SOA is approximately 10% in January and 5% in July. Most of the additional increases are due to increases in SOA from ARO2 (Figure S2-9 in Appendix A). Figure 2-7 shows the predicted changes in monthly average ARO1 and ARO2 SOA components for January 2013, as Case C1 – Case C0. Approximately 0.3-0.5 µg m^{-3} of the increase can be attributed to increased ARO1 oxidation products from the 2p model (Figure 7a). The GLY and MGLY surface uptake is decreased by less than 0.2 µg m^{-3}. The decrease is likely caused by the reduction of GLY and MGLY product yields in the ARO1 + OH mechanism due to the reduction of peroxy radical pathway branching ratio. The oligomers are increase by 0.4-0.7 µg m^{-3} following the increase of semi-volatile products from the 2p model. The changes in ARO2 SOA are represented by less than 0.04 µg m^{-3} increases in the south of China, and slight decreases (less than 0.01 µg m^{-3}) in the north. The increases of ARO2 SOA are likely due to slightly more semi-volatile products and their oligomers, and slightly more surface uptake of glyoxal and methylglyoxal. And the slight decreases of ARO2 SOA are related to the lower OH concentrations.

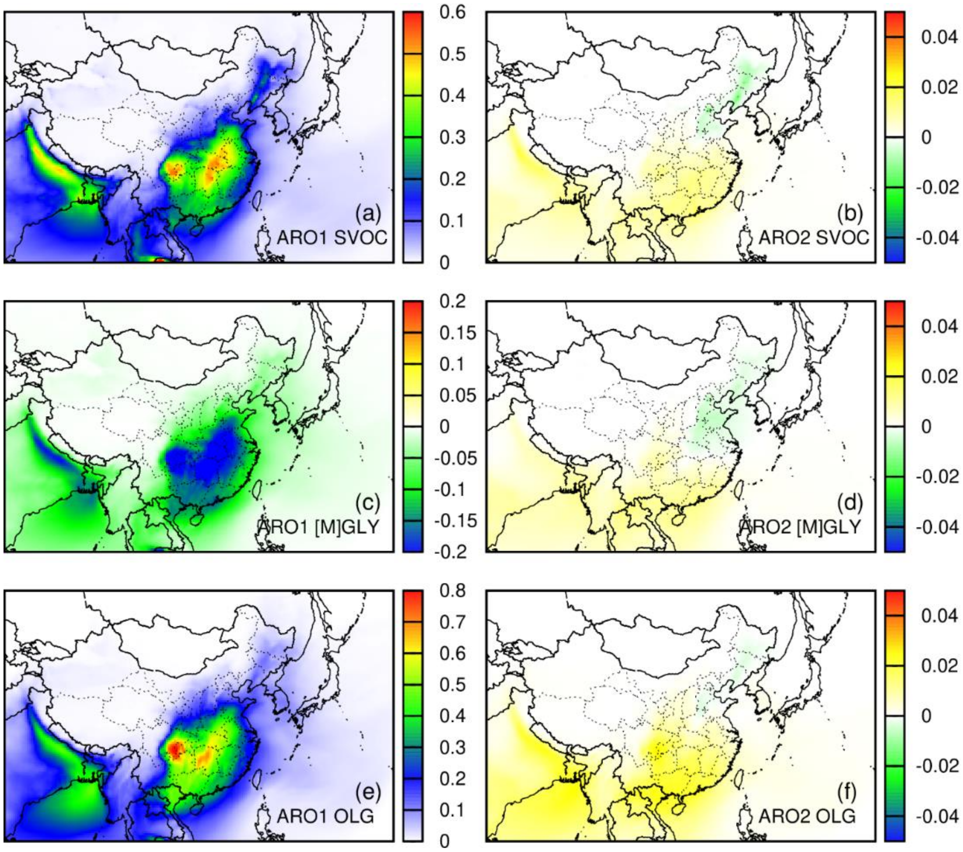


Figure 2-7 Predicted increase in monthly average ARO1 (a-c) and ARO2 (d-f) SOA components for January 2013 using the modified SOA yields (Case C1- Case C0). Units are µg m^{-3}.

Figure 2-8 and Figure S2-10 in Appendix A illustrate the hourly concentrations of ARO1 SOA in the major cities of China in January and July 2013, respectively. The higher ARO1 SOA yields used in Case C1 lead to significant increases in the hourly concentrations up to 2 times higher than those from Case C0. In most cities, the maximum peak hour concentration of ARO1 SOA increases ~1-2 µg m^{-3} for January days. Among the cities, Zhengzhou, located in central China, has the highest wintertime ARO1 SOA with a peak hourly concentration reaching 5.0 µg m^{-3} in Case C1. In July, the ARO1 SOA concentrations are generally lower than those in January. Concentrations in the coastal cities such as Guangzhou and Shanghai show larger day-to-day variations in the ARO1 SOA concentrations in July, which was likely due to strong land-sea breeze circulation in the summer. The maximum hourly concentration of ARO1 SOA in Shanghai reached approximately 5 µg m^{-3} in July, and in nearby areas even reached as high as 10 µg m^{-3}. Predictions from Case 0 were significantly lower, approximately two thirds of those predicted with higher yields.

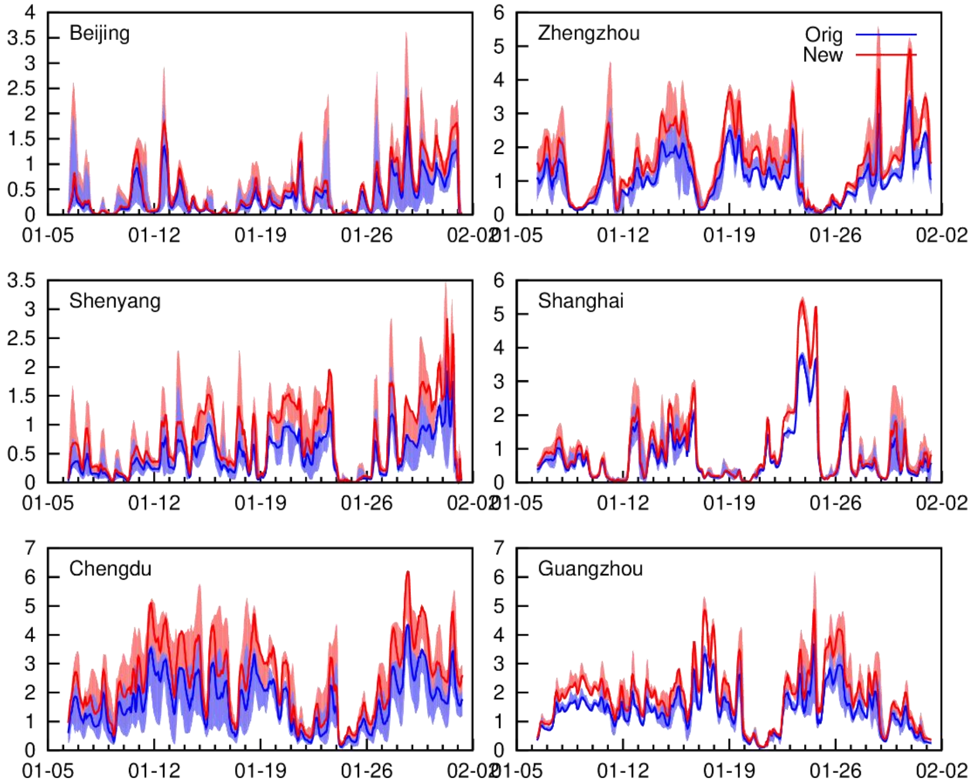


Figure 2-8 Predicted hourly concentrations of SOA from ARO1 in several urban areas for January 2013 using the original and modified SOA yields. Units are µg m^{-3}. Shaded areas represent the range of concentrations within the 3×3 grid cells with the urban in the center. The solid lines are the average concentrations in the 3×3 grids.

While it appears that the MCM box model with modified branching ratios of toluene + OH reaction can simulate the high-NO$_x$ chamber experiments by Hildebrandt et al.[9] and the predicted ambient toluene SOA concentrations are reasonable when compared with estimated SOA using specific tracers, these results do not provide a strict confirmation of the validity of the high cresol branching ratio suggested by Ji et al.[12] because the differences in the predictions are relatively small and large uncertainties exist in the estimated model parameters and input data.

2.4 Discussion

2.4.1 Impact on SOA formation due to changes in the gas phase chemistry only

The change of regional SOA in Case C1 is due to the changes in gas phase ARO1 + OH reaction and the differences in the ARO1 SOA yields. In Case C2, the modified gas phase ARO1 + OH mechanism used in Case C1 and original SOA yield used in Case C0 is applied to evaluate the impact on the regional toluene SOA due to gas phase chemistry changes only. Figure S2-11 in Appendix A shows that the ARO1 SOA concentrations in Case C2 are lower than those in Case C0 by approximately 0.05 – 0.1 μg m^{-3} in winter and less than 0.05 μg m^{-3} in the NCP region in summer, which is generally consistent with the GLY/MGLY SOA changes in Case C1 (as shown in Figure 2-7b). These decreases account for approximately 5% of the original ARO1 SOA concentrations, and there are no significant changes in concentrations of the semi-volatile products and their oligomers. As the HO_2 and NO_x concentrations in Case C2 are not significantly different from those in Case C0, the gas phase ARO1 + OH changes do not affect the low-NO_x pathway contributions to aromatic SOA. The relative contributions of the low-NO_x pathway to ARO1 SOA are approximately 1%~5% in winter and 5%~15% in summer.

2.4.2 Uncertainty of saturation vapor pressure of MCM species on the modeled SOA yields

The simulated SOA mass concentrations, thus the SOA yields, in the chamber experiments are affected by the vapor pressure of individual species in the MCM mechanism, estimated using the Estimation Program Interface Suite (EPI Suite) from US EPA [76]. The uncertainty in the predicted vapor pressure by the EPI Suite is estimated using the reference compounds included in the EPI's database (a total of 3707 species). As shown in Figure S2-12 in Appendix A, the error in the estimated vapor pressure increases with lower vapor pressure and can be as large as several orders of magnitude. After grouping the vapor pressure of the species into log-scale bins with a bin width of 1.0, the standard deviation of the ratio of the experimental vapor pressure to the EPI-estimated vapor pressure (in logarithmic scale) is determined as a function of the EPI-estimated saturation concentration C^*_{EPI}, as shown in Figure S2-13 in Appendix A.

A Monte Carlo technique is used to assess the uncertainty in the MCM-modeled SOA yield and major components of SOA due to uncertainties in the saturation vapor pressure. For each chamber experiment listed in Table S2-1 in Appendix A, 200 simulations were performed. In each simulation, the saturation concentrations of each semi-volatile species in the MCM mechanism based-on the EPI Suite (C^*_{EPI}) were modified using Equation (2.2),

$$C^* = 10^{\Delta logp}\ C^*_{EPI} \qquad (2.2)$$

where $\Delta logp$ is randomly generated to follow a normal distribution with zero mean and standard based on the C^*_{EPI} (equation in Figure S2-13). All simulations were conducted at 298 K and were run long enough so that the amount of SOA produced in the simulations exceeds at least 100 µg m^{-3}. Figure S2-14 in Appendix A shows that uncertainties in the saturation vapor pressure estimation could lead to large uncertainties in the estimated SOA yields. For the high-NO$_x$ condition, SOA yields at C$_{OM}$ = 10 µg m^{-3} has a 90th percentile range of [0.09, 0.45], and for the low-NO$_x$ condition, the 90th percentile range is [0.02,0.22]. The uncertainty in the SOA yields from the existing chamber experiments was equally large, as shown in Figure S2-14 in Appendix A. Additional studies are needed to reduce the uncertainties in the SOA yield estimations and quantify the impact on regional SOA estimations.

2.4.3 Uncertainty in SOA yields due to problems in the MCM predictions of later generation products

Another large uncertainty in the predicted SOA yields is from the MCM predictions of later generation oxidation products from the toluene oxidation. For example, recent studies reported detection of tri-, tetra-, and pentahydroxy toluene, and hydroxy and dihydroxyl methyl benzoquinone from cresol oxidation under both low and high-NOx conditions, which imply successive addition of the OH group to the aromatic ring after the formation of methyl catechol.[10, 29, 99] However, the MCM mechanism assumes that OH reaction with methyl catechol follows the H-abstract pathway entirely to form nitro methyl catechol (MNCATECH), suggesting that it likely misses important SOA precursors and lead to underpredictions in the SOA yields. It is possible that when the additional later generation oxidation pathways are included, good predictions of SOA yields can be obtained without increasing the cresol pathway yield. However, the extent of missing these later OH addition products on SOA yield simulations also depends on their vapor pressure, which have large uncertainties. The inaccuracies in MCM representation of the later generation oxidation processes on the SOA yield may have been partially compensated by the uncertainties in the estimated saturation vapor pressure, leading to reasonable high-NO_x SOA yields reported in this study. These species may be more important under low-NO_x conditions, and properly incorporating them in a modified MCM can potentially improve the modeled SOA yields. Studies are needed to further assess the uncertainties in SOA yields and their impact on regional SOA estimations with the improved toluene oxidation chemistry for later generation oxidation produces while considering the uncertainties in the estimated vapor pressures.

3 EVALUATION OF REGIONAL MODEL PREDICTIONS OF WINTERTIME SECONDARY ORGANIC AEROSOL FROM AROMATIC COMPOUNDS AND MONOTERPENES WITH PRECURSOR-SPECIFIC TRACERS

The community multiscale air quality (CMAQ), with modifications to track precursor-specific SOA, was applied to model SOA formation from aromatic compounds and monoterpenes in Shanghai in November 2018. The modeled total aromatic SOA showed a strong correlation with measured 2,3-dihydroxy-4-oxopentanoic acid (DHOPA) concentrations in the ambient aerosols ($R>0.5$ for hourly data and $R>0.75$ for daily average data). The ratios of observed DHOPA and modeled aromatic SOA with all components included is around $0.5\text{-}1.6\times10^{-3}$, lower than the commonly used ratio of 4×10^{-3} determined for toluene in a series of smog chamber experiments. This suggests that aromatic SOA could be underestimated when directly using the chamber-derived ratios. The predicted monoterpene SOA shows a stronger correlation with the sum of two α-pinene tracers (α-pinT), pinic acid and 3-MBTCA, with $R>0.6$ and $R>0.8$ for hourly and daily data, respectively. The α-pinT to modeled monoterpene SOA ratios are 0.13-0.25, which generally match the ratio of 0.168 ± 0.081 reported in chamber studies. However, since the current model does not treat α-pinene and its SOA explicitly, future modeling studies should include a more detailed treatment of monoterpene emissions and reactions to predict SOA from these important precursors and compare with the ambient precursor-specific SOA-tracers.

3.1 Introduction

Carbonaceous aerosols generally account for 20-50% of total ambient aerosols globally [100-103]. Cao, et al. [104] reported that elemental carbon (EC) and organic matter combined accounted for 39-44% of $PM_{2.5}$ in 14 Chinese cities, and in Shanghai, 40% of the $PM_{2.5}$ were carbonaceous aerosols [105]. Carbonaceous aerosol can have significant impacts on atmospheric visibility [106], regional and global climate [107], and human health [108]. While the EC aerosols are directly emitted from fuel combustion sources, organic aerosols are generated from direct emissions (i.e., POA) and gas-to-particle partitioning of the semi-volatile oxidation products from parent VOCs. SOA contributions to total organic aerosol loading vary from 20% to 80% with significant spatial and seasonal variation [1-4].

Chemical transport models have been widely used to quantitatively study the regional and global impacts of carbonaceous aerosols [5-8]. Correctly predicting SOA in these models is challenging, especially in polluted urban areas, as many precursors contribute to SOA formation, and the ability of each precursor to form SOA is different. Aromatic compounds [9-11], isoprene [2, 12-14], monoterpenes [15, 16], and sesquiterpenes [17, 18] are some of the major contributors to SOA. In addition, aqueous [20] and heterogeneous processes [19] of dicarbonyls such as GLY and MGLY have been shown to contribute significantly to SOA formation. Predicted SOA concentrations are affected by the model representation of the emission, photochemical oxidation, gas-to-particle partitioning, and the multiphase reaction processes [21-23]. Many of these physical and chemical processes remain uncertain due to an incomplete understanding of the SOA formation mechanisms and large differences between the atmospheric conditions in the ambient environment and the chamber conditions under which the SOA formation experiments were conducted to determine parameters used in the models [24, 25].

Predicted SOA concentrations are not well constrained by observations. Techniques to apportion the observed total OA concentrations to POA and SOA include the minimum OC/EC ratio method [109] and its extensions [110, 111], and the positive matrix factorization (PMF) analysis of the aerosol mass spectrums [35, 36]. Wu and Yu [112] improved the minimum OC/EC method. Instead of using the minimal OC/EC ratio as the primary OC/EC ratio, they proposed to choose a representative primary OC/EC ratio so that the correlation between SOC and EC is minimized. The OC/EC ratio methods provide estimations of the total SOA, and the PMF based analysis typically provide estimations of several groups of oxygenated OA (which are considered as SOA) based on the level of oxidation and the correlation with other secondary species [113, 114]. The PMF-resolved SOA factors, however, are not directly related to specific precursors. Most of the regional model evaluations of SOA predictions were carried out by comparing model predictions with these bulk observation-based SOA estimations [21, 37]. Few studies are reported to evaluate the precursor-specific SOA predictions with the corresponding observations.

In 2007, a tracer method to estimate the contributions of different precursors to ambient SOA concentrations was established by Kleindienst, Jaoui, Lewandowski, Offenberg, Lewis, Bhave and Edney [38]. In this technique, the mass fraction of the identified precursor-specific tracers to SOA formed from the precursor was determined in smog chamber experiments [38, 39]. Among the tracers identified, 2,3-dihydroxy-4-oxopentanoic acid (DHOPA) is a widely used one for SOA from monoaromatic hydrocarbons. Pinic acid, pinonic acid, and 3-methyl-1,2,3-butanetricarboxylic acid (3-MBTCA) are tracers for SOA from α-pinene. The concentrations of the tracers in ambient particles are measured and used to calculate the total SOA formed from the corresponding precursor with chamber-determined mass fractions. Since then, ambient concentrations of these tracers have been quantified and applied to estimate the precursor contributions to SOA under various atmospheric environments, assuming that these ratios are applicable under atmospheric conditions [11, 38-42]. However, since the ambient conditions are different from those in the chamber experiments, it remains unclear whether the tracer concentrations are correlated with their target SOA concentrations and whether the SOA estimated using chamber-based tracer mass fractions can match these from air quality model predictions.

In this study, we applied a regional air quality model with precursor-resolved SOA representation to simulate SOA in Shanghai, China, in November 2018. We compared the model predicted SOA derived from aromatic compounds and monoterpenes with the hourly and daily average concentrations of the specific SOA tracers and estimated the tracer-to-SOA ratio based on two different sets of model results using two different emission inventories. This study is the first evaluation of model-predicted precursor-resolved SOA with source-specific tracers, to the best of the authors' knowledge.

3.2 Methods

3.2.1 Modeling precursor-specific SOA

A precursor-specific SOA scheme was implemented in the CMAQ model v5.0.1 [115]. A complete description of the scheme is available in Ying et al. [84], so only a brief description of the model is provided below. The gas phase photochemical mechanism is based on SAPRC-11 [81] as discussed in Section 2.2.2. In SAPRC-11, the aromatic compounds were lumped into two species, ARO1 and ARO2. ARO1 represents aromatics with lower OH reactivity, and ARO2 represents aromatics with higher OH reactivity. Biogenic emissions are handled in the SAPRC-11 mechanism with monoterpene (TERP), sesquiterpene (SESQ), and isoprene (ISOP). Since GLY and MGLY are formed from the oxidation reactions of multiple precursors in SARPC-11, the gas phase mechanism was modified to track the GLY and MGLY from different precursors with extra tagged species. For example, GLY generated from ARO1 and ARO2 are tracked with GLY_A1 and GLY_A2, respectively.

The SOA module is based on the aerosol AERO6. In the module, SOA was formed in three pathways: (1) Equilibrium gas-to-particle partitioning of semi-volatile products from the oxidation of precursors, represented by the classical Odum two-product model [116]. The SOA yields are the same as those used in Hu et al. (2017). (2) Oligomerization of the particle phase semi-volatile products, which are assumed to form oligomers through first-order decay reactions with a half-life of 20 hours [117]. The CMAQ original AERO6 was modified to tracked the oligomers from each specific precursor by introducing extra precursor-tagged species. (3) Irreversible surface uptake of isoprene epoxides, GLY, and MGLY on wet aerosols of cloud droplets, with uptake coefficients parameterized according to Li, Cleveland, Ziemba, Griffin, Barsanti, Pankow and Ying [73]. As tagged GLY and MGLY were used in the gas phase, the SOA module was modified to track their contributions to the secondary GLY and MGLY from each precursor with tagged aerosol species.

3.2.2 Model application

The modified CMAQ model was used to simulate air quality in China during November 2018. The predictions were used to compare with the observations from the Shanghai Academy of Environmental Sciences (SAES, 31.17°N, 121.43°E), located in the southwest of the central urban area of Shanghai, China. Details of the measurements and the observation data analyses have been documented elsewhere [41, 118]. The CMAQ model has 197×127 grid cells in each layer and has 18 vertical layers to reach the model top of approximately 20 km. The horizontal grid resolution is 36×36 km^2. The model uses stretching vertical layers with the first layer height of approximately 35 m. Initial and boundary conditions for the CMAQ model were generated using the CMAQ default vertical profiles that represent clean continental conditions. Simulation results from the first five days were treated as spin-up and were not included in the analyses reported below.

The meteorological inputs were generated using WRFv4.2 with initial and boundary conditions from the NCEP GDAS/FNL 0.25 Degree Global Tropospheric Analyses and Forecast Grids (available at https://rda.ucar.edu/datasets/ds083.3/). The land use/land cover and topographical data were based on the 30 s resolution default WRF input dataset. Reanalysis nudging was enabled to improve the agreement between predicted and observed meteorological parameters [119]. The major physics options for the WRF simulations were described by Zhang, et al. [120].

Two sets of emissions were applied in this study to represent the anthropogenic emissions from China, the Regional Emission inventory in Asia v3.1 (REAS3) [121], and the 2017 Multiscale Emission Inventory of China (MEIC). Emissions from other countries were always based on the REAS3 inventory. The MEIC and REAS3 inventories were in 0.25°×0.25° grids and reprojected to the Lambert conformal coordinates. Windblown dust emissions in the entire domain were generated inline [86]. The MEIC inventory already includes speciated nonmethane hydrocarbons (NMHC). For the REAS3 emissions, selected speciation profiles from the SPECAITE database developed by the US EPA were used to estimate emissions of model-ready VOCs [122]. Biogenic emissions were generated by the MEGAN v2.10 [123].

3.3 Results

3.3.1 Model performance evaluation

The WRF model predicted meteorological inputs significantly affect the accuracy of the chemical transport model predictions. Predicted temperature (TEMP) and relative humidity (RH) at 2 m above the surface, and wind speed (WSPD) and wind direction (WDIR) at 10 m above the surface were compared with the observation data from the National Climate Data Center (NCDC) within the Yangtze River Delta (YRD) area. There were 54 observation sites available in the YRD region. The average observation, average prediction, mean bias (MB), gross error (GE), and root mean square error (RMSE) were calculated for TEMP, RH, WSPD, and WDIR, as shown in Table 3-1. TEMP is overpredicted with an MB of 0.95 K, which is slightly higher than the recommended benchmark (≤ ±0.5 K) by Emery, et al. [124], but the GE of 2.1 K is close to the benchmark value of <2.0 K. WSPD is well predicted with the MB and RMSE lower than the benchmarks (MB ≤ ±0.5 m s^{-1}, RMSE < 2 m s^{-1}), but the MB and GE of WDIR are approximately 22% and 16% higher than their respective benchmarks (MB ≤ ±10°, GE < 30°). In general, the WRF performance statistics in this study is comparable to other studies using WRF in China simulations [120, 125-127].

Table 3-1 Meteorology model performance.

		TEMP (K)	RH (%)	WSPD (m s^{-1})	WDIR (°)
Average Observation		287.31	80.20	3.20	159.72
Average Prediction		288.26	73.16	3.66	123.03
Mean Bias	$MB = \frac{1}{N}\sum_{i=1}^{N}(pre_i - obs_i)$	0.950	-7.036	0.452	12.166
Gross Error	$GE = \frac{1}{N}\sum_{i=1}^{N}\lvert pre_i - obs_i\rvert$	2.092	11.977	1.282	34.914
Root Mean Square Error	$MB = \frac{1}{N}\sum_{i=1}^{N}(pre_i - obs_i)$	2.719	15.744	1.706	49.676

The hourly $PM_{2.5}$ mass was measured with an online beta attenuation particulate monitor (FH 72 C14 series, Thermo Fisher Scientific). Water-soluble inorganic ions (SO_4^{2-}, NO_3^-, NH_4^+, Cl^-, and K^+) with an online Monitor for AeRosols and Gases in the ambient Air (MARGA, Model ADI 2080, Applikon Analytical B.V.). Organic and elemental carbon (OC and EC) were monitored by a semicontinuous OC–EC analyzer (model RT-4, Sunset Laboratory, Tigard, OR, USA). Molecular tracers were measured through the Thermal desorption Aerosol Gas chromatography

(TAG) every two hours from November 9 to December 3, 2018. Each sampling started at the odd hour and last for 1 hour. A total of 270 valid samples and 11 blank samples were collected and analyzed [41]. Hourly model predictions with the MEIC and REAS3 emission inventory were evaluated by comparing with the observations, as shown in Figure 3-1.

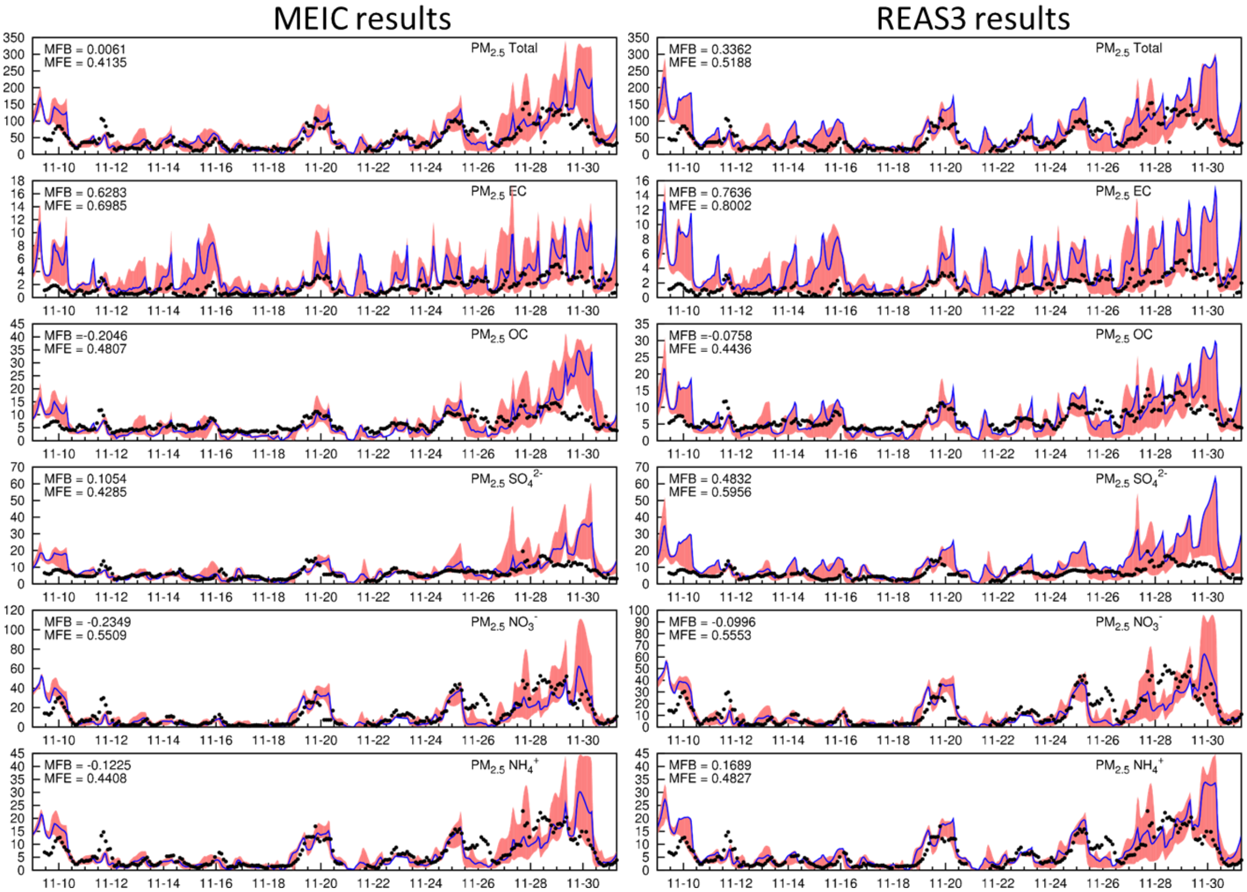


Figure 3-1 Time series and model performance statistics of $PM_{2.5}$ mass and major chemical components in Shanghai, based on the MEIC (left column) and REAS3 (right column) emission inventories from November 9 to December 1, 2018. Black dots are the observed concentrations; solid lines are predicted concentrations in the grid where the observation site is located; shaded areas represent the concentration ranges from 3×3 grids with the observation site in the center grid. Units are μg m^{-3}.

Generally, the predicted $PM_{2.5}$ and its components with either emission inventories agree with the observations with mean fractional bias (MFB) values of 0.00 and 0.34 for the MEIC and

REAS3, respectively. The MEIC predictions also have a slightly lower MFE than REAS3 (0.41 vs. 0.52). The high concentration of total $PM_{2.5}$ over 100 μg m^{-3} on November 20, 25, and 27-30 are well captured. However, the model overpredicts the $PM_{2.5}$ concentrations on November 30 for both simulations. The overprediction is associated with the underprediction of wind speeds during calm conditions around Shanghai, which causes the over-accumulation of pollutants.

The sulfate aerosols are well-predicted using the MEIC emission inventory without a significant bias (MFB=0.11) but are over-predicted with the REAS3 inventory (MFB=0.48). This is because the REAS3 inventory SO_2 emissions are 20-30% higher than those based on MEIC inventory in the Shanghai urban area, and the SO_2 emissions in the surrounding areas are 1.4-2 times those estimated in the MEIC inventory. The higher SO_2 emissions from REAS in China were reported in the previous studies [128, 129]. The observed concentrations of nitrate and ammonium secondary aerosols are well reproduced by both MEIC and REAS3 based emissions.

Observed $PM_{2.5}$ EC concentrations are between 0-4 μg m^{-3}. Both MEIC and REAS3 inventories lead to over-predictions of EC with MFB larger than 0.6. EC emissions were likely overestimated in both inventories. Predicted EC concentrations show strong spatial gradients as indicated by the large ranges based on the predictions within the 3x3 grid cells with the monitor station grid cell at the center, as it is primarily from vehicle emissions in urban areas. Uncertainties in the predicted wind speed and direction could also cause large errors in the predicted concentrations. The $PM_{2.5}$ OC predictions also compare well with the observations with MFB values of -0.20 and -0.08 based on the MEIC and REAS3 emissions, respectively, and the MFE values are less than 0.5 for both predictions. Over-predictions of EC are not expected to affect model predictions of SOA.

The positive matrix factorization (PMF) analysis of the AMS data collected during this period provided an estimation of the primary organic carbon (POC) and secondary organic carbon (SOC). The details of the sampling and data analysis were reported by He et al. [41]. As shown in Figure 3-2, both MEIC and REAS3 emissions lead to reasonable predictions of POC, as indicated by the MFB (-0.33 for MEIC and 0.17 for REAS3) and MFE (0.58 for MEIC and 0.54 for REAS3) values. Predicted SOC using MEIC does not have an overall bias using data for the entire month (MFB=-0.06). However, it significantly overpredicts the SOC by more than 25 µg m^{-3} at the end of November. The model with REAS3 emissions lead to a lower predicted SOC with the MFB of -0.52 but better captured the high concentrations at the end of November.

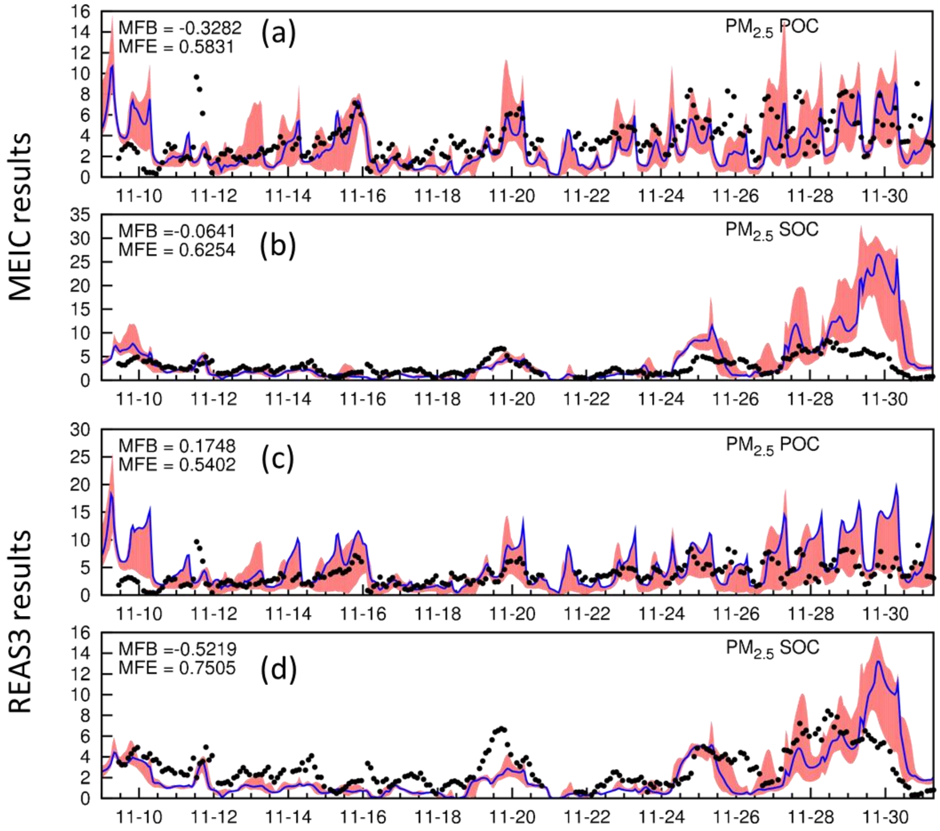


Figure 3-2 Time series of predicted POC and SOC using the (a,b) MEIC emissions (lines), (c,d) REAS3 emissions (lines), and PMF-based observations (dots) in Shanghai, from November 9 to December 1, 2018. Shaded areas represent the concentration ranges from the 3×3 grids with the observation site in the center grid. Units: µg m^{-3}.

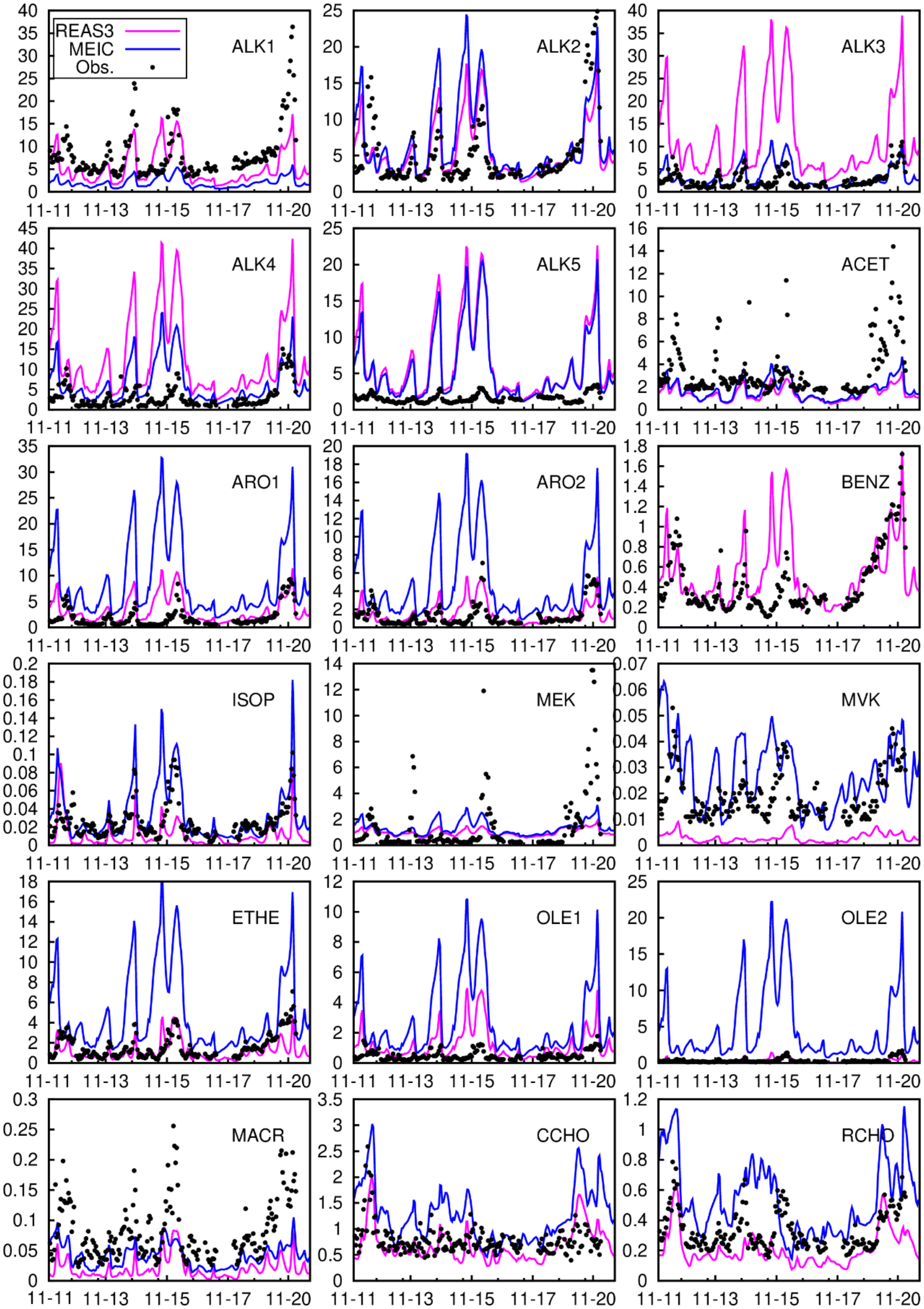


Figure 3-3 Time series of predicted VOCs using the MEIC and REAS3 emissions in Shanghai, from November 11 to 20, 2018. Observations are based on hourly individual VOCs lumped into SARPC11 model species. Units are ppb.

In summary, $PM_{2.5}$ predictions from both inventories reasonably agree with observations. The good agreement between the predicted and observed secondary inorganic aerosols suggests that the model can reproduce the oxidation capacity of the urban atmosphere in this region. The better agreement between the PMF resolved SOC and the predicted SOC with the MEIC inventories could imply that the MEIC inventory is more appropriate for SOA predictions, but it is necessary to compare the precursor predictions to confirm this.

Volatile organic compounds are direct precursors to SOA and affect the OH radical concentrations. Hourly concentrations of 115 individual VOCs monitored at SAES station for November 11-20 were obtained to evaluate the model performance. The measured VOC species were lumped to match the species in the SARPC-11 and compared with predictions using REAS3 and MEIC emissions, as shown in Figure 3-3. While the predicted VOC concentrations using both inventories have some success in matching the observations and reproducing the day-to-day variations of the concentrations, neither emission inventory gives satisfactory results for all species.

The ALK1-5 species represent the alkanes and other non-aromatic compounds with increasingly higher OH reaction rate constants. The predicted ALK2, which mostly includes the less reactive short-chain alkane species, shows relatively good agreement with observations. Both MEIC and REAS3 lead to similar higher predictions of ALK4-5, mainly long-chain alkanes, than the observations by 2-5 times. This overprediction might partially be caused by the fact that the measurements did not have all the species included in the emission inventories. For the other two groups of ALK species, REAS3 predicts better for ALK1 (ethane), but MEIC predicts better for ALK3. Ethene (ETHE) and OLE1-2 (olefin species with increasingly higher OH reactivities) are better predicted with REAS3 emissions, but all are significantly overpredicted with MEIC.

For the lumped aromatics species ARO1 (mostly toluene) and ARO2 (mostly xylene), the predicted concentrations with REAS3 show much better agreement with observations (between 70% and two times) than those from MEIC, which are several times higher than the observations. Large over predictions of aromatics in MEIC was not expected as previous modeling studies using an earlier version of MEIC showed relatively good agreement with observations in Nanjing in August 2013 (MFB=-0.63-0.77) [1] and from June to August 2014 (NMB=0.2) [130]. Since previous studies were for the summer months, it is possible that the seasonal variations in the emissions were not properly captured in the MEIC inventory.

At the grid of SAES site location, isoprene (ISOP) is mostly from anthropogenic emissions, as shown in

Table S3-1 in Appendix B, and is well predicted with the MEIC inventory. The VOC profiles used to speciate REAS3 emissions might have used lower isoprene emission factors. Consequently, methacrolein (MACR) and methyl vinyl ketone (MVK), which are major oxidation products of isoprene, are better predicted with MEIC. Methyl ethyl ketone (MEK) is also better predicted with MEIC as it is an oxidation product from several VOCs, including MVK and MACR. The other oxygenated species, acetone (ACET), acetaldehyde (CCHO), and higher aldehydes (RCHO), which have both primary emissions and secondary formations, are reasonably predicted with both inventories.

Although neither inventory generates perfect estimations of the VOCs, the REAS3 inventory's better predictions of the gas phase aromatics provide more confidence in the SOA predictions than the SOA predicted by the MEIC inventory. The MEIC-predicted SOA should be considered as an upper limit of the SOA from the aromatic compounds.

3.3.2 Compare observed SOA tracers with modeled SOA

The concentrations of SOA tracers (i.e., pinic acid, 3-MBTCA, and DHOPA) were measured hourly at SAES from November 9 through December 1, 2018[41]. The total concentrations of the two α-pinene tracers (α-pinT) were compared with the predicted monoterpene SOA, and the DHOPA concentrations were compared with the total SOA (including semi-volatile components, oligomers, and surface uptake products from GLY and MGLY) from ARO1 and ARO2, as shown in Figure 3-4. Generally, the predicted SOA has similar day-to-day variations with the SOA tracers. High concentrations of DHOPA in the range of 0.015-0.02 μg m^{-3} occurred on November 20 and the last several days of November. High concentrations of the model predicted aromatic-SOA (~10-30 μg m^{-3}) were also predicted for these days. The α-pinT concentrations were in the range of 0.001-0.08 μg m^{-3}, while the modeled monoterpene SOA concentrations reached 0.35 μg m^{-3} and 0.6 μg m^{-3} with REAS3 and MEIC emissions, respectively.

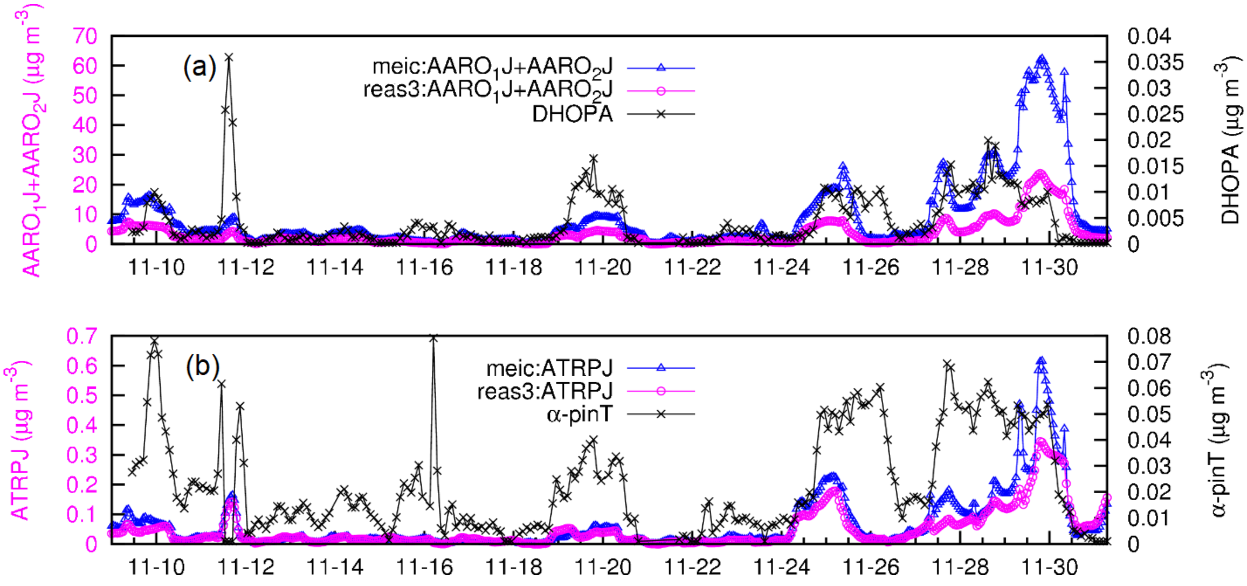


Figure 3-4 Time series of (a) model predicted aromatic SOA (AARO1J + AARO2J) and observed aromatic-derived SOA tracer (DHOPA), and (b) model predicted monoterpene SOA (ATRP1J+ATRP2J) and observed α-pinene-derived SOA tracer (α-pinT).

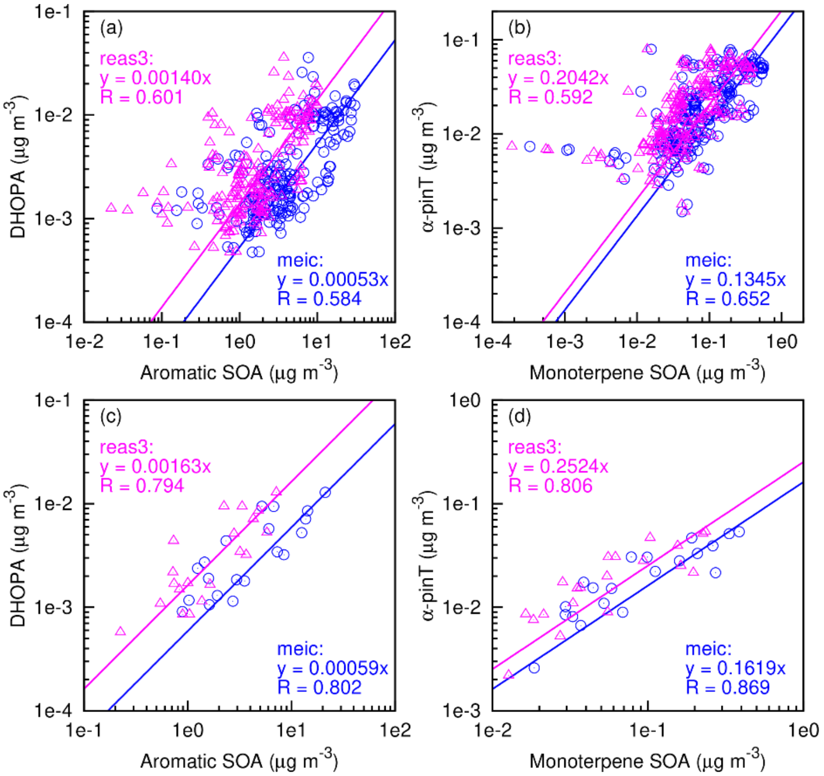


Figure 3-5 Correlation between model predicted aromatic SOA (all components) and observed aromatic-derived SOA tracer (DHOPA) (a,c), and model predicted monoterpene SOA (all components) and observed α-pinene-derived SOA tracer (α-pinT) (b,d), using hourly (a,b) and daily (c,d) average concentrations.

The precursor-specific SOA were much higher and likely overestimated on November 29 and 30, along with other PM species such as SOC and POC (see Figure 3-1 and Figure 3-2). These data points may significantly affect the analysis of the linear correlations between observed SOA tracers and modeled specific SOA and were excluded for a proper estimation of the relationship between modeled SOA and observed tracer concentrations shown in Figure 3-5. The predictions with REAS3 and MEIC emissions show strong correlations with the detected SOA tracers. The correlations between hourly predicted monoterpene SOA concentrations and measured corresponding tracer α-pinT concentrations (R=0.6~0.65) are slightly higher than those between monoaromatic SOA and DHOPA (R~0.6). The correlations are increased for daily averaged SOA predictions and the corresponding SOA tracers, with R~0.8 for aromatic SOA and R > 0.8 for monoterpene SOA. The daily average correlations are improved likely by smoothing out the difference in the formation timescales of the tracers and other major SOA components.

Table 3-2 Correlation between model predicted SOA mass concentrations and observed SOA tracers.

Precursor	Mass fraction	Correlation Coefficient	Comment
ARO1+ARO2[*]	0.00161 ± 0.00014	0.523	This study; Hourly MEIC
	0.00180 ± 0.00016	0.758	This study; Daily MEIC
	0.00553 ± 0.00032	0.571	This study; Hourly REAS3
	0.00614 ± 0.00060	0.765	This study; Daily REAS3
ARO1+ARO2[£]	0.00053 ± 0.00003	0.584	This study; Hourly MEIC
	0.00059 ± 0.00004	0.802	This study; Daily MEIC
	0.00140 ± 0.00006	0.601	This study; Hourly REAS3
	0.00163 ± 0.00019	0.794	This study; Daily REAS3
Aromatics[^]	0.00198 ± 0.0016		He et al. [41]
Toluene	0.0040 ± 0.0013		Kleindienst et al. [38]
Toluene-NO_x	0.0032 ± 0.0004		Al-Naiema et al. [39]
Toluene-H_2O_2	0.0068 ± 0.0008		Al-Naeima et al. [39]
o/m/p-Xylene	0.0033 ± 0.00024		Al-Naeima et al. [39]
TERP[#]	0.1345 ± 0.0095	0.652	This study; Hourly MEIC
	0.1619 ± 0.0137	0.869	This study; Daily MEIC
	0.2042 ± 0.0193	0.592	This study; Hourly REAS3
	0.2524 ± 0.0365	0.806	This study; Daily REAS3
α-Pinene	0.168 ± 0.081		Kleindienst et al. [38] Applied in He et al. [41]

[*] SOA formed from ARO1 and ARO2, including the semi-volatile SOA, oligomers, and glyoxal and methylglyoxal SOA products.

[£] SOA from ARO1 and ARO2, only including the semi-volatile SOA.

[^] Based on the average DHOPA-to-SOA ratio of benzene, toluene, ethylbenzene, o/m/p-xylenes and 1,3,5- and 1,2,4-trimethylbenzene with NO_x, as reported in Al-Naiema et al. (2020)

[#] SOA formed from monoterpenes, including SSOA, oligomers, and glyoxal and methylglyoxal SOA products.

The tracer-to-SOA mass ratios, representing the mass fraction of the precursor-specific SOA tracers in the SOA derived from the precursors, were determined using linear regression with forced zero intercepts (Figure 3-5). A robust linear regression method was used to reduce the impact of outliers, and the bootstrap technique [131] was used to determine the uncertainties in the slopes. For aromatic SOA, the tracer-to-SOA mass ratio is 0.00140 ± 0.00006 based on hourly SOA from REAS3. The slope derived from the data with MEIC emissions is significantly lower (0.00053 ± 0.00003) caused by higher SOA predictions due to the overestimation of ARO1 and ARO2 concentrations. The linear regression slopes between the model predicted monoterpene-SOA with REAS3 and MEIC emissions and the measured α-pinT tracers are 0.2042 ± 0.0193 and 0.1345 ± 0.0095, respectively, which are both close to the mass fraction of α-pinT to the α-pinene-SOA of 0.1680 ± 0.0081, as suggested from the previous chamber study [38] and applied by He et al. [41]. The detailed regression slope and uncertainties are summarized in Table 3-2.

3.4 Discussion

3.4.1 Impact of non-volatile SOA components to tracer-to-SOA ratio

The predicted SOA used in the previous analyses includes semi-volatile components based on equilibrium gas-to-particle partitioning, oligomers from semi-volatile products, and SOA from irreversible surface uptake of GLY and MGLY. The significance of the surface uptake of GLY and MGLY on the SOA formation was discussed in previous studies [132-134]. As the relative humidity was quite high in winter, significant contributions of GLY and MGLY to SOA were predicted.

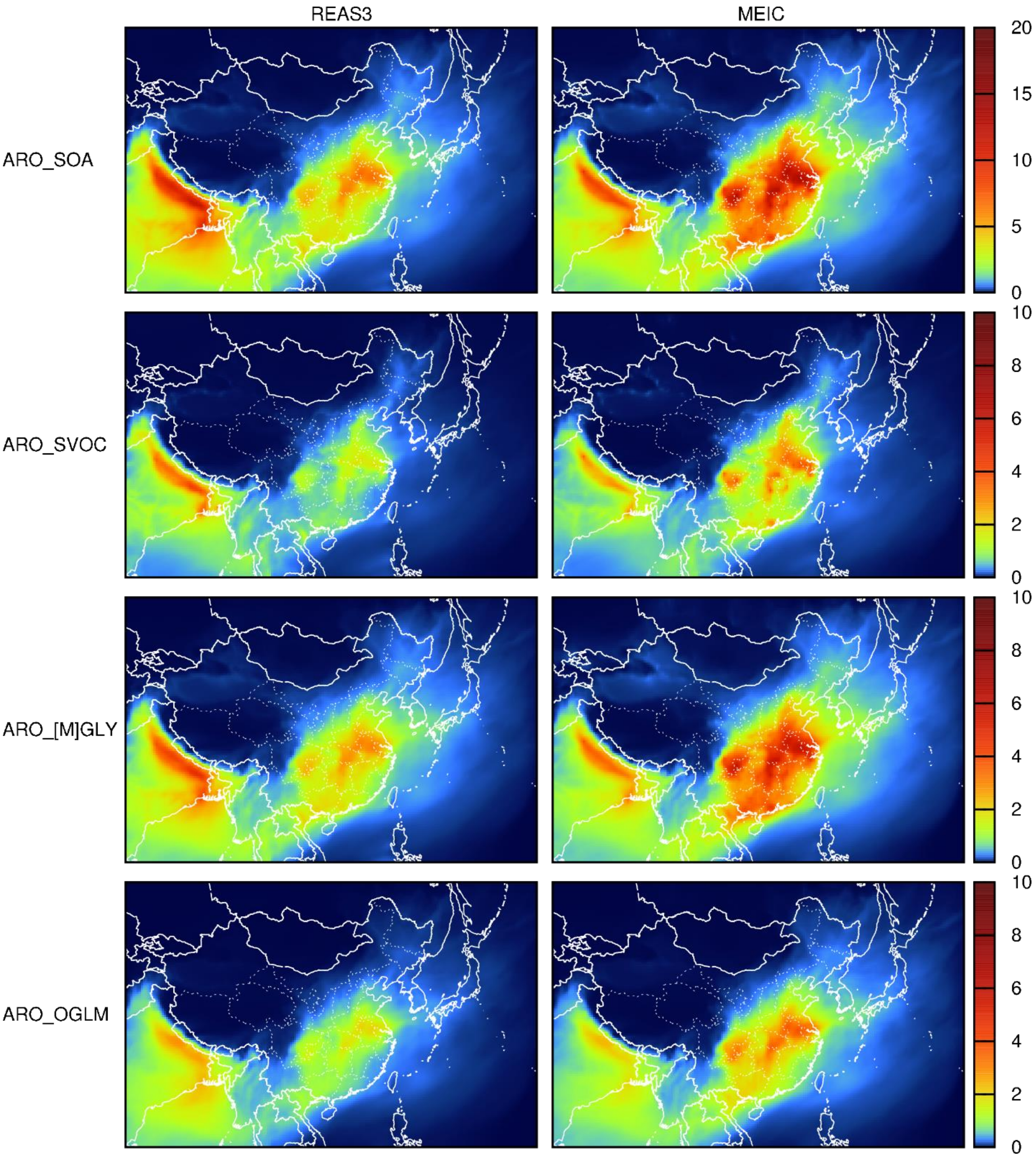


Figure 3-6 Monthly averaged concentration of total aromatics SOA with REAS3 (left column) and MEIC (right column) inventories, and contributions to total SOA from semi-volatile components (ARO_SVOC), surface uptake of glyoxal and methylglyoxal SOA (ARO_[M]GLY), and oligomers from semi-volatile components (ARO_OLGM). Units are μg m^{-3}.

Figure 3-6 shows the high monthly averaged total aromatic SOA concentrations in China for November 2018. The model with MEIC emissions predicts total aromatic SOA concentrations to be approximately 10-15 μg m^{-3} in central and eastern China, and the REAS3 emission inventory, with lower emissions of aromatics, predicts lower aromatic SOA, approximately 5-10 μg m^{-3}. However, the fraction of GLY and MGLY SOA in total SOA predicted by the two emission inventories is similar. Figure 3-6 also shows that GLY and MGLY SOA has the highest contributions to total aromatic SOA. At the grid cell where the SAES monitor is located, semi-volatile SOA and its oligomers combined have concentrations of 1.48 and 4.01 μg m^{-3} for REAS3 and MEIC emissions, respectively. The GLY and MGLY SOA at the same grid cell is 1.51 μg m^{-3} based on REAS3 and 3.62 μg m^{-3} based on MEIC, as high as the SOA predicted from the traditional pathways. However, the chamber experiments used to measure the tracer-to-SOA ratio were typically operated under much lower RH e.g., [39], and the GLY and MGLY contributions to SOA in these chambers were expected to be very small. In addition to GLY and MGLY contributions, the oligomers formed from semi-volatile products contribute as much as the semi-volatile products to the aromatic SOA. However, oligomer formation in the chamber experiments was usually small due to a short detention time of several hours. Thus, the chamber determined ratio might only be good for the estimation of the semi-volatile aromatic SOA components.

The DHOPA tracer-to-SOA ratios were recalculated using predicted aromatic SOA without the oligomers and GLY and MGLY components to evaluate the predicted semi-volatile aromatic SOA with DHOPA. As shown in Figure 3-7, excluding the non-volatile SOA components does not significantly influence the correlation coefficients between the predicted aromatics SOA and measured DHOPA concentrations. However, the mass ratio of DHOPA to aromatic SOA with MEIC and REAS3 emissions are increased to 0.00161 ± 0.00014 and 0.00553 ± 0.00032, respectively. The REAS3 ratio is closer to the mass fraction reported by Kleindienst et al. [38], which has been more broadly used in most aromatic SOA estimation [61, 135, 136]. It is also in better agreement with those reported by Al-Naiema et al. [39] for toluene. Since the predicted precursor ARO1 and ARO2 concentrations with the REAS3 inventory also generally agree with the observations, this suggests that the semi-volatile aromatic-SOA can be reasonably predicted by the SOA mechanism in the regional model if precursor emissions were estimated correctly.

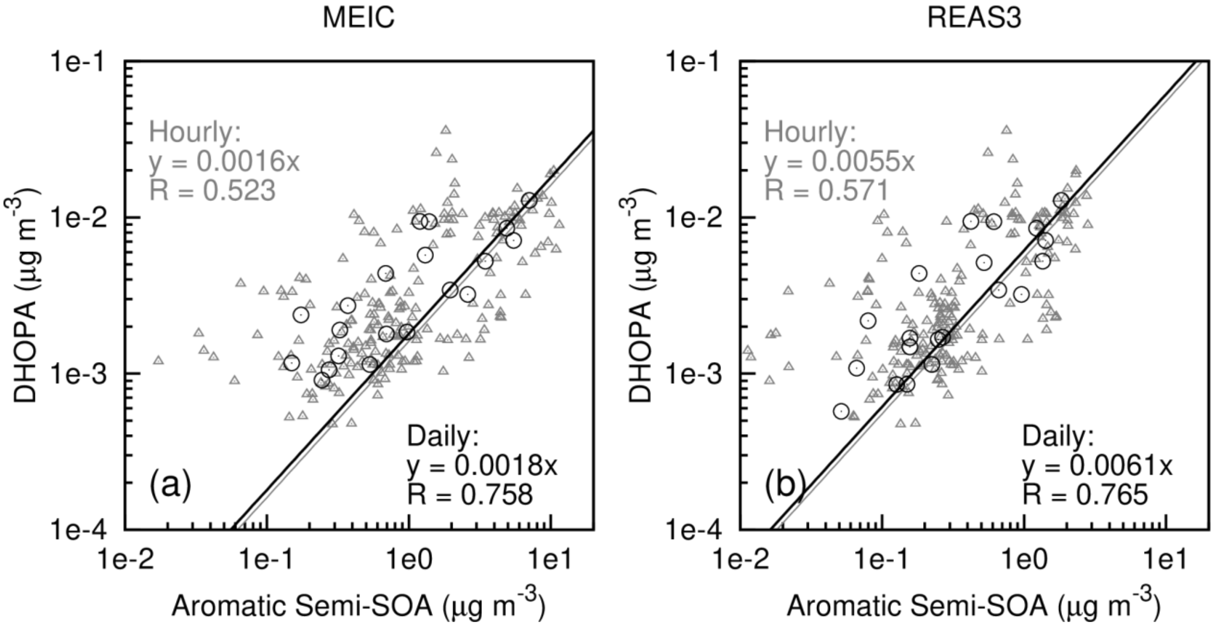


Figure 3-7 Linear correlation between model predicted aromatic-SOA (excluding oligomers and GLY/MGLY products) and observed aromatic-derived SOA tracer (DHOPA) with (a) REAS3 emissions, and (b) MEIC emissions.

3.4.2 Uncertainties in the tracer-to-terpene SOA ratio

The AERO6 module in CMAQ uses the traditional Odum 2-product model for monoterpene SOA predictions [116]. The newly released AERO7 module replaced the original monoterpene-SOA yield parameters with the volatility basis set (VBS) fit based on the recent experimental study by Saha and Grieshop [137]. The semi-volatile products were lumped into seven log-10-spaced bins based on saturation mass concentration (C*) from 10^{-2} to 10^{4} μg m^{-3}. The enthalpy of the VBS products ($\Delta H_{vap,i}$) was estimated using $\Delta H_{vap,i} = 80 - 11(\log_{10} C_i^*)$, which is based on linear regression of the chamber data. As seen in Figure S3-1 in Appendix B, at the standard temperature of 298K, the monoterpene-SOA yield is higher than the 2-product representation when the total organic aerosol concentration (C_{OA}) is less than ~27 μg m^{-3}, and it is lower than the 2-product yield with higher C_{OA}.

To check if this new model representation of monoterpene SOA can lead to significant changes in the estimated SOA and the tracer-to-SOA ratios, the CMAQ model was modified to include this new representation, and an additional simulation was conducted. The results were compared with the results from the original AERO6 module. As shown in Figure S3-2 in Appendix B, the updated VBS-style monoterpene SOA parameterization led to slightly higher SOA under low concentrations but lower SOA under high concentrations. As a result, the difference in the ratio of α-pinT to predicted TERP SOA is negligible.

While the α-pinT to predicted monoterpene SOA ratio is in good agreement with the reported tracer-to-α-pinene-SOA ratio, two additional factors should be further discussed. First, the tracer mass fraction reported in the previous chamber study of Kleindienst, Jaoui, Lewandowski, Offenberg, Lewis, Bhave and Edney [38] was calculated based on the ratio of the sum of 7 α-pinene SOA tracers to α-pinene SOA, but only two tracers (pinic acid and 3-MBTCA) were measured in this study. Thus, the calculated α-pinT to SOA ratio in this study should be increased to directly compare with the data from Kleindienst et al. [38], but the exact amount of adjustment is difficult to determine. Second, only a fraction of the monoterpene SOA is α-pinene-SOA concentration, as α-pinene is lumped with other monoterpenes in the model and the SOA yields of these individual monoterpene species are not the same [138]. Consider the two factors, the α-pinT to modeled α-pinene SOA ratio should be higher than α-pinT to modeled monoterpene SOA ratio, suggesting that the SOA from monoterpenes was likely underestimated in the model, either due to underestimation of emissions or the SOA yields.

3.4.3 Separate DHOPA from ARO1 and ARO2

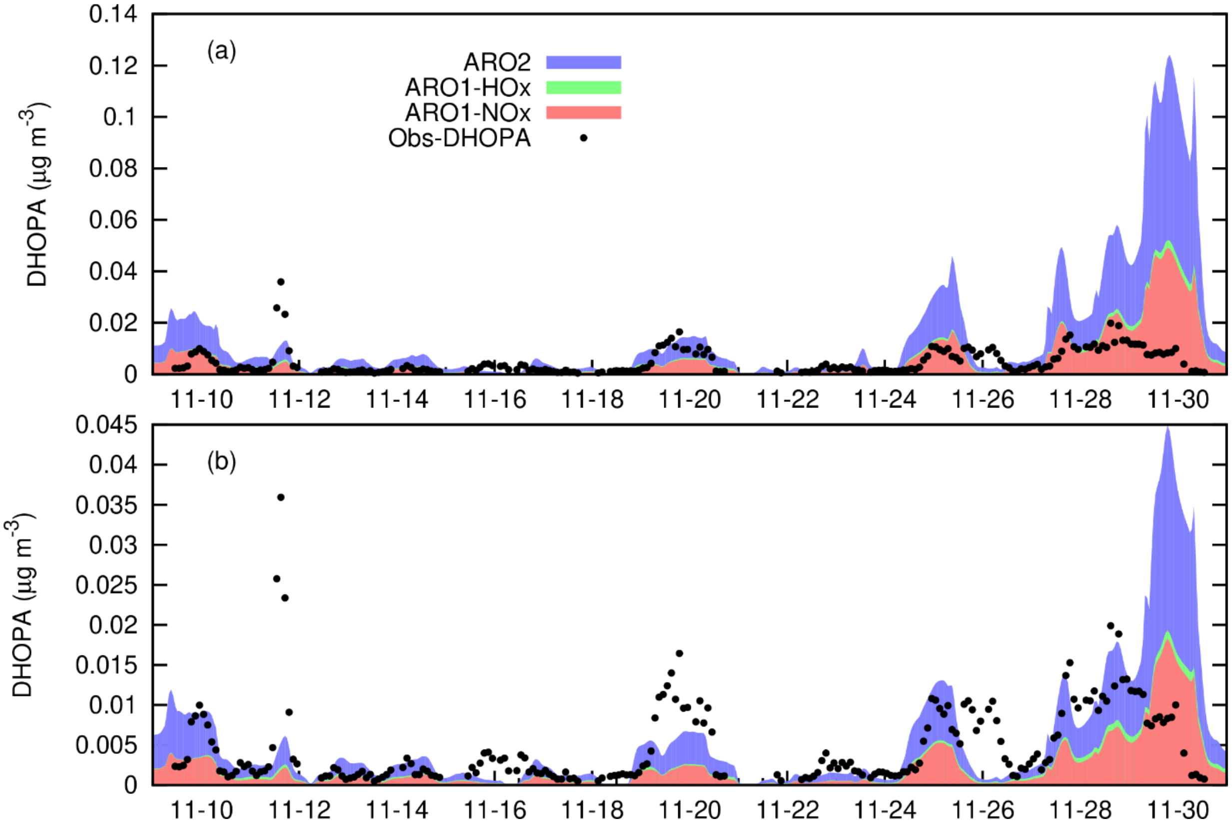


Figure 3-8 Estimated DHOPA concentrations from ARO1 under high-NO_x conditions (ARO1-NO_x) and low-NO_x conditions (ARO1-HO_x), and from ARO2 based on (a) MEIC and (b) REAS3, and the observed DHOPA in Shanghai.

In a recent chamber study on the aromatic SOA tracers, Al-Naiema, Offenberg, Madler, Lewandowski, Kettler, Fang and Stone[39] reported the mass fraction of DHOPA to the SOA from major aromatic precursors. The mass fraction of DHOPA to the toluene SOA under high and low NO_x conditions are $f_{tol\text{-}nox} = 0.0032 \pm 0.0004$ and $f_{tol\text{-}HOx} = 0.0068 \pm 0.0008$, respectively. The isomers of xylenes (o/m/p-xylenes) were tested individually in the chamber with NO_x present, and the average DHOPA mass fraction for the three xylene isomers is $f_{xyl} = 0.0033 \pm 0.00024$. Using predicted high-$NO_x$ and low-NO_x SOA from ARO1 and total SOA from ARO2, and the literature reported DHOPA to SOA ratio for toluene and xylene, which are the two most abundant species in ARO1 and ARO2, respectively, the amount of DHOPA from ARO1 and ARO2 were estimated, as shown in Figure 3-8. The estimated DHOPA based on predictions with REAS3 emissions generally agrees with the observed hourly data from the SAES site (Figure 3-8b), while the predictions using MEIC emissions are significantly higher (Figure 3-8a). Approximately half of the predicted DHOPA is from ARO1 under high-NO_x conditions, and the remaining is from ARO2.

3.5 Conclusions

The predicted hourly aromatic and monoterpene SOA show strong correlations with the hourly tracers DHOPA (R~0.6) and α-pinT (R~0.6-0.65). The correlations become stronger when daily average concentrations are considered, R~0.8 for aromatic SOA and R>0.8 for monoterpene SOA. The mass fraction of hourly and daily DHOPA is in the range of $5\text{-}6\times10^{-3}$ when SOA from oligomers, glyoxal (GLY), and methylglyoxal (MGLY) is excluded, close to the toluene mass fraction of DHOPA in aromatic SOA reported in the literature. This suggests that the CMAQ model can predict the semi-volatile aromatic SOA reasonably well. The mass fractions of hourly and daily α-pinT to the monoterpene SOA with REAS3 and MEIC emissions fall in a range of 0.13~0.25, similar to the reported α-pinT to α-pinene SOA mass fraction of 0.168. However, since α-Pinene is only one of the lumped monoterpenes, and the α-pinT used in this study did not include all tracer species measured in the chamber experiments from which the ratio was determined, a future study should individually track the emissions of major monoterpene species and the SOA formation from them so that a more detailed evaluation of the modeled biogenic SOA can be performed.

4 ESTIMATION OF AROMATIC SECONDARY ORGANIC AEROSOL USING A MOLECULAR TRACER – A CHEMICAL TRANSPORT MODEL ASSESSMENT [*]

A modified community multiscale air quality (CMAQ) model, which can simulate the regional distributions of 2,3-dihydroxy-4-oxopentanoic acid (DHOPA), a marker species for monoaromatic secondary organic aerosol (SOA), was applied to assess the applicability of using the DHOPA to aromatic SOA mass ratio (f_{SOA}) from smog chamber experiments to estimate aromatic SOA during a three-week wintertime air quality campaign in urban Shanghai. The modeled daily DHOPA concentrations based on the chamber-derived mass yields agree well with the organic marker field measurements (R = 0.79; MFB = 0.152; MFE = 0.440). Two-thirds of the DHOPA is from the oxidation of ARO1 (lumped less reactive aromatic species; mostly toluene), with the rest from ARO2 (lumped more reactive aromatic species; mostly xylenes). Modeled DHOPA is mainly in the particle phase under ambient organic aerosol (OA) loading but could exhibit significant gas-particle partitioning when a higher estimation of the DHOPA vapor pressure is used. The modeled f_{SOA} shows a strong dependence on the OA loading when only semivolatile aromatic SOA components are included in the f_{SOA} calculation. However, this OA dependence becomes weaker when non-volatile oligomers and dicarbonyl SOA products are considered. A constant f_{SOA} value of ~0.002 is determined when all aromatic SOA components are included, which is a factor of 2 smaller than the commonly applied chamber-based f_{SOA} value of 0.004 for toluene. This model-derived f_{SOA} value does not show much spatial variation and is not sensitive to alternative estimates of DHOPA vapor pressures and SOA yields, and thus

---

[*] Reprinted with permission from "Estimation of Aromatic Secondary Organic Aerosol Using a Molecular Tracer – A Chemical Transport Model" by Zhang, J., He, X., Gao, Y., Zhu, S., Jing, S., Wang, H., Yu, J. Z., and Ying, Q., 2021. Environ. Sci. Technol. Accepted, Copyright [2021] by American Chemical Society.

provides an appropriate scaling factor to assess aromatic SOA from DHOPA measurements. This result helps refine the quantification of SOA attributable to monoaromatic hydrocarbons in urban environments and thereby facilitates the evaluation of control measures targeting these specific precursors.

4.1 Introduction

Kleindienst et al.[38] developed a widely used organic tracer method to estimate the contributions of different precursors to ambient SOA concentrations. The mass fraction of the quantified precursor-specific organic tracers to total SOA formed from the precursor (f_{SOA}) was determined in smog chamber experiments.[38, 39] Based on the measured concentrations of the precursor-specific tracers in ambient aerosols, the total SOA formed from the precursor could be determined using the lab-determined mass factions, with the assumption that f_{SOA} values under ambient conditions are similar to those determined in the chamber experiments.[38, 40, 139-142] Among the tracers identified, 2,3-dihydroxy-4-oxopentanoic acid (DHOPA; C5H8O5) is a widely used tracer for SOA formed from monoaromatic hydrocarbons. Numerous studies have measured concentrations of DHOPA in various atmospheric environments and used it to estimate aromatic SOA concentrations.[11, 38-42] However, since the ambient conditions are different from those in the chamber experiments,[24, 25] the f_{SOA} determined in chamber experiments might not be directly applicable to estimate SOA in the ambient environment.

Chemical transport models (CTMs) have been widely used to quantitatively study the impacts of OA in polluted urban areas,[5-8] including SOA from aromatic compounds.[9-11] However, most regional models do not separate the various SOA components from a specific precursor, such as non-volatile oligomers and SOA from later generation dicarbonyl products such as GLY and MGLY. The lack of precursor-specific SOA representation and organic markers in the CTMs makes it difficult to evaluate SOA model predictions for target precursors against ambient measurements and assess whether a constant f_{SOA} could be used to estimate precursor-specific SOA using ambient measurements of the tracer concentrations.

In this study, the DHOPA yields from the oxidation of toluene and xylenes under high and low-NO_x conditions were determined using published chamber data and a theoretical equilibrium partitioning framework. A regional air quality model with precursor-resolved SOA representation was modified to include the organic marker and was applied to simulate DHOPA and aromatic SOA during a three-week wintertime air quality campaign in urban Shanghai when hourly concentrations of the organic marker species and major aerosol components were measured. This is the first study that directly models DHOPA and compares the modeled concentrations with ambient measurements. The modeled DHOPA and aromatic SOA concentrations were used to calculate f_{SOA} and the results revealed that the organic aerosol loading significantly influences f_{SOA}. Directly using the chamber measured f_{SOA} could lead to overestimating the traditional semivolatile components of the aromatic SOA but underestimating total SOA from aromatic compounds.

4.2 Methods

4.2.1 Modeling precursor-specific aromatic SOA

The CMAQ[80] v5.0.1 was used as a base model. The gas phase chemical mechanism was based on the SAPRC-11[143] same as the mechanism in the simulation of Section 3, which includes reactions of two lumped aromatic species, ARO1 and ARO2, among other SOA precursors. ARO1 represents the less reactive aromatic compounds, with toluene as the major species and including other monoalkylbenzenes. ARO2 represents the more reactive species such as xylenes and other di- and polysubstituted alkylbenzenes with xylenes as the dominant species. The SOA module in the AERO6 was used for SOA predictions. The SOA module includes SOA formation from three pathways. The first is the equilibrium partitioning of semivolatile products based on the classical Odum two-product model representations of SOA formation.[79] The ARO1 yields under high-NO_x conditions are based on the 4-bin VBS parameters reported by Hildebrandt et al.,[144] fitted to the Odum two-product model by Ying et al. [83] and then slightly increased in Hu et al.[88] to adjust for potential vapor wall-loss. The SOA yields for ARO2 high-NO_x oxidation products are based on the 5-bin VBS parameters reported by Ahlberg et al.,[145] fitted to the Odum two-product representation. The two-product model parameters are summarized in Table S4-3. The second is the formation of oligomers. The semivolatile SOA products from equilibrium partitioning are assumed to oligomerize according to a first-order decay with a half-life of 20 hours.[82] The third process is the irreversible surface uptake of GLY and MGLY. The uptake process is considered to occur on wet aerosols or in cloud droplets, where the GLY and MGLY molecules undergo further reactions to allow continued uptake. The uptake coefficients are based on those used by Li et al.[73]

Several changes to the SAPRC-11 and AERO6 were made to quantify SOA from ARO1 and ARO2 separately as discussed in Section 3.2.1. First, since GLY and MGLY can be formed from multiple precursors in SAPRC-11, the gas phase mechanism was modified so that GLY and MGLY from different precursors are tracked with separate tagged species. The SOA code in the AERO6 module was modified to allow additional tagged GLY and MGLY species. Second, the original AERO6 only includes two oligomer species to differentiate those from biogenic and anthropogenic precursors. Extra species were introduced to represent oligomers from different precursors. More details of the precursor-resolved SOA module can be found in Ying et al.[84] A summary schematic of the processes and species involved in the ARO1 SOA formation process is shown in Figure 4-1.

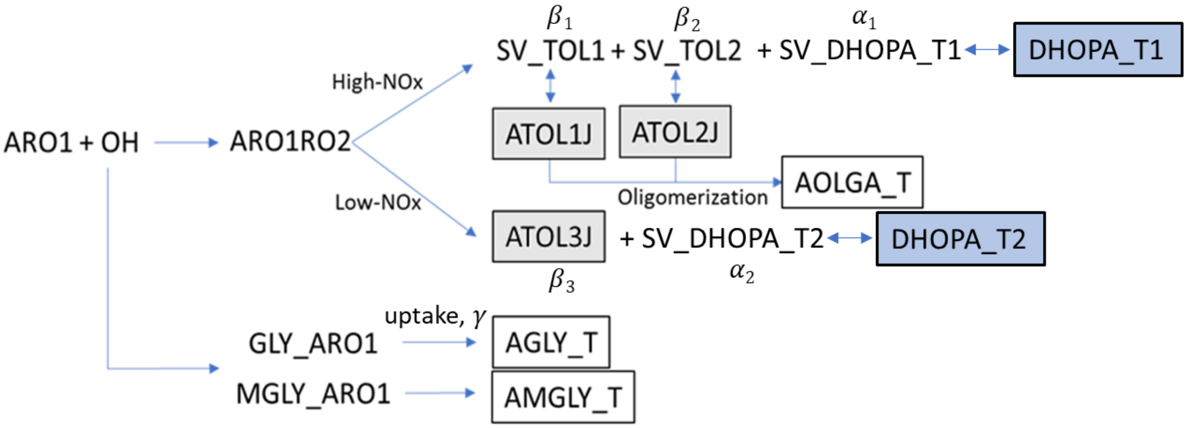


Figure 4-1 The precursor-tracking scheme to model SOA and DHOPA formation from ARO1 in the CMAQ model. The gray shaded boxes represent lumped SOA products from the traditional 2-product model. These three components are termed Semi-SOA in the paper, even though ATOL3J is non-volatile. AGLY_T and AMGLY_T are SOA products from irreversible surface uptake of glyoxal and methylglyoxal, respectively. SV_DHOPA_T1 and SV_DHOPA_T2 are gas phase DHOPA species, and the corresponding DHOPA_T1 and DHOPA_T2 are particle phase species. β_1 and β_2 are the mass yields of the semivolatile SOA species for the high-NO_x pathway, and β_3 is the SOA yield for the low-NO_x oxidation pathway. α_1 and α_2 are the mass yields of the DHOPA species for the high-NO_x and low-NO_x oxidation pathways. SOA and DHOPA formation from ARO2 are also tracked similarly.

4.2.2 *Estimation of DHOPA mass yield*

The mass yield of DHOPA (α), defined as the ratio of the amount of DHOPA produced to the amount of precursor VOC reacted, is one of the key parameters to model DHOPA in a chemical transport model. Based on the equilibrium absorptive partitioning between the gas and organic matter (OM) in the particle phase,[146] α in a chamber experiment from a specific aromatic precursor can be described by Eq. (4.1),[147]

$$\alpha = \frac{f_{SOA}}{F_{p,t}} \sum_{i=1}^{N} \beta_i \cdot F_{p,i} \tag{4.1}$$

where f_{SOA} is the measured particle phase DHOPA to SOA mass ratio; β_i's are the mass yields of an individual semivolatile product i (i could be a lumped SOA species). N is the number of SOA products used in fitting the chamber data. The chamber SOA formation data are typically described by the Odum two-product model [79] or the VBS approach.[148] $F_{p,t}$ and $F_{p,i}$ are the fraction of DHOPA and semivolatile products in the absorbing OM phase, respectively. $F_{p,t}$ depends on the organic aerosol concentration (C_{OA}), as shown in Eq. (4.2),

$$F_{p,t} = \left(1 + \frac{1}{K_{OM}C_{OA}}\right)^{-1} \tag{4.2}$$

K_{OM} is the absorptive gas/particle partitioning coefficient of DHOPA estimated using Eq (4.3) [146],

$$K_{OM} = \frac{RT}{10^6 p_L^0 \gamma \overline{MW}_{OM}} \tag{4.3}$$

where R is the ideal gas constant (m^3 Pa K^{-1} mol^{-1}); T is the temperature (K); p_L^0 is the subcooled vapor pressure (Pa); γ is the activity coefficient of the compound in the absorbing OM phase (assumed to be unity); $\overline{MW}_{OM}$ is the mean molecular weight of the OM phase assumed to be 200 g mol^{-1}, according to William et al.[149] Several different estimations of p_L^0, which vary by three-orders of magnitude, are considered in this study (Table 4-1). Method 1 in the vapor pressure estimation module from the Extended AIM aerosol thermodynamics model, hereafter E-AIM Method 1 (http://www.aim.env.uea.ac.uk/aim/aim.php) estimates p_L^0 to be 1.33×10^{-5} Pa using the group contribution method of Moller et al.[150] and the boiling point estimation method of Nannoolal et al.[151] Adopting this p_L^0 value, we estimate that the K_{OM} for DHOPA is approximately 0.930 m^3 µg^{-1} (or $C^* = 1/K_{OM} = 1.08$ µg m^{-3}). This result suggests that DHOPA is almost entirely partitioned to the organic aerosol phase under smog chamber OA conditions ($F_{p,t} \sim 0.986$; see Table 4-2). This agrees with the finding from a previous study that DHOPA was only detected in the particle phase.[152] The E-AIM Method 1 is also the preferred method in a study that assessed the capability of a group of 12 methods specifically for low volatility multifunctional groups.[153] E-AIM Method 3 and the US EPA's EPI Suite,[154] both providing much higher vapor pressure estimations, lead to very small K_{OM}'s of 7×10^{-4} and 3.9×10^{-3}, respectively. These K_{OM} estimates are unrealistic as most DHOPA would be in the gas phase, so vapor pressures from E-AIM Method 3 and the EPI Suite are not used in subsequent analyses.

Table 4-1 Estimated thermodynamic properties of DHOPA and mass yields from toluene and xylene oxidation under high and low-NO_x conditions

	DHOPA thermodynamic properties			DHOPA mass yield (α)		
	p_L^0 (Pa)	ΔH_{vap} (kJ mol^{-1})	K_{OM} (m^3 μg^{-1})	Toluene High-NO_x	Xylene High-NO_x	Toluene Low-NO_x
Non-volatile	-	-	-	0.00187	0.00088	0.00388
E-AIM Method 1*	1.33×10^{-5}	151.12	0.9295	0.00189	0.00090	0.00391
E-AIM Method 2^	9.51×10^{-5}	119.88	0.1303	0.00205	0.00100	0.00408
E-AIM Method 3$	1.74×10^{-2}	88.92	0.0007	0.03428	0.02024	0.04176
EPI Suite#	3.14×10^{-3}	82.53	0.0039	0.00774	0.00424	0.01073

* based on the normal boiling point estimation method of Nannoolal et al.[151] and vapor pressure equation of Moller et al. [150]

^ based on the normal boiling point estimation method of Nannoolal et al. [151] and the vapor pressure equation of Nannoolal et al. [155]

$ based on the normal boiling point estimation method of Stein and Brown[156] and the vapor pressure equation of Myrdal and Yalkowsky[157]

based on the normal boiling point estimation method of Stein and Brown[156] and the vapor pressure equation of Grain-Watson [153]

$F_{p,i}$, which describes the volatility of the lumped semivolatile products for toluene and xylenes, is calculated by Eq. (4.4),

$$F_{p,i} = \left(1 + \frac{C_i^*}{C_{OA}}\right)^{-1} \tag{4.4}$$

where C_i^* is the saturation mass concentration of the lumped products. The β_i and C_i^* values for toluene and xylene are based on the VBS parameters from Hildebrandt et al. [158] and Ahlberg et al.,[145] respectively. The Hildebrandt et al. parameters generate similar SOA yields to those generated using the two-product parameters used in the CMAQ model.

Table 4-2 Detailed calculation of DHOPA yields (α) in the photooxidation of toluene and xylene, with DHOPA subcooled vapor pressure estimated using E-AIM Method 1 (1.33×10^{-5} Pa).

SOA (μg m^{-3})	f_{SOA}	F_{pt}	$Y_{SOA}{}^{a}$	α
		Toluene High-NO_x [b]		
28.9	0.00450	0.964	0.309	0.0014
77.8	0.00415	0.986	0.470	0.0020
73.5	0.00245	0.986	0.460	0.0011
97.6	0.00307	0.989	0.512	0.0016
103.4	0.00242	0.990	0.523	0.0013
94.7	0.00296	0.989	0.506	0.0015
125.6	0.00462	0.992	0.561	0.0026
116.9	0.00419	0.991	0.547	0.0023
88.1	0.00636	0.988	0.493	0.0032
Average	0.00385			0.00189 ± 0.0007
		Xylenes High-NO_x [c]		
222.7	0.0024	0.9884	0.376	0.00091
103.2	0.0036	0.9752	0.256	0.00095
70.5	0.0039	0.9641	0.211	0.00085
Average	0.0033			0.00090 ± 0.00004
		Toluene Low-NO_x [d]		
143.2	0.0068	0.993	0.570 [d]	0.00388

a. SOA mass yield $Y_{SOA} = \sum_{i=1}^{N} \beta_i \cdot F_{p,i}$. The toluene and xylene SOA yield under high NOx conditions were estimated by the VBS model from the chamber results of Hildebrandt et al. [158] and Ahlberg et al., respectively. For toluene, the saturation concentrations (C*) for the four VBS bins are 1,10,100 and 1000 μg m^{-3}, and the corresponding mass yields (β) are 0.01, 0.24, 0.7, and 0.7. For xylene, the saturation concentrations (C*) span from 0.1 to 1000 μg m^{-3}, with five log-spaced bins. The corresponding β values are 1.09×10^{-4}, 3.37×10^{-3}, 9.64×10^{-2}, 0.180, 1.04.

b. Toluene oxidation by OH under high-NO_x conditions is based on data reported by Kleindienst et al. [38] Chamber temperature was 25 °C.

c. Xylene oxidation by OH under high-NO_x conditions and toluene oxidation under low-NO_x conditions are based on data reported by Al-Naiema et al. [39]. Chamber temperature was not reported but assumed to be 29.1°C based on the cited references [159, 160]. The α value for xylene under low-NO_x conditions is assumed to be the same as the high NO_x value.

d. SOA products under low NO_x conditions are non-volatile. The fixed yield of 0.570 is estimated using the yield of 0.300 reported by Ng et al. [30] and scaled up using the adjustment factor of 1.9 suggested by Zhang et al.[161] to account for vapor wall loss.

The DHOPA mass yields derived in this study using Eq. (4.1) are listed in Table 4-1 based on the DHOPA chamber data reported by Kleindienst et al.[38] for toluene and Al-Naiema et al.[39] for xylenes. The DHOPA mass yields for the toluene and xylene oxidation reaction under high-NO_x conditions using the E-AIM Method 1 estimated p_L^0 is estimated to be 0.0019 ± 0.0007 and 0.00090 ± 0.00007, respectively. The estimated mass yields with different vapor pressure estimations are summarized in Table 4-1. With p_L^0 set at 1.33×10^{-5} Pa, the detailed calculations for the mass yields of DHOPA at various SOA concentrations are listed in Table 4-2. As shown in Table 4-1, the DHOPA mass fractions based on the E-AIM Method 1 and 2 are similar and are very close to those estimated assuming non-volatile DHOPA, even though the estimated p_L^0's differ by almost seven times. While the DHOPA yields for ARO1 and ARO2 are based on data for toluene and xylene, more detailed estimation with data for ethylbenzene, trimethylbenzene, and naphthalene does not lead to significantly different estimations (see Section S4.1 in Appendix C).

Since a non-negligible fraction of DHOPA can still partition to the gas phase based on the K_{OM} values estimated using E-AIM Method 1 and 2, especially at low OA loadings, DHOPA is added to the SOA module as a semivolatile product from the oxidation reactions of the lumped aromatic compounds, with the mass yields listed in Table 4-1. The temperature dependence of p_L^0 is calculated using the Clausius-Clapeyron equation with an estimated ΔH_{vap} of 151 kJ mol^{-1} for E-AIM Method 1 and 120 kJ mol^{-1} for E-AIM Method 2. These values agree with the estimation of ΔH_{vap} (129 kJ mol^{-1}) using a simple empirical equation proposed by Epstein et al. [162], which uses C^* as the sole predictor. DHOPA from ARO1 and ARO2 under high-NO_x and low-NO_x conditions are tracked with four gas phase and four organic phase species. An illustration of the DHOPA scheme for ARO1 is shown in Figure 4-1. The impact of organic water uptake under ambient relative humidity conditions [73, 163, 164] is not considered in the DHOPA partitioning and adjustment of chamber-derived f_{SOA}, and will be explored in a future study.

4.2.3 CMAQ simulation

The modified CMAQ model is used to simulate DHOPA and SOA formation in China for November 2018. The CMAQ model has the same domain setup as discussed in Section 3.2.2. Initial and boundary conditions for the CMAQ model are generated using the vertical profiles distributed with the CMAQ model. The first five days of the simulation results are treated as spin-up and excluded from the final analysis. The base case simulation describes in Section 4.3 uses DHOPA yields estimated using thermodynamic properties from E-AIM Method 1. Several additional simulations are also performed, as listed in Table S4-4.

Anthropogenic emissions are based on the REAS3.[121] Selected speciation profiles from the SPECAITE database developed by the US EPA are used to estimate emissions of model-ready VOCs.[122] Windblown dust emissions in the entire domain are generated inline.[86] Biogenic emissions are produced by the MEGAN v2.10.[123] The meteorological inputs are generated using WRFv4.2 with initial and boundary conditions from the NCEP GDAS/FNL 0.25 Degree Global Tropospheric Analyses and Forecast Grids. The land use/land cover and topographical data are based on the 30 s resolution default WRF input dataset. Reanalysis nudging is enabled to improve the agreement between predicted and observed meteorological parameters.[119] The major physics options for the WRF simulations are described by Zhang, Li, Ying, Yu, Wu, Cheng, He and Jiang

[120] WRF performance in this study is discussed in Section 3.3.1, which is comparable to other studies using WRF in China simulations.[120, 125-127]

4.3 Results

4.3.1 General model performance evaluation

The predicted hourly $PM_{2.5}$ mass and major components are compared with the observations from SAES. Details of the measurements and the observation data analyses have been documented elsewhere.[41, 118] As discussed in Section 3.3.1, the $PM_{2.5}$ mass concentrations with REAS3 emission inventory are slightly overestimated, with a mean fractional bias (MFB) of 0.34 and a mean fractional error (MFE) of 0.52. Ammonium (MFB = 0.17, MFE = 0.48), nitrate (MFB = -0.10, MFE = 0.56), and total organic carbon (MFB = -0.08, MFE = 0.44) agree well with the observations, which provides confidence that the model correctly represents the oxidation capacity of the atmosphere and provides the amount of organic matter needed for DHOPA and SOA partitioning calculations. The over-prediction of $PM_{2.5}$ is mostly caused by over-predictions of sulfate (MFB = 0.48, MFE = 0.60) and elemental carbon (MFB = 0.76, MFE = 0.80), and sometimes nitrate, which are not expected to influence the SOA calculation significantly.

4.3.2 Evaluation of DHOPA and its precursors

Predicted precursor concentrations of ARO1 (mostly toluene) and ARO2 (mostly xylenes) agree well with observations of major aromatic compounds lumped to these two species, which are available for November 11-21, 2018 (Figure S4-7). Over-predictions are mostly due to higher predictions on November 15, when both ARO1 and ARO2 are overestimated significantly. However, since large spatial gradients around the monitor sites were predicted, the over-predictions could be caused by slight inaccuracy in the wind fields.

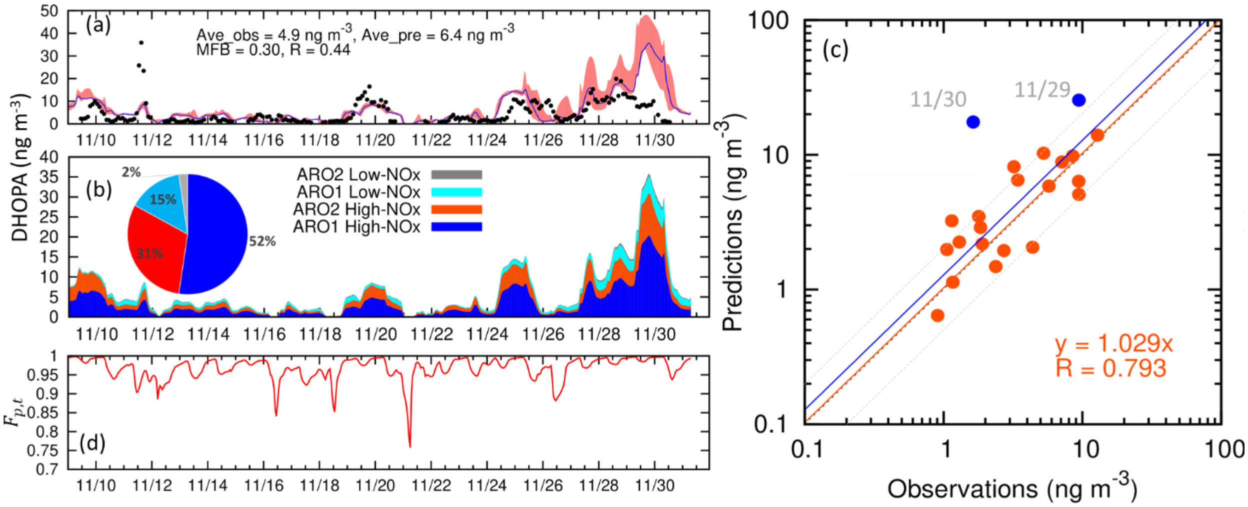


Figure 4-2 (a) Predicted and observed hourly concentrations of DHOPA in Shanghai, from November 10 to November 30, 2018, (b) contributions to predicted DHOPA from ARO1 and ARO2 high-NO_x and low-NO_x pathways, (c) predicted and observed daily DHOPA, and (d) fraction of DHOPA in the particle phase. The dashed line in (c) is the 1:1 line. Blue dots are data for November 29 and 30 when all PM components are over-predicted. The solid blue line is from linear regression using all data points. The solid red line is from linear regression, excluding the two blue data points. Both lines are determined using least-squares regression by forcing a zero intercept. Dashed lines are the 1:2 and 2:1 lines. R is the Pearson correlation coefficient. Shaded areas in (a) represent the concentration ranges from the 3×3 grids with the observation site in the center grid.

The base case model successfully captures the concentrations and temporal variations of hourly DHOPA at the monitor site, as shown in Figure 4-2(a), with an MFB value of 0.30. Average predicted and observed concentrations are 6.4 and 4.9 ng m^{-3}, respectively. The higher prediction is mostly caused by significant over-prediction on the last two days associated with the underprediction of wind speeds during calm conditions around Shanghai, which causes the over-accumulation of all pollutants. Predicted and observed daily average DHOPA concentrations show a strong correlation (R = 0.793) and low bias and error (MFB = 0.152, MFE = 0.440) when the two over-prediction days are excluded, as shown in Figure 4-2 (c). In Shanghai, most of the DHOPA is from the ARO1 (52%) and ARO2 (31%) high-NO_x pathways, while the low NO_x pathways only account for 17% of the DHOPA (Figure 4-2b). While the majority of the DHOPA is from ARO1 (67%), a significant fraction is from ARO2 (33%), indicating that it was likely incorrect to attribute all DHOPA-based aromatic SOA estimation to toluene, as did in several previous studies.[41, 61, 165] The regional distributions of monthly average DHOPA from these four pathways are shown in Figures S3 and S4. In most areas, the ARO1 high-NO_x pathway contributes to 50-60% of the DHOPA. The ARO2 high-NO_x pathway is also important, with contributions ranging from 20-45%. The low-NOx pathways have lower contributions but are much more important in remote areas and over the oceans where NO_x concentrations are low. The predicted DHOPA is almost entirely in the particle phase, with $F_{p,t}$ mostly above 0.95, as shown in Figure 4-2 (d).

4.3.3 Impact of OA loading on DHOPA-based estimation of ambient aromatic SOA

Since the DHOPA mass yields are independently derived based on smog chamber experimental data, the good agreement of predicted DHOPA with ambient measurement strongly supports that the method described in Section 4.2.2 appropriately describes the formation and partitioning of DHOPA from precursors. It also implies that f_{SOA}, the DHOPA to SOA mass ratio, should be adjusted to account for the differences in the OA loadings and temperature between the chamber and ambient conditions. The adjustments are done using Eq. (4.5),

$$f_{SOA} = \frac{\alpha F_{p,t}}{\sum_{i=1}^{N} \beta_i \cdot F_{p,i}} \tag{4.5}$$

where $F_{p,t}$ and $F_{p,i}$ are estimated using Eqs. (4.2) and (4.4), respectively, with temperature corrected K_{OM} and C^* values and the measured C_{OA}. The amount of ambient aromatic SOA is estimated by splitting the measured DHOPA into four formation pathways and divide the DHOPA from each pathway with the corresponding f_{SOA}.

Figure 4-3 shows the comparison of estimated aromatic SOA using adjusted f_{SOA} and fixed f_{SOA}. Using a single value of f_{SOA} based on the toluene data alone does not lead to significantly different estimations than those using split DHOPA and individual f_{SOA} values (Figure 4-3a). However, estimated SOA using adjusted f_{SOA} is consistently lower than those estimated using constant values. The relative difference between the two approaches increases from about 50% at the OA loading of approximately 15 μg m^{-3} to 140-200% when OA is ~5 μg m^{-3}.

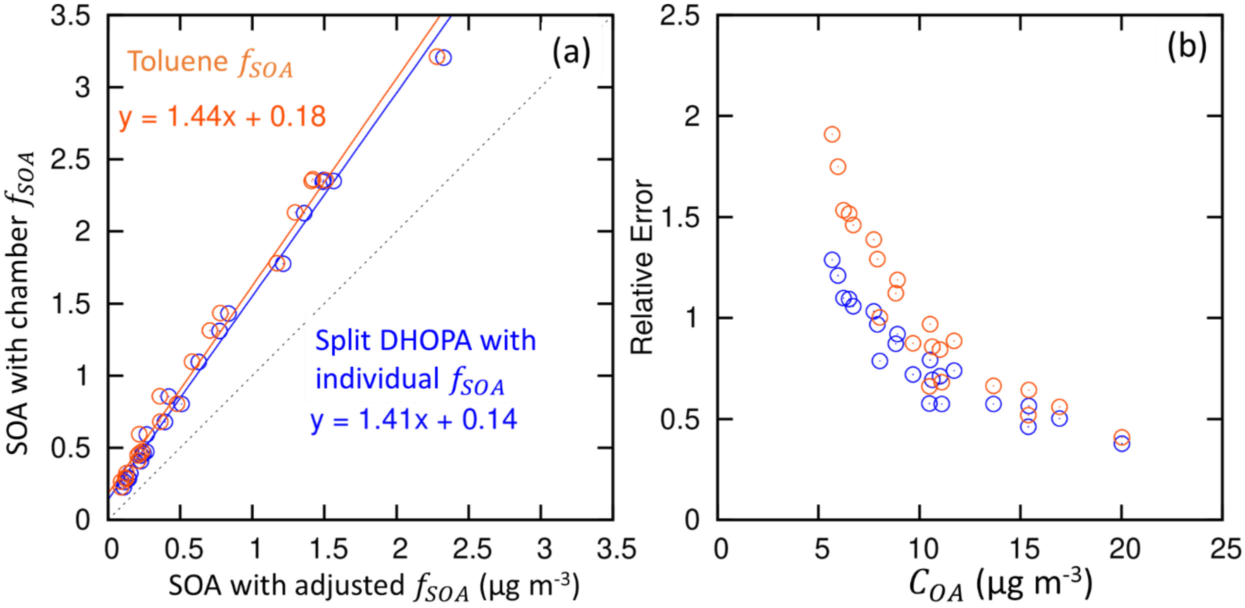


Figure 4-3 (a) Difference in the estimated daily aromatics SOA using constant chamber reported $\boldsymbol{f_{SOA}}$ (C_c) and adjusted $\boldsymbol{f_{SOA}}$ (C_{adj}); (b) relative error in the estimated aromatics SOA as a function of the organic aerosol loading. Orange dots show estimations using high-NO_x toluene $\boldsymbol{f_{SOA}}$ only and blue dots show estimations using split DHOPA based on the modeled DHOPA fractions in Figure 4-1 and $\boldsymbol{f_{SOA}}$ for individual pathways. Relative difference is defined as (C_c-C_{adj})/C_{adj}. Constant $\boldsymbol{f_{SOA}}$ values used in the calculation are listed in Table 2.

4.3.4 *Modeled relationships between DHOPA and aromatic SOA*

The smog chamber experiments were performed under low relative humidity so that little SOA was from heterogeneous uptake of glyoxal and methylglyoxal. The duration of the experiments was usually not long enough to have significant oligomer formation. Thus, it is expected that using ambient DHOPA and the adjusted f_{SOA} would lead to underestimation of the actual amount of SOA from aromatic compounds.

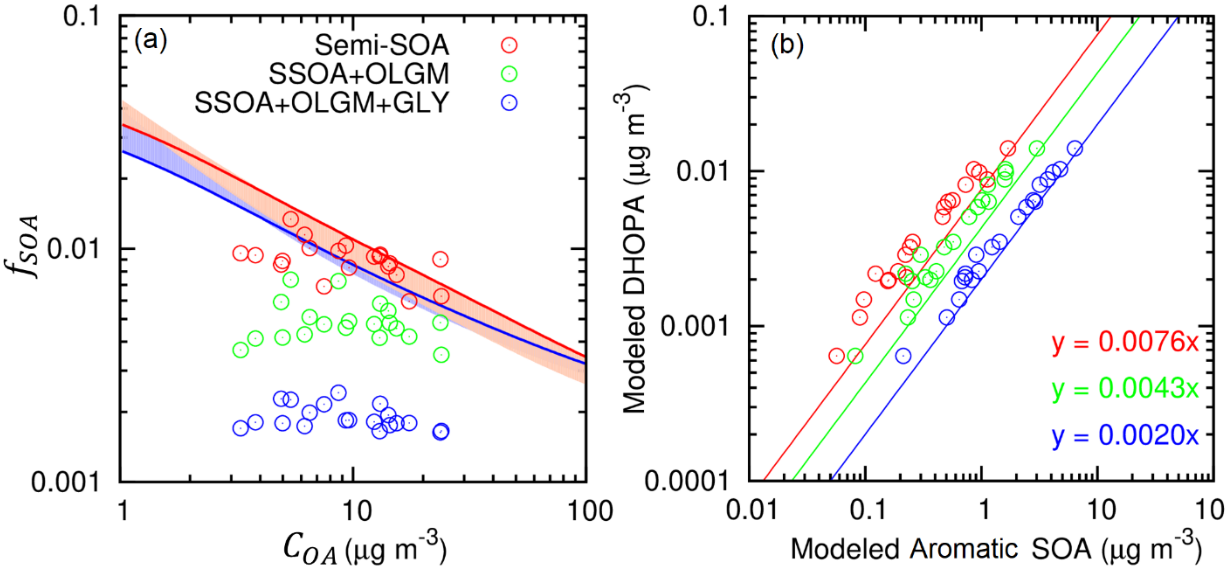


Figure 4-4 (a) The modeled DHOPA mass fraction $\boldsymbol{f_{SOA}}$ (DHOPA/aromatic SOA) as a function of total OA concentrations. The shaded areas represent the theoretical calculations of mass fraction for temperatures between 283 and 295 K. The solid lines are calculated mass fractions under the standard temperature of 298 K. Blue and red represent toluene and xylene, respectively. (b) Average ratio of DHOPA/aromatic SOA based on linear regression with a zero intercept. Red, green and blue dots represent semivolatile SOA (semi-SOA, SSOA), SSOA + oligomers (OLGM), and SSOA+ OLGM + glyoxal and methylglyoxal SOA products (GLY), respectively.

Figure 4-4(a) shows that at the SAES site, the CMAQ modeled f_{SOA} considering only semivolatile components (i.e., the mass ratio of DHOPA to semivolatile constituents) increases with decreasing total OA concentrations. The trend is well represented by the theoretical lines for toluene and xylene generated using Eq. (4.5). However, the modeled f_{SOA} values are significantly lower when oligomers and dicarbonyl SOA products from the oxidation of aromatic compounds are included, as these products account for a large fraction of the SOA generated from the aromatics (Figure S4-10). Since these products are treated as non-volatile in the model, f_{SOA} values show less dependence on the OA concentrations. As shown in Figure 4-4(b), the appropriate f_{SOA} to estimate aromatic SOA from DHOPA, including oligomers and dicarbonyls SOA, is approximately 0.0020 at this location, based on a linear regression of modeled DHOPA and modeled aromatics SOA with a forced zero intercept. Thus, while the common practice of using f_{SOA}~0.0040 leads to an overestimation of the semivolatile aromatic SOA (see Figure 4-3), it might still underestimate the total amount of SOA from the aromatic compounds by a factor of 2.

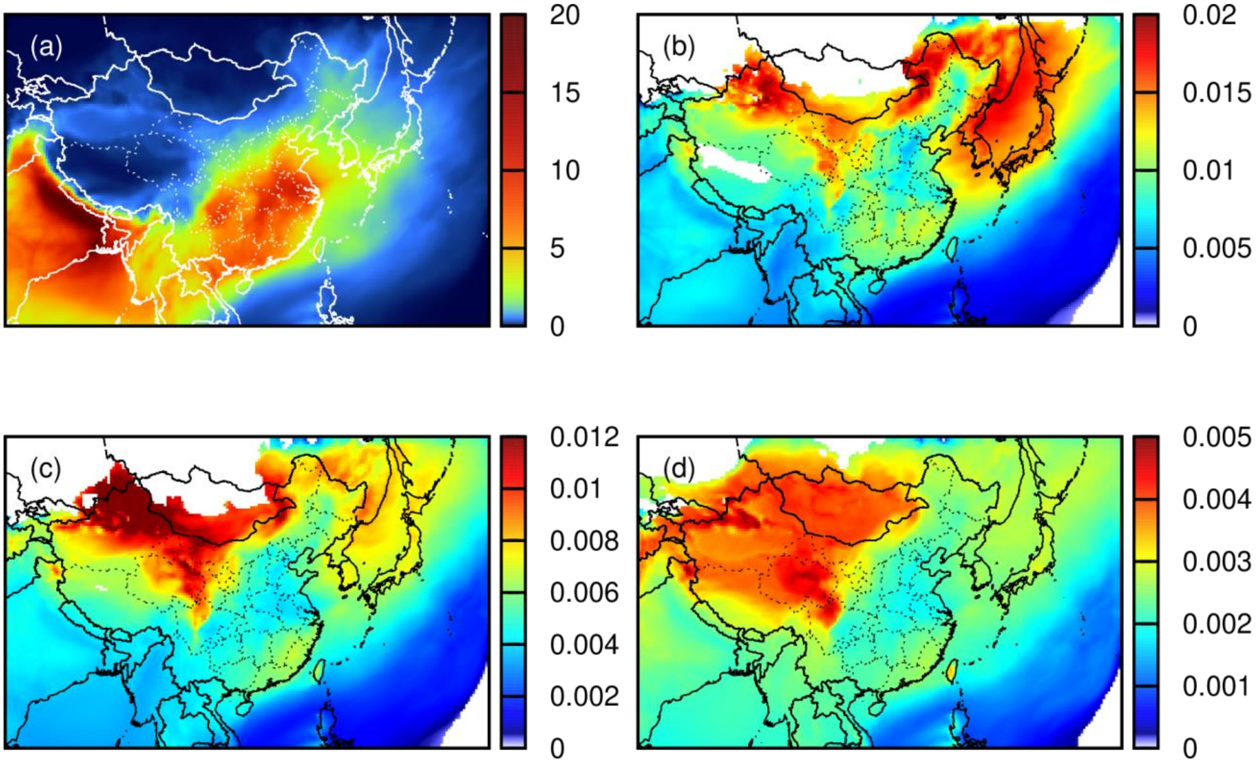


Figure 4-5 Regional distribution of monthly average (a) DHOPA (ng m^{-3}) and f_{SOA} (b-d). The f_{SOA} represents the mass fraction of DHOPA in the aromatic (b) semi-SOA; (c) semi-SOA and oligomers; (d) semi-SOA, oligomers, and glyoxal and methylglyoxal products with a threshold of SOA larger than 0.01 μg m^{-3}.

Figure 4-5 shows that the modeled monthly average DHOPA concentrations vary between 5-20 ng m^{-3} in central, southern, and eastern China. f_{SOA} values calculated using the monthly average SOA with semivolatile components only show more significant regional variations due to spatial variations in the OA loading and temperature. f_{SOA} values for semivolatile components plus oligomers and all aromatic SOA components show less significant spatial differences. Table S4-6 shows the mean and standard deviation of f_{SOA} at different DHOPA concentration ranges. The f_{SOA} value of 0.002 derived in Shanghai is generally applicable in most polluted areas. For all grid cells with DHOPA concentration between 2-5, 5-10, and >10 μg m^{-3}, the mean and standard deviation of f_{SOA} are 0.00231 ± 0.00025, 0.00225 ± 0.00021 and 0.00222 ± 0.00012, respectively.

4.4 Discussion

4.4.1 Uncertainty in DHOPA thermodynamic data

The modeled particle phase DHOPA concentrations are affected by its mass yields and the gas-to-particle partitioning of DHOPA. The mass yields of DHOPA from toluene and xylene based on the E-AIM Method 1 and 2 only differ by 4-8% (Table 4-1), with Method 2 giving slightly higher yields. However, under ambient temperatures of 281-294 K and OA loading of less than 20 μg m^{-3} (most less than 10 μg m^{-3}) in Shanghai, the fraction of DHOPA in the particle phase, F_{pt}, can be less than 70% (Figure S4-11), potentially leading to differences in the predicted DHOPA concentrations. An additional CMAQ simulation (Sensitivity Case 1; See Table S4-4) was conducted using the saturation vapor pressure estimated using the E-AIM Method 2 along with the slightly higher DHOPA mass yields. Predicted DHOPA concentrations are lower than those from the base case (Figure S4-12a, 12c). When the last two days were excluded, the predictions are slightly lower than the observations by about 16% on average because F_{pt} values frequently fall below 0.8, reaching as low as 0.4 (Figure S4-12b). The model estimated f_{SOA} values with all aromatic SOA components included, however, are different from those in the base case by only 5% (0.0019 vs. 0.0020).

As the model results are affected by many other input parameters, it is impossible to conclude whether the DHOPA thermodynamic data based on the E-AIM Method 1 are more realistic than those based on the E-AIM Method 2. Future in-situ observations with higher ambient temperature and simultaneously measuring gas and particle DHOPA can provide the information needed to further evaluate the DHOPA related parameters.

4.4.2 Uncertainty in Xylene SOA yields

The estimation of DHOPA mass yields is also be affected by the SOA yields. In several previous studies, the SOA yields for xylene were the same as those of toluene to provide an upper limit estimation of aromatic SOA. The toluene SOA yields are significantly higher than the xylene yields used in the base case (see Figure S4-13). To evaluate the sensitivity of DHOPA predictions and the modeled f_{SOA} to this upper-limit xylene yield estimation, two additional simulations were conducted (Sensitivity Case 2 and 3; Table S4-4), using DHOPA thermodynamic data from E-AIM Method 1 and 2, respectively. In both simulations, the xylene SOA yields are the same as those for toluene. As shown in Figures S9 and S10, simulated DHOPA concentrations are higher than those in the base case mainly because the DHOPA mass yields from xylene are increased by more than a factor of 3.

The modeled ratio of DHOPA to semivolatile aromatic SOA show large variations among the four simulations (0.0073 ± 0.0005). However, f_{SOA} values based on all aromatic SOA components are rather stable (0.0020 ± 0.0001), which suggests that $f_{SOA} = 0.002$ is broadly appropriate to estimate the amount of total aromatic SOA in the ambient environment based on DHOPA measurements.

4.4.3 Sensitivity to temperature

As most of the DHOPA is in the particle phase in the base case simulation, lower temperatures would not lead to significant differences. We explore how higher temperature can affect gas-to-particle partitioning and the predicted particle phase DHOPA concentrations. In two additional sensitivity simulations (Sen. Cases 4 and 5; see Table S4-4), temperatures in the entire domain are increased uniformly by 5 K and 10 K while other parameters remain the same as in the base case. Compared to the base case, the predicted average particle phase DHOPA concentrations in Shanghai decrease from 6.4 ng m^{-3} in the base case to 6.0 ng m^{-3} and 5.3 ng m^{-3} with the temperature increasing by 5 K and 10 K, respectively, as shown in Figure S4-16(a). Due to higher temperatures, more DHOPA is partitioned to the gas phase. Figure S4-16(b) shows that the $F_{p,t}$ values from Sen. Case 4 are mostly between 0.7 and 1.0 and could be as low as approximately 0.5. While responding to the 10 K increase of temperature, the $F_{p,t}$ values from Sen. Case 5 are mostly between 0.5 and 0.9. Figure S4-17(a) shows the regional distribution of f_{SOA} responses to the uniform temperature increase in China. The regional plots show higher temperature sensitivity in the areas with larger f_{SOA} in the base case, which is likely caused by the lower OA loadings. In the high OA loading areas, the f_{SOA} changes are usually less than 10% when the temperature is increased by 5 K, which means the f_{SOA} calculated in the base case is still appropriate to be applied for aromatic SOA estimation responding to the temperature uncertainties within 5 K. While with the uniform increase of temperature by 10 K, the f_{SOA} changes are slightly larger but mostly lower than 20%. However, the f_{SOA} based on the total aromatic SOA is relatively stable in the responses of temperature increase. The f_{SOA} values in the southern coastal areas show reducing trends due to significantly higher temperatures. In those areas, the monthly average temperatures in the base case are as high as 20~25°C and reach 30-35°C in Sen. Case 5. Because the enthalpy of vaporization for DHOPA (151 kJ mol^{-1}) used in the model is larger than the CMAQ default values for the aromatic SOA (18 kJ mol^{-1} for toluene SOA and 32 kJ mol^{-1} for xylene SOA), more DHOPA is evaporated into the gas phase than the aromatic SOA, leading to lower f_{SOA} values.

4.4.4 *Summertime DHOPA and* f_{SOA}

To further evaluate the model's ability to predict DHOPA and its gas-particle partitioning, the modified CMAQ model was applied to simulate DHOPA in China from June to August 2012 and compared with observation data from 15 sites across China from the literature.[91, 136, 166-168] The modeled DHOPA concentrations show good agreement with the observations (MFB = -0.03, MFE = 0.76, R=0.43). Most of the DHOPA is still from the ARO1 high-NO$_x$ pathway, although the contributions of the ARO1 low-NO$_x$ pathway are higher in the summer due to lower NO$_x$ concentrations. The modeled f_{SOA}, when all aromatic SOA components are included, remains at approximately 0.002, without significant spatial and temporal variations. These results suggest that the DHOPA is semivolatile, and the mass yields and thermodynamic parameters used in the model are applicable in different environmental conditions. A detailed description of this modeling exercise is included in Section S4.1 of Appendix C.

5 A MODELING STUDY OF ISOPRENE SECONDARY ORGANIC AEROSOL TRACERS CONSIDERING VOLATILITY AND POTENTIAL CONTRIBUTIONS FROM NON-HETEROGENEOUS FORMATION PATHWAY

The isoprene-SOA scheme in the community multiscale air quality (CMAQ) model is expanded to simulate the unique isoprene markers 2-methyltetrols (2-MT) and 2-methylglyceric acid (2-MG) by treating them as semivolatile species and including a non-heterogeneous formation pathway. Predictions from the expanded model are evaluated against field measurements from 4 sites in the Pearl River Delta (PRD) region and 14 sites across China during the summertime. The predicted 2-MT and 2-MG from the modified model agree with the PRD region observations (MFB = -0.007~0.002 and R = 0.6~0.8) after adjusting the gas-particle partitioning using observed organic aerosol (OA) concentrations. Predictions of 2-MT at the other sites also show reasonable agreement with observations (MFB = -0.286 and R = 0.534) after adjusting for potential errors in the model OA but 2-MG concentrations are under-predicted (MFB = -0.989 and R = 0.192). The modified scheme leads to consistently improved predictions of the two tracers with lower biases and errors and much stronger correlations with observations than the original scheme, in which the two tracers are non-volatile and are only produced in the aerosol liquid water after the surface uptake of their precursors, isoprene epoxide and methacrylic acid epoxide. The modeled mass fraction (f_{SOA}) of the sum of 2-MG and 2-MT in the total isoprene-SOA varies gently, between ~0.01-0.02 in polluted areas, suggesting that the chamber-derived f_{SOA} of 0.063 may lead to large underestimations of overall isoprene SOA. This work provides new insights into the formation of the key isoprene SOA tracers and would allow better assessments of the contributions of biogenic emissions to regional and global aerosol burden.

5.1 Introduction

Isoprene is the main biogenic nonmethane volatile organic compound emitted into the atmosphere, with estimated annual global emissions in the range of 500-750 Tg.[123] It is highly reactive and can be oxidized by OH,[14, 169] O_3,[170] and NO_3[171] to generate semivolatile and low volatile organic aerosol products.[12, 45, 172-174] It is estimated that 20% of the global SOA is formed from isoprene oxidation products in cloud water.[44] In areas with larger isoprene emissions, such as the Eastern United States, isoprene contributes to more than 45% of the summertime SOA.[84]

The chemical composition of isoprene SOA has been extensively investigated in modeling studies and chamber experiments.[12, 43-45] Among the identified isoprene SOA species, 2-methyltetrols (2-MT, including 2-methylthreitol and 2-methylerythritol) and 2-methylglyceric acid (2-MG) are considered as unique tracer compounds,[38, 46] and are used to estimate the overall ambient isoprene-derived SOA,[47, 48] with the assumption that the mass fractions of the tracers in the ambient isoprene SOA (f_{SOA}, or f_{SOC} for secondary organic carbons, SOC) measured in the chamber experiments are similar to those in the ambient air. For example, Ding et al.[48] estimated the isoprene SOC across China in an entire year using the chamber-derived f_{SOA} value of 0.063 reported by Kleindienst et al.[38] However, the applicability of the chamber-derived f_{SOA} or f_{SOC} values under ambient conditions needs to be verified using a chemical transport model with a detailed chemical mechanism of isoprene to determine the isoprene SOA and the associated unique tracers.

The formation of the isoprene tracers is affected by the abundance of NO_x in the atmosphere. Under low-NO_x conditions, isoprene is oxidized to form isoprene hydroxyhydroperoxides (ISOPOOH) and subsequently generates isoprene epoxydiols (IEPOX) in the gas phase,[12, 175] which is transferred to the aerosol phase and produces 2-MT through multiphase chemistry.[12] Under high-NO_x conditions, the photooxidation of isoprene generates methacryloylperoxynitrate (MPAN), which further reacts with OH to generate methacrylic acid epoxide (MAE) in the gas phase. MAE is subsequently transported to the particle phase and generates 2-MG.[45] In some lab experiments,[12, 45] formations of 2-MT and 2-MG are enhanced in the presence of acidic seed aerosols.

Pye et al.[49] implemented a multiphase reaction scheme in a regional chemical transport model to explicitly predict the acidity-dependent surface uptake of IEPOX and MAE and the reactions that generate 2-MG and 2-MT in the aerosol liquid water, which are considered to be non-volatile.[49, 50] The modeled tracer concentrations in the Eastern US show improved agreement with the observations than the estimations based on chamber-derived f_{SOA} and the modeled Odum 2-product[116] semivolatile isoprene SOA. Budisulistiorini et al.[50] implemented similar reactions into a box model and reaffirmed the importance of aqueous phase chemistry in the isoprene SOA formation based on field measurements during the 2013 Southern Oxidant and Aerosol Study. In China, Qin et al.[176] applied the Pye et al. algorithm to study summertime biogenic SOA, and found that modeled isoprene concentrations were mediated by aerosol pH.

2-MG and 2-MT have also been detected in the aerosol phase in chamber experiments under dry conditions, and non-acidic seed aerosols.[38, 46, 51-53] For example, Kleindienst et al.[51] detected the formation of 2-MT and 2-MG in chamber experiments where RH was lower than 3%. Nestorowicz et al. reported that significant amounts of 2-MT and 2-MG were produced in the oxidation of isoprene using ammonium sulfate as seed particles under RH of 9-49%. These experiments suggest that formation in the aerosol water might not be the sole important pathway in the formation of the tracers. For instance, the formation of the tracers in the gas phase,[43, 51] followed by gas-to-particle partitioning, may also contribute to the tracers in the aerosol phase.

Al-Naiema and Stone[11] detected 2-MT and 2-MG in Iowa City in gas and particle phases. Even though the data were collected in November, approximately 37% of 2-MT and 15% of 2-MG were found in the gas phase. In addition, Xie et al.[177] also measured gaseous and particle-phase 2-MT in summer Denver and determined that the variation in gas-particle partitioning of 2-MT was driven by the changes in OA concentrations. Furthermore, the saturation vapor pressures of 2-MT and 2-MG were estimated to be 1.49×10^{-4} Pa and 1.87×10^{-3} Pa, respectively, by Couvidat and Seigneur,[43] which also suggests that they are semivolatile under typical ambient OA loadings.

In this study, we expanded the isoprene SOA model of Pye et al.[49] to simulate the formation of 2-MT and 2-MG by treating these two species as semivolatile and including a non-aqueous formation pathway. The model predictions of the two tracers were compared with summertime field measurements from multiple sites in China to evaluate the capability of the expanded model. To the best of the authors' knowledge, this is the first regional chemical transport modeling study to evaluate the potential volatility and the non-aqueous formation of the two isoprene markers in relatively polluted atmospheres. The modeled tracer concentrations and the overall isoprene SOA also provide an evaluation of the appropriateness of using lab-derived f_{SOA} to estimate the concentrations of ambient isoprene SOA.

5.2 Methods

5.2.1 Isoprene SOA and unique tracer formation

Figure 5-1 shows the updated isoprene SOA and tracer formation scheme. The isoprene SOA scheme used in this study is mainly based on that described by Pye et al.,[49] which includes the formation of lumped semivolatile isoprene SOA and the oligomers based on the traditional Odum 2-product model, and the heterogeneous formation of 2-MG, 2-MT, and organosulfate products in the aerosol water from surface uptake of IEPOX and MAE and the subsequent aqueous-phase reactions. A summary of the scheme and the major changes made in this study are described as follows.

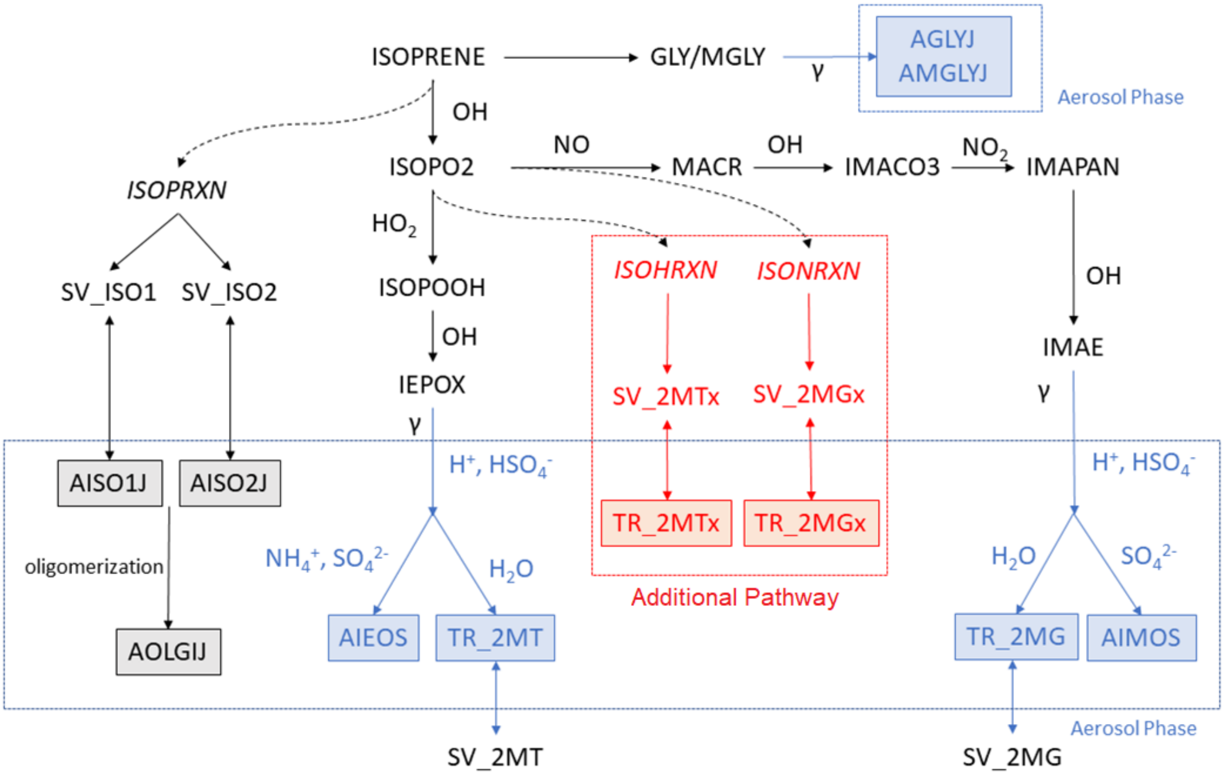


Figure 5-1 Schematic of the formation of isoprene SOA and its organic tracers in the revised CMAQ model. AISO1J and AISO2J represent the lumped isoprene SOA from the equilibrium partitioning of semivolatile products, and AOLGIJ is their oligomerization product. The blue boxes represent the formation of 2-MT (TR_2MT), 2-MG (TR_2MG), two organosulfate compounds (AIEOS and AIMOS), and glyoxal and methylglyoxal SOA (AGLYJ and AMGLYJ) from the heterogeneous chemistry. The formation of organic tracers from the additional pathway (TR_2MTx and TR_2MGx) is shown in the red boxes. 2-MT and 2-MG are assumed to be semivolatile and in equilibrium with the corresponding gas-phase species (SV_2MT, SV_2MG, SV_2MTx, and SV_2MGx), and the equilibrium gas-particle partitioning is based on their estimated saturation vapor pressure and the organic matter concentration in the aerosol phase.

The CMAQ model v5.0.1[115] with aerosol module AERO6 is used as a based model to implement the isoprene SOA scheme. The gas-phase chemical mechanism is based on SAPRC-11.[81] The isoprene oxidation chemistry in the original SAPRC-11 is replaced with the expanded isoprene mechanism described by Xie et al.[178] and Lin et al.[45] to simulate the formation of IEPOX and MAE. The yields of the two lumped semivolatile isoprene SOA products are updated to account for vapor wall-loss, according to Zhang et al.[161] Moreover, additional species are added to track isoprene-specific oligomers. Glyoxal (GLY) and methylglyoxal (MGLY) from isoprene oxidation are also determined separately, and the SOA from the two species are modeled as irreversible surface uptake.[134]

The reactive uptake of IEPOX and MAE into the aqueous phase is also treated as irreversible surface-controlled uptake processes in the heterogeneous chemistry module. The uptake coefficients (γ) depend on the pseudo-first-order reaction rates of IEPOX and MAE in the aerosol water ($k_{particle}$, s^{-1}), their Henry's Law constants (H^*, M atm^{-1}), and the effective radius of the particle (r_p, m), as shown in Eqs. (5.1-5.3),[49, 50, 179, 180]

$$\frac{1}{\gamma}=\frac{1}{\alpha}+\frac{\nu}{4H^{*}RT\sqrt{D_a k_{particle}}\,(\coth(q)-1/q)} \tag{5.1}$$

$$q=r_p\sqrt{\frac{k_{particle}}{D_a}} \tag{5.2}$$

$$k_{particle}=\sum_{i=1}^{N_{nuc}}\sum_{j=1}^{N_{acid}}k_{i,j}[nuc_i][acid_j] \tag{5.3}$$

where α is the accommodation coefficient (0.02),[181] R is the ideal gas constant (0.082 L atm mol^{-1} K^{-1}), T is the ambient temperature (K), D_a is the diffusivity in the aerosol phase (taken as 1.0×10^{-9} m^2 s^{-1} for both IEPOX and MAE),[182] N_{nuc} is the number of nucleophiles (water, or SO_4^{2-}) and N_{acid} is the number of acids (H^+, NH_4^+ or HSO_4^-). $[nuc_i]$ and $[acid_j]$ are the molar concentration of nucleophiles and acids (M), respectively, determined based on the predicted inorganic aerosol composition and the inorganic thermal dynamics module ISORROPIA II.[183] $k_{i,j}$ is the third-order rate constants for the particle-phase reactions (M^{-2} s^{-1}), and the values are listed in Table S5-1. The $k_{particle}$ includes the formation of the tracers and the organosulfate products in the aerosol aqueous phase. 2-MT and 2-MG are formed with water as the nucleophile. Thus, the molar fraction of 2-MT in total IEPOX SOA and 2-MG in the total MAE SOA newly formed at each time step can be determined using Eq. (5.4),

$$\beta = \frac{\sum_{i=1}^{N_{acid}} [water][acid_i]}{k_{particle}} \tag{5.4}$$

The formation of 2-MT dimers is not included in the study, as a previous study shows that dimer concentrations are much lower than the 2-MT monomer.[184]

In this study, 2-MT and 2-MG are treated as semivolatile, and their equilibrium gas-to-particle partitioning is described by their saturation mass concentrations (C^*, μg m^{-3}), calculated using Eq. (5.5),

$$C^* = \frac{1}{K_{OM}} = \frac{10^6 p_L^0 \gamma \overline{MW_{OM}}}{RT} \tag{5.5}$$

where K_{OM} is the absorptive gas/particle partitioning coefficient (m^3 μg^{-1});[146] p_L^0 is the subcooled vapor pressure (Pa); γ is the activity coefficient of the compound in the absorbing OM phase (assumed to be unity); $\overline{MW_{OM}}$ is the mean molecular weight of the OM phase assumed to be 200 g mol^{-1}, according to Williams et al.[149] p_L^0 values for 2-MT and 2-MG are estimated from the E-AIM aerosol thermodynamics model, which includes several methods to estimate vapor pressure and boiling points[150, 151, 155]. The p_L^0 for 2-MG (3.76×10^{-4} Pa) and 2-MT (1.81×10^{-4} Pa) used in this study are based on the lowest of the three estimations provided by E-AIM. The enthalpy of vaporization (ΔH_{vap}) of the two species are 130.0 (2-MG) and 117.7 kJ mol^{-1} (2-MT), also from E-AIM estimations.

In addition to the heterogeneous formation of 2-MG and 2-MT, an additional pathway to form these two organic tracers is included. In this pathway, the formation of 2-MG and 2-MT is assumed to occur in the gas phase and is represented by an empirical mass yield. Since there is no chamber study revealing the gas-phase mass yield of 2-MG or 2-MT from isoprene oxidation reactions, the values are constrained by the field measurement data from the Pearl River Delta (PRD) in southern China, as described in Section 5.2.2. The estimated mass yields are shown in Table 5-1. 2-MG and 2-MT from this additional pathway are tracked separately using a different set of gas and particle phase species. The uncertainties caused by the yield will be analyzed in the discussion section.

Table 5-1 Mass yield (α), saturation mass concentration (C^*), and enthalpy of vaporization for 2-MG and 2-MT in the additional formation pathway.

Species	Produced From	$\alpha^\$$	C^* (μg m^{-3})	ΔH_{vap} (kJ mol^{-1})
2-MG	High-NO_x pathway	0.0026	30.3	130.0
2-MT	Low-NO_x pathway	0.0840	14.6	117.7

$^\$$ The yields are selected so that they lead to optimal predictions of total 2-MT and 2-MG in the base case simulation in the PRD region.

5.2.2 *Observation data and model simulation*

Observations of 24-hour average concentrations of $PM_{2.5}$ mass, major chemical components, and organic marker compounds, including 2-MT and 2-MG, are available from four monitors in the PRD region in 2012 every six days. The four monitors are Guangzhou (23.13°N, 113.30°E), Nansha (22.75°N, 113.60°E), Dongguan (22.97°N, 113.74°E), and Nanhai (23.06°N, 113.15°E). Guangzhou and Dongguan sites are located in urban areas; Nansha is a suburban site, and Nanhai is an industrial site. In addition, daily 2-MT and 2-MG data collected at 14 sites (approximately two days per month) across China[136] are used to further assess the model's capability in predicting isoprene SOA tracers.

The modified CMAQ model is applied to simulate isoprene SOA and the two isoprene SOA tracers in China in summer 2012 (June – August). The CMAQ simulation has the same domain as previous simulations in Section 2-4. Initial and boundary conditions are generated using the CMAQ default vertical profiles. The first five days of the simulation are excluded in the final analysis as spin-up. In addition to the base case, three sensitivity simulations are performed, as listed in Table S5-2. Details of the sensitivity simulations are discussed in Section 5.4.

The meteorological inputs are generated using WRF v4.2 with initial and boundary conditions from the ERA5 Reanalysis 0.25 Degree Latitude-Longitude Grid (available at https://rda.ucar.edu/datasets/ds633.0/). The land use/land cover and topographical data are based on the 30 s resolution default WRF input dataset. Reanalysis nudging is enabled to improve the agreement between the predicted and observed meteorological parameters.[119] Other major physics options for the WRF simulations have been described by Zhang et al.[120] The model performance of major meteorological parameters in the PRD region is shown in Table S5-3. The model performance statistics are comparable with the same resolution previous modeling studies in China.[185]

Anthropogenic emissions are based on the REAS3[121] using an in-house emission processor. The selected detailed speciation profiles from the SPECIATE database developed by the US EPA are used to estimate emissions of CMAQ-ready VOCs. Windblown dust emissions in the entire domain are generated by the inline module [86]. Biogenic emissions are produced by the MEGAN v2.10 [123].

5.3 Results

5.3.1 General model performance evaluation

As shown in Figure S5-1, the $PM_{2.5}$ mass concentrations are slightly underpredicted, with a mean fractional bias (MFB) of approximately -0.3 and a mean fractional error (MFE) of 0.4, but the daily variations are well captured. The underprediction of $PM_{2.5}$ is mainly caused by the underestimation of $PM_{2.5}$ nitrate (MFB = -1.4) and organic carbon (OC; MFB = -0.6). Ammonium sulfate is the dominant $PM_{2.5}$ secondary inorganic aerosol component and is reasonably captured by the model (MFB = -0.36 ~ -0.43, MFE = 0.5 ~ 0.64). When $PM_{2.5}$ nitrate is underpredicted, measured concentrations at the four sites show significant differences, suggesting that the model does not capture some local-scale processes. Elemental carbon concentrations are also well captured (MFB = -0.35. MFE = 0.58).

The primary and secondary organic carbon (POC and SOC) observations are based on the PMF analysis of the observed aerosol composition with marker compounds for primary and secondary organic aerosols [34]. The SOC concentrations are slightly underestimated with an MFB of -0.4 and MFE of 0.6. On the other hand, the POC concentrations show significant under-predictions with an MFB of -1.0. The under-prediction of POC is likely due to emission under-estimation in the PRD and may have contributed to the under-prediction of semivolatile SOC. However, the generally good agreement in the secondary inorganic and organic $PM_{2.5}$ components suggests that the oxidation capacity in the PRD region is reasonably captured and thus provides confidence in the organic tracer predictions.

The gas-phase isoprene concentrations are compared with daily observations from 20 sites across the country in June – August 2012.[186] The model predicted isoprene concentrations are unbiased and are mostly within a factor of 3 of the observations with MFB = -0.05 and MFE =

0.70, as shown in Figure S5-2. The isoprene model performance is comparable to those reported in Zhang et al. using 3D-REAM/MEGAN v2.1 and is also similar to that in the eastern United States using CMAQ/MEGAN v2.1 with improved isoprene emission factors.[187]

5.3.2 2-MG and 2-MT in the PRD region

In addition to comparing raw base case tracer predictions with observations, repartitioning of the tracers in the gas and aerosol phases are performed based on the observed OA concentrations using Eq. (5.6),

$$C_{p,rep} = \frac{C_{tot}}{1 + C^{*}/C_{OA}} \tag{5.6}$$

where the $C_{p,rep}$ is the repartitioned concentration in the aerosol phase (μg m^{-3}); C_{tot} is the total tracer concentrations in the gas and aerosol phase (μg m^{-3}); C_{OA} is the observed OA loading, which is estimated from OC measurements assuming an OA/OC ratio of 1.6;[188, 189] C^{*} is the saturation tracer mass concentration (μg m^{-3}) calculated using Eq. (5.5), with an adjusted p_L^0 based on the WRF modeled daily-average temperature, as shown in Eq. (5.7),

$$p_L^0 = p_{L,ref}^0 \exp\left[\frac{\Delta H_{vap}}{R}\left(\frac{1}{T_{ref}} - \frac{1}{T}\right)\right] \tag{5.7}$$

where the $p_{L,ref}^0$ is the subcooled vapor pressure (Pa) at reference temperature (298.15 K); ΔH_{vap} is the enthalpy of vaporization (J mol^{-1}); R is the ideal gas constant (8.314 J K^{-1} mol^{-1}); T_{ref} is the reference temperature (K), and T is the modeled ambient temperature (K).

The average daily temperature in the PRD region is ~300-308 K, and most of the tracers are in the gas phase based on the estimated saturation vapor pressure and the enthalpy of vaporization (Figure 5-2b and Figure 5-3b). Since OC concentrations are under-predicted in the PRD region, repartitioning leads to more tracers in the aerosol phase. The fractions 2-MT in the aerosol phase increase from 13-25% to 25~33% after repartitioning. Likewise, the aerosol fractions of 2-MG increase from ~5% to ~15% due to repartitioning with the observed OA. The daily model predictions of the tracers include formation from heterogeneous chemistry and the additional pathway shown in Figure 5-1. Contributions from the additional pathway account for 74-87% of 2-MT and 54-78% of 2-MG. The contributions are highest at the rural site (Nansha) and lowest at the industrial site (Nanhai).

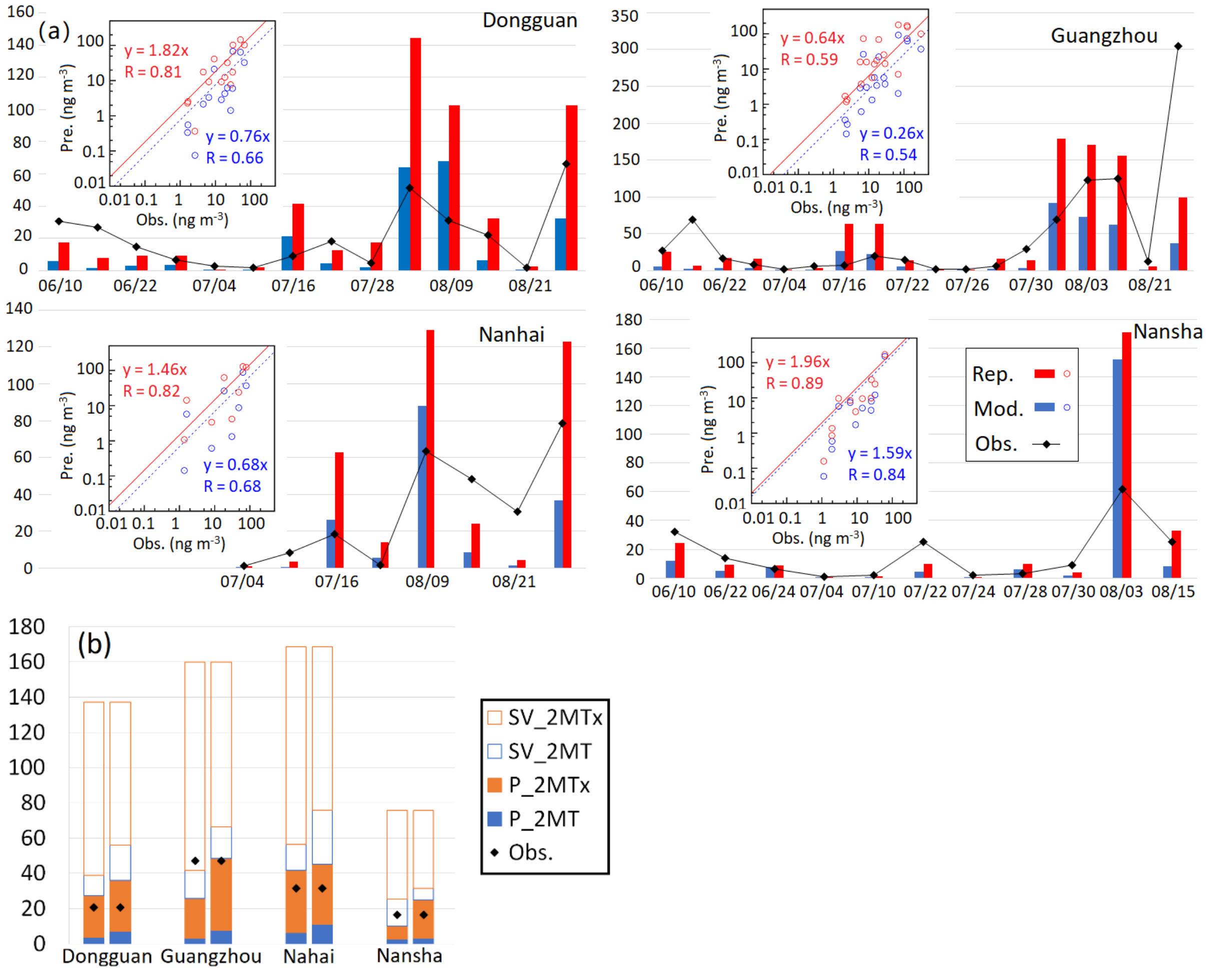


Figure 5-2 (a) Predicted and observed daily 2-MT in the PRD region. The blue bars show the base case model predictions. The red bars show the adjusted predictions by repartitioning based on the measured OA. The black lines represent field measurements. (b) Components of site average 2-MT concentrations with observations. The left bars represent the raw model predictions, and the right bars represent the repartitioned concentrations. SV_2MTx and SV_2MT represent the gas-phase 2-MT from the additional pathway and the evaporation of the tracer formed in the heterogeneous pathway; P_2MTx and P_2MT represent the aerosol phase 2-MT in equilibrium with the corresponding SV species. Units are ng m^{-3}.

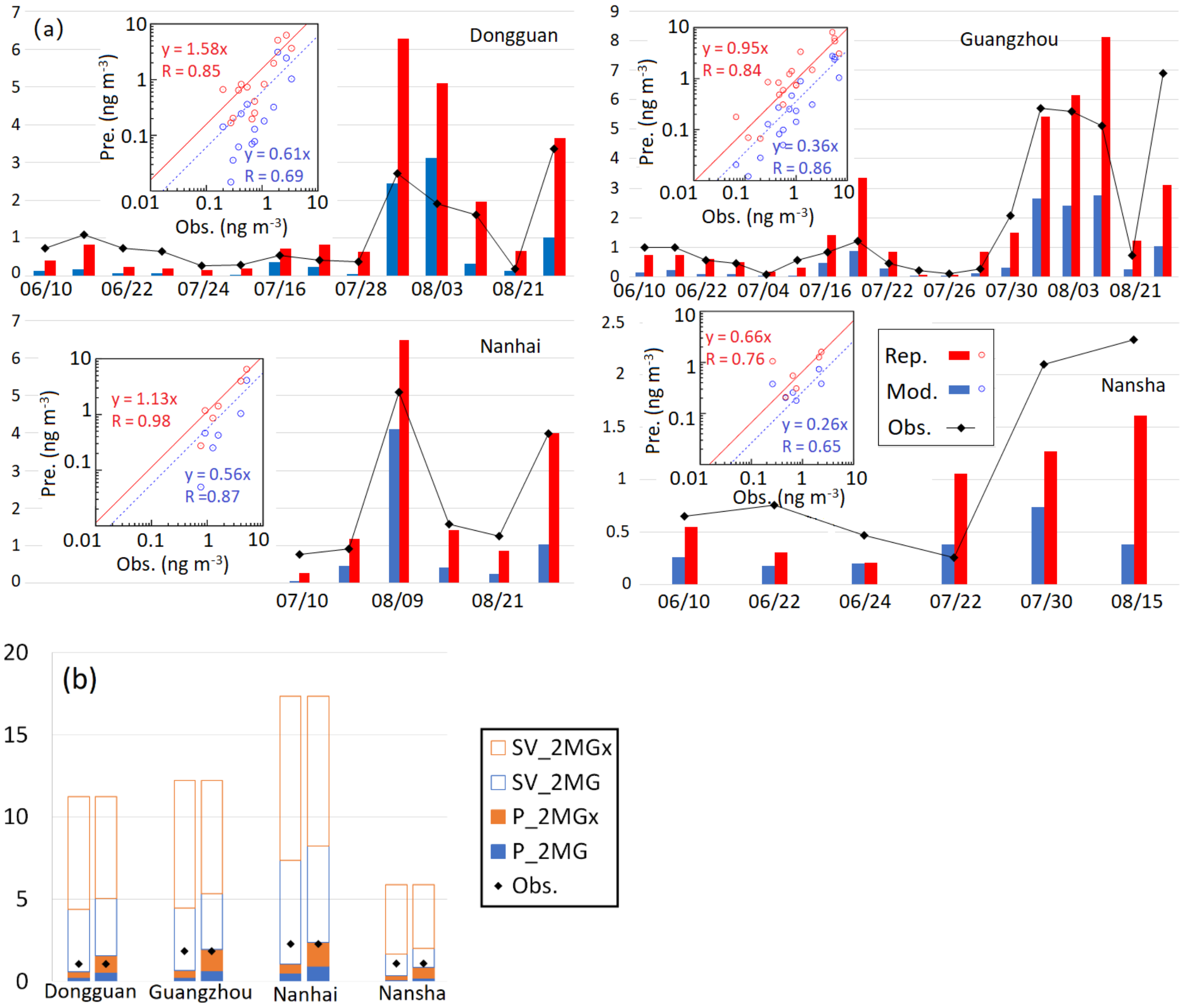


Figure 5-3 Same as Figure 5-2 but for 2-MG.

The repartitioned 2-MT and 2-MG aerosol-phase concentrations agree better with the observations than the original base case predictions. Figure 5-2a shows that the raw model predicted 2-MT daily concentrations are generally lower than the observations except for the Nansha site, where the 2-MT concentrations are overpredicted by approximately 60%. The Pearson correlation coefficients between observations and repartitioned predictions are approximately 0.6 ~ 0.9, improved by 0.05 ~ 0.15 from the raw predictions. Figure 5-3a shows that the 2-MG concentrations are lower than those of 2-MT by a factor of 10, which is consistent with most field measurements.[38, 48, 136, 141] The repartitioned model predictions of 2-MT at individual sites are highly correlated with the daily measurements (R = 0.75-0.98), and the correlations are generally stronger than those based on raw predictions (R = 0.65-0.87).

Table 5-2 Summary of 2-MG and 2-MT model performance in the PRD region.

Average observation (ng m^{-3})	Case	Average prediction[\$] (ng m^{-3})	MFB[^]	MFE	R[%]
2-MG: 1.6	Base case	0.7	-1.022	1.061	0.763
	Repartitioned base case	1.7	-0.007	0.522	0.831
	NV/Het. Chem.	6.8	0.678	1.003	0.467
	Sens. 1[#] (lower ΔH_{vap})	4.0	0.587	0.735	0.841
	Sens. 2[#] (higher yields)	4.5	0.847	0.862	0.877
2-MT: 30.7	Base case	18.0	-0.843	1.107	0.470
	Repartitioned base case	39.3	0.002	0.723	0.605
	NV/Het. Chem.	41.3	0.057	0.782	0.355
	Sens. 1	80.0	0.501	0.809	0.624
	Sens. 2	90.2	0.666	0.868	0.613

[\$] The average predictions were calculated with the predicted concentrations with corresponding observations available.

[^] Mean fractional bias, $\mathrm{MFB} = 2\frac{C_{pre}-C_{obs}}{C_{pre}+C_{obs}}$; Mean fractional error, $\mathrm{MFE} = 2\frac{|C_{pre}-C_{obs}|}{C_{pre}+C_{obs}}$. C_{pre} and C_{obs} are daily predictions and observations, respectively.

[%] Pearson correlation coefficient.

[#] Case Sens. 1 and Sens. 2 results shown in this table were adjusted based on the observed OA loadings in the PRD region.

The model performance statistics of 2-MT and 2-MG for the raw base case and repartitioning-adjusted results for each site in the PRD region are shown in Table S5-4, and the overall model performance statistics are summarized in Table 5-2. The base case model underpredicted 2-MG and 2-MT with the MFB approximately -1.0 and -0.8, respectively. The error in the predictions is also quite significant, with an MFE of ~1.1 for both tracers. The repartitioning process generally removes the under-predictions with MFB close to zero and significantly reduces the MFE (0.522 for 2-MG and 0.732 for 2-MT). The adjusted predictions also reveal higher correlations with observations (R = 0.83 for 2-MG and R = 0.60 for 2-MT) compared to the raw model predictions (R = 0.76 for 2-MG and R = 0.47 for 2-MT). In summary, the model performance statistics show that the base case model under-predicts particle phase 2-MG and 2-MT, but the agreement with the observations improves significantly after repartitioning using observed OA.

5.3.3 2-MG and 2-MT across the country

Table 5-3 Summary of 2-MG and 2-MT model performance in the 14 sites across China.

Average observation (ng m^{-3})	Case	Average prediction (ng m^{-3})	MFB	MFE	R
2-MG: 12.4	Base case	1.2	-1.622	1.663	0.082
	Repartitioned base case	5.5	-0.989	1.179	0.196
	NV/Het. Chem.	8.1	-0.875	1.239	0.099
	Sens. 1[#] (lower ΔH_{vap})	6.1	-0.970	1.176	0.316
	Sens. 2[#] (higher yields)	11.5	-0.485	1.028	0.281
2-MT: 105.6	Base case	20.7	-1.253	1.316	0.393
	Repartitioned base case	86.2	-0.286	0.824	0.543
	NV/Het. Chem.	45.1	-0.842	1.104	0.122
	Sens. 1	91.2	-0.355	0.878	0.644
	Sens. 2	152.0	0.168	1.194	0.596

[$] The average predictions were calculated with the predicted concentrations with corresponding observations available.

[#] Case Sens. 1 and Sens. 2 results shown in this table were adjusted based on the OA loadings increased by a factor of 3.

To further evaluate the updated isoprene SOA scheme, the 2-MG and 2-MT model performance at the 14 monitoring sites is determined and summarized in Table 5-3. The base case model results show significant underprediction of the two tracers (MFB = -1.62, MFE = 1.66 for 2-MG and MFB = -1.25, MFE = 1.32 for 2-MT) and weak correlations between the predictions and observations (R = 0.08 for 2-MG and R = 0.39 for 2-MT).

Measurements of OC are not available at these locations, but it is expected that OC might be under-predicted, similar to those in the PRD region, by a factor of 2-3, which could cause under-predictions of the tracers. For example, the predicted summertime average OC concentrations at Xianghe, a suburban site in northern China, is 7.5 μg m^{-3}, which is approximately a factor of 3 lower than the observed $PM_{2.1}$ OC concentration of 25.2 μg m^{-3} in summer 2012.[190] To assess this potential under-prediction of OC on the modeled tracer concentrations, the total tracer concentrations at these locations are repartitioned by increasing the predicted OC concentrations by a factor of 3, using the same technique described in the previous section.

As shown in Figure S5-3, the repartitioned tracer concentrations show better agreement with the observations in both MFB (-0.99 for 2-MG and -0.29 for 2-MT), MFE (1.18 for 2-MG and 0.82 for 2-MT), and correlations (R = 0.20 for 2-MG and R = 0.54 for 2-MT). The model performance for individual sites is shown in Table S5-5. The adjusted concentrations are still lower than observations, especially for 2-MG formed under high-NO_x conditions. Since the anthropogenic emissions could be significantly underestimated, especially in less economically developed western and southwestern regions [125], NO_x emissions might have been underestimated, causing under-predictions of 2-MG.

Figure S5-4 shows the spatial distribution of the average model predicted 2-MG and 2-MT concentrations from June to August. The highest concentrations of 2-MG (~30 ng m^{-3}) and 2-MT (~400 ng m^{-3}) are in southwest China near the foothill of the west rim of the Sichuan Basin, where the heterogeneous pathway has significant contributions to both 2-MG (~20 ng m^{-3}) and 2-MT (~200 ng m^{-3}) due to higher acidity (~1, see Figure S5-4c), which leads to higher surface uptake coefficients (Figure S5-5a and b). In the polluted regional in central and eastern China, predicted pH is in the range of 3-4, which agree with those reported by Zhang et al.[191] The formation of the two tracers from the additional pathway has wider spatial distributions (Figure S5-4e and f) than the heterogeneous pathway and is more closely related to gas-phase IEPOX and MAE.

The spatial distributions of observed 2-MT and 2-MG are similar to the precursor gas of IEPOX and MAE, respectively (Figure S5-6c and d). As shown in Figure S5-7, the correlations between observed aerosol 2-MG and 2-MT with the modeled heterogeneous reaction products are low (R = -0.024 for 2-MG and 0.165 for 2-MT). In contrast, stronger correlations are obtained between the observed aerosol tracers and the modeled gas phase precursors (R = 0.443 for 2-MG with IMAE, and 0.745 for 2-MT with IEPOX) and between the tracers and the modeled semivolatile isoprene SOA products (R = 0.321 for 2-MG and 0.732 for 2-MT). The strong correlations of the observed tracers with modeled gas-phase precursors and the semivolatile SOA products further support the potential contributions of an additional gas-phase formation pathway for these tracers, particularly for 2-MT.

5.3.4 Mass fraction of isoprene tracers in the SOA

The mass fractions (f_{SOA}) of organic tracers (2-MG + 2-MT) in the isoprene SOA are determined based on the modeled daily tracer concentrations and the isoprene SOA concentrations. and the total OA concentrations at the PRD sites and 8 other sites across China (sites with summertime average OA < 5 μg m^{-3} are excluded). Significant scattering is observed when fsoa is plotted against OA, especially at low OA concentrations, as shown in Figure 5-4(a), but at higher OA concentrations (OA > ~10 μg m^{-3}), fsoa approaches a constant. To provide an estimation of the fsoa, a Michaelis-Menten type equation is used to fit fsoa against OA. For example, based on the fitted equation, fsoa approaches 0.013 when all isoprene SOA components (semivolatile products, oligomers, 2-MT, 2-MG, organosulfates, and SOA related to the uptake of GLY and MGLY from isoprene oxidation) are included. If SOA from the surface uptake of dicarbonyls are excluded, f_{SOA} approaches 0.040, which is ~30% lower than the average f_{SOA} of 0.063 determined by Kleindienst et al.[38]. At OA ~ 4 μg m^{-3}, f_{SOA} is reduced to half of the asymptotic value for both estimations.

An alternative direct approach to estimate isoprene SOA based on measured 2-MT + 2-MT concentrations is shown in Figure 5-4(b). The modeled isoprene SOA concentrations and tracer concentrations are fitted empirically with power-law equations. For example, the total isoprene SOA can be calculated using $C_{SOA} = 32.5C_{tracer}^{0.68}$, where the concentrations are in μg m^{-3}. For the model data, these equations lead to lower overall errors and biases. They are also more convenient to use as no OC or OA data are needed.

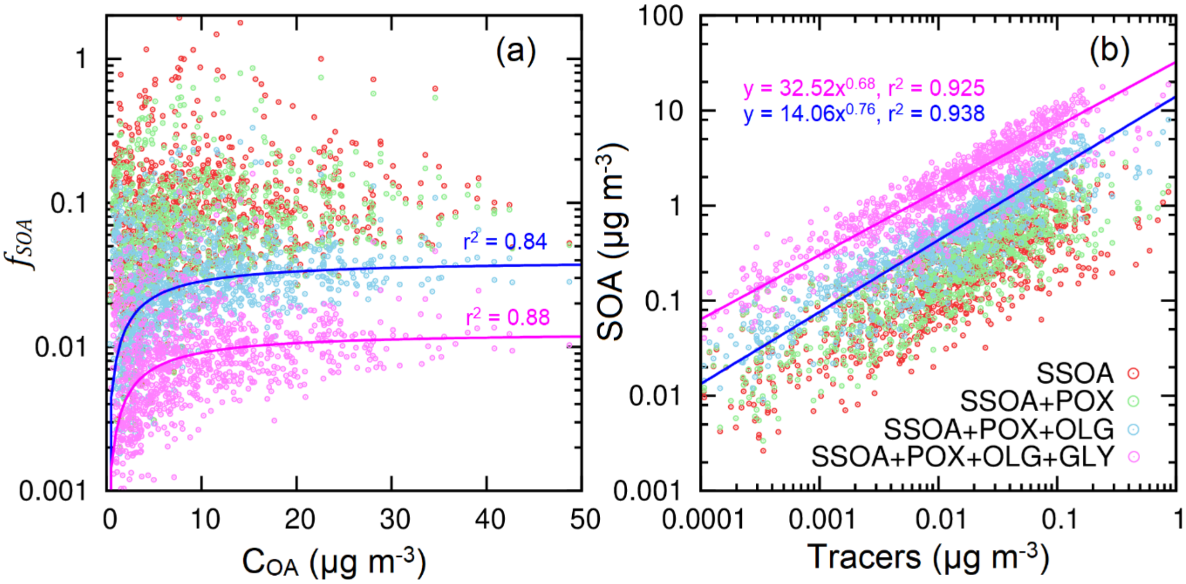


Figure 5-4 (a) The daily tracer mass fraction (f_{SOA}) as a function of predicted OA concentrations (C$_{OA}$) in 11 sites across the country, based on model predictions. (b) The linear relationships between the tracer predictions and isoprene SOA. SSOA: semivolatile isoprene SOA (which includes 2-MG and 2-MT from the additional pathway); POX: SOA from surface uptake of IEPOX and MAE; OLG: oligomers from SSOA; GLY: dicarbonyl SOA from surface uptake of glyoxal and methylglyoxal from isoprene oxidation. R: Pearson correlation coefficient.

Figure 5-5 shows that although emissions of isoprene peaks in southern and southeastern China (Figure S5-6a), long-term average f_{SOA} values vary slightly in the more polluted central, eastern, southern, and southeastern China due to regional transport of isoprene oxidation products. However, the f_{SOA} is increased dramatically in southwest China, where isoprene emissions are very low. The f_{SOA} values used to estimate total isoprene SOA are within 0.01 ~ 0.02 in the polluted area. When SOA from dicarbonyls is excluded, f_{SOA} in the polluted region is approximately 0.03-0.04, consistent with the values derived in Figure 5-4(a) ~ 0.15. The range of f_{SOA} becomes narrower when more isoprene SOA components are considered. Table S5-7 lists the f_{SOA} values at the 14 monitor locations.

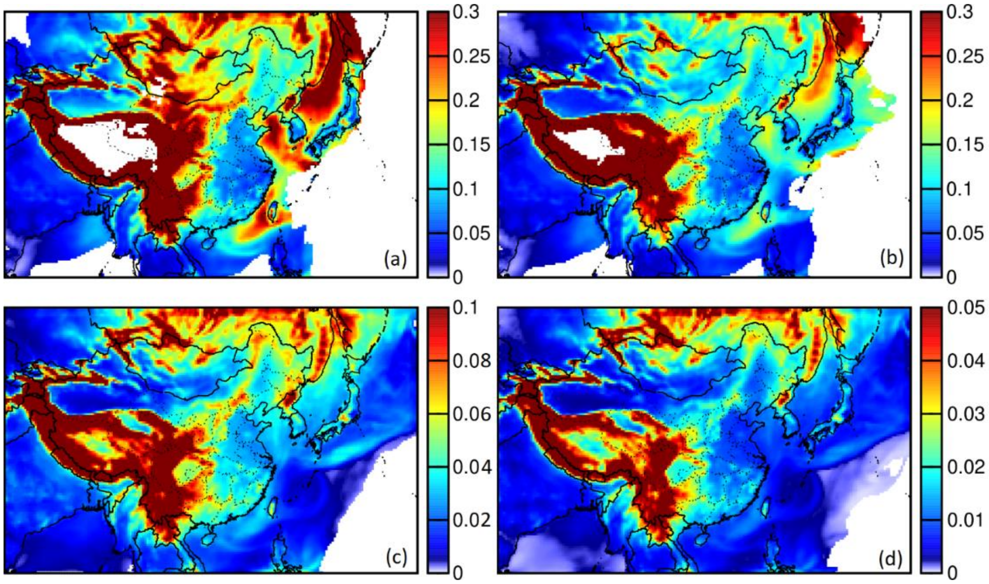


Figure 5-5 Spatial distribution of f_{SOA} calculated based on average base case model predictions from June to August 2012 when (a) only SSOA; (b) SSOA and POX; (c) SSOA, POX, and OLG; (d) SSOA, POX, OLG, and GLY are considered with a threshold of 0.1 μg m^{-3} isoprene SOA.

5.4 Discussion

5.4.1 Non-volatile 2-MG and 2-MT without the additional pathway

An additional simulation (NV/Het. Chem.) is performed by turning off the additional pathway and setting the 2-MG and 2-MT to be non-volatile to assess if treating 2-MG and 2-MT as non-volatile without the additional pathway could also lead to reasonable estimations of the two tracers. In this simulation, since the tracers are non-volatile, no repartitioning is necessary when comparing with observations. The 2-MG is over-estimated (MFB = 0.57) with moderate correlations with observations (R = 0.467) in the PRD region (Table 5-2) but underestimated significantly at other locations (MFB = -0.96) without apparent correlation with observations (R = 0.099; see Table 5-3). For 2-MT, the predicted average concentration in the PRD region is similar to that of the repartitioned base case results but with a much weaker correlation (R = 0.355). In the other locations, 2-MT from this simulation is much lower than observations (MFB = -0.96), and the correlation (R = 0.122) is weaker than that of the repartitioned base case (R = 0.520). These results suggest that it might not be accurate to treat the 2-MT and 2-MG as non-volatile and assume that they are only generated from the heterogeneous process that involves surface uptake and aqueous-phase reactions.

5.4.2 Uncertainties in 2-MG and 2-MT partitioning due to the enthalpy of vaporization

The ΔH_{vap} values for 2-MG and 2-MT in the base case simulation (~120-130 kJ mol^{-1}) are estimated based on the E-AIM model. Significantly lower estimations (~40 kJ mol^{-1}) were used by Couvidat and Seigneur[43] based on ΔH_{vap} for semivolatile isoprene SOA derived from chamber experiments. With high ambient temperatures in the summer, higher ΔH_{vap} values lead to higher saturation vapor pressure and lower fractions of the tracers in the aerosol phase. For example, as shown in Figure S5-8, for typical OA loading between 10 and 50 μg m^{-3}, the fraction of 2-MT in the particle phase ($F_{p,t}$) is sensitive to temperature with a high ΔH_{vap} of 117.7 kJ mol^{-1}. When the temperatures are above 300 K, 2-MT is mostly in the gas phase with high ΔH_{vap}. However, the $F_{p,t}$ of 2-MT becomes higher with a lower ΔH_{vap} estimation of 38.4 kJ mol^{-1}.

To assess the influence of lower ΔH_{vap} values on the model predictions for aerosol phase 2-MG and 2-MT, a sensitivity simulation (Sens. 1) has been conducted with ΔH_{vap} of 43.2 kJ mol^{-1} and 38.4 kJ mol^{-1} for 2-MG and 2-MT, respectively, while keeping the other parameters the same as the base model. In this case, 2-MG and 2-MT aerosol fractions are significantly higher than those in the base case (Figure S5-9). Since the average daily temperature in the PRD region is generally above 300 K, the average 2-MG and 2-MT concentrations (repartitioned with observed OC) in the region are increased by approximately a factor of 2 compared to the base case (Table 5-2), leading to over predictions (MFB = 0.50 for 2-MG and 0.58 for 2-MT). However, there are no significant differences in the correlation coefficients with observations.

The lower ΔH_{vap} has less influence on 2-MG and 2-MT predictions in other places (Table 5-3), where the temperature is relatively lower than the PRD region. The overall repartitioned 2-MG and 2-MT predictions are increased by approximately 11% and 6% from the repartitioned base case results, and the correlations between observations and model predictions become stronger, with an increase of approximately 0.1 for the overall correlation coefficients. While this sensitivity simulation shows that the model predictions can be sensitive to the estimation of ΔH_{vap}, more modeling and field studies are needed to determine the appropriate ΔH_{vap} for use in the models.

5.4.3 Uncertainties in 2-MG and 2-MT formation due to the yield in the additional pathway

The 2-MG and 2-MT yields from the additional pathway in the base case simulation are selected so that predicted concentrations agree with field measurements in the PRD region. Next, the uncertainties of the model predictions to this important parameter are tested in an additional sensitivity simulation case (Sens. 2). Since the model under-predicts the observations in other places across China, the 2-MG and 2-MT yields increase by a factor of 3 and 2, respectively, in this simulation. In the PRD region, the repartitioned 2-MG and 2-MT predictions are increased by a factor of 2.6 and 2.3, respectively, as shown in Table 5-2 (individual site performance is shown in Table S5-4), but the correlation with the observations increased slightly. The overall average predictions of 2-MG and 2-MT in other places across China, as shown in Table 5-3, are increased by 145% and 76%, respectively.

6 MODELING SECONDARY ORGANIC AEROSOL TRACERS AND TRACER-TO-SOA RATIOS FOR MONOTERPENES AND SESQUITERPENES USING A CHEMICAL TRANSPORT MODEL

The community multiscale air quality (CMAQ) model was modified to simulate secondary organic aerosol (SOA) formation from five explicit (α-pinene, β-pinene, d-limonene, Δ^3-carene, and sabinene) and one lumped monoterpene (MT) species and sesquiterpenes (SQTs). The contribution of each oxidation pathway (including OH, O_3, NO_3, and $O(^3P)$) was explicitly tracked in the SOA module. Three MT SOA tracers (pinic acid, PA; pinonic acid, PNA; and 3-methyl-1,2,3-butanetricarboxylic acid, MBTCA) and one SQT SOA tracer (β-caryophyllinic acid, BCARYA) were modeled to assess the tracer-to-SOA ratios (f_{SOA}) for ambient SOA estimation. Good model performance for BCARYA and MBTCA and reasonable agreement between model predictions and observations of PA and PNA were achieved. The modeled daily f_{SOA} showed significant variations, suggesting that using chamber-derived constant f_{SOA} could lead to large errors in estimating terpene SOA. Among the four tracers, MBTCA and BCARYA were more appropriate for tracking MT and SQT SOA due to their nonvolatility. Their f_{SOA} values mainly depend on the organic aerosol loading and could be approximated using simple power-law equations. In addition, equations directly linking the tracer concentrations to the corresponding SOA concentrations were proposed and could lead to good SOA estimations. This work provides new insights into the formation of the key MT and SQT SOA tracers and would allow better assessments of the biogenic emissions to regional and global aerosol burden.

6.1 Introduction

Monoterpenes (MTs, $C_{10}H_{16}$) and sesquiterpenes (SQTs, $C_{15}H_{24}$) emitted from natural sources are diverse mixtures of highly reactive compounds and important sources of secondary organic aerosol (SOA). The major species of the naturally emitted MTs are α-pinene, β-pinene, d-limonene, Δ^3-carene, and sabinene. Model estimations of global average SOA production from MTs and SQTs are both on the order of 10-30 Tg a^{-1},[192-194] which is significant considering the estimated range of global isoprene SOA production (10-100 Tg a^{-1}).[193]

SOA from different precursors in ambient organic aerosol samples are commonly determined using ambient concentrations of precursor-specific molecular tracer compounds and the chamber-determined mass fraction (f_{SOA}) of the tracers in the SOA generated from the precursor.[11, 38-42] Among many products identified from the MT SOA,[195] pinic acid (PA) and pinonic acid (PNA) are often considered as molecular tracers.[16, 38, 55, 195] 3-methyl-1,2,3-butanetricarboxylic acid (MBTCA) is also considered a unique tracer for α-pinene and β-pinene.[56] MBTCA is mainly generated from gas-phase oxidation of cis-PNA,[57] and aqueous oxidation of PA may also contribute to it.[58] For SQTs, the most commonly used molecular tracer is β-caryophyllinic acid (BCARYA),[59] generated from the photochemical oxidation of β-caryophyllene, one of the major biogenic SQTs.[60] Although this source apportionment method is easy to apply and has the distinct advantage of specific linkage to individual VOC precursors, the validity of the assumption that the chamber-derived f_{SOA} could be applied under ambient conditions has been questioned.[11, 61]

CTMs are widely used to quantify the terpene SOA in the ambient air. For example, Tsigaridis et al.[196] modeled SOA from α-pinene and β-pinene separately in a global CTM using the classical Odum 2-product (2-p) representation[79] but only considered their SOA formation from O_3 oxidation. The Community Earth System Model version 2 (CESM2)[197] was developed to model SOA from isoprene, MTs, and SQTs from OH, O_3, and NO_3 oxidation pathways using a 5-bin Volatility Basis Set. However, the MTs and SQTs were modeled as a lumped species, and SOA yields for MTs with O_3 and OH were not differentiated. Qin et al.[198] developed a detailed treatment of MT chemistry in the community multiscale air quality model (CMAQ) to explicitly model SOA from α-pinene, β-pinene, d-limonene, Δ^3-carene, and sabinene from OH, O_3, NO_3, and $O(^3P)$ with 2-p representation. The detailed representation led to higher MT SOA estimations than those using a lumped MT, and NO_3 oxidation was more important than other pathways.

0-D box model simulations of PA and PNA and their gas-particle partitioning from different MTs have been performed using detailed gas-phase chemical mechanisms,[199, 200] along with equilibrium partitioning into an absorbing organic phase.[201-203] However, while the chamber formation of the tracers can be reasonably reproduced after optimizing model parameters such as the gas-organic matter partitioning coefficient (K_{om})[201] or including additional formation pathways,[203] these mechanisms are often too computationally intensive to be implemented in 3-D CTMs. In addition, there is no reported modeling of MBTCA even in 0-D box models. As a result, no 3-D CTMs currently consider SOA from SQTs and explicit MTs, and their unique tracers simultaneously.

In this study, the CMAQ model is modified to include a detailed representation of the gas phase chemistry of 5 individual MTs, one lumped MT, and one lumped SQT species. The formation of SOA from these explicit and lumped species via different oxidation pathways is tracked separately. In addition, the formation of PA, PNA, MBTCA, and BCARYA in the gas phase and their gas-particle partitioning are also considered in the model. The model is applied to simulate the concentrations of the tracers along with MT and SQT SOA in China for a three-month summer period. To the best of the authors' knowledge, this is the first time that the capability of a 3-D CTM in reproducing the observed organic tracer concentrations from MTs and SQTs, and the spatial and temporal variability of the mass fraction f_{SOA} are evaluated against ambient measurement data over a large geographical area impacted by different levels of anthropogenic emissions.

6.2 Methods

6.2.1 The detailed SAPRC-11 mechanism and SOA modeling

The community multiscale air quality (CMAQ) model v5.0.1[115] is applied as a base model to implement a detailed chemical mechanism and the associated SOA and SOA tracer calculations. The gas-phase photochemistry is based on a detailed SAPRC-11 mechanism (S11D).[81] The S11D, with separate reactions for more than 700 different explicit and lumped VOCs, is the basis of more condensed versions of S11 used in several previous modeling studies.[84, 204] The S11D implemented in this study has 1261 reactions and 513 species. The number of reactions and species is significantly lower than some other detailed mechanisms such as the Master Chemical Mechanism (MCM) because a small number of species is used to represent the common organic products and the intermediate radicals leading to these products.

The explicit representation of the emitted VOCs allows separate tracking of SOA and SOA tracers from α-pinene, β-pinene, d-limonene, Δ^3-carene, and sabinene. The rest of the MTs are included as a lumped model species. In addition, the SQTs are also modeled as a lumped species. The semivolatile products in the gas phase and the corresponding SOA products from OH, NO_3, $O(^3P)$, and O_3 oxidation are tracked in the model with different species to allow quantitative analyses of the contributions of different oxidants and precursors to MT and SQT SOA.

The SOA module in the aerosol module version 6 (AERO6) is modified to link with the detailed S11D to predict SOA from different precursors. The SOA module includes SOA formation from equilibrium partitioning of semivolatile products based on the classical Odum 2-p representations.[79] Oligomerization of condensed semivolatile SOA products from equilibrium partitioning is modeled as a first-order decay process with a half-life of 20 hours.[117] Oligomers are assumed to be nonvolatile. The precursor and oxidant-specific 2-p model parameters are compiled from the literature[205,206, 138, 207] and summarized in Section S6.1. The SOA module also includes irreversible surface uptake of glyoxal and methylglyoxal on wet aerosols or cloud droplets.[45] The multiphase reactions of isoprene epoxydiol and methacrylic acid epoxide and the formation of 2-methyltetrols, 2-methylglyceric acid, and organosulfate are based on Pye et al.[49] In addition, SOA formation from long-chain alkanes and major aromatic compounds (toluene, xylenes, ethylbenzene, naphthalene, and methylnaphthalene) is also included in the model using the 2-p representation.

6.2.2 *Monoterpene and sesquiterpene SOA tracer formation*

The updated CMAQ model includes three major MT tracers (PA, PNA, and MBTCA) and one SQT tracer (BCARYA). The tracers are formed in the gas phase oxidation of the precursors by OH and O_3 using mass yields (α) calculated from published chamber data. The formation of these marker compounds with NO_3 is generally smaller due to the enhanced formation of organonitrate compounds,[206] and thus is not considered. A total of 13 unique model species are added to the CMAQ model to separately track the tracers from different precursors and the oxidation pathways (Table 6-1). Equilibrium partitioning is applied to these semivolatile tracer species to determine their concentrations in the gas and absorbing organic matter (OM) phases.

The subcooled saturation vapor pressure (p_L^0) and enthalpy of vaporization (ΔH_{vap}) data are used to calculate the saturation concentration (C^*) of the tracers, which is needed for the tracer yield and partitioning calculations. p_L^0 for PA is estimated using Method 1 of the vapor pressure estimation module from the Extended AIM (E-AIM) aerosol thermodynamics model (http://www.aim.env.uea.ac.uk/aim/aim.php). A previous study evaluated 12 different approaches of vapor pressure estimation and found that the E-AIM Method 1 yielded the best estimations, particularly for low volatility multifunctional groups.[153] The E-AIM estimated PA p_L^0 is 1.98×10^{-4} Pa at 298K (saturation concentration C^* = 16 μg m^{-3}), which is consistent with the estimations from the literature (Table S6-2). However, the p_L^0 of PNA estimated by E-AIM is approximately 100 times higher than the Pankow et al.[55] estimation of 7.19×10^{-5} Pa (C^* = 5.8 μg m^{-3}), which is close to the estimation reported by Bilde and Pandis.[208] Since the large E-AIM estimation of p_L^0 would unrealistically leave all pinonic acid in the gas phase, Pankow et al.'s estimation is adopted in this study. The MBTCA (C^* = 0.002 μg m^{-3})[209] and BCARYA (C^* = 0.02 μg m^{-3}, based on E-AIM) are essentially non-volatile under ambient conditions. The low C^* of MBTCA is also demonstrated by Zhang et al. [210] that the temperature dependence of MBTCA in the particle phase was due to changes in OH and unrelated to gas-particle partitioning. ΔH_{vap} values of the tracers are either from the literature or based on E-AIM, as listed in Table 6-1.

The α values of MT SOA tracers based on reported values or determined from chamber data from the literature (Table 6-1). The PA and PNA yields from ozonolysis reactions are based on those reported by Yu et al.[54] The α values of PA, PNA, and MBTCA from OH reactions with α-pinene and β-pinene are calculated based on the experimental data of Mutzel et al.[16] using Eq. (6.1),[61] which is based on the equilibrium absorptive partitioning of the tracer and other semivolatile products between the gas phase and OM in the particle phase [146].

$$\alpha = \frac{f_{SOA}}{F_{p,t}} Y_{SOA} \tag{6.1}$$

Table 6-1 Thermodynamic properties and mass yields (α) of monoterpene and sesquiterpene SOA tracers from different formation pathways

SOA Tracers	Model species	Formation Pathway	α^a
Pinic acid Base case C^* = 16.0 μg m^{-3}[b], $\Delta H_{vap} = 124$ kJ mol^{-1} Optimized C^* = 33.4 μg m^{-3}, $\Delta H_{vap} = 111$ kJ mol^{-1}	PA_1	α-pinene + OH	0.0062 (base case) 0.0071 (optimized [g])
	PA_2	β-pinene + OH	0.0114 (base case) 0.0125 (optimized)
	PA_3	α-pinene + O_3	0.0661
	PA_4	β-pinene + O_3	0.0431
	PA_5	Δ^3-carene + O_3	0.0164
	PA_6	sabinene + O_3	0.0191
Pinonic acid Base case C^* = 5.8 μg m^{-3}[c], $\Delta H_{vap} = 108$ kJ mol^{-1} Optimized C^* = 34.0 μg m^{-3}, $\Delta H_{vap} = 111$ kJ mol^{-1}	PNA_1	α-pinene + OH	0.0053 (base case) 0.0067 (optimized)
	PNA_2	β-pinene + OH	0.0013 (base case) 0.0016 (optimized)
	PNA_3	α-pinene + O_3	0.0568
	PNA_4	β-pinene + O_3	0.0085
MBTCA [d] C^* = 0.0018 μg m^{-3}[e], $\Delta H_{vap} = 150$ kJ mol^{-1}	MBTCA_1	α-pinene + OH	0.0129 (base case) 0.0021 (optimized)
	MBTCA_2	β-pinene + OH	0.0065 (base case) 0.0011 (optimized)
β-Caryophyllinic acid C^* = 0.017 μg m^{-3}, $\Delta H_{vap} = 168$ kJ mol^{-1}	BCARYA	sesquiterpenes + OH/NO_3/O_3	0.0146 [f]

[a] The α values for the reactions with OH were derived from Mutzel et al.[16], and values for the reactions with O_3 were obtained from Yu et al.[54]

[b] The C^* values for pinic acid and β-caryophyllinic acid were estimated based E-AIM Method 1.

[c] C^* for pinonic acid is derived based on the vapor pressure estimated by UNIFAC according to Pankow et al.[55]

[d] MBTCA: 3-methyl-1,2,3-butanetricarboxylic acid.

[e] C^* for MBTCA is obtained from Kostenidou et al.[211]

[f] The α value for β-caryophyllinic acid is derived from Kleindienst et al.[38] chamber experiments using Eqs.(6.1)-(6.4).

[g] See section 4.1 for discussion of the optimized values.

f_{SOA} is the measured particle-phase organic tracer to SOA mass ratio; Y_{SOA} is the measured mass yield of total SOA; $F_{p,t}$ is the fraction of the organic tracer in the absorbing OM phase, calculated using Eq. (6.2) with measured organic aerosol concentration (C_{OA}) in the chamber,

$$F_{p,t} = \left(1 + \frac{1}{K_{OM} C_{OA}}\right)^{-1} \tag{6.2}$$

K_{OM} is the absorptive gas/particle partitioning coefficient of organic tracer estimated using Eq (6.3)[146]:

$$K_{OM} = \frac{RT}{10^6 p_L^0 \gamma \overline{MW}_{OM}} \tag{6.3}$$

where R is the ideal gas constant (m^3 Pa K^{-1} mol^{-1}); T is the temperature (K); p_L^0 is the subcooled vapor pressure (Pa); γ is the activity coefficient of the compound in the absorbing OM phase (assumed to be unity); $\overline{MW}_{OM}$ is the mean molecular weight of the OM phase assumed to be 200 g mol^{-1}, according to William et al.[149] Assuming that the reported Y_{SOA} did not account for vapor wall loss, the solved α values were adjusted by using the correction factor of α-pinene (1.3) suggested by Zhang et al.[161]

The α value of BCARYA is calculated similarly based on data from Kleindienst et al.[38] The Y_{SOA} value is not reported so it is estimated by the Odum 2-product representations of SOA formation from β-caryophyllene, as shown in Eq. (6.4):

$$Y_{SOA} = \sum_{i=1}^{2} \beta_i \left(1 + \frac{C_i^*}{C_{OA}}\right)^{-1} \tag{6.4}$$

where β_i's are the mass yields of a lumped semivolatile product i; $C_i^*$is the saturation mass concentration of the lumped product. The β_i and C_i^* are obtained from Carlton et al.[212] and corrected for the vapor wall-loss.[161] The average α value for BCARYA is 0.0146 and is applied in OH, NO_3, and O_3 oxidation pathways.

6.2.3 CMAQ simulation

The modified CMAQ model is applied to simulate organic tracers listed in Table 6-1 and SOA formation in China in summer 2012 (Jun. – Aug.). The CMAQ model domain has 197×127 grid cells in each layer with a horizontal grid resolution of 36×36 km^2. The 18 stretching vertical layers reach a model top of approximately 20 km, with the first layer height of roughly 35 m. Initial and boundary conditions are generated using the CMAQ default vertical profiles. The first five days of the simulation are excluded in the final analysis as spin-up. In the base case simulation, the organic tracer yields and saturation mass concentrations are listed in Table 6-1. Details of the preparation of the meteorology and emission fields are described in Section S6.2.

6.3 Results

6.3.1 General model performance evaluation

The daily $PM_{2.5}$ mass concentration, major chemical components, and organic marker compounds were measured in the Pearl River Delta (PRD) region from four sites, i.e., Guangzhou (23.13°N, 113.30°E), Nansha (22.75°N, 113.60°E), Dongguan (22.97°N, 113.74°E), and Nanhai (23.06°N, 113.15°E), every six days during the summertime of 2012. The detailed description of PRD sites and observation data can be found from Wang et al.[213] and He.[214] Additional MT and SQT tracer measurements (approximately two days per month) from 14 sites across China[136] are used to further evaluate the model capability to simulate the tracers in broad areas. Locations of the monitor sites are shown in Figure S6-17.

Table S6-3 shows that total $PM_{2.5}$ is slightly underpredicted with a mean fractional bias (MFB) of -0.3 and a mean fractional error (MFE) of 0.4. However, the total organic carbon (OC) is underpredicted by a factor of 2. As shown in Figure S6-4, the underprediction of OC is mainly caused by low prediction of primary OC (POC; MFB = -1.0, MFE = 1.0), which is likely due to the underestimation of emissions in the PRD region and may lead to the small under-prediction of secondary OC (SOC; MFB = -0.34, MFE = 0.61). Nevertheless, the general agreement in the secondary inorganic and organic $PM_{2.5}$ components suggests that the oxidation capacity in the PRD region is reasonably captured and thus provides confidence in the SOA tracer predictions.

As there is no direct measurement of gas-phase MT mixing ratios during the simulation period, the seasonal average concentrations are compared with historical observation data[215-218] (Table S6-4). The predicted MT shows a broad range with lower concentrations in the urban centers and higher concentrations in the surrounding areas. However, the observed α-pinene and β-pinene concentrations are generally within the range of model predictions. Note that these two MT compounds contribute to approximately half of all the modeled MTs.

6.3.2 *Organic tracer evaluation in the PRD region*

As mentioned in Section 6.3.1, the total OA in the PRD region is under-predicted by approximately a factor of 2 (MFB = -0.6), leading to lower organic tracers partitioned into the aerosol phase. To remediate this under-prediction of OA on the assessment of the model performance of the organic tracers, repartitioning of the tracers between gas and particle phases are performed based on the OA measurements using Eq. (6.5):

$$C_{p,rep} = \frac{C_{tot}}{1 + C^*/C_{OA}} \tag{6.5}$$

where $C_{p,rep}$ is the repartitioned tracer concentrations in the aerosol phase (μg m^{-3}); C_{tot} is the total tracer concentrations predicted in both gas and particle phase (μg m^{-3}); C_{OA} is the observed OA loadings, which is estimated using SOC multiplied by OA/OC ratio of 1.6 [188, 189]; C^* is the saturation mass concentration of organic tracers (μg m^{-3}), which is adjusted based on the WRF modeled daily average temperature, as shown in Eq. (6.6):

$$C^* = C^*_{ref} \frac{T_{ref}}{T} \exp\left[\frac{\Delta H_{vap}}{R}\left(\frac{1}{T_{ref}} - \frac{1}{T}\right)\right] \tag{6.6}$$

where the C^*_{ref} is the saturation mass concentration (μg m^{-3}) at reference temperature (298 K); ΔH_{vap} is the enthalpy of vaporization (J mol^{-1}); R is the ideal gas constant (8.314 J K^{-1} mol^{-1}); T_{ref} is the reference temperature (K), and T is the modeled ambient temperature (K). The repartitioning process only affects semivolatile tracers (PA and PNA) concentrations but has little effect on the low-volatile tracers (MBTCA and BCARYA).

Figure 6-1 and Figure S6-5 show the evaluation of repartitioned organic tracer predictions against the measurements from 4 sites in the PRD region. The day-to-day variations of the tracers are well captured at each monitor site. PA is slightly overpredicted with an MFB of approximately 0.4, and it shows a moderate correlation between observations and predictions (Pearson correlation coefficient R = 0.45). The fraction of PA in the particle phase ($F_{p,t}$) is generally below 0.3 due to high temperature during the summertime in the PRD region. α-pinene ozonolysis is the most significant pathway to form PA, accounting for approximately 68% on average, followed by the oxidation of β-pinene by O_3 and OH (~10% each). Chen and Yu reported an average OA/OC ratio 2.1 in a suburban site in Hong Kong.[219] Another set of repartitioning calculations is done with this higher OA/OC ratio, leading to slightly higher particle phase PA, but it does not significantly alter the correlations between predictions and observations (Figure S6-5).

PNA is also overpredicted (MFB ~ 0.9, MFE ~ 1.16, and R = 0.27). The predicted fraction of PNA in the particle phase is generally between 0.2 and 0.6. Approximately 70% ~ 90% of PNA is formed from α-pinene ozonolysis, and 10-20% is from α-pinene oxidation by OH. As the C^* values of PA and PNA have large uncertainties, the overprediction of PA and PNA can be caused by the underestimation of C^*, which leads to more acids partitioned into the particle phase. An additional simulation that reduces the C^* values is performed and discussed in Section 1.4.1.

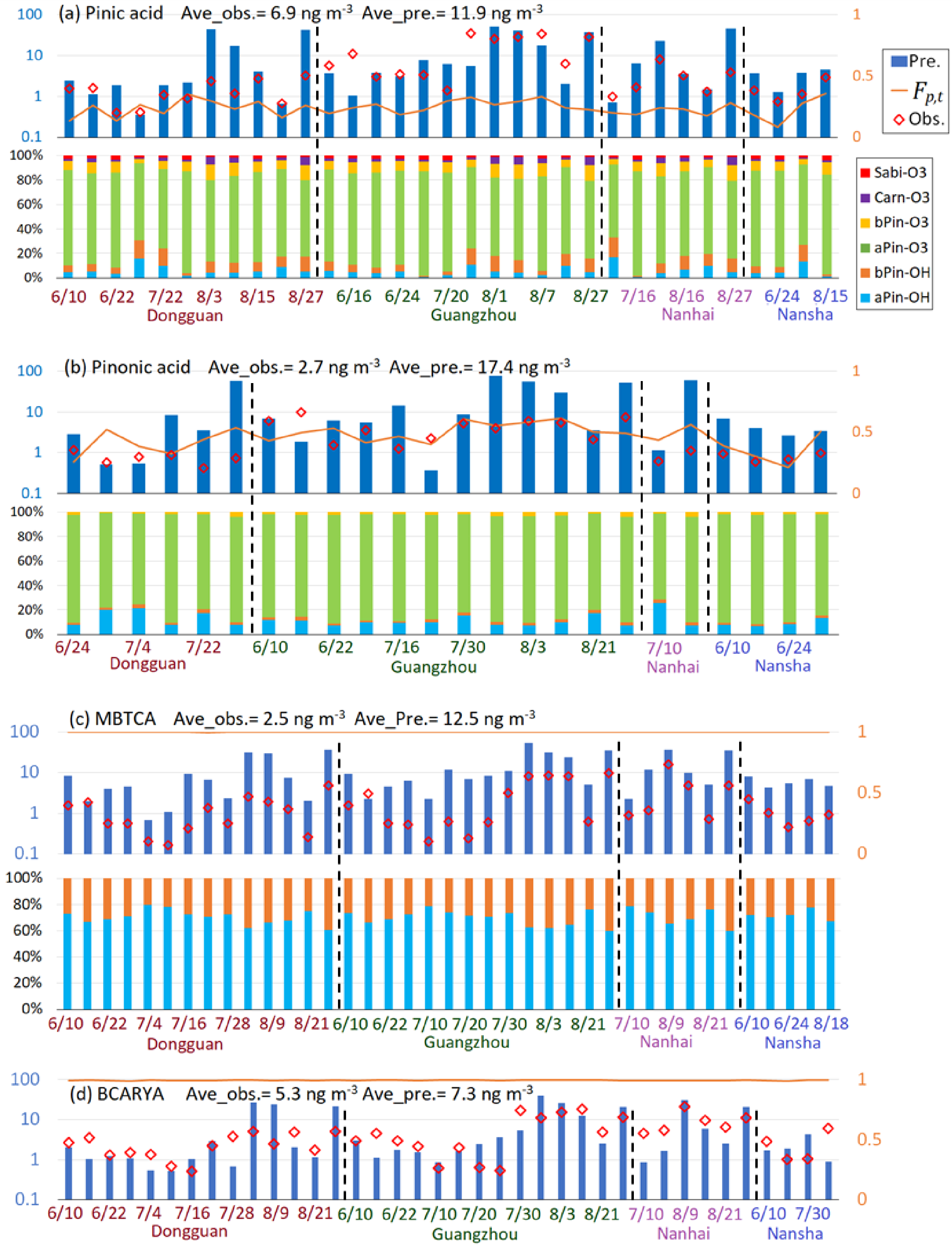


Figure 6-1 Observed and predicted organic tracer concentrations, their mass fraction in the particle phase ($F_{p,t}$), and the relative contributions from different formation pathways in the PRD region.

The model significantly overpredicts MBTCA (MFB and MFE ~ 1.3), but the predictions show a high correlation with the measurements (R ~ 0.8). MBTCA is formed from α-pinene (66%) and β-pinene (34%) reactions with OH. Since previous studies showed that the MBTCA was nearly nonvolatile,[209, 211] the over-prediction of MBTCA is likely caused by the overestimation of the mass yields and not C^*. Therefore, in the additional simulation discussed in Section 1.4.1, the mass yields of MBTCA used in the base case are reduced to obtain better predictions.

The predicted BCARYA agrees well with the measurements in the PRD region (MFB = -0.06, MFE = 0.75) with a high correlation (R = 0.62). The fractions of model predictions of BCARYA in the particle phase are close to 1.0.

6.3.3 Organic tracer evaluation across the country

To further evaluate the model's capability in simulating MT and SQT SOA tracers, the predictions are compared with ambient measurements at 14 sites across China.[136] The overall model performance is generally consistent with that in the PRD region, as shown in Figure S6-6. The model overpredicts PA (MFB = 0.70, R = 0.17), PNA (MFB = 0.47, R = 0.52), and MBTCA (MFB = 1.3, R = 0.65) but BCARYA predictions agree well with the measurements (MFB = -0.16, R = 0.67).

The OC measurements are not available at the monitor sites. However, it is expected that the OC may also be underpredicted according to OC evaluation in the PRD region (Table S6-3) and comparison with historical data (Figure S6-7). The average ratio of the observations to the predictions based on the historical data is 1.6. Therefore, the tracers are repartitioned based on OA loadings increased by a factor of 1.6 (Figure S6-6) to assess the impacts of potential underprediction of OA loadings on the tracer predictions, especially for PA and PNA. The repartitioning leads to slightly higher over predictions of PA and PNA, and the correlations between predictions and observations are not changed. As expected, the MBTCA and BCARYA concentrations are not influenced by the repartitioning.

The contributions of each MT tracer formation pathway are shown in Figure S6-8. The α-pinene ozonolysis is the most important pathway to form PA and PNA. For MBTCA, the contributions are mostly from α-pinene + OH, but β-pinene + OH is also important.

6.3.4 Monoterpene SOA formation

In this study, the semivolatile SOA formations from individual MTs due to reactions with OH, O_3, NO_3, and $O(^3P)$ are tracked explicitly in the CMAQ model. Figure 6-2 shows the regional concentrations of total MT SOA formed from each oxidation pathway. The NO_3 oxidation pathway (Figure 6-2c) has the highest contribution to MT SOA. The concentrations are above 1 $\mu g\ m^{-3}$ around the Sichuan Basin in southwest China, and the highest concentrations reach approximately 1.8 $\mu g\ m^{-3}$. The contributions of OH and O_3 pathways (Figure 6-2a and b) are approximately 0.3 $\mu g\ m^{-3}$ and 0.6 $\mu g\ m^{-3}$ in southern China. The relative contributions from each formation pathway agree with the predictions in the United States by Qin et al.,[198] in which the SOA from the NO_3 oxidation is more significant than that from O_3 and OH combined. The $O(^3P)$ oxidation pathway (Figure 6-2d) has negligible contributions to MT SOA. In addition, the oligomerized SOA contributes significantly to the total MT SOA (Figure 6-2e). The MT SOA oligomer concentrations in southern China are generally above 1 $\mu g\ m^{-3}$. Since the half-life of semivolatile products to generate oligomers is set to 20 hours, the oligomers show much broader spatial distributions than the semivolatile SOA. The oligomer concentrations in this study are similar to the aged SOA concentrations (from stepwise decreasing of the volatility of the lumped semivolatile gas-phase products) in the southeastern US estimated by Qin et al.[198]

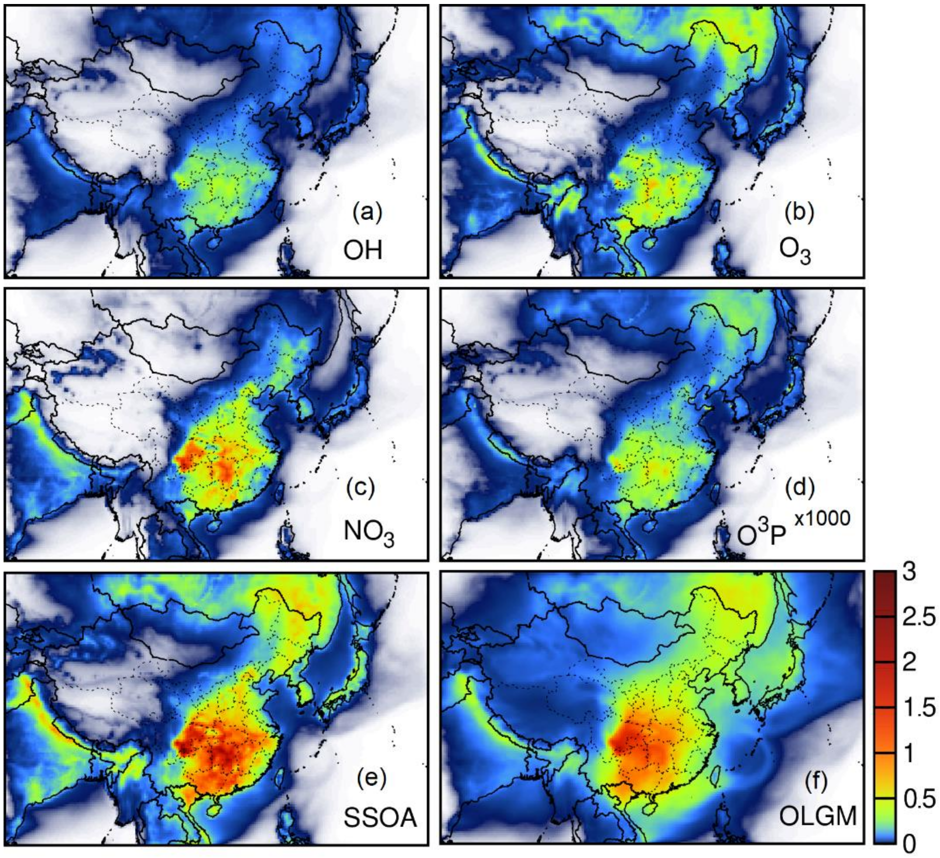


Figure 6-2 Regional contributions to the total monoterpene semivolatile SOA (SSOA) from individual oxidation pathways, (a) OH, (b) O_3, (c) NO_3, and (d) O^3P, and (e) total SSOA from all pathways, and (f) oligomers. Units: μg m^{-3}. The contributions of O^3P are scaled up by 1000 times to illustrate the regional distributions.

Figure S6-9 shows that β-pinene and Δ^3-carene contribute most to MT SOA among the MTs, followed by α-pinene, d-limonene, and sabinene. The oligomers are as important as the corresponding semivolatile SOA formed from all oxidation pathways but have broader spatial distributions. The detailed contributions from each formation pathway of individual MT and the corresponding oligomer are shown in Figure S6-10. Even though the major contribution pathways of MT SOA tracers (mainly from ozonolysis) and MT SOA (mainly from reactions with NO_3 radical) are different, their spatial distributions are similar (shown in Figure S6-11), and the daily

SOA concentrations from OH, O_3, and NO_3 show high correlations (Figure S6-12), implying that the tracer concentrations can still be used to estimate SOA concentrations.

6.4 Discussion

6.4.1 Improving the model performance of PA, PNA, MBTCA

PA and PNA are treated as semivolatile species, and their C^* values have been estimated in wide ranges in the literature (3.2 – 57 μg m^{-3} for PA and 5.8 – 781 μg m^{-3} for PNA).[55, 201, 208, 209] The overprediction of PA and PNA could be caused by low C^* values used in the base case. Thus, the C^* of the two tracers are adjusted using the measurements in the PRD region to improve the model performance of PA and PNA. The optimized C^* at 300 K (C^*_{300}) is found by minimizing the MFB of the repartitioned tracer predictions in the particle phase. The ΔH_{vap} values are estimated based on the semiempirical equation of Epstein et al.[162], as shown in Eq. (6.7),

$$\Delta H_{vap} = -11 \log_{10} C^*_{300} + 129 \tag{6.7}$$

where C^*_{300} is the saturation mass concentration under 300 K. After optimization, C^* at 298 K (C^*_{298}) is calculated from C^*_{300} and ΔH_{vap}. As shown in Figure S6-13, C^*_{298} is 33.4 μg m^{-3} (ΔH_{vap} = 111 kJ mol^{-1}), which is between the estimations based on E-AIM Method 1 and Method 2 (see **Table S6-2**). The optimized C^*_{298} for PNA is 34.9 μg m^{-3} (ΔH_{vap} = 111 kJ mol^{-1}), which is close to the estimation by Jenkin.[201] To be consistent, the α values of PA (updated to 0.0071 for APIN and 0.0125 for BPIN) and PNA (updated to 0.0067 for APIN and 0.0016 for BPIN) from APIN and BPIN photooxidation with OH radical[16] are updated based on the optimized C^*_{298} values using Eq.(1). The yields from the O_3 reactions are not adjusted as they were directly measured by Yu et al. based on simultaneous measurements of PA and PNA in the gas and particle phases.[54]

The MBTCA overprediction is likely caused by overestimating its yields from the photooxidation of α-pinene and β-pinene with OH because it is essentially nonvolatile and thus not affected by gas-particle partitioning. The MFB of MBTCA in the PRD region and the 14 sites across China is consistently ~1.3, corresponding to an overall overprediction by a factor of 6. Because there are no other chamber studies reported the yield of MBTCA, we reduced the mass yields by a factor of 6 for both species (α = 0.0021 for α-pinene + OH and α = 0.0011 for β-pinene + OH).

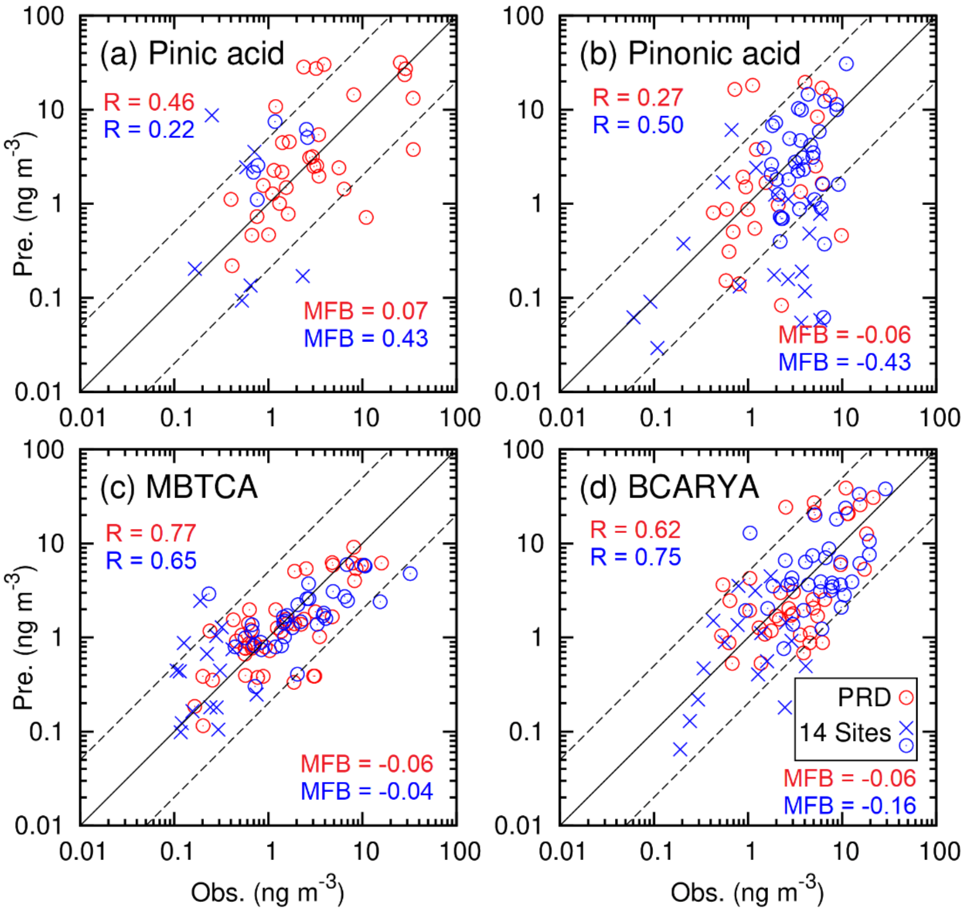


Figure 6-3 Predicted and observed (a) pinic acid, (b) pinonic acid, (c) MBTCA, and (d) BCARYA at all monitor locations using optimized saturation concentrations for pinic acid and pinonic acid and optimized mass yield for MBTCA. R is the Pearson correlation coefficient. The cross markers represent data from the sites in the clean areas of west and southwest China (Linzhi, Namco, Dunhuang, and Xishuangbanna). Solid line is 1:1, and dashed lines represent 1:5 and 5:1.

An additional simulation is conducted with the updated mass yields and C^* (Table 6-1), and the results are evaluated against the measurements in the PRD region and at the 14 sites across China (Figure 6-3). In this simulation, the predicted PA and PNA concentrations agree well with the PRD region measurements after repartitioning using observed OC concentrations, with a much-reduced MFB of 0.07 and -0.06, respectively, and the correlation between predictions and observations does not change much. The model slightly overpredicts PA comparing with the measurements from 14 sites across the country with an MFB = 0.4 and MFE = 1.1, and the overprediction becomes higher by repartitioning PA with 1.6×OA loadings (Figure S6-14; MFB = 0.5, MFE = 1.2). The PNA is slightly underpredicted with an MFB = -0.4 but the agreement with the observations becomes better after repartitioning with 1.6×OA (Figure S6-14; MFB = -0.3, MFE = 0.9). The MBTCA predictions show good agreements with the observations in PRD (MFB = -0.06, MFE = 0.55) and the 14 sites (MFB = -0.04, MFE = 0.62), with high correlations between the model predictions and measurements (R = 0.65 ~ 0.77). The improved PA, PNA, and MBTCA predictions are used in the subsequent analyses.

6.4.2 *Tracer-to-SOA ratio (f_{SOA}) used to estimate terpene SOA*

The tracer-to-SOA ratios (f_{SOA}) are determined based on the improved daily average concentrations extracted from the four sites in the PRD region and 14 sites across China (all days), as shown in Figure 6-4. The relationship between daily f_{SOA} and OA loadings are fitted by power-law functions, which allows a simple way to adjust f_{SOA} to estimate MT and SQT SOA based on measured ambient OA loading in field campaigns. The f_{SOA} values for MT SOA based on all three tracers (PA, PNA and MBTCA) show larger variabilities at low OA loadings because the gas-particle partitioning of PA and PNA are sensitive to temperature changes due to large ΔH_{vap} values, especially when the OA loading is low (Figure S6-15). In addition, the uncertainty in $C*$ can also lead to additional uncertainties in the modeled f_{SOA}. Instead, MBTCA is estimated to be nearly nonvolatile and thus is less affected by the C^* or ΔH_{vap}. Thus, the MBTCA alone is potentially a better molecular tracer to estimate MT SOA (Figure 6-4b and 4e). OA dependence of MBTCA-based f_{SOA} (f_{SOA}^{MBTCA}) can be approximated with Eq. (6.8),

$$f_{SOA} = 0.007C_{OA}^{-0.49} \tag{6.8}$$

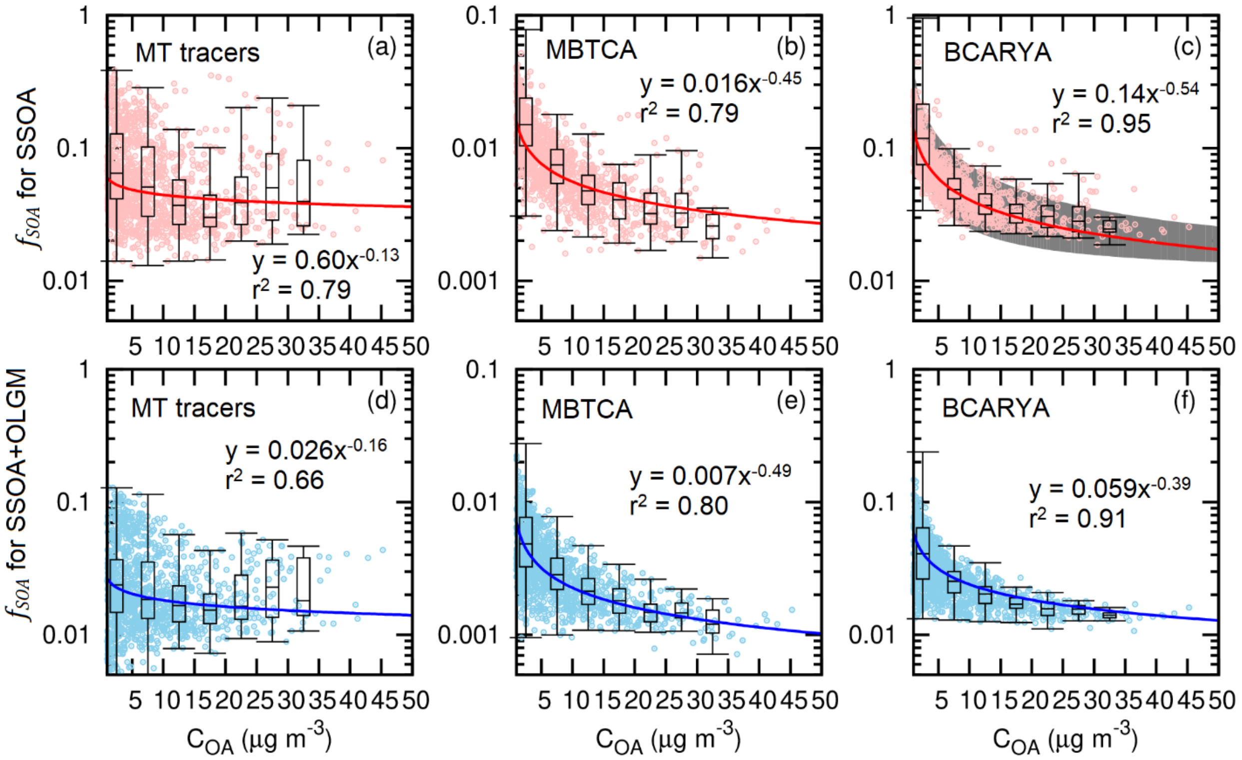


Figure 6-4 Correlation between daily SOA tracers-to-SOA ratio (f_{SOA}) and organic aerosol concentrations (C_{OA}) for semivolatile SOA (SSOA) only (a-c) and total SOA (SSOA + oligomers). f_{SOA} in (a,d) are based on the sum of all three MT tracers. The shaded areas in panel (c) are the theoretical variations of f_{SOA} due to OA loading, determined using $f_{SOA} = \alpha \frac{F_{p,t}}{Y_{SOA}}$, for the temperature range 15-35°C. The MATLAB curve fitting tool (robust regression with bi-square weighing function) is used to find the regression lines. r^2 is the weighted coefficient of determination. The box-whisker plots are for COA bins with a bin width of 5 μg m^{-3}. The whiskers are 1.5 IQR (interquartile range). Dashed lines in panels (c) and (d) are chamber-derived f_{SOA} of 0.0109 by Kleindienst et al.[38]

The MBTCA-based f_{SOA} values decrease significantly with the increase of OA loadings due to the changes in the gas-particle partitioning of MT SOA, and show a stronger correlation with the OA loading than f_{SOA} calculated from all three tracers. A similar trend is predicted for the BCARYA-based f_{SOA} for total SQT SOA (f_{SOA}^{BCARYA}), with the power-law function shown in Eq. (6.9),

$$f_{SOA} = 0.059C_{OA}^{-0.39} \tag{6.9}$$

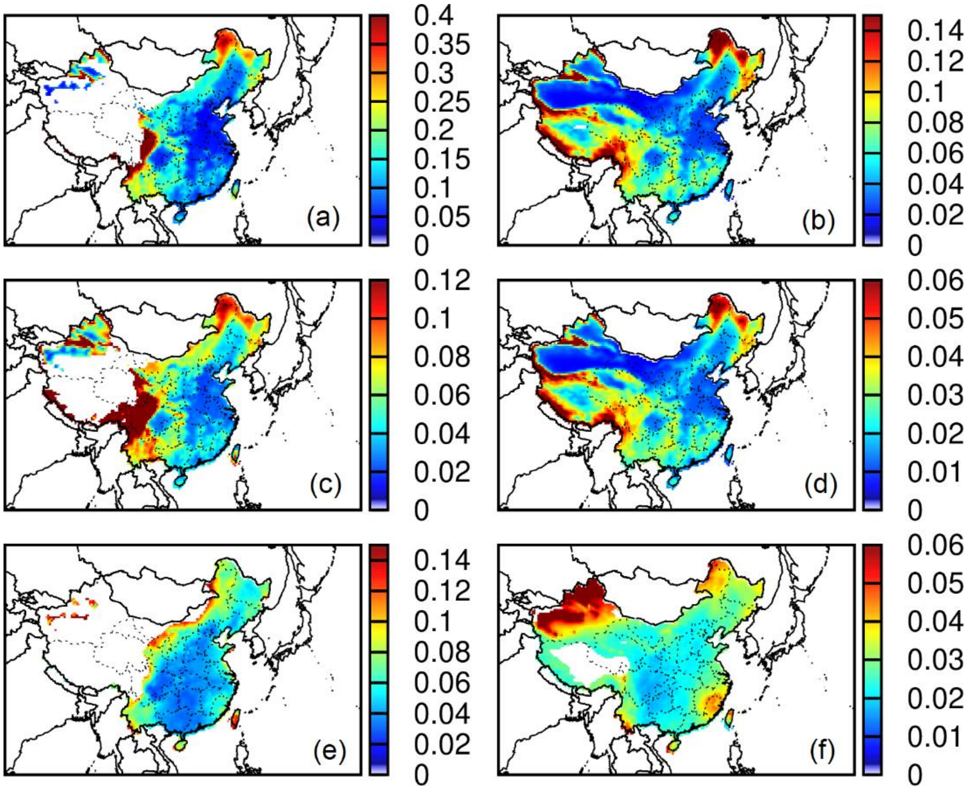


Figure 6-5 Regional distribution of three-month average tracer-to-SOA ratio ($\boldsymbol{f_{SOA}}$) based on (a,b) MT tracer and MT SOA, (c,d) MBTCA and MT SOA, and (e,f) BCARYA and SQT SOA. Panels (a,c,e) are for semivolatile SOA, and (b,d,f) are for total SOA (SSOA + oligomers). MT tracer is the sum of the mass concentrations of PA, PNA, and MBTCA.

Figure 6-5 shows the regional distribution of f_{SOA} calculated from the average concentrations of organic tracers and corresponding SOA concentrations during the summertime.

The spatial distribution of the f_{SOA} values generally anticorrelate with the OA loadings (*c.f.*, Figure S6-17). In areas where the OA concentrations are relatively high, the f_{SOA} values are low and show small spatial variations. The f_{SOA} determined by the seasonal average MBTCA and MT SOA (0.001-0.003), and BCARYA and SQT SOA (0.02-0.03) are generally uniform, while the one based on all three MT tracers show larger spatial variations.

6.4.3 Simple relationship between organic tracer and SOA

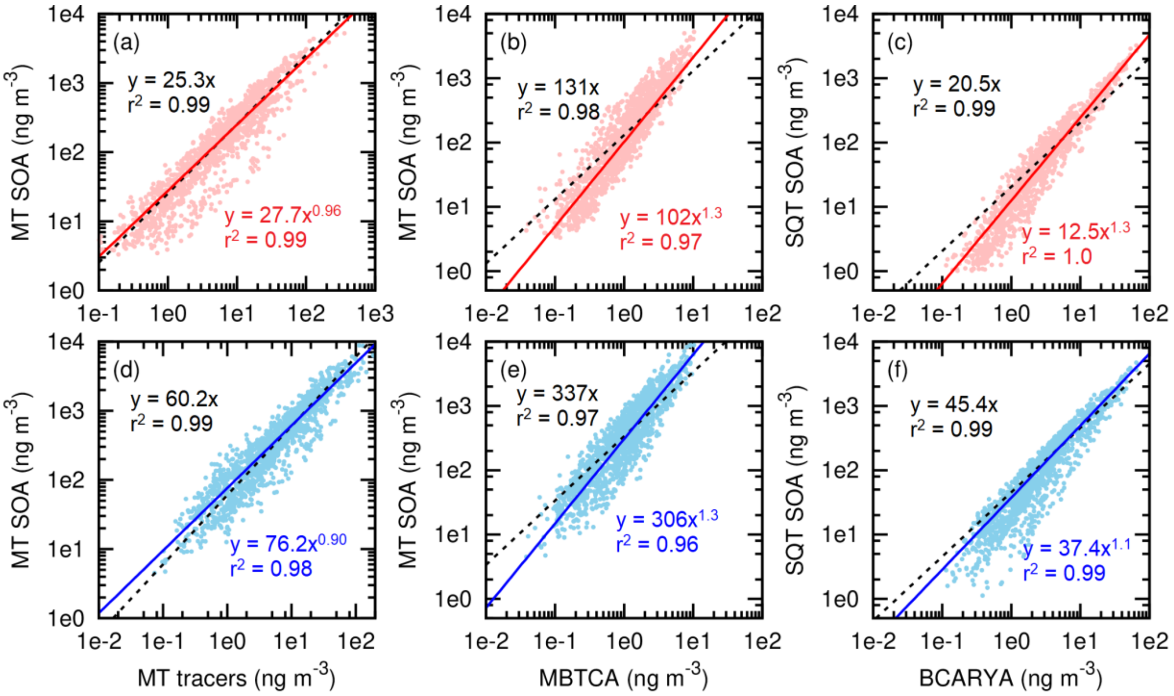


Figure 6-6 Linear and power-law equations to directly estimate MT and SQT SOA (semivolatile SOA, a-c; total SOA, d-f) using tracer concentrations. “MT tracers” is the sum of PA, PNA, and MBTCA). The Matlab curve fitting tool (robust regression with bi-square weighting function) is used to find the regression lines. r^2 is the weighted coefficient of determination.

An alternative method to estimate SOA from MT and SQT is to develop direct relationships between the tracer and SOA concentrations. Figure 6-6 shows two sets of equations derived from modeled tracer and SOA concentrations. One set of equations assumes a linear relationship between the tracer and SOA concentrations, and the other set uses a power-law equation. The power-law relationship implies that the log-transformed tracer and SOA concentrations are linearly correlated. The MFB and MFE in the calculated MT and SQT SOA concentrations from the two sets of equations are compared in Figure S6-18. The MFB between predicted SOA and estimations using power functions is close to zero, and the corresponding MFE's are all lower than 0.5. The power-law equations show better predictions of the MT and SQT SOA concentrations using MBTCA and BCARYA. The linear equations lead to slightly better results when the sum of the three MT tracers is used to estimate MT SOA.

The direct power-law predictions of MT and SQT SOA described above are also compared to the predictions based on f_{SOA}, as shown in Eqs. (6.8) and (6.9), using predicted OA and tracers (MBTCA and BCARYA). Both methods lead to good SOA estimations when compared to the CMAQ-predicted SOA (MFB < ±0.1 and MFE < 0.5 for MBTCA; MFB < ±0.2 and MFE ~ 0.4 for BCARYA). Figure S6-19 shows that the f_{SOA} estimations are slightly better in matching the CMAQ-predicted MT and SQT SOA at higher concentrations, but the power-law equations lead to slightly better estimated SOA in lower concentrations. Overall, both methods can be applied in future field measurement studies to estimate MT SOA and SQT SOA.

6.4.4 Diurnal variation of MT SOA and its molecular tracers

MT SOA and tracers are generated from distinct precursor and formation pathways. Since OH, NO_3 and O_3 have different diurnal variations, it is expected that f_{SOA} also demonstrates diurnal variations and the f_{SOA} derived based on daily concentrations may not be applicable for the hourly SOA estimation. For example, Figure S6-20 shows that the MT SOA concentrations are higher at night and lower during the day in Beijing and positively correlate with the sum of all three MT tracers (PA, PNA, and MBTCA) (R=0.97). However, the SOA concentrations are strongly anticorrelated with MBTCA (R = -0.85), which is only generated from OH oxidation. The negative correlation of hourly MBTCA with SOA leading to large diurnal variation of MBTCA-based f_{SOA}. In contrast, the f_{SOA} based on all three tracers has much less variation and thus is more appropriate for estimating hourly MT SOA. However, since PA and PNA are volatile and their predictions are sensitive to C^* values, future experimental studies are challenged to identify new low-volatile monoterpene tracers from the NO_3 and O_3 oxidation pathways.

7 CONCLUSIONS AND FUTURE RESEARCH

7.1 Summary of major conclusions

The overall objective of this study is to develop a regional chemical transport model (CTM) to simulate organic markers and precursor-specific SOA to evaluate model representations of precursor-specific SOA formation and reassess the tracer-to-SOA ratios (f_{SOA}) used in the field measurement study to estimate SOA concentrations in ambient air.

In Section 3, the model predicted aromatic SOA and monoterpene SOA show strong correlations with the daily measured organic tracers DHOPA (R~0.8) and α-pinT (R>0.8), which affirms that the tracer-based method is a good approach to evaluate model representations of precursor-specific SOA formation. The tracer-to-SOA ratios derived from the SOA predictions excluding nonvolatile components (oligomers, and secondary glyoxal GLY and methylglyoxal SOA) are close to the values reported in the literatures[38, 39], suggesting that the chamber-derived f_{SOA} values may need to be adjusted to account for all SOA components from a specific precursor. The proper form of such adjustments must be determined using regional precursor and tracer-resolved chemical transport models.

The Community Multiscale Air Quality (CMAQ) model was modified to simulate organic tracers, including 2,3-dihydroxy-4-oxo-pentatonic acid (DHOPA) for aromatic SOA, 2-methyltetrols (2-MT) and 2-methylglyceric acid (2-MG) for isoprene SOA, pinic acid (PA), pinonic acid (PNA), and MBTCA for monoterpene (MT) SOA, and β-caryophyllinic acid (BCARYA) for sesquiterpenes (SQT) SOA. Generally, the model predictions agree with the observations, and modeling results are reliable for the f_{SOA} reassessment in the ambient atmosphere.

The DHOPA is well predicted by the modified CMAQ model comparing with the field measurements (R = 0.8; MFB = 0.15; MFE = 0.44) during the wintertime in Section 4. The modeled f_{SOA} shows a strong dependence on the OA loading due to the partitioning effects on the DHOPA and SOA, when only semivolatile aromatic SOA components are included. The OA dependence becomes weaker when the oligomers and dicarbonyl SOA products are considered, and f_{SOA} value of ~0.002 is determined to estimate total aromatic SOA. The chamber-determined f_{SOA} (0.004) leads to an underestimation of total aromatic SOA by a factor of 2.

The isoprene-SOA scheme in the CMAQ was expanded to simulate 2-MT and 2-MG by treating them as semivolatile species and including a non-heterogeneous formation pathway in Section 5. The predictions were evaluated against field measurements across China during the summertime. The modified scheme leads to consistently improved predictions of the two tracers with lower biases and errors and much stronger correlations with observations than the original scheme, in which the two tracers are treated as nonvolatile species and are only formed in the aqueous phase after surface uptake of their precursors, isoprene epoxide and methacrylic acid epoxide.

The CMAQ model was modified to simulate SOA formation from five explicit and one lumped MT species and SQT in Section 6. Among each explicitly tracked photooxidation pathway, including OH, O_3, NO_3, and O^3P, the NO_3 oxidation pathway contributes most significantly to the MT SOA formation. MBTCA and BCARYA predictions well agree with the observations from 18 sites across China, and they are good tracers to estimate MT SOA and SQT SOA. The power-law equations directly link the tracer concentrations to the corresponding SOA concentrations were proposed and could lead to good SOA estimations.

7.2 Recommendations for future research

This research leads to a chemical transport model that can simulate seven important organic tracers to cover SOA from all major precursors, including aromatic compounds, isoprene, monoterpenes, and sesquiterpenes. The formation pathways of each tracer species were discussed in detail in Section 4-6. However, the current understanding of the SOA formation pathways are far from complete. More chamber, field, and modeling studies are needed to resolve the detailed SOA chemical components and their formation pathways.

In Section 5, the 2-MT and 2-MG were simulated considering partitioning between gas and organic matters, and the partitioning coefficient (K_{om}) was only applied to get the equilibrium between gas and aerosol phase. However, these two tracers can be formed in the aqueous phase chemistry following surface uptake of their precursors IEPOX and MAE. Some studies applied Raoult's or Henry's Law to estimate gas-particle partitioning the SOA products from IEPOX and MAE [220, 221]. The gas-OM-aqueous three-phase partitioning equilibrium involving the tracers needs to be developed in the regional CTMs model and evaluated against ambient measurements.

The gas-particle partitioning coefficients for each organic tracer in this research were calculated by assuming the activity coefficient of the tracers in the OM mixture is unity, as in Eq. (4.3). Few studies have been quantitively analyzed the impacts of activity coefficient on the SOA formation. A recent study from Li et al. [163] suggested that the activity coefficients of some semivolatile SOA species in a mixture of POA and SOA components are greater than 1.0 based on the widely used UNIFAC model, leading to a decrease of SOA concentrations in the CMAQ model in hot and humid environments. A UNIFAC activity coefficient estimation scheme, such as that used in Pankow et al.[164], needs to be implemented to improve model predictions of SOA and SOA tracers and the f_{SOA} evaluations need to be updated accordingly.

The organic tracers were treated as nonreactive products from the precursor oxidation chemistry without additional chemical loss. However, in reality, these tracers are not chemically inert. For example, the PNA oxidation reaction is responsible for the low-volatile organic tracer MBTCA[209], and DHOPA and 2-MT were found to be heterogeneously oxidized by O_3 in the OM phase[222]. The impacts on the tracer predictions and f_{SOA} values due to chemical loss reactions should be quantified in the future work.

The f_{SOA} values are determined based on daily average concentrations of organic tracers and SOA in the simulations. While the f_{SOA} diurnal variation has not been studied in this study. For example, in Section 6, the MBTCA is formed through the photooxidation of α-pinene and β-pinene with OH radical, which is mainly occurred during the daytime. However, according to the modeling results in this study and previous work [198], the oxidation pathway with NO_3 radical contributes most significantly to the total MT SOA, and it mainly occurs during the nighttime. Theoretically the f_{SOA} values vary significantly in day and night. The variation can be revealed by comparing modeled organic tracers and SOA against hourly measurements.

REFERENCES

1. Hu, J.; Wang, P.; Ying, Q.; Zhang, H.; Chen, J.; Ge, X.; Li, X.; Jiang, J.; Wang, S.; Zhang, J., Modeling biogenic and anthropogenic secondary organic aerosol in China. *1foldr Import 2019-10-08 Batch 1* **2017**.

2. Carlton, A.; Wiedinmyer, C.; Kroll, J., A review of Secondary Organic Aerosol (SOA) formation from isoprene. **2009**.

3. Dechapanya, W.; Russell, M.; Allen, D. T., Estimates of anthropogenic secondary organic aerosol formation in Houston, Texas special issue of aerosol science and technology on findings from the fine particulate matter supersites program. *Aerosol science and technology* **2004,** *38*, (S1), 156-166.

4. Yu, S.; Bhave, P. V.; Dennis, R. L.; Mathur, R., Seasonal and regional variations of primary and secondary organic aerosols over the continental United States: Semi-empirical estimates and model evaluation. *Environmental science & technology* **2007,** *41*, (13), 4690-4697.

5. Bessagnet, B.; Menut, L.; Curci, G.; Hodzic, A.; Guillaume, B.; Liousse, C.; Moukhtar, S.; Pun, B.; Seigneur, C.; Schulz, M., Regional modeling of carbonaceous aerosols over Europe—focus on secondary organic aerosols. *Journal of Atmospheric Chemistry* **2008,** *61*, (3), 175-202.

6. Huang, Y.; Wu, S.; Dubey, M.; French, N., Impact of aging mechanism on model simulated carbonaceous aerosols. *Atmospheric Chemistry & Physics* **2013,** *13*, (13).

7. Zhang, H.; Wang, Z.; Guo, P.; Wang, Z., A modeling study of the effects of direct radiative forcing due to carbonaceous aerosol on the climate in East Asia. *Advances in Atmospheric sciences* **2009,** *26*, (1), 57-66.

8. Liousse, C.; Penner, J.; Chuang, C.; Walton, J.; Eddleman, H.; Cachier, H., A global three-dimensional model study of carbonaceous aerosols. *Journal of Geophysical Research: Atmospheres* **1996,** *101*, (D14), 19411-19432.

9. Emanuelsson, E. U.; Hallquist, M.; Kristensen, K.; Glasius, M.; Bohn, B.; Fuchs, H.; Kammer, B.; Kiendler-Scharr, A.; Nehr, S.; Rubach, F., Formation of anthropogenic secondary organic aerosol (SOA) and its influence on biogenic SOA properties. *Atmospheric Chemistry and Physics* **2013,** *13*, (5), 2837.

10. Nakao, S.; Liu, Y.; Tang, P.; Chen, C.-L.; Zhang, J.; Cocker Iii, D., Chamber studies of SOA formation from aromatic hydrocarbons: observation of limited glyoxal uptake. *Atmospheric Chemistry and Physics* **2012,** *12*, (9), 3927-3937.

11. Al-Naiema, I. M.; Stone, E. A., Evaluation of anthropogenic secondary organic aerosol tracers from aromatic hydrocarbons. *Atmospheric Chemistry & Physics* **2017,** *17*, (3).

12. Surratt, J. D.; Chan, A. W.; Eddingsaas, N. C.; Chan, M.; Loza, C. L.; Kwan, A. J.; Hersey, S. P.; Flagan, R. C.; Wennberg, P. O.; Seinfeld, J. H., Reactive intermediates revealed in

secondary organic aerosol formation from isoprene. *Proceedings of the National Academy of Sciences* **2010,** *107*, (15), 6640-6645.

13. Henze, D. K.; Seinfeld, J. H., Global secondary organic aerosol from isoprene oxidation. *Geophysical Research Letters* **2006,** *33*, (9).

14. Kroll, J. H.; Ng, N. L.; Murphy, S. M.; Flagan, R. C.; Seinfeld, J. H., Secondary organic aerosol formation from isoprene photooxidation. *Environmental science & technology* **2006,** *40*, (6), 1869-1877.

15. Friedman, B.; Farmer, D. K., SOA and gas phase organic acid yields from the sequential photooxidation of seven monoterpenes. *Atmospheric Environment* **2018,** *187*, 335-345.

16. Mutzel, A.; Rodigast, M.; Iinuma, Y.; Böge, O.; Herrmann, H., Monoterpene SOA–contribution of first-generation oxidation products to formation and chemical composition. *Atmospheric environment* **2016,** *130*, 136-144.

17. Varutbangkul, V.; Brechtel, F.; Bahreini, R.; Ng, N.; Keywood, M.; Kroll, J.; Flagan, R.; Seinfeld, J.; Lee, A.; Goldstein, A., Hygroscopicity of secondary organic aerosols formed by oxidation of cycloalkenes, monoterpenes, sesquiterpenes, and related compounds. **2006**.

18. Xu, L.; Pye, H. O.; He, J.; Chen, Y.; Murphy, B. N.; Ng, L. N., Experimental and model estimates of the contributions from biogenic monoterpenes and sesquiterpenes to secondary organic aerosol in the southeastern United States. *Atmospheric chemistry and physics* **2018,** *18*, (17), 12613.

19. Ervens, B.; Volkamer, R., Glyoxal processing by aerosol multiphase chemistry: towards a kinetic modeling framework of secondary organic aerosol formation in aqueous particles. *Atmospheric Chemistry and Physics* **2010,** *10*, (17), 8219-8244.

20. Carlton, A. G.; Turpin, B. J.; Altieri, K. E.; Seitzinger, S.; Reff, A.; Lim, H.-J.; Ervens, B., Atmospheric oxalic acid and SOA production from glyoxal: Results of aqueous photooxidation experiments. *Atmospheric Environment* **2007,** *41*, (35), 7588-7602.

21. Li, J.; Zhang, M.; Wu, F.; Sun, Y.; Tang, G., Assessment of the impacts of aromatic VOC emissions and yields of SOA on SOA concentrations with the air quality model RAMS-CMAQ. *Atmospheric Environment* **2017,** *158*, 105-115.

22. Lannuque, V.; Camredon, M.; Couvidat, F.; La, Y. S.; Valorso, R.; Hodzic, A.; Madronich, S.; Bessagnet, B.; Aumont, B. In *Representation of SOA formation in air quality models: a new parameterization developed on explicit simulations*, 2016; 2016.

23. Jo, D. S.; Hodzic, A.; Emmons, L. K.; Marais, E. A.; Peng, Z.; Nault, B. A.; Hu, W.; Campuzano-Jost, P.; Jimenez, J. L., A simplified parameterization of isoprene-epoxydiol-derived secondary organic aerosol (IEPOX-SOA) for global chemistry and climate models: a case study with GEOS-Chem v11-02-rc. **2019**.

24. Deng, W.; Liu, T.; Zhang, Y.; Situ, S.; Hu, Q.; He, Q.; Zhang, Z.; Lü, S.; Bi, X.; Wang, X., Secondary organic aerosol formation from photo-oxidation of toluene with NOx and SO2: Chamber simulation with purified air versus urban ambient air as matrix. *Atmospheric Environment* **2017,** *150*, 67-76.

25. Jorga, S. D.; Kaltsonoudis, C.; Liangou, A.; Pandis, S. N., Measurement of formation rates of secondary aerosol in the ambient urban atmosphere using a dual smog chamber system. *Environmental Science & Technology* **2019,** *54*, (3), 1336-1343.

26. Heiden, A.; Kobel, K.; Komenda, M.; Koppmann, R.; Shao, M.; Wildt, J., Toluene emissions from plants. *Geophysical Research Letters* **1999,** *26*, (9), 1283-1286.

27. Schnitzhofer, R.; Beauchamp, J.; Dunkl, J.; Wisthaler, A.; Weber, A.; Hansel, A., Long-term measurements of CO, NO, NO2, benzene, toluene and PM10 at a motorway location in an Austrian valley. *Atmospheric Environment* **2008,** *42*, (5), 1012-1024.

28. Beauregard, D.; Branch, E. I., Locating and Estimating Air Emissions from Sources of Toluene. *US Environmental Protection Agency* **1993**.

29. Schwantes, R. H.; Schilling, K. A.; McVay, R. C.; Lignell, H.; Coggon, M. M.; Zhang, X.; Wennberg, P. O.; Seinfeld, J. H., Formation of highly oxygenated low-volatility products from cresol oxidation. *Atmospheric Chemistry and Physics* **2017,** *17*, (5), 3453-3474.

30. Ng, N.; Kroll, J.; Chan, A.; Chhabra, P.; Flagan, R.; Seinfeld, J., Secondary organic aerosol formation from m-xylene, toluene, and benzene. **2007**.

31. Hildebrandt, L.; Donahue, N. M.; Pandis, S. N., High formation of secondary organic aerosol from the photo-oxidation of toluene. *Atmos. Chem. Phys.* **2009,** *9*, (9), 2973-2986.

32. Gery, M. W.; Fox, D. L.; Jeffries, H. E.; Stockburger, L.; Weathers, W. S., A continuous stirred tank reactor investigation of the gas-phase reaction of hydroxyl radicals and toluene. **1985,** *17*, (9), 931-955.

33. Jenkin, M. E.; Saunders, S. M.; Wagner, V.; Pilling, M. J., Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part B): tropospheric degradation of aromatic volatile organic compounds. *Atmos. Chem. Phys.* **2003,** *3*, (1), 181-193.

34. Zhang, Y.; Sheesley, R. J.; Schauer, J. J.; Lewandowski, M.; Jaoui, M.; Offenberg, J. H.; Kleindienst, T. E.; Edney, E. O., Source apportionment of primary and secondary organic aerosols using positive matrix factorization (PMF) of molecular markers. *Atmospheric Environment* **2009,** *43*, (34), 5567-5574.

35. Feng, J.; Li, M.; Zhang, P.; Gong, S.; Zhong, M.; Wu, M.; Zheng, M.; Chen, C.; Wang, H.; Lou, S., Investigation of the sources and seasonal variations of secondary organic aerosols in PM2. 5 in Shanghai with organic tracers. *Atmospheric environment* **2013,** *79*, 614-622.

36. Yuan, Z.; Yu, J.; Lau, A. K.-H.; Louie, P. K. K.; Fung, J., Application of positive matrix factorization in estimating aerosol secondary organic carbon in Hong Kong and its relationship with secondary sulfate. **2006**.

37. Woody, M. C.; Baker, K. R.; Hayes, P. L.; Jimenez, J. L.; Koo, B.; Pye, H. O., Understanding sources of organic aerosol during CalNex-2010 using the CMAQ-VBS. *Atmospheric Chemistry and Physics* **2016,** *16*, (6), 4081.

38. Kleindienst, T. E.; Jaoui, M.; Lewandowski, M.; Offenberg, J. H.; Lewis, C. W.; Bhave, P. V.; Edney, E. O., Estimates of the contributions of biogenic and anthropogenic hydrocarbons to secondary organic aerosol at a southeastern US location. *Atmospheric Environment* **2007,** *41*, (37), 8288-8300.

39. Al-Naiema, I. M.; Offenberg, J. H.; Madler, C. J.; Lewandowski, M.; Kettler, J.; Fang, T.; Stone, E. A., Secondary organic aerosols from aromatic hydrocarbons and their contribution to fine particulate matter in Atlanta, Georgia. *Atmospheric Environment* **2020,** *223*, 117227.

40. Kleindienst, T. E.; Lewandowski, M.; Offenberg, J. H.; Edney, E. O.; Jaoui, M.; Zheng, M.; Ding, X.; Edgerton, E. S., Contribution of primary and secondary sources to organic aerosol and PM2. 5 at SEARCH network sites. *Journal of the Air & Waste Management Association* **2010,** *60*, (11), 1388-1399.

41. He, X.; Wang, Q.; Huang, X. H.; Huang, D. D.; Zhou, M.; Qiao, L.; Zhu, S.; Ma, Y.-g.; Wang, H.-l.; Li, L., Hourly measurements of organic molecular markers in urban Shanghai, China: Observation of enhanced formation of secondary organic aerosol during particulate matter episodic periods. *Atmospheric Environment* **2020,** *240*, 117807.

42. Lyu, X.; Guo, H.; Cheng, H.; Wang, X.; Ding, X.; Lu, H.; Yao, D.; Xu, C., Observation of SOA tracers at a mountainous site in Hong Kong: chemical characteristics, origins and implication on particle growth. *Science of The Total Environment* **2017,** *605*, 180-189.

43. Couvidat, F.; Seigneur, C., Modeling secondary organic aerosol formation from isoprene oxidation under dry and humid conditions. *Atmospheric Chemistry and Physics* **2011,** *11*, (2), 893-909.

44. Lamkaddam, H.; Dommen, J.; Ranjithkumar, A.; Gordon, H.; Wehrle, G.; Krechmer, J.; Majluf, F.; Salionov, D.; Schmale, J.; Bjelić, S., Large contribution to secondary organic aerosol from isoprene cloud chemistry. *Science Advances* **2021,** *7*, (13), eabe2952.

45. Lin, Y.-H.; Zhang, H.; Pye, H. O.; Zhang, Z.; Marth, W. J.; Park, S.; Arashiro, M.; Cui, T.; Budisulistiorini, S. H.; Sexton, K. G., Epoxide as a precursor to secondary organic aerosol formation from isoprene photooxidation in the presence of nitrogen oxides. *Proceedings of the National Academy of Sciences* **2013,** *110*, (17), 6718-6723.

46. Jaoui, M.; Corse, E. W.; Lewandowski, M.; Offenberg, J. H.; Kleindienst, T. E.; Edney, E. O., Formation of organic tracers for isoprene SOA under acidic conditions. *Atmospheric Environment* **2010,** *44*, (14), 1798-1805.

47. Wang, Y.; Pavuluri, C. M.; Fu, P.; Li, P.; Dong, Z.; Xu, Z.; Ren, H.; Fan, Y.; Li, L.; Zhang, Y.-L., Characterization of secondary organic aerosol tracers over Tianjin, North China during summer to autumn. *ACS Earth and Space Chemistry* **2019,** *3*, (10), 2339-2352.

48. Ding, X.; He, Q.-F.; Shen, R.-Q.; Yu, Q.-Q.; Zhang, Y.-Q.; Xin, J.-Y.; Wen, T.-X.; Wang, X.-M., Spatial and seasonal variations of isoprene secondary organic aerosol in China: Significant impact of biomass burning during winter. *Scientific reports* **2016,** *6*, (1), 1-10.

49. Pye, H. O.; Pinder, R. W.; Piletic, I. R.; Xie, Y.; Capps, S. L.; Lin, Y.-H.; Surratt, J. D.; Zhang, Z.; Gold, A.; Luecken, D. J., Epoxide pathways improve model predictions of isoprene markers and reveal key role of acidity in aerosol formation. *Environmental science & technology* **2013,** *47*, (19), 11056-11064.

50. Budisulistiorini, S. H.; Nenes, A.; Carlton, A. G.; Surratt, J. D.; McNeill, V. F.; Pye, H. O., Simulating aqueous-phase isoprene-epoxydiol (IEPOX) secondary organic aerosol production during the 2013 Southern Oxidant and Aerosol Study (SOAS). *Environmental science & technology* **2017,** *51*, (9), 5026-5034.

51. Kleindienst, T.; Lewandowski, M.; Offenberg, J.; Jaoui, M.; Edney, E., The formation of secondary organic aerosol from the isoprene+ OH reaction in the absence of NO x. *Atmospheric Chemistry and Physics* **2009,** *9*, (17), 6541-6558.

52. Zhang, H.; Surratt, J.; Lin, Y.; Bapat, J.; Kamens, R., Effect of relative humidity on SOA formation from isoprene/NO photooxidation: enhancement of 2-methylglyceric acid and its corresponding oligoesters under dry conditions. *Atmospheric Chemistry and Physics* **2011,** *11*, (13), 6411-6424.

53. Wang, W.; Iinuma, Y.; Kahnt, A.; Ryabtsova, O.; Mutzel, A.; Vermeylen, R.; Van der Veken, P.; Maenhaut, W.; Herrmann, H.; Claeys, M., Formation of secondary organic aerosol marker compounds from the photooxidation of isoprene and isoprene-derived alkene diols under low-NO x conditions. *Faraday discussions* **2013,** *165*, 261-272.

54. Yu, J.; Griffin, R. J.; Cocker III, D. R.; Flagan, R. C.; Seinfeld, J. H.; Blanchard, P., Observation of gaseous and particulate products of monoterpene oxidation in forest atmospheres. *Geophysical Research Letters* **1999,** *26*, (8), 1145-1148.

55. Pankow, J. F.; Seinfeld, J. H.; Asher, W. E.; Erdakos, G. B., Modeling the formation of secondary organic aerosol. 1. Application of theoretical principles to measurements obtained in the α-pinene/, β-pinene/, sabinene/, Δ3-carene/, and cyclohexene/ozone systems. *Environmental science & technology* **2001,** *35*, (6), 1164-1172.

56. Szmigielski, R.; Surratt, J. D.; Gómez-González, Y.; Van der Veken, P.; Kourtchev, I.; Vermeylen, R.; Blockhuys, F.; Jaoui, M.; Kleindienst, T. E.; Lewandowski, M., 3-methyl-1, 2, 3-butanetricarboxylic acid: An atmospheric tracer for terpene secondary organic aerosol. *Geophysical Research Letters* **2007,** *34*, (24).

57. Donahue, N. M.; Henry, K. M.; Mentel, T. F.; Kiendler-Scharr, A.; Spindler, C.; Bohn, B.; Brauers, T.; Dorn, H. P.; Fuchs, H.; Tillmann, R., Aging of biogenic secondary organic aerosol via gas-phase OH radical reactions. *Proceedings of the National Academy of Sciences* **2012,** *109*, (34), 13503-13508.

58. Amorim, J. V.; Guo, X.; Gautam, T.; Fang, R.; Fotang, C.; Williams, F. J.; Zhao, R., Photo-oxidation of pinic acid in the aqueous phase: a mechanistic investigation under acidic and basic pH conditions. *Environmental Science: Atmospheres* **2021**.

59. Jaoui, M.; Lewandowski, M.; Kleindienst, T. E.; Offenberg, J. H.; Edney, E. O., β-caryophyllinic acid: An atmospheric tracer for β-caryophyllene secondary organic aerosol. *Geophysical Research Letters* **2007,** *34*, (5).

60. Sakulyanontvittaya, T.; Duhl, T.; Wiedinmyer, C.; Helmig, D.; Matsunaga, S.; Potosnak, M.; Milford, J.; Guenther, A., Monoterpene and sesquiterpene emission estimates for the United States. *Environmental science & technology* **2008,** *42*, (5), 1623-1629.

61. Gao, Y.; Wang, H.; Zhang, X.; Jing, S. a.; Peng, Y.; Qiao, L.; Zhou, M.; Huang, D. D.; Wang, Q.; Li, X., Estimating secondary organic aerosol production from toluene photochemistry in a megacity of China. *Environmental science & technology* **2019,** *53*, (15), 8664-8671.

62. Sato, K.; Nakashima, Y.; Morino, Y.; Imamura, T.; Kurokawa, J.-i.; Kajii, Y., Total OH reactivity measurements for the OH-initiated oxidation of aromatic hydrocarbons in the presence of NOx. *Atmospheric Environment* **2017,** *171*, 272-278.

63. Seinfeld, J. H.; Pandis, S. N., *Atmospheric chemistry and physics: from air pollution to climate change*. John Wiley & Sons: 2016.

64. Zhang, R.; Wang, G.; Guo, S.; Zamora, M. L.; Ying, Q.; Lin, Y.; Wang, W.; Hu, M.; Wang, Y., Formation of urban fine particulate matter. *Chemical reviews* **2015,** *115*, (10), 3803-3855.

65. Klotz, B.; Sørensen, S.; Barnes, I.; Becker, K. H.; Etzkorn, T.; Volkamer, R.; Platt, U.; Wirtz, K.; Martín-Reviejo, M., Atmospheric oxidation of toluene in a large-volume outdoor photoreactor: In situ determination of ring-retaining product yields. *The Journal of Physical Chemistry A* **1998,** *102*, (50), 10289-10299.

66. Smith, D.; McIver, C.; Kleindienst, T., Primary product distribution from the reaction of hydroxyl radicals with toluene at ppb NO x mixing ratios. *Journal of Atmospheric Chemistry* **1998,** *30*, (2), 209-228.

67. Noda, J.; Volkamer, R.; Molina, M. J., Dealkylation of alkylbenzenes: A significant pathway in the toluene, o-, m-, p-xylene+ OH reaction. *The Journal of Physical Chemistry A* **2009,** *113*, (35), 9658-9666.

68. Atkinson, R.; Aschmann, S. M.; Arey, J.; Carter, W. P., Formation of ring-retaining products from the OH radical-initiated reactions of benzene and toluene. *International Journal of Chemical Kinetics* **1989,** *21*, (9), 801-827.

69. Baltaretu, C. O.; Lichtman, E. I.; Hadler, A. B.; Elrod, M. J., Primary atmospheric oxidation mechanism for toluene. *The Journal of Physical Chemistry A* **2009,** *113,* (1), 221-230.

70. Ji, Y.; Zhao, J.; Terazono, H.; Misawa, K.; Levitt, N. P.; Li, Y.; Lin, Y.; Peng, J.; Wang, Y.; Duan, L.; Pan, B.; Zhang, F.; Feng, X.; An, T.; Marrero-Ortiz, W.; Secrest, J.; Zhang, A. L.; Shibuya, K.; Molina, M. J.; Zhang, R., Reassessing the atmospheric oxidation mechanism of toluene. **2017,** *114,* (31), 8169-8174.

71. Qi, X.; Zhu, S.; Zhu, C.; Hu, J.; Lou, S.; Xu, L.; Dong, J.; Cheng, P., Smog chamber study of the effects of NOx and NH3 on the formation of secondary organic aerosols and optical properties from photo-oxidation of toluene. *Science of The Total Environment* **2020,** *727,* 138632.

72. Sarwar, G.; Appel, K.; Carlton, A.; Mathur, R.; Schere, K.; Zhang, R.; Majeed, M., Impact of a new condensed toluene mechanism on air quality model predictions in the US. *Geoscientific Model Development* **2011,** *4,* (1), 183.

73. Li, J.; Cleveland, M.; Ziemba, L. D.; Griffin, R. J.; Barsanti, K. C.; Pankow, J. F.; Ying, Q., Modeling regional secondary organic aerosol using the Master Chemical Mechanism. *Atmospheric Environment* **2015,** *102,* 52-61.

74. Bloss, C.; Wagner, V.; Jenkin, M. E.; Volkamer, R.; Bloss, W. J.; Lee, J. D.; Heard, D. E.; Wirtz, K.; Martin-Reviejo, M.; Rea, G.; Wenger, J. C.; Pilling, M. J., Development of a detailed chemical mechanism (MCMv3.1) for the atmospheric oxidation of aromatic hydrocarbons. *Atmos. Chem. Phys.* **2005,** *5,* (3), 641-664.

75. Jenkin, M. E.; Saunders, S. M.; Pilling, M. J., The tropospheric degradation of volatile organic compounds: a protocol for mechanism development. *Atmospheric Environment* **1997,** *31,* (1), 81-104.

76. USEPA *Estimation Programs Interface Suite™ for Microsoft® Windows* United States Environmental Protection Agency: Washington, DC, USA, 2009.

77. Fredenslund, A.; Jones, R. L.; Prausnitz, J. M., Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE Journal* **1975,** *21,* (6), 1086-1099.

78. Chacon-Madrid, H.; Donahue, N., Fragmentation vs. functionalization: chemical aging and organic aerosol formation. *Atmospheric Chemistry and Physics* **2011,** *11,* (20), 10553-10563.

79. Odum, J. R.; Jungkamp, T.; Griffin, R. J.; Forstner, H.; Flagan, R. C.; Seinfeld, J. H., Aromatics, reformulated gasoline, and atmospheric organic aerosol formation. *Environmental Science & Technology* **1997,** *31,* (7), 1890-1897.

80. 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. *Applied Mechanics Reviews* **2006,** *59,* (2), 51-77.

81. Carter, W. P.; Heo, G., Development of revised SAPRC aromatics mechanisms. *Atmospheric environment* **2013,** *77*, 404-414.

82. Morris, R. E.; Koo, B.; Guenther, A.; Yarwood, G.; McNally, D.; Tesche, T. W.; Tonnesen, G.; Boylan, J.; Brewer, P., Model sensitivity evaluation for organic carbon using two multi-pollutant air quality models that simulate regional haze in the southeastern United States. *Atmospheric Environment* **2006,** *40*, (26), 4960-4972.

83. Ying, Q.; Cureño, I. V.; Chen, G.; Ali, S.; Zhang, H.; Malloy, M.; Bravo, H. A.; Sosa, R., Impacts of Stabilized Criegee Intermediates, surface uptake processes and higher aromatic secondary organic aerosol yields on predicted PM2.5 concentrations in the Mexico City Metropolitan Zone. *Atmospheric Environment* **2014,** *94*, 438-447.

84. Ying, Q.; Li, J.; Kota, S. H., Significant contributions of isoprene to summertime secondary organic aerosol in eastern United States. *Environmental science & technology* **2015,** *49*, (13), 7834-7842.

85. Hu, J.; Chen, J.; Ying, Q.; Zhang, H., One-year simulation of ozone and particulate matter in China using WRF/CMAQ modeling system. *Atmos. Chem. Phys.* **2016,** *16*, (16), 10333-10350.

86. Tong, D.; Bowker, G.; He, S.; Byun, D.; Mathur, R.; Gillette, D., Development of a Windblown Dust Module within the Community Multi-scale Air Quality (CMAQ) Model: Description and Preliminary Applications in the Continental United States, submitted. *J. Geophys. Res* **2011**.

87. 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). *Atmospheric Chemistry and Physics* **2006,** *6*, 3181-3210.

88. Hu, J.; Wang, P.; Ying, Q.; Zhang, H.; Chen, J.; Ge, X.; Li, X.; Jiang, J.; Wang, S.; Zhang, J.; Zhao, Y.; Zhang, Y., Modeling biogenic and anthropogenic secondary organic aerosol in China. *Atmos. Chem. Phys.* **2017,** *17*, (1), 77-92.

89. Shi, Z.; Li, J.; Huang, L.; Wang, P.; Wu, L.; Ying, Q.; Zhang, H.; Lu, L.; Liu, X.; Liao, H.; Hu, J., Source apportionment of fine particulate matter in China in 2013 using a source-oriented chemical transport model. *Science of The Total Environment* **2017,** *601-602*, 1476-1487.

90. Li, J.; Zhang, H.; Ying, Q.; Wu, Z.; Zhang, Y.; Wang, X.; Li, X.; Sun, Y.; Hu, M.; Zhang, Y.; Hu, J., Impacts of water partitioning and polarity of organic compounds on secondary organic aerosol over eastern China. *Atmos. Chem. Phys.* **2020,** *20*, (12), 7291-7306.

91. Guo, S.; Hu, M.; Guo, Q.; Zhang, X.; Zheng, M.; Zheng, J.; Chang, C. C.; Schauer, J. J.; Zhang, R., Primary sources and secondary formation of organic aerosols in Beijing, China. *Environmental science & technology* **2012,** *46*, (18), 9846-9853.

92. Gao, Y.; Zhang, Z.; Yao, W.; Ying, Q.; Long, C.; Fu, X., Forecasting the cumulative number of COVID-19 deaths in China: a Boltzmann function-based modeling study. *Infection Control & Hospital Epidemiology* **2020,** *41*, (7), 841-843.

93. He, X.; Wang, Q.; Huang, X. H. H.; Huang, D. D.; Zhou, M.; Qiao, L.; Zhu, S.; Ma, Y.-g.; Wang, H.-l.; Li, L.; Huang, C.; Xu, W.; Worsnop, D. R.; Goldstein, A. H.; Yu, J. Z., Hourly measurements of organic molecular markers in urban Shanghai, China: Observation of enhanced formation of secondary organic aerosol during particulate matter episodic periods. *Atmospheric Environment* **2020,** *240*, 117807.

94. Feng, J.; Li, M.; Zhang, P.; Gong, S.; Zhong, M.; Wu, M.; Zheng, M.; Chen, C.; Wang, H.; Lou, S., Investigation of the sources and seasonal variations of secondary organic aerosols in PM2.5 in Shanghai with organic tracers. *Atmospheric Environment* **2013,** *79*, 614-622.

95. Ding, X.; Wang, X.-M.; Gao, B.; Fu, X.-X.; He, Q.-F.; Zhao, X.-Y.; Yu, J.-Z.; Zheng, M., Tracer-based estimation of secondary organic carbon in the Pearl River Delta, south China. **2012,** *117*, (D5).

96. Tang, R.; Wu, Z.; Li, X.; Wang, Y.; Shang, D.; Xiao, Y.; Li, M.; Zeng, L.; Wu, Z.; Hallquist, M.; Hu, M.; Guo, S., Primary and secondary organic aerosols in summer 2016 in Beijing. *Atmos. Chem. Phys.* **2018,** *18*, (6), 4055-4068.

97. Wang, Y.; Pavuluri, C. M.; Fu, P.; Li, P.; Dong, Z.; Xu, Z.; Ren, H.; Fan, Y.; Li, L.; Zhang, Y.-L.; Liu, C.-Q., Characterization of Secondary Organic Aerosol Tracers over Tianjin, North China during Summer to Autumn. *ACS Earth and Space Chemistry* **2019,** *3*, (10), 2339-2352.

98. Wang, P.; Chen, Y.; Hu, J.; Zhang, H.; Ying, Q., Source apportionment of summertime ozone in China using a source-oriented chemical transport model. *Atmospheric Environment* **2019,** *211*, 79-90.

99. Finewax, Z.; de Gouw, J. A.; Ziemann, P. J., Identification and Quantification of 4-Nitrocatechol Formed from OH and NO3 Radical-Initiated Reactions of Catechol in Air in the Presence of NO x: Implications for Secondary Organic Aerosol Formation from Biomass Burning. *Environmental science & technology* **2018,** *52*, (4), 1981-1989.

100. Contini, D.; Vecchi, R.; Viana, M., Carbonaceous aerosols in the atmosphere. In Multidisciplinary Digital Publishing Institute: 2018.

101. Putaud, J.-P.; Van Dingenen, R.; Alastuey, A.; Bauer, H.; Birmili, W.; Cyrys, J.; Flentje, H.; Fuzzi, S.; Gehrig, R.; Hansson, H.-C., A European aerosol phenomenology–3: Physical and chemical characteristics of particulate matter from 60 rural, urban, and kerbside sites across Europe. *Atmospheric Environment* **2010,** *44*, (10), 1308-1320.

102. Kanakidou, M.; Seinfeld, J.; Pandis, S.; Barnes, I.; Dentener, F. J.; Facchini, M. C.; Dingenen, R. V.; Ervens, B.; Nenes, A.; Nielsen, C., Organic aerosol and global climate modelling: a review. *Atmospheric Chemistry and Physics* **2005,** *5*, (4), 1053-1123.

103. Novakov, T.; Hegg, D. A.; Hobbs, P. V., Airborne measurements of carbonaceous aerosols on the East Coast of the United States. *Journal of Geophysical Research: Atmospheres* **1997,** *102*, (D25), 30023-30030.

104. Cao, J.; Lee, S.; Chow, J. C.; Watson, J. G.; Ho, K.; Zhang, R.; Jin, Z.; Shen, Z.; Chen, G.; Kang, Y., Spatial and seasonal distributions of carbonaceous aerosols over China. *Journal of Geophysical Research: Atmospheres* **2007,** *112,* (D22).

105. Ye, B.; Ji, X.; Yang, H.; Yao, X.; Chan, C. K.; Cadle, S. H.; Chan, T.; Mulawa, P. A., Concentration and chemical composition of PM2. 5 in Shanghai for a 1-year period. *Atmospheric Environment* **2003,** *37*, (4), 499-510.

106. Hand, J.; Copeland, S.; Day, D.; Dillner, A.; Indresand, H.; Malm, W.; McDade, C.; Moore, C.; Pitchford, M.; Schichtel, B., Spatial and Seasonal Patterns and Temporal Variability of Haze and its Constituents in the United States: Report V. June 2011. *IMPROVE Reports* **2011**.

107. Goldstein, A. H.; Koven, C. D.; Heald, C. L.; Fung, I. Y., Biogenic carbon and anthropogenic pollutants combine to form a cooling haze over the southeastern United States. *Proceedings of the National Academy of Sciences* **2009,** *106*, (22), 8835-8840.

108. Highwood, E. J.; Kinnersley, R. P., When smoke gets in our eyes: The multiple impacts of atmospheric black carbon on climate, air quality and health. *Environment international* **2006,** *32*, (4), 560-566.

109. Turpin, B. J.; Huntzicker, J. J., Secondary formation of organic aerosol in the Los Angeles Basin: a descriptive analysis of organic and elemental carbon concentrations. *Atmospheric Environment. Part A. General Topics* **1991,** *25*, (2), 207-215.

110. Cao, J.; Lee, S.; Ho, K.; Zou, S.; Fung, K.; Li, Y.; Watson, J. G.; Chow, J. C., Spatial and seasonal variations of atmospheric organic carbon and elemental carbon in Pearl River Delta Region, China. *Atmospheric Environment* **2004,** *38*, (27), 4447-4456.

111. Castro, L.; Pio, C.; Harrison, R. M.; Smith, D., Carbonaceous aerosol in urban and rural European atmospheres: estimation of secondary organic carbon concentrations. *Atmospheric Environment* **1999,** *33*, (17), 2771-2781.

112. Wu, C.; Yu, J. Z., Determination of primary combustion source organic carbon-to-elemental carbon (OC/EC) ratio using ambient OC and EC measurements: secondary OC-EC correlation minimization method. *Atmospheric Chemistry and Physics* **2016,** *16*, (8), 5453-5465.

113. Hu, D.; Bian, Q.; Lau, A. K.; Yu, J. Z., Source apportioning of primary and secondary organic carbon in summer PM2. 5 in Hong Kong using positive matrix factorization of secondary and primary organic tracer data. *Journal of Geophysical Research: Atmospheres* **2010,** *115*, (D16).

114. Yuan, B.; Shao, M.; de Gouw, J.; Parrish, D. D.; Lu, S.; Wang, M.; Zeng, L.; Zhang, Q.; Song, Y.; Zhang, J., Volatile organic compounds (VOCs) in urban air: How chemistry affects the

interpretation of positive matrix factorization (PMF) analysis. *Journal of Geophysical Research: Atmospheres* **2012,** *117*, (D24).

115. 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. **2006**.

116. Odum, J. R.; Hoffmann, T.; Bowman, F.; Collins, D.; Flagan, R. C.; Seinfeld, J. H., Gas/particle partitioning and secondary organic aerosol yields. *Environmental Science & Technology* **1996,** *30*, (8), 2580-2585.

117. Morris, R. E.; Koo, B.; Guenther, A.; Yarwood, G.; McNally, D.; Tesche, T.; Tonnesen, G.; Boylan, J.; Brewer, P., Model sensitivity evaluation for organic carbon using two multi-pollutant air quality models that simulate regional haze in the southeastern United States. *Atmospheric Environment* **2006,** *40*, (26), 4960-4972.

118. Wang, Q.; He, X.; Zhou, M.; Huang, D. D.; Qiao, L.; Zhu, S.; Ma, Y.-g.; Wang, H.-l.; Li, L.; Huang, C., Hourly Measurements of Organic Molecular Markers in Urban Shanghai, China: Primary Organic Aerosol Source Identification and Observation of Cooking Aerosol Aging. *ACS Earth and Space Chemistry* **2020,** *4*, (9), 1670-1685.

119. Hu, J.; Ying, Q.; Chen, J.; Mahmud, A.; Zhao, Z.; Chen, S.-H.; Kleeman, M. J., Particulate air quality model predictions using prognostic vs. diagnostic meteorology in central California. *Atmospheric Environment* **2010,** *44*, (2), 215-226.

120. Zhang, H.; Li, J.; Ying, Q.; Yu, J. Z.; Wu, D.; Cheng, Y.; He, K.; Jiang, J., Source apportionment of PM2. 5 nitrate and sulfate in China using a source-oriented chemical transport model. *Atmospheric environment* **2012,** *62*, 228-242.

121. Ohara, T.; Akimoto, H.; Kurokawa, J.-i.; Horii, N.; Yamaji, K.; Yan, X.; Hayasaka, T., An Asian emission inventory of anthropogenic emission sources for the period 1980? 2020. **2007**.

122. Wang, P.; Ying, Q.; Zhang, H.; Hu, J.; Lin, Y.; Mao, H., Source apportionment of secondary organic aerosol in China using a regional source-oriented chemical transport model and two emission inventories. *Environ. Pollut.* **2018,** *237*, 756-766.

123. Guenther, A.; Karl, T.; Harley, P.; Wiedinmyer, C.; Palmer, P.; Geron, C., Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature). **2006**.

124. Emery, C.; Tai, E.; Yarwood, G., Enhanced meteorological modeling and performance evaluation for two Texas ozone episodes. *Prepared for the Texas natural resource conservation commission, by ENVIRON International Corporation* **2001**.

125. Hu, J.; Chen, J.; Ying, Q.; Zhang, H., One-year simulation of ozone and particulate matter in China using WRF/CMAQ modeling system. *1foldr Import 2019-10-08 Batch 1* **2016**.

126. Hu, J.; Wu, L.; Zheng, B.; Zhang, Q.; He, K.; Chang, Q.; Li, X.; Yang, F.; Ying, Q.; Zhang, H., Source contributions and regional transport of primary particulate matter in China. *Environ. Pollut.* **2015,** *207*, 31-42.

127. Wang, L.; Jang, C.; Zhang, Y.; Wang, K.; Zhang, Q.; Streets, D.; Fu, J.; Lei, Y.; Schreifels, J.; He, K., Assessment of air quality benefits from national air pollution control policies in China. Part I: Background, emission scenarios and evaluation of meteorological predictions. *Atmospheric Environment* **2010,** *44,* (28), 3442-3448.

128. Wang, T.; Wang, P.; Theys, N.; Tong, D.; Hendrick, F.; Zhang, Q.; Van Roozendael, M., Spatial and temporal changes in SO 2 regimes over China in the recent decade and the driving mechanism. *Atmospheric Chemistry and Physics* **2018,** *18*, (24), 18063-18078.

129. Saikawa, E.; Kim, H.; Zhong, M.; Avramov, A.; Zhao, Y.; Janssens-Maenhout, G.; Jurokawa, J.; Klimont, Z.; Wagner, F.; Naik, V., Comparison of emissions inventories of anthropogenic air pollutants and greenhouse gases in China. *Atmospheric Chemistry and Physics* **2017,** *17*, (10), 6393-6421.

130. Huang, X.; Zhou, L.; Ding, A.; Qi, X.; Nie, W.; Wang, M.; Chi, X.; Petäjä, T.; Kerminen, V.-M.; Roldin, P., Comprehensive modelling study on observed new particle formation at the SORPES station in Nanjing, China. *Atmospheric Chemistry and Physics* **2016**.

131. Hesterberg, T., Bootstrap. *Wiley Interdisciplinary Reviews: Computational Statistics* **2011,** *3*, (6), 497-526.

132. 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. *Journal of geophysical research: atmospheres* **2008,** *113*, (D15).

133. Qiu, X.; Wang, S.; Ying, Q.; Duan, L.; Xing, J.; Cao, J.; Wu, D.; Li, X.; Chengzhi, X.; Yan, X., Importance of Wintertime Anthropogenic Glyoxal and Methylglyoxal Emissions in Beijing and Implications for Secondary Organic Aerosol Formation in Megacities. *Environmental Science & Technology* **2020,** *54*, (19), 11809-11817.

134. Ying, Q.; Cureño, I. V.; Chen, G.; Ali, S.; Zhang, H.; Malloy, M.; Bravo, H. A.; Sosa, R., Impacts of Stabilized Criegee Intermediates, surface uptake processes and higher aromatic secondary organic aerosol yields on predicted PM2. 5 concentrations in the Mexico City Metropolitan Zone. *Atmospheric environment* **2014,** *94*, 438-447.

135. Shen, R.-Q.; Ding, X.; He, Q.-F.; Cong, Z.-Y.; Wang, X.-M., Seasonal variation of secondary organic aerosol tracers in Central Tibetan Plateau. *Atmospheric Chemistry and Physics* **2015,** *15*, (15), 8781-8793.

136. Ding, X.; He, Q. F.; Shen, R. Q.; Yu, Q. Q.; Wang, X. M., Spatial distributions of secondary organic aerosols from isoprene, monoterpenes, β-caryophyllene, and aromatics over China during summer. *Journal of Geophysical Research: Atmospheres* **2014,** *119*, (20), 11,877-11,891.

137. Saha, P. K.; Grieshop, A. P., Exploring divergent volatility properties from yield and thermodenuder measurements of secondary organic aerosol from α-pinene ozonolysis. *Environmental Science & Technology* **2016,** *50*, (11), 5740-5749.

138. Griffin, R. J.; Cocker III, D. R.; Flagan, R. C.; Seinfeld, J. H., Organic aerosol formation from the oxidation of biogenic hydrocarbons. *Journal of Geophysical Research: Atmospheres* **1999,** *104*, (D3), 3555-3567.

139. Ding, X.; Wang, X. M.; Gao, B.; Fu, X. X.; He, Q. F.; Zhao, X. Y.; Yu, J. Z.; Zheng, M., Tracer-based estimation of secondary organic carbon in the Pearl River Delta, south China. *Journal of Geophysical Research: Atmospheres* **2012,** *117*, (D5).

140. Lewandowski, M.; Jaoui, M.; Offenberg, J. H.; Kleindienst, T. E.; Edney, E. O.; Sheesley, R. J.; Schauer, J. J., Primary and secondary contributions to ambient PM in the midwestern United States. *Environmental Science & Technology* **2008,** *42*, (9), 3303-3309.

141. Hu, D.; Bian, Q.; Li, T. W.; Lau, A. K.; Yu, J. Z., Contributions of isoprene, monoterpenes, β-caryophyllene, and toluene to secondary organic aerosols in Hong Kong during the summer of 2006. *Journal of Geophysical Research: Atmospheres* **2008,** *113*, (D22).

142. Ding, X.; Wang, X.-M.; Zheng, M., The influence of temperature and aerosol acidity on biogenic secondary organic aerosol tracers: Observations at a rural site in the central Pearl River Delta region, South China. *Atmospheric Environment* **2011,** *45*, (6), 1303-1311.

143. Carter, W. P. L.; Heo, G., Development of revised SAPRC aromatics mechanisms. *Atmospheric Environment* **2013,** *77*, 404-414.

144. Hildebrandt, L.; Donahue, N. M.; Pandis, S. N., High formation of secondary organic aerosol from the photo-oxidation of toluene. *Atmospheric Chemistry and Physics* **2009,** *9*, (9), 2973-2986.

145. Ahlberg, E.; Falk, J.; Eriksson, A.; Holst, T.; Brune, W. H.; Kristensson, A.; Roldin, P.; Svenningsson, B., Secondary organic aerosol from VOC mixtures in an oxidation flow reactor. *Atmospheric Environment* **2017,** *161*, 210-220.

146. Pankow, J. F., An absorption model of gas/particle partitioning of organic compounds in the atmosphere. *Atmospheric Environment* **1994,** *28*, (2), 185-188.

147. Gao, Y.; Wang, H.; Zhang, X.; Jing, S. a.; Peng, Y.; Qiao, L.; Zhou, M.; Huang, D. D.; Wang, Q.; Li, X.; Li, L.; Feng, J.; Ma, Y.; Li, Y., Estimating Secondary Organic Aerosol Production from Toluene Photochemistry in a Megacity of China. *Environmental Science & Technology* **2019,** *53*, (15), 8664-8671.

148. Donahue, N. M.; Robinson, A.; Stanier, C.; Pandis, S., Coupled partitioning, dilution, and chemical aging of semivolatile organics. *Environmental science & technology* **2006,** *40*, (8), 2635-2643.

149. Williams, B. J.; Goldstein, A. H.; Kreisberg, N. M.; Hering, S. V., In situ measurements of gas/particle-phase transitions for atmospheric semivolatile organic compounds. *Proceedings of the National Academy of Sciences* **2010,** *107*, (15), 6676-6681.

150. Moller, B.; Rarey, J.; Ramjugernath, D., Estimation of the vapour pressure of non-electrolyte organic compounds via group contributions and group interactions. *Journal of Molecular Liquids* **2008,** *143*, (1), 52-63.

151. Nannoolal, Y.; Rarey, J.; Ramjugernath, D.; Cordes, W., Estimation of pure component properties: Part 1. Estimation of the normal boiling point of non-electrolyte organic compounds via group contributions and group interactions. *Fluid Phase Equilibria* **2004,** *226*, 45-63.

152. Al-Naiema, I. M.; Stone, E. A., Evaluation of anthropogenic secondary organic aerosol tracers from aromatic hydrocarbons. *Atmospheric Chemistry and Physics* **2017,** *17*, (3), 2053-2065.

153. Barley, M. H.; McFiggans, G., The critical assessment of vapour pressure estimation methods for use in modelling the formation of atmospheric organic aerosol. *Atmos. Chem. Phys.* **2010,** *10*, (2), 749-767.

154. US EPA *Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.11*, United States Environmental Protection Agency: Washington, DC, USA, 2012.

155. Nannoolal, Y.; Rarey, J.; Ramjugernath, D., Estimation of pure component properties: Part 3. Estimation of the vapor pressure of non-electrolyte organic compounds via group contributions and group interactions. *Fluid Phase Equilibria* **2008,** *269*, (1), 117-133.

156. Stein, S. E.; Brown, R. L., Estimation of normal boiling points from group contributions. *Journal of Chemical Information and Computer Sciences* **1994,** *34*, (3), 581-587.

157. Myrdal, P. B.; Yalkowsky, S. H., Estimating Pure Component Vapor Pressures of Complex Organic Molecules. *Industrial & Engineering Chemistry Research* **1997,** *36*, (6), 2494-2499.

158. Hildebrandt, L.; Donahue, N. M.; Pandis, S. N., High formation of secondary organic aerosol from the photo-oxidation of toluene. *Atmospheric Chemistry & Physics Discussions* **2009,** *9*, (1).

159. Edney, E. O.; Kleindienst, T. E.; Jaoui, M.; Lewandowski, M.; Offenberg, J. H.; Wang, W.; Claeys, M., Formation of 2-methyl tetrols and 2-methylglyceric acid in secondary organic aerosol from laboratory irradiated isoprene/NOX/SO2/air mixtures and their detection in ambient PM2.5 samples collected in the eastern United States. *Atmospheric Environment* **2005,** *39*, (29), 5281-5289.

160. Xie, M.; Chen, X.; Hays, M. D.; Lewandowski, M.; Offenberg, J.; Kleindienst, T. E.; Holder, A. L., Light Absorption of Secondary Organic Aerosol: Composition and Contribution of Nitroaromatic Compounds. *Environmental Science & Technology* **2017,** *51*, (20), 11607-11616.

161. Zhang, X.; Cappa, C. D.; Jathar, S. H.; McVay, R. C.; Ensberg, J. J.; Kleeman, M. J.; Seinfeld, J. H., Influence of vapor wall loss in laboratory chambers on yields of secondary organic aerosol. *Proceedings of the National Academy of Sciences* **2014,** *111*, (16), 5802-5807.

162. Epstein, S. A.; Riipinen, I.; Donahue, N. M., A semiempirical correlation between enthalpy of vaporization and saturation concentration for organic aerosol. *Environmental science & technology* **2010,** *44*, (2), 743-748.

163. Li, J.; Zhang, H.; Ying, Q.; Wu, Z.; Zhang, Y.; Wang, X.; Li, X.; Sun, Y.; Hu, M.; Zhang, Y., Impacts of water partitioning and polarity of organic compounds on secondary organic aerosol over eastern China. *Atmospheric Chemistry and Physics* **2020,** *20*, (12), 7291-7306.

164. Pankow, J. F.; Marks, M. C.; Barsanti, K. C.; Mahmud, A.; Asher, W. E.; Li, J.; Ying, Q.; Jathar, S. H.; Kleeman, M. J., Molecular view modeling of atmospheric organic particulate matter: Incorporating molecular structure and co-condensation of water. *Atmospheric Environment* **2015,** *122*, 400-408.

165. Ding, X.; Zhang, Y.-Q.; He, Q.-F.; Yu, Q.-Q.; Wang, J.-Q.; Shen, R.-Q.; Song, W.; Wang, Y.-S.; Wang, X.-M., Significant increase of aromatics-derived secondary organic aerosol during fall to winter in China. *Environmental science & technology* **2017,** *51*, (13), 7432-7441.

166. Yuan, Q.; Lai, S.; Song, J.; Ding, X.; Zheng, L.; Wang, X.; Zhao, Y.; Zheng, J.; Yue, D.; Zhong, L., Seasonal cycles of secondary organic aerosol tracers in rural Guangzhou, Southern China: The importance of atmospheric oxidants. *Environmental Pollution* **2018,** *240*, 884-893.

167. Tang, R.; Wu, Z.; Li, X.; Wang, Y.; Shang, D.; Xiao, Y.; Li, M.; Zeng, L.; Wu, Z.; Hallquist, M., Primary and secondary organic aerosols in summer 2016 in Beijing. *Atmospheric Chemistry and Physics* **2018,** *18*, (6), 4055-4068.

168. Liu, T.; Hu, B.; Xu, X.; Hong, Y.; Zhang, Y.; Wu, X.; Xu, L.; Li, M.; Chen, Y.; Chen, X., Characteristics of PM2. 5-bound secondary organic aerosol tracers in a coastal city in Southeastern China: Seasonal patterns and pollution identification. *Atmospheric Environment* **2020,** *237*, 117710.

169. Kroll, J. H.; Ng, N. L.; Murphy, S. M.; Flagan, R. C.; Seinfeld, J. H., Secondary organic aerosol formation from isoprene photooxidation under high-NOx conditions. *Geophysical Research Letters* **2005,** *32*, (18).

170. Kleindienst, T. E.; Lewandowski, M.; Offenberg, J. H.; Jaoui, M.; Edney, E. O., Ozone-isoprene reaction: Re-examination of the formation of secondary organic aerosol. *Geophysical research letters* **2007,** *34*, (1).

171. Ng, N.; Kwan, A.; Surratt, J.; Chan, A.; Chhabra, P.; Sorooshian, A.; Pye, H. O.; Crounse, J.; Wennberg, P.; Flagan, R., Secondary organic aerosol (SOA) formation from reaction of

isoprene with nitrate radicals (NO 3). *Atmospheric Chemistry and Physics* **2008,** *8*, (14), 4117-4140.

172. Song, M.; Zhang, C.; Wu, H.; Mu, Y.; Ma, Z.; Zhang, Y.; Liu, J.; Li, X., The influence of OH concentration on SOA formation from isoprene photooxidation. *Science of the Total Environment* **2019,** *650*, 951-957.

173. Brownwood, B.; Turdziladze, A.; Hohaus, T.; Wu, R.; Mentel, T. F.; Carlsson, P. T.; Tsiligiannis, E.; Hallquist, M.; Andres, S.; Hantschke, L., Gas-particle partitioning and SOA yields of organonitrate products from NO3-initiated oxidation of isoprene under varied chemical regimes. *ACS Earth and Space Chemistry* **2021,** *5*, (4), 785-800.

174. Marais, E. A.; Jacob, D. J.; Jimenez, J. L.; Campuzano-Jost, P.; Day, D. A.; Hu, W.; Krechmer, J.; Zhu, L.; Kim, P. S.; Miller, C. C., Aqueous-phase mechanism for secondary organic aerosol formation from isoprene: application to the southeast United States and co-benefit of SO 2 emission controls. *Atmospheric Chemistry and Physics* **2016,** *16*, (3), 1603-1618.

175. Paulot, F.; Crounse, J. D.; Kjaergaard, H. G.; Kürten, A.; Clair, J. M. S.; Seinfeld, J. H.; Wennberg, P. O., Unexpected epoxide formation in the gas-phase photooxidation of isoprene. *science* **2009,** *325*, (5941), 730-733.

176. Qin, M.; Wang, X.; Hu, Y.; Ding, X.; Song, Y.; Li, M.; Vasilakos, P.; Nenes, A.; Russell, A. G., Simulating biogenic secondary organic aerosol during summertime in China. *Journal of Geophysical Research: Atmospheres* **2018,** *123*, (19), 11,100-11,119.

177. Xie, M.; Hannigan, M. P.; Barsanti, K. C., Gas/particle partitioning of 2-methyltetrols and levoglucosan at an urban site in Denver. *Environmental science & technology* **2014,** *48*, (5), 2835-2842.

178. Xie, Y.; Paulot, F.; Carter, W.; Nolte, C.; Luecken, D.; Hutzell, W.; Wennberg, P.; Cohen, R.; Pinder, R., Understanding the impact of recent advances in isoprene photooxidation on simulations of regional air quality. *Atmospheric Chemistry and Physics* **2013,** *13*, (16), 8439-8455.

179. Schwartz, S. E., Mass-transport considerations pertinent to aqueous phase reactions of gases in liquid-water clouds. In *Chemistry of multiphase atmospheric systems*, Springer: 1986; pp 415-471.

180. Hanson, D. R.; Ravishankara, A.; Solomon, S., Heterogeneous reactions in sulfuric acid aerosols: A framework for model calculations. *Journal of Geophysical Research: Atmospheres* **1994,** *99*, (D2), 3615-3629.

181. McNeill, V. F.; Woo, J. L.; Kim, D. D.; Schwier, A. N.; Wannell, N. J.; Sumner, A. J.; Barakat, J. M., Aqueous-phase secondary organic aerosol and organosulfate formation in atmospheric aerosols: a modeling study. *Environmental science & technology* **2012,** *46*, (15), 8075-8081.

182. Gharagheizi, F.; Eslamimanesh, A.; Mohammadi, A. H.; Richon, D., Representation and prediction of molecular diffusivity of nonelectrolyte organic compounds in water at infinite dilution using the artificial neural network-group contribution method. *Journal of Chemical & Engineering Data* **2011,** *56*, (5), 1741-1750.

183. Fountoukis, C.; Nenes, A., ISORROPIA II: a computationally efficient thermodynamic equilibrium model for K-Ca 2–Mg 2–NH. **2007**.

184. Lin, Y.-H.; Zhang, Z.; Docherty, K. S.; Zhang, H.; Budisulistiorini, S. H.; Rubitschun, C. L.; Shaw, S. L.; Knipping, E. M.; Edgerton, E. S.; Kleindienst, T. E., Isoprene epoxydiols as precursors to secondary organic aerosol formation: acid-catalyzed reactive uptake studies with authentic compounds. *Environmental science & technology* **2012,** *46*, (1), 250-258.

185. Hu, J.; Chen, J.; Ying, Q.; Zhang, H., One-year simulation of ozone and particulate matter in China using WRF/CMAQ modeling system. *Atmospheric Chemistry and Physics* **2016,** *16*, (16), 10333-10350.

186. Zhang, Y.; Zhang, R.; Yu, J.; Zhang, Z.; Yang, W.; Zhang, H.; Lyu, S.; Wang, Y.; Dai, W.; Wang, Y., Isoprene mixing ratios measured at twenty sites in China during 2012–2014: Comparison with model simulation. *Journal of Geophysical Research: Atmospheres* **2020,** *125*, (24), e2020JD033523.

187. Wang, P.; Schade, G.; Estes, M.; Ying, Q., Improved MEGAN predictions of biogenic isoprene in the contiguous United States. *Atmospheric Environment* **2017,** *148*, 337-351.

188. Aiken, A. C.; Decarlo, P. F.; Kroll, J. H.; Worsnop, D. R.; Huffman, J. A.; Docherty, K. S.; Ulbrich, I. M.; Mohr, C.; Kimmel, J. R.; Sueper, D., O/C and OM/OC ratios of primary, secondary, and ambient organic aerosols with high-resolution time-of-flight aerosol mass spectrometry. *Environmental science & technology* **2008,** *42*, (12), 4478-4485.

189. Yu, X.-Y.; Cary, R.; Laulainen, N. S., Primary and secondary organic carbon downwind of Mexico City. *Atmospheric Chemistry and Physics* **2009,** *9*, (18), 6793-6814.

190. Xin, J.; Wang, Y.; Pan, Y.; Ji, D.; Liu, Z.; Wen, T.; Wang, Y.; Li, X.; Sun, Y.; Sun, J., The campaign on atmospheric aerosol research network of China: CARE-China. *Bulletin of the American Meteorological Society* **2015,** *96*, (7), 1137-1155.

191. Zhang, B.; Shen, H.; Liu, P.; Guo, H.; Hu, Y.; Chen, Y.; Xie, S.; Xi, Z.; Skipper, T. N.; Russell, A. G., Significant contrasts in aerosol acidity between China and the United States. *Atmospheric Chemistry and Physics* **2021,** *21*, (10), 8341-8356.

192. Goto, D.; Takemura, T.; Nakajima, T., Importance of global aerosol modeling including secondary organic aerosol formed from monoterpene. *Journal of Geophysical Research: Atmospheres* **2008,** *113*, (D7).

193. Kelly, J. M.; Doherty, R. M.; O'Connor, F. M.; Mann, G. W., The impact of biogenic, anthropogenic, and biomass burning volatile organic compound emissions on regional and

seasonal variations in secondary organic aerosol. *Atmospheric Chemistry and Physics* **2018,** *18*, (10), 7393-7422.

194. Khan, M.; Jenkin, M.; Foulds, A.; Derwent, R.; Percival, C.; Shallcross, D., A modeling study of secondary organic aerosol formation from sesquiterpenes using the STOCHEM global chemistry and transport model. *Journal of Geophysical Research: Atmospheres* **2017,** *122*, (8), 4426-4439.

195. Yu, J.; Cocker, D. R.; Griffin, R. J.; Flagan, R. C.; Seinfeld, J. H., Gas-phase ozone oxidation of monoterpenes: Gaseous and particulate products. *Journal of Atmospheric Chemistry* **1999,** *34*, (2), 207-258.

196. Tsigaridis, K.; Krol, M.; Dentener, F.; Balkanski, Y.; Lathiere, J.; Metzger, S.; Hauglustaine, D.; Kanakidou, M., Change in global aerosol composition since preindustrial times. *Atmospheric Chemistry and Physics* **2006,** *6*, (12), 5143-5162.

197. Tilmes, S.; Hodzic, A.; Emmons, L.; Mills, M.; Gettelman, A.; Kinnison, D.; Park, M.; Lamarque, J. F.; Vitt, F.; Shrivastava, M., Climate forcing and trends of organic aerosols in the Community Earth System Model (CESM2). *Journal of Advances in Modeling Earth Systems* **2019,** *11*, (12), 4323-4351.

198. Qin, M.; Hu, Y.; Wang, X.; Vasilakos, P.; Boyd, C. M.; Xu, L.; Song, Y.; Ng, N. L.; Nenes, A.; Russell, A. G., Modeling biogenic secondary organic aerosol (BSOA) formation from monoterpene reactions with NO3: A case study of the SOAS campaign using CMAQ. *Atmospheric Environment* **2018,** *184*, 146-155.

199. Jenkin, M. E.; Shallcross, D. E.; Harvey, J. N., Development and application of a possible mechanism for the generation of cis-pinic acid from the ozonolysis of α-and β-pinene. *Atmospheric Environment* **2000,** *34*, (18), 2837-2850.

200. Colville, C. J.; Griffin, R. J., The roles of individual oxidants in secondary organic aerosol formation from Δ3-carene: 1. gas-phase chemical mechanism. *Atmospheric Environment* **2004,** *38*, (24), 4001-4012.

201. Jenkin, M., Modelling the formation and composition of secondary organic aerosol from α- and β-pinene ozonolysis using MCM v3. *Atmospheric Chemistry and Physics* **2004,** *4*, (7), 1741-1757.

202. Chen, J.; Griffin, R. J., Modeling secondary organic aerosol formation from oxidation of α-pinene, β-pinene, and d-limonene. *Atmospheric Environment* **2005,** *39*, (40), 7731-7744.

203. Capouet, M.; Müller, J. F.; Ceulemans, K.; Compernolle, S.; Vereecken, L.; Peeters, J., Modeling aerosol formation in alpha-pinene photo-oxidation experiments. *Journal of Geophysical Research: Atmospheres* **2008,** *113*, (D2).

204. Yan, Y.; Cabrera-Perez, D.; Lin, J.; Pozzer, A.; Hu, L.; Millet, D. B.; Porter, W. C.; Lelieveld, J., Global tropospheric effects of aromatic chemistry with the SAPRC-11 mechanism

implemented in GEOS-Chem version 9-02. *Geoscientific model development* **2019,** *12*, (1), 111-130.

205. Takekawa, H.; Takasu, S., Estimation of Parameters for Two-Product Model of Secondary Organic Aerosol (SOA) Generation by Monoterpene Ozonolysis under Lower SOA Concentration. *Earozoru Kenkyu* **2017,** *32*, (3), 176-187.

206. Mutzel, A.; Zhang, Y.; Böge, O.; Rodigast, M.; Kolodziejczyk, A.; Wang, X.; Herrmann, H., Importance of secondary organic aerosol formation of α-pinene, limonene, and m-cresol comparing day-and nighttime radical chemistry. *Atmospheric Chemistry and Physics* **2021,** *21*, (11), 8479-8498.

207. Hoffmann, T.; Odum, J. R.; Bowman, F.; Collins, D.; Klockow, D.; Flagan, R. C.; Seinfeld, J. H., Formation of organic aerosols from the oxidation of biogenic hydrocarbons. *Journal of Atmospheric Chemistry* **1997,** *26*, (2), 189-222.

208. Bilde, M.; Pandis, S. N., Evaporation rates and vapor pressures of individual aerosol species formed in the atmospheric oxidation of α-and β-pinene. *Environmental science & technology* **2001,** *35*, (16), 3344-3349.

209. Müller, L.; Reinnig, M.-C.; Naumann, K.; Saathoff, H.; Mentel, T.; Donahue, N.; Hoffmann, T., Formation of 3-methyl-1, 2, 3-butanetricarboxylic acid via gas phase oxidation of pinonic acid–a mass spectrometric study of SOA aging. *Atmospheric chemistry and physics* **2012,** *12*, (3), 1483-1496.

210. Zhang, Y.; Müller, L.; Winterhalter, R.; Moortgat, G.; Hoffmann, T.; Pöschl, U., Seasonal cycle and temperature dependence of pinene oxidation products, dicarboxylic acids and nitrophenols in fine and coarse air particulate matter. *Atmospheric Chemistry and Physics* **2010,** *10*, (16), 7859-7873.

211. Kostenidou, E.; Karnezi, E.; Kolodziejczyk, A.; Szmigielski, R.; Pandis, S. N., Physical and Chemical Properties of 3-Methyl-1,2,3-butanetricarboxylic Acid (MBTCA) Aerosol. *Environ. Sci. Technol.* **2017**, (52), 1150-1155.

212. Carlton, A. G.; Bhave, P. V.; Napelenok, S. L.; Edney, E. O.; Sarwar, G.; Pinder, R. W.; Pouliot, G. A.; Houyoux, M., Model representation of secondary organic aerosol in CMAQv4. 7. *Environmental science & technology* **2010,** *44*, (22), 8553-8560.

213. Wang, Q.; He, X.; Huang, X. H.; Griffith, S. M.; Feng, Y.; Zhang, T.; Zhang, Q.; Wu, D.; Yu, J. Z., Impact of secondary organic aerosol tracers on tracer-based source apportionment of organic carbon and PM2. 5: A case study in the Pearl River Delta, China. *ACS Earth and Space Chemistry* **2017,** *1*, (9), 562-571.

214. He, X. Speciation of atmospheric organic carbon aerosol: From offline to online measurement. Ph.D Dissertation, Hong Kong University of Science & Technology, 2020.

215. Zhang, H.; Zhang, Y.; Huang, Z.; Acton, W. J. F.; Wang, Z.; Nemitz, E.; Langford, B.; Mullinger, N.; Davison, B.; Shi, Z., Vertical profiles of biogenic volatile organic compounds as observed online at a tower in Beijing. *Journal of Environmental Sciences* **2020,** *95*, 33-42.

216. Barletta, B.; Meinardi, S.; Simpson, I. J.; Zou, S.; Rowland, F. S.; Blake, D. R., Ambient mixing ratios of nonmethane hydrocarbons (NMHCs) in two major urban centers of the Pearl River Delta (PRD) region: Guangzhou and Dongguan. *Atmospheric Environment* **2008,** *42*, (18), 4393-4408.

217. Li, K.; Li, J.; Tong, S.; Wang, W.; Huang, R.-J.; Ge, M., Characteristics of wintertime VOCs in suburban and urban Beijing: concentrations, emission ratios, and festival effects. *Atmospheric Chemistry and Physics* **2019,** *19*, (12), 8021-8036.

218. Li, K.; Li, J.; Wang, W.; Tong, S.; Liggio, J.; Ge, M., Evaluating the effectiveness of joint emission control policies on the reduction of ambient VOCs: Implications from observation during the 2014 APEC summit in suburban Beijing. *Atmospheric Environment* **2017,** *164*, 117-127.

219. Chen, X.; Yu, J. Z., Measurement of organic mass to organic carbon ratio in ambient aerosol samples using a gravimetric technique in combination with chemical analysis. *Atmospheric Environment* **2007,** *41*, (39), 8857-8864.

220. Lopez-Hilfiker, F.; Mohr, C.; D'ambro, E.; Lutz, A.; Riedel, T.; Gaston, C.; Iyer, S.; Zhang, Z.; Gold, A.; Surratt, J., Molecular composition and volatility of organic aerosol in the Southeastern US: implications for IEPOX derived SOA. *Environmental science & technology* **2016,** *50*, (5), 2200-2209.

221. Fahey, K. M.; Carlton, A. G.; Pye, H. O.; Baek, J.; Hutzell, W. T.; Stanier, C. O.; Baker, K. R.; Appel, K. W.; Jaoui, M.; Offenberg, J. H., A framework for expanding aqueous chemistry in the Community Multiscale Air Quality (CMAQ) model version 5.1. *Geoscientific model development* **2017,** *10*, (4), 1587-1605.

222. Wang, R.; Huang, Y.; Cao, G., Heterogeneous oxidation of isoprene SOA and toluene SOA tracers by ozone. *Chemosphere* **2020,** *249*, 126258.

223. Li, J.; Xie, S. D.; Zeng, L. M.; Li, L. Y.; Li, Y. Q.; Wu, R. R., Characterization of ambient volatile organic compounds and their sources in Beijing, before, during, and after Asia-Pacific Economic Cooperation China 2014. *Atmos. Chem. Phys.* **2015,** *15*, (14), 7945-7959.

224. Song, M.; Tan, Q.; Feng, M.; Qu, Y.; Liu, X.; An, J.; Zhang, Y., Source Apportionment and Secondary Transformation of Atmospheric Nonmethane Hydrocarbons in Chengdu, Southwest China. *Journal of Geophysical Research: Atmospheres* **2018,** *123*, (17), 9741-9763.

225. Han, D.; Wang, Z.; Cheng, J.; Wang, Q.; Chen, X.; Wang, H., Volatile organic compounds (VOCs) during non-haze and haze days in Shanghai: characterization and secondary organic

aerosol (SOA) formation. *Environmental Science and Pollution Research* **2017,** *24*, (22), 18619-18629.

226. Zhang, Y.; Wang, X.; Barletta, B.; Simpson, I. J.; Blake, D. R.; Fu, X.; Zhang, Z.; He, Q.; Liu, T.; Zhao, X.; Ding, X., Source attributions of hazardous aromatic hydrocarbons in urban, suburban and rural areas in the Pearl River Delta (PRD) region. *Journal of Hazardous Materials* **2013,** *250-251*, 403-411.

227. Xia, L.; Cai, C.; Zhu, B.; An, J.; Li, Y.; Li, Y., Source apportionment of VOCs in a suburb of Nanjing, China, in autumn and winter. *Journal of Atmospheric Chemistry* **2014,** *71*, (3), 175-193.

228. Yun, H.; Wang, W.; Wang, T.; Xia, M.; Yu, C.; Wang, Z.; Poon, S. C.; Yue, D.; Zhou, Y., Nitrate formation from heterogeneous uptake of dinitrogen pentoxide during a severe winter haze in southern China. *Atmospheric chemistry and physics* **2018**.

229. Jia, L.; Xu, Y.; Ge, M.; Du, L.; Zhuang, G., Smog chamber studies of ozone formation potentials for isopentane. *Chinese Science Bulletin* **2009,** *54*, (24), 4624-4632.

230. Li, L.; Tang, P.; Nakao, S.; Cocker III, D. R., Impact of molecular structure on secondary organic aerosol formation from aromatic hydrocarbon photooxidation under low-NO x conditions. *Atmospheric Chemistry and Physics* **2016,** *16*, (17), 10793-10808.

231. Chan, A. W. H.; Kautzman, K.; Chhabra, P.; Surratt, J.; Chan, M.; Crounse, J.; Kürten, A.; Wennberg, P.; Flagan, R.; Seinfeld, J., Secondary organic aerosol formation from photooxidation of naphthalene and alkylnaphthalenes: implications for oxidation of intermediate volatility organic compounds (IVOCs). *Atmospheric Chemistry and Physics* **2009,** *9*, (9), 3049-3060.

232. Capouet, M.; Müller, J.-F., A group contribution method for estimating the vapour pressures of α-pinene oxidation products. *Atmospheric Chemistry and Physics* **2006,** *6*, (6), 1455-1467.

233. Zhang, X.; McVay, R. C.; Huang, D. D.; Dalleska, N. F.; Aumont, B.; Flagan, R. C.; Seinfeld, J. H., Formation and evolution of molecular products in α-pinene secondary organic aerosol. *Proceedings of the National Academy of Sciences* **2015,** *112*, (46), 14168-14173.

234. Li, Y.; Chen, Q.; Guzman, M.; Chan, C. K.; Martin, S., Second-generation products contribute substantially to the particle-phase organic material produced by β-caryophyllene ozonolysis. *Atmospheric Chemistry and Physics* **2011,** *11*, (1), 121-132.

235. Alfarra, M.; Hamilton, J.; Wyche, K.; Good, N.; Ward, M.; Carr, T.; Barley, M.; Monks, P.; Jenkin, M.; Lewis, A., The effect of photochemical ageing and initial precursor concentration on the composition and hygroscopic properties of β-caryophyllene secondary organic aerosol. *Atmospheric Chemistry and Physics* **2012,** *12*, (14), 6417-6436.

236. Cheng, Y.; Chow, J. C.; Watson, J. G.; Zhou, J.; Liu, S.; Cao, J., Decreasing concentrations of carbonaceous aerosols in China from 2003 to 2013. *Scientific reports* **2021,** *11*, (1), 1-10.

APPENDIX A

Table S2-1 Initial conditions of the toluene oxidation chamber experiments simulated in this study

Experiments[^]	Toluene (ppb)	NO_x (ppb)	H_2O_2 (ppm)	OH (10^6 cm^3)	Seed[#] (μg m^{-3})	T (^{o}C)	Lights
H-1	940	1300	49	3.1	61	20	30%
H-2	380	720	72	3.5	64	20	100%
H-3	950	570	91	2.6	113	20	100%
H-4	190	320	87	1.9	50	18	100%
H-5	180	270	57	1.1	42	12	100%
H-6	200	430	85	2.4	36	31	100%
L-1	380	<5[$]	88	1.6	59	20	100%
L-2	270	<5	40	1	57	11	100%
L-3	180	<5	46	0.6	45	11	100%
L-4	200	<5	42	1.2	60	32	100%
L-5	570	<5	87	1.4	61	21	100%
L-6	570	<5	45	1.6	47	21	100%

[^] H-1 to H-6 are high-NO_x experiments and L1 to L6 are low-NO_x experiments.

[*] Experiment H-1 was conducted with ~30% of the black lights and the optimal J_{NO2} was scaled accordingly to verify the MCM mechanism. The SOA mass yields at C_{OM} = 10 and 20 μg m^{-3} with the new mechanism (0.04 and 0.07) were slightly higher than those from the original mechanism (0.03 and 0.06), and both are slightly lower than the measured yields in Hildebrandt et al.'s study (0.08 and 0.12). This experiment was not included in the simulations to estimate J_{NO2} or in the 2-product parameter fitting.

[$] Use 3 ppb in the model simulations

[#] included in the model as $(NH_4)_2SO_4$.

Table S2-2 Mean fractional bias and mean fraction error of predicted hourly $PM_{2.5}$ at 5 United States Consulates in China in January and July 2013

	January			July		
Site	No. of data	MFB*	MFE*	No. of data	MFB	MFE
Beijing	611	-0.528	0.641	609	-0.545	0.688
Shanghai	596	-0.564	0.611	593	-0.255	0.455
Guangzhou	613	-0.310	0.484	556	-0.298	0.565
Chengdu	606	-0.158	0.408	592	0.061	0.439
Shenyang	-	-	-	514	0.054	0.605

[*] MFE: Mean fractional bias. $\boldsymbol{MFE = \frac{2}{N}\sum\left|\frac{O_i - P_i}{O_i + P_i}\right|}$. MFB: Mean fractional bias. $\boldsymbol{MFB = \frac{2}{N}\sum\left(\frac{O_i - P_i}{O_i + P_i}\right)}$. P and O represent predictions and observations, respectively.

Table S2-3 2p model parameters for ARO1 SOA yield under high-NOx and low-NOx conditions used in this study

		Parameters	
		Case C0, C2	Case C1
High-NO$_x$ conditions	α_1	0.239	0.770
	C_1^* (μg m^{-3})	10.15	21.10
	α_2	0.738	-
	C_2^* (μg m^{-3})	2147	-
Low-NO$_x$ conditions	α_1	0.100	0.179
	C_1^* (μg m^{-3})	9.592	-

Table S2-4 Observed average concentrations of major aromatic compounds in Chinese cities. Units are ppb.

City	Toluene	Ethylbenzene	Xylenes	TEX*	Time	Reference
Beijing (u[#])	2.42	0.79	1.24	4.44	Nov. 2014	223
Beijing (s)	0.92	0.37	0.46	1.75		
Beijing(s)	0.82	0.20	0.75	1.77	Nov.2017-Feb. 2018	224
Chengdu (u)	1.80	0.83	1.71	4.34	Oct.2016 – Sep.2017	
Shanghai (u)	11.10	3.76	4.98	19.84	Nov. 2013 (Non-H^)	225
	14.20	5.23	8.23	27.66	Nov. 2013 (H)	
Guangzhou (u)	4.64	0.79	0.74	6.17	Nov.-Dec. 2009	226
Guangzhou (s)	2.97	0.57	0.58	4.11		
Nanjing (s)	1.67	1.01	1.05	3.73	Sep. 2011-Feb.2012	227
Nanjing (s)	3.23	1.21	1.86	6.30	Aug. 2013	88
Guangzhou (u)	5.45	1.31	3.27	10.03	Jan. 2017 (daytime)	228
	6.08	1.81	5.52	13.41	Jan. 2017 (nighttime)	
Zhengzhou (u)	1.12	0.31	1.27	2.70	Jul.-Sep. 2019	Unpublished

* Sum of toluene, ethylbenzene and xylenes (o + m/p). These are the major compounds included in the SAPRC model species ARO1 (toluene+ethylbenzene) and ARO2 (xylenes).

[^] Non-H stands for non-hazy days, and H stands for hazy days.

[#] (u) stands for urban monitor sites, and (s) stands for suburban monitor sites.

Table S2-5 Predicted concentrations of ARO1 and ARO2 concentrations from the base case simulation. Units are ppb.

	January			July		
Site	ARO1	ARO2	Sum	ARO1	ARO2	Sum
Beijing	4.12	2.67	6.79	1.63	0.87	2.49
	(1.23,10.02)	(0.69,6.87)	(1.92,16.88)	(0.45,4.58)	(0.21,2.77)	(0.67,7.35)
Shenyang	3.54	1.84	5.39	1.63	0.80	2.44
	(1.96,7.94)	(1.04,4.22)	(3.01,12.15)	(0.81,3.85)	(0.39,1.97)	(1.20,5.82)
Nanjing	4.32	2.19	6.52	1.39	0.70	2.09
	(3.15,6.49)	(1.53,3.38)	(4.67,9.86)	(0.94,3.76)	(0.47,1.92)	(1.42,5.68)
Chengdu	3.99	2.25	6.24	2.67	1.38	4.05
	(1.10,6.16)	(0.34,3.59)	(1.44,9.75)	(0.32,4.52)	(0.11,2.43)	(0.43,6.96)
Guangzhou	3.92	1.77	5.69	2.88	1.36	4.24
	(1.44,3.92)	(0.51,1.82)	(1.95,5.73)	(0.29,3.34)	(0.14,1.72)	(0.43,5.07)
Shanghai	2.79	1.38	4.18	2.03	0.92	2.96
	(2.59,5.00)	(1.12,2.65)	(3.71,7.64)	(1.54,3.40)	(0.68,1.66)	(2.22,5.05)
Zhengzhou	4.67	2.43	7.10	2.00	1.00	3.00
	(3.21,4.67)	(1.54,2.43)	(4.75,7.10)	(0.92,2.00)	(0.42,1.00)	(1.34,3.00)

Note: For each city the concentrations are from the grid cell where the urban center is located. The numbers in the parenthesis represent the minimum and maximum concentrations with the 9 grid cells with the urban center grid cell in the middle.

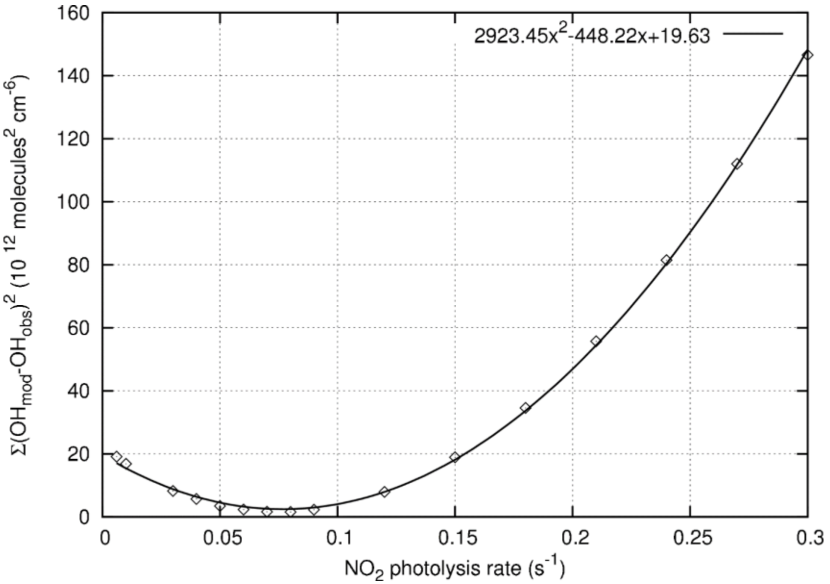


Figure S2-1 The sum of squared errors (SSE) (units: 10^{12} molecules cm^{-3}) of predicted OH for the high-NO_x chamber experiments with different NO_2 photolysis rate coefficients (J_{NO2}). A quadratic function was used to fit the SSE as a function of j_{NO2}. The minimum SSE occurs at J_{NO2}=0.0767 min^{-1}.

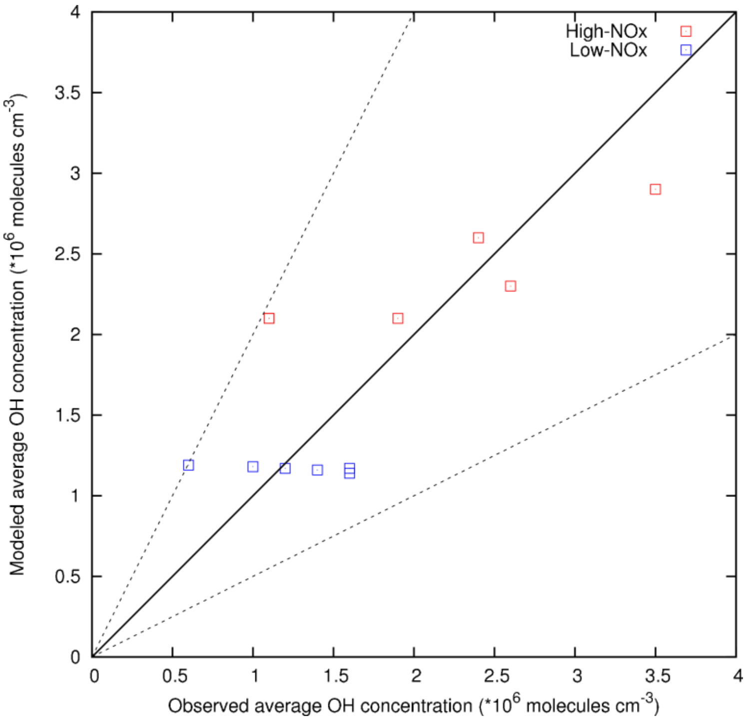


Figure S2-2 Predicted and reported OH for all chamber experiments simulated in this study with the optimal photolysis rate coefficient (J_{NO2}) of 0.0767 min^{-1}. The optimized J_{NO2} was found based on simulations for the high-NO_x conditions.

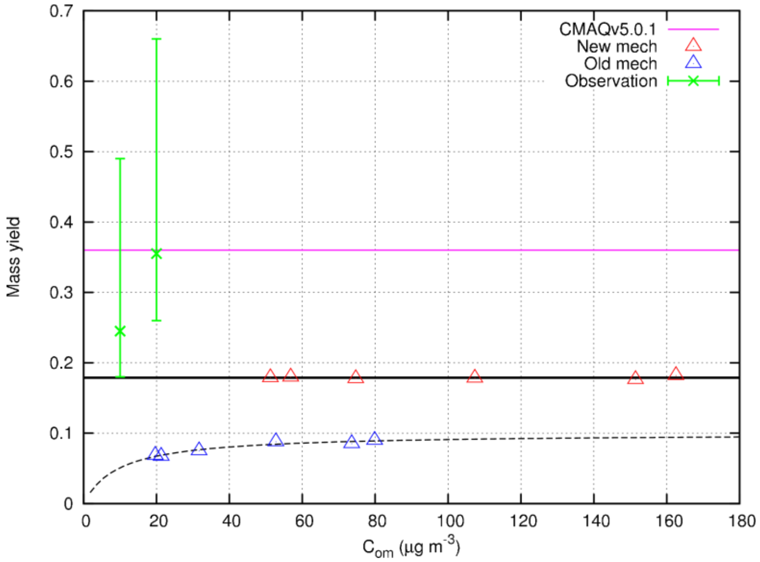


Figure S2-3 SOA yields under low-NO_x conditions used in the original CMAQ model (pink line) and those derived based on the photochemical box model simulations using the original MCM3.2 (Old mech) and the modified MCM3.2 mechanism with a higher branching ratio for the o-cresol pathway (New mech). Triangles show the yields at the end of each simulation. The green error bars show the minimum, maximum and mean SOA yields measured by Hildebrandt et al. at C_{om} = 10 and 20 µg m^{-3}.

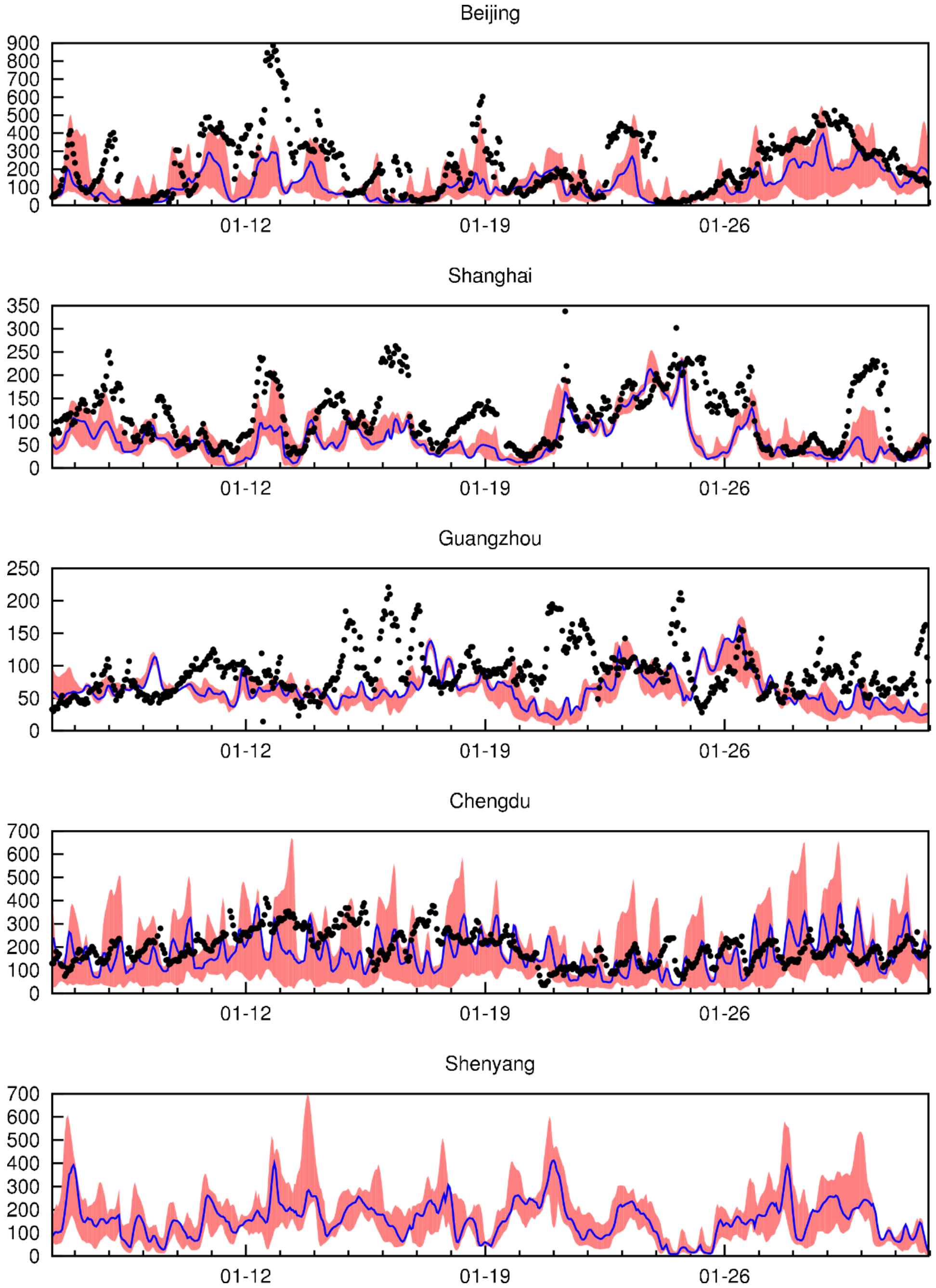


Figure S2-4 Predicted (blue line) and observed (black dots) hourly $PM_{2.5}$ concentrations in 5 Chinese cities in January 2013. The shaded areas represent the range of concentrations within the 9 grid cells (3×3) with the urban center in the middle. Units are $\mu g\ m^{-3}$.

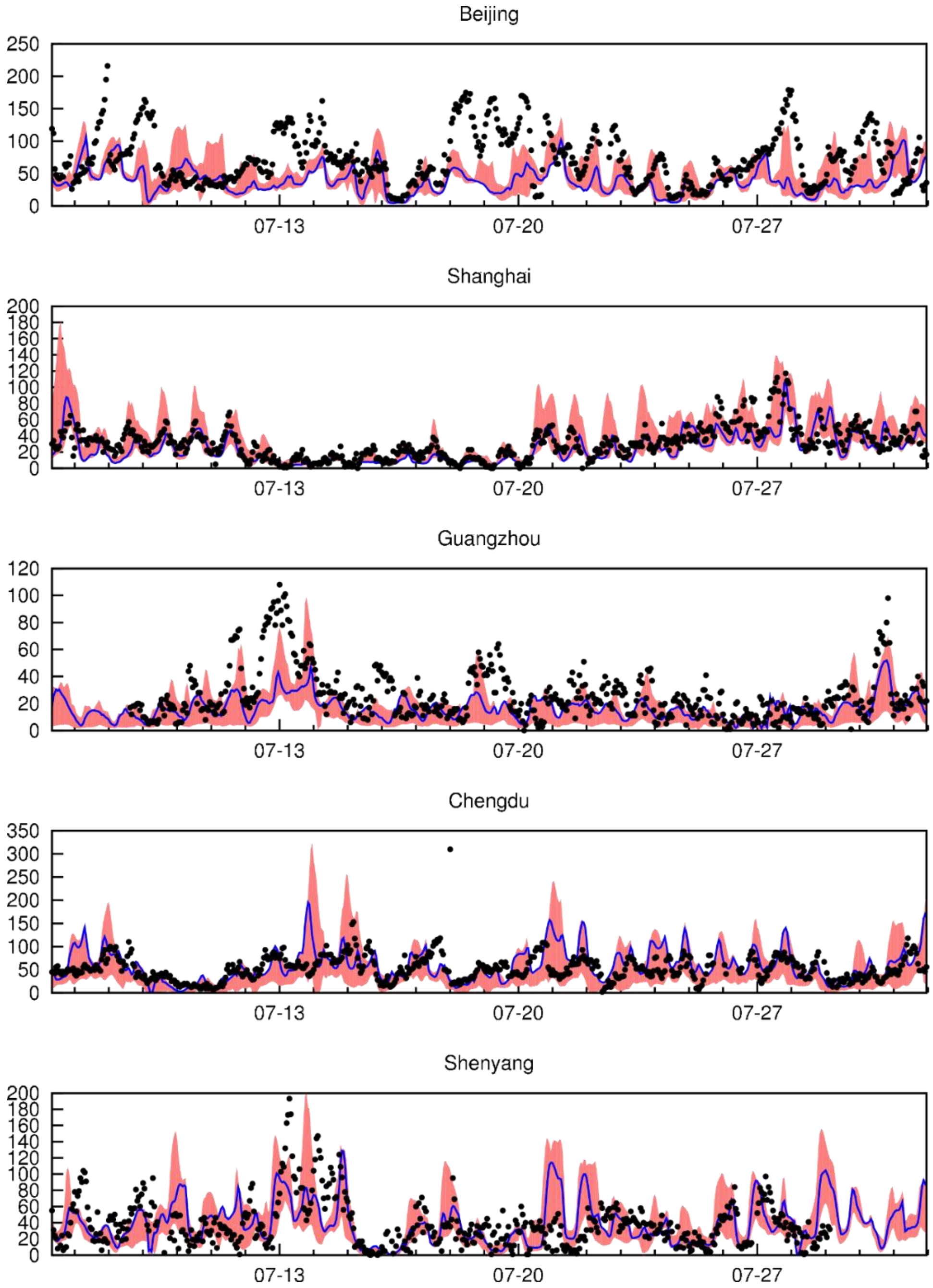


Figure S2-5 Predicted (blue line) and observed (black dots) hourly $PM_{2.5}$ concentrations in 5 Chinese cities in July 2013. The shaded areas represent the range of concentrations within the 9 grid cells (3x3) with the urban center in the middle. Units are μg m^{-3}.

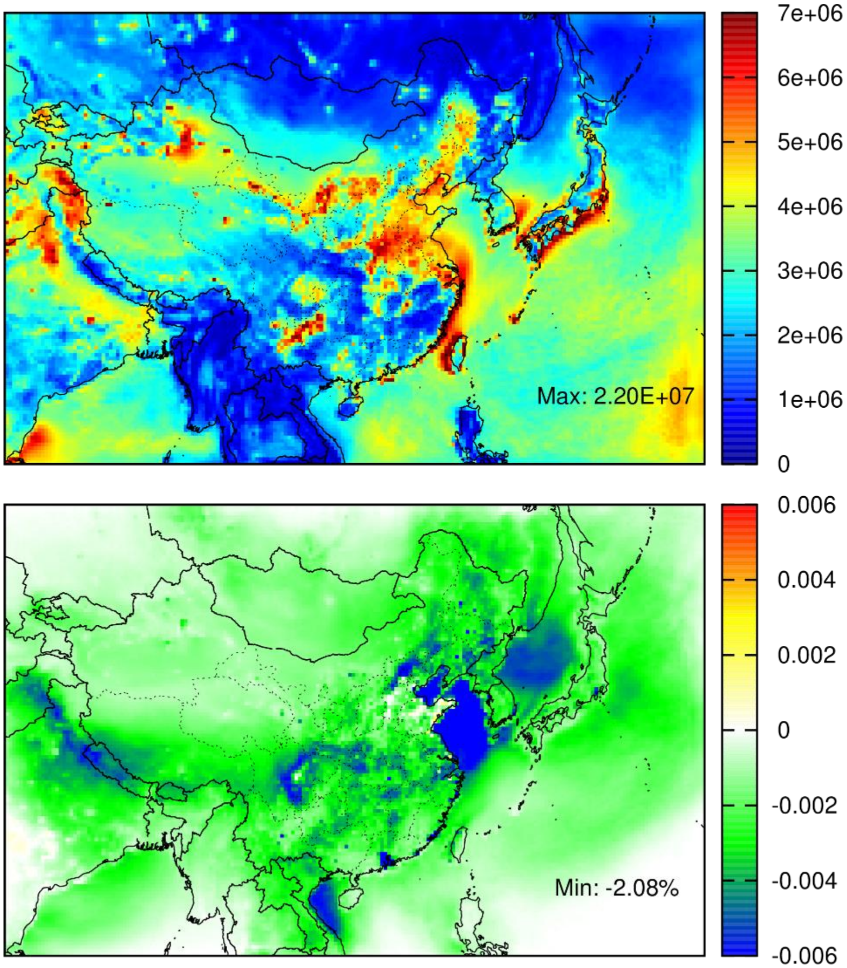


Figure S2-6 Monthly average OH radical concentrations (molecules cm^{-3}) from Case C1 for July 2013 (a), and the relative difference (Case C1-C0)/C0 (b).

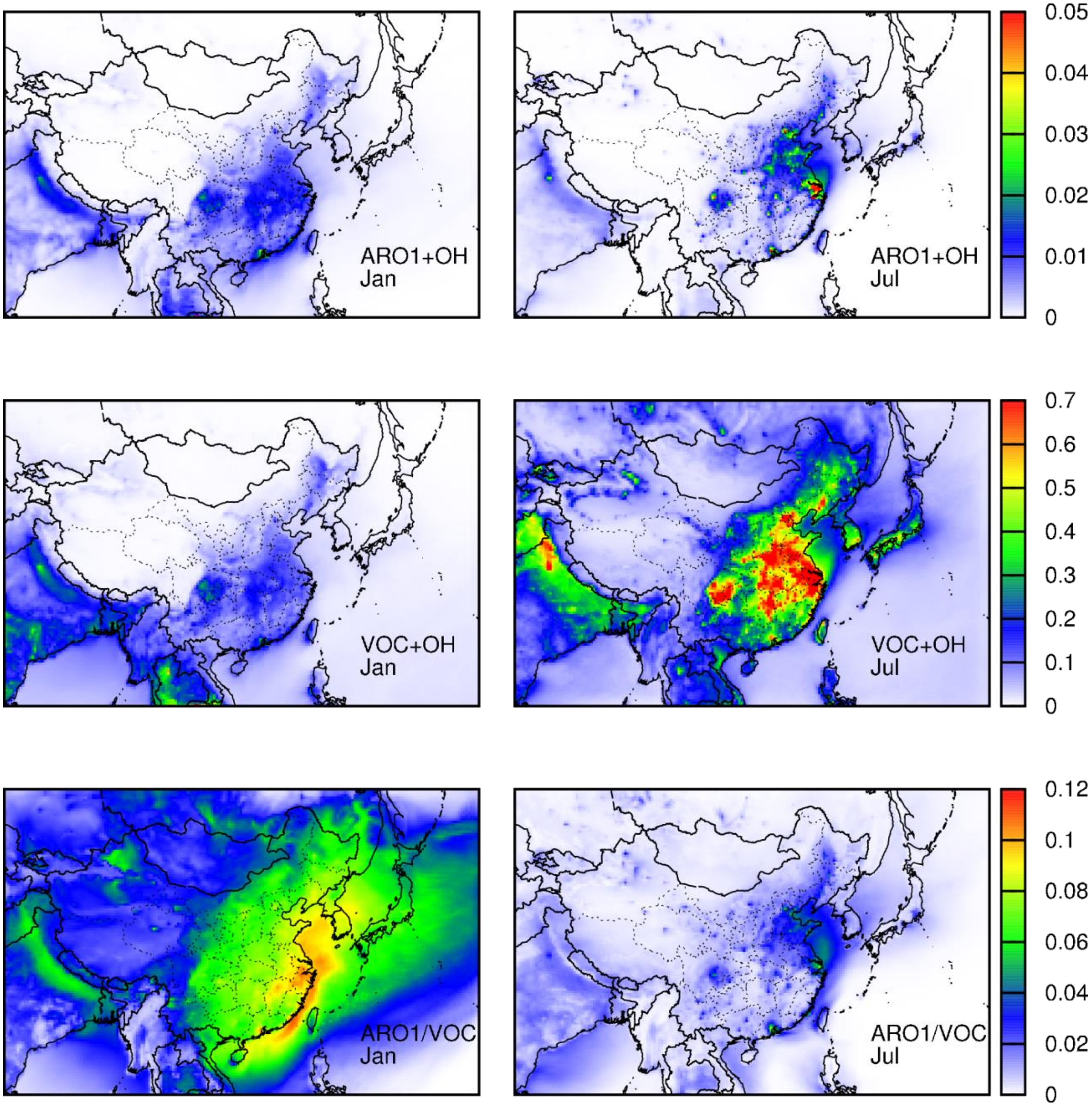


Figure S2-7 The amount of ARO1 and VOCs reacted with OH radical in one hour (ppb hr^{-1}) averaged for the entire month of January and July, and the fractional contribution of ARO1 in OH consumption by the VOCs.

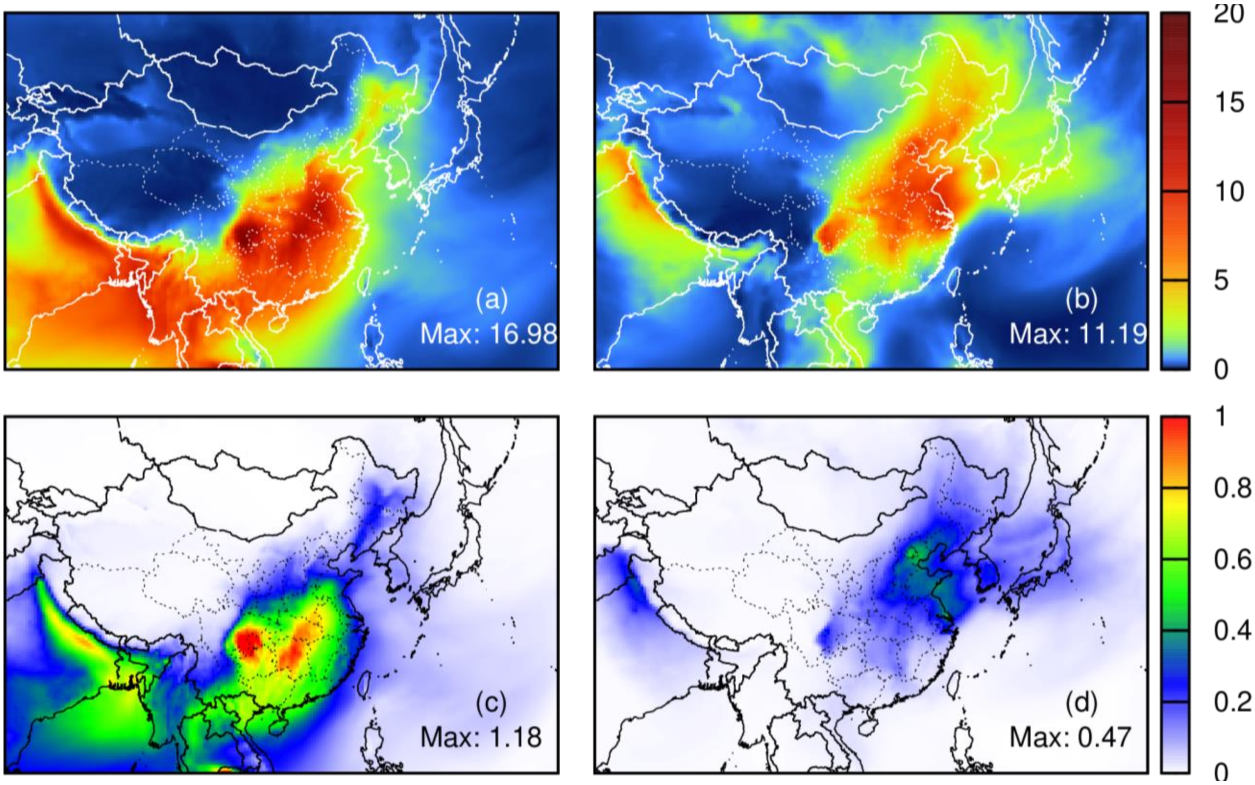


Figure S2-8 Predicted monthly average total SOA for January 2013 (a) and July 2013 (b) from Case C1, and the increase in total SOA (Case C1- Case C0) for January (c) and July 2013 (d). Units are µg m^{-3}.

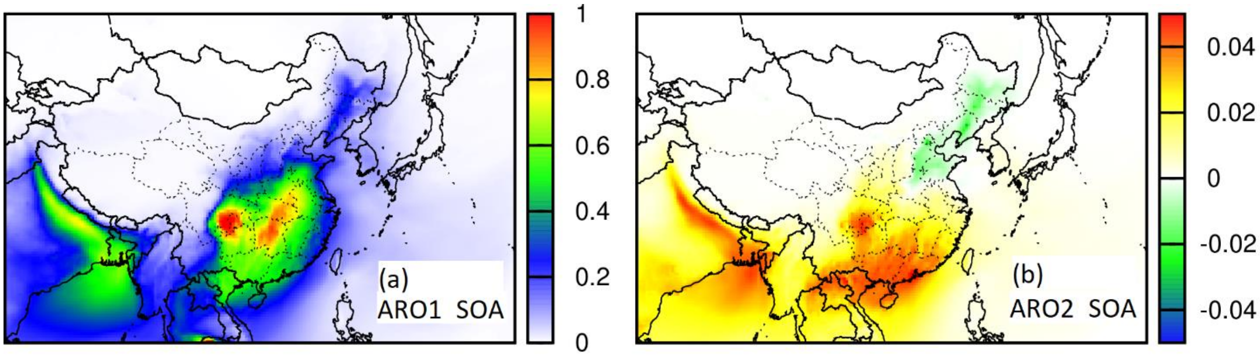


Figure S2-9 Increase in total SOA from (a) ARO1 and (b) ARO2 in January 2013 (Case C1 – Case C0). Units are µg m^{-3}.

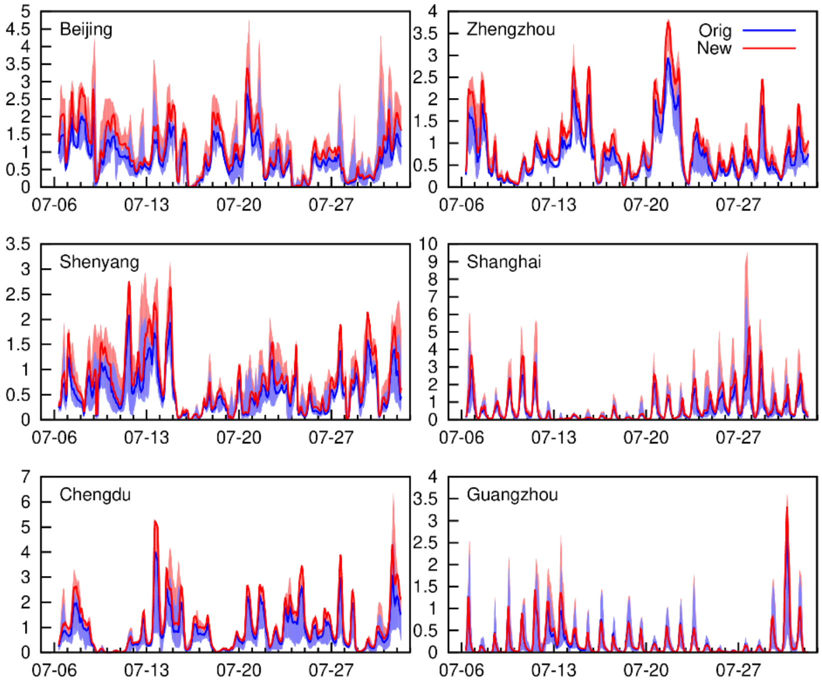


Figure S2-10 Predicted hourly concentrations of ARO1 SOA in several urban areas for July 2013 using the original and modified SOA yields. Units are μg m^{-3}. Shaded area represents the range of concentrations within the 3×3 grids with the urban center in the center grid. The solid lines are the average concentrations in the 3×3 grids.

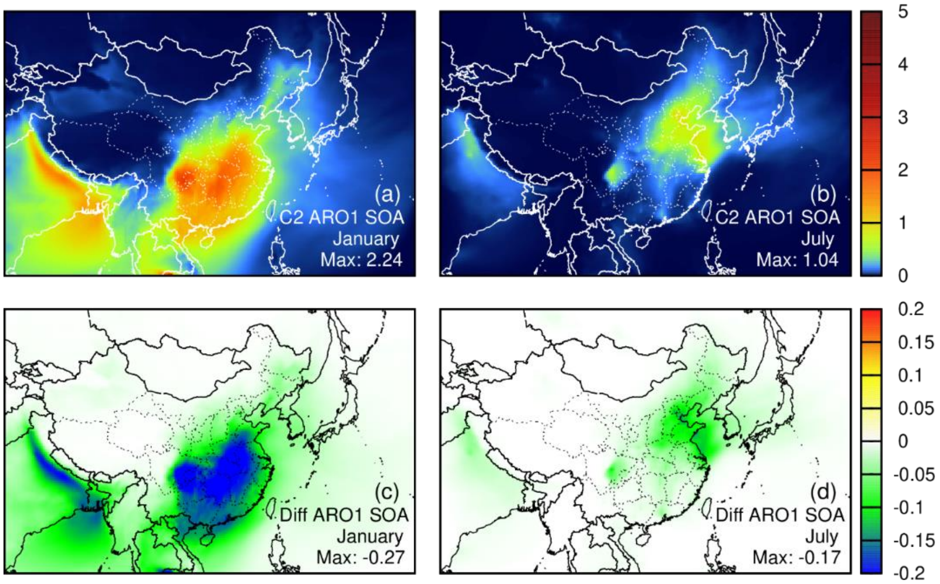


Figure S2-11 Predicted monthly average ARO1 SOA (Case C2) in January (a) and July (b) 2013. The monthly average ARO1 SOA changes (Case C2 – Case C0) for January (c) and July (d) 2013. Units are μg m^{-3}.

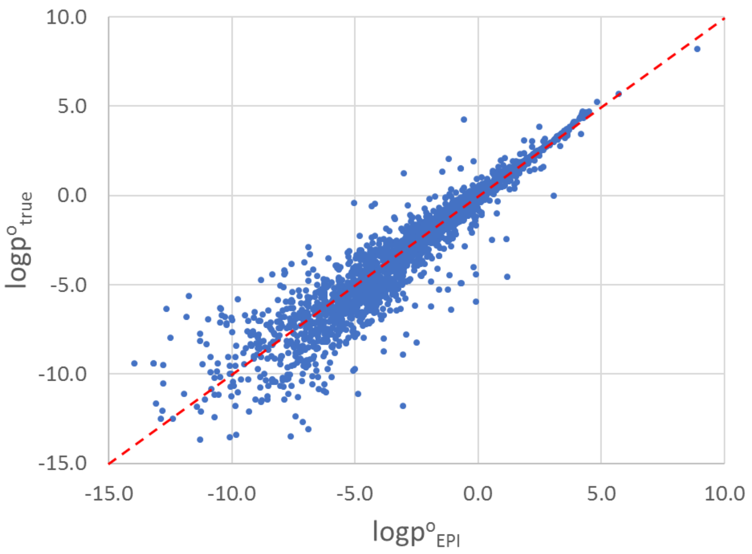


Figure S2-12 Predicted and experimental saturation vapor pressure for 3707 reference compounds in the EPI suite database.

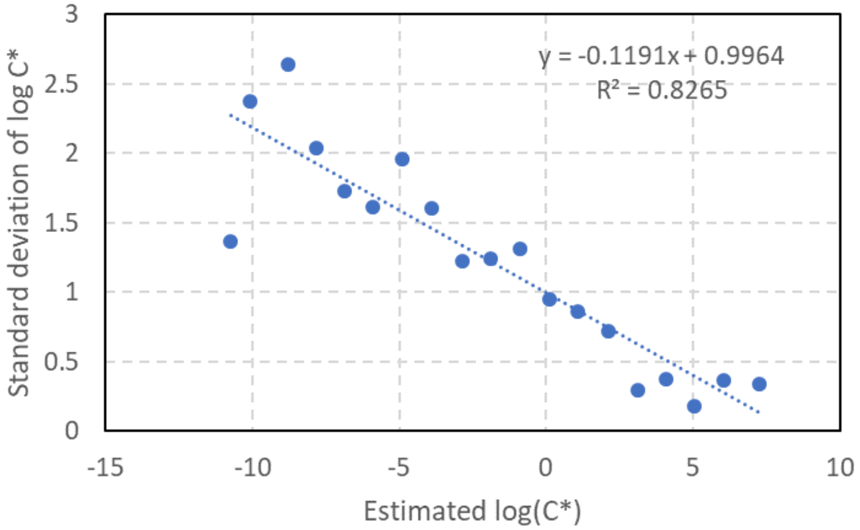


Figure S2-13 Standard deviation of estimated C^* as a function of $\boldsymbol{C^*_{EPI}}$.

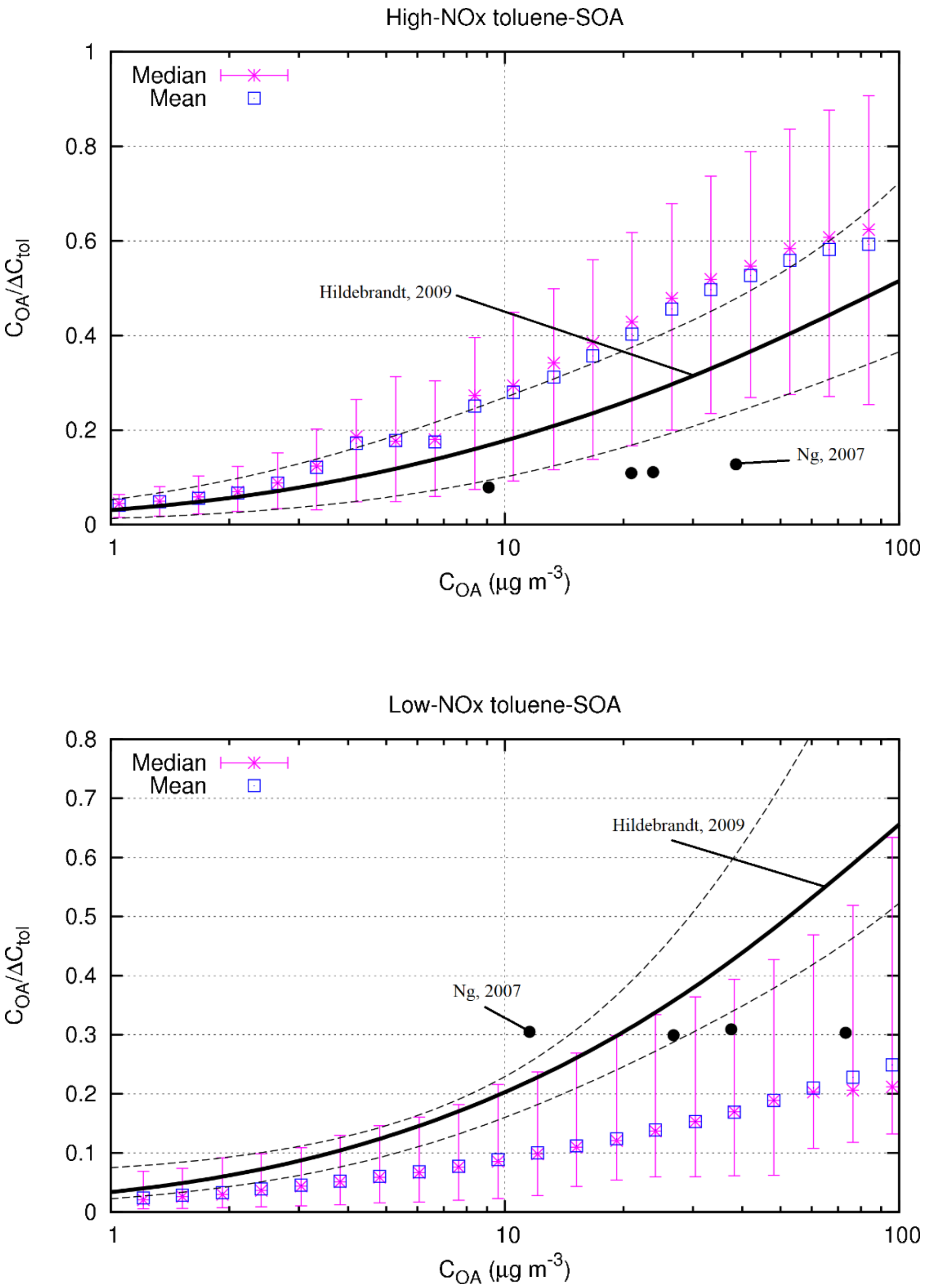


Figure S2-14 Uncertainty in the predicted SOA yields using the MCM mechanism due to uncertainty in the saturation vapor pressure of the semi-volatile products. The error bars are standard deviations calculated using 100 Monte Carlo simulations with random perturbed saturation vapor pressures. The dashed lines show the range of the SOA yield based on the experimental data (based on Figure 10 in Hildebrandt et al.[158])

Section S2.1 NO_2 photolysis rate in the MCM box model

NO_2 photolysis rate in the chamber experiments ($J_{NO_2,chamber}$) is subscribed in the simulation. Since the detailed spectrum of black lights in the chamber experiments are not reported, the photolysis rate of other reactions in the chamber ($J_{other,chamber}$) are calculated by adjusting the clear sky photolysis rate using the following equation,

$$J_{other,chamber} = \frac{J_{other,clear}}{J_{NO_2,clear}} J_{NO_2,chamber} \tag{S1}$$

Solar spectrum is used to calculate the clear sky photolysis rates.

In the H_2O_2 experiments, the photolysis rate of H_2O_2 is also calculated using the above equation, with $J_{NO_2,chamber}$ taken to be the same as the one determined in the high-NO_x experiments because they are performed in the same chambers with the same light source.

The zero solar zenith angle (θ_s) used in the calculation is chosen without a specific reason, as we assume that $\frac{J_{other,clear}}{J_{NO_2,clear}}$ is not a strong function of solar zenith angle (see Table S2-6 below). The potential error in the estimation of the HNO_2 photolysis rate using the above equation at other solar zenith angles, when the ratio $\frac{J_{o,clear}}{J_{NO_2,clear}}$ is determined using a solar zenith angle of 0, is also shown in Table S2-6.

As shown in Table S2-6 the relative error in the estimated J_{HNO_2} is less than 10% for solar zenith angles between 0 and 90 degrees. A better choice of solar zenith angle for our calculations would be ~80 degrees as it gives the NO_2 photolysis rate close to the reported value in the chamber, but it should not lead to significant differences in predicting SOA in the experiments.

The ratio $\frac{J_{other,clear}}{J_{NO_2,clear}}$ with black light might be different from those based on the solar spectrum. In another chamber study illuminated with black lights[229], the photolysis rate of NO_2

was determined from the photo-stationary relationship between NO_2, NO, and O_3, and the HNO_2 photolysis was calculated using the measured black light spectrum. J_{HNO_2}/J_{NO_2} was reported to be 0.145, close to the ratio of 0.161 used in our study. Furthermore, the modeled OH radical concentrations in the chamber experiments are close to the reported OH based on the decay of the precursor, further suggesting that the photolysis rates used in this study are reasonable.

Table S2-6 Clear sky photolysis of NO_2 and HNO_2 at different solar zenith angles.

θ_s(°)	J_{NO_2} (min^{-1})	J_{HNO_2} (min^{-1})	J_{HNO_2}/J_{NO_2}	$J_{HNO_2,est}$*	Error^
0	0.656	0.107	0.163	0.107	0.0%
10	0.650	0.106	0.163	0.106	0.0%
20	0.632	0.103	0.163	0.103	0.3%
30	0.599	0.097	0.162	0.098	0.8%
40	0.548	0.088	0.161	0.089	1.5%
50	0.473	0.075	0.159	0.077	2.6%
60	0.370	0.058	0.156	0.060	4.4%
70	0.235	0.036	0.152	0.038	7.2%
80	0.091	0.014	0.149	0.015	9.2%
90	0.008	0.001	0.153	0.001	6.6%

* Estimated HNO_2 photolysis rates are based on the NO_2 photolysis rate at the specific θ_s and the constant rate of $J_{HNO_2}/J_{NO_2} = 0.163$ at $\theta_s = 0$.

^ Error is calculated as $(J_{HNO_2,est} - J_{HNO_2})/J_{HNO_2}$

Section S2.2 Full reactions of original, upper-limit and modified ARO1+OH used in the study

The original ARO1 + OH reaction in SAPRC-11 mechanism is shown as the Reaction RS1, which is used in the simulation of Case C0.

ARO1 + OH = 0.089 RO2XC + 0.622 RO2C + 0.209 HO2 + 0.612 xHO2 + 0.089 zRNO3 + 0.14 yR6OOH + 0.007 xMEO2 + 0.049 xBALD + 0.064 xPROD2 + 0.003 xCCHO + 0.006 xRCHO + 0.135 CRES + 0.032 XYNL + 0.268 xGLY + 0.231 xMGLY + 0.255 xAFG1 + 0.244 xAFG2 + 0.567 yRAOOH + 0.084 OH + 0.084 AFG3 + 0.042 AFG5 (RS1)

The upper-limit ARO1 + OH reaction assuming the OH + ring reactions generate cresol (from toluene) and phenolic compounds (from other monoalkylbenzenes) only.

ARO1 + OH = 0.014 RO2XC + 0.099 RO2C + 0.887 HO2 + 0.065 xHO2 + 0.0015 zRNO3 + 0.14 yR6OOH + 0.049 xBALD + 0.064 xPROD2 + 0.717 CRES + 0.170 XYNL (RS2)

The modified ARO1 + OH reaction in SAPRC-11 mechanism is shown as the Reaction RS3, which is used in the simulation of Case C1 and C2.

ARO1 + OH = 0.059 RO2XC + 0.415 RO2C + 0.477 HO2 + 0.396 xHO2 + 0.0727 zRNO3 + 0.14 yR6OOH + 0.0042 xMEO2 + 0.049 xBALD + 0.064 xPROD2 + 0.0018 xCCHO + 0.0036 xRCHO + 0.365 CRES + 0.087 XYNL + 0.162 xGLY + 0.140 xMGLY + 0.154 xAFG1 + 0.147 xAFG2 + 0.342 yRAOOH + 0.051 OH + 0.051 AFG3 + 0.025 AFG5 (RS3)

Section S2.3 Evaluation of the modified SAPRC-11 ARO1 mechanism on O_3 formation using smog chamber data

The SAPRC modeling program for chamber experiments was used to simulate 16 EPA smog chamber experiments, selected from the collection of chamber experiments used to evaluate the SAPRC mechanisms, to understand how the higher cresol branching ratio of toluene-OH initial oxidation changes the O_3 formation simulation in box models. The programs and the input data for the chamber experiments were downloaded from Dr. William P.L. Carter's SAPRC mechanism website (https://intra.engr.ucr.edu/~carter/SAPRC/SAPRCfiles.htm). The details of the selected experiments, including the toluene and NO_x initial concentrations and the smog chamber conditions, are shown in Table S2-7. Each of the smog chamber experiments lasted for 6-10 hours. For each chamber experiment, three cases were simulated, 1) with the original ARO1 mechanism in the lumped SAPRC-11, 2) with the modified ARO1 mechanism, as shown in Table 2, and 3) the explicit toluene mechanism in the detailed version of the SAPRC-11. The original ARO1 mechanism and the detailed toluene mechanism simulate the chamber data better as the SAPRC mechanism is optimized based on the chamber data. The modified ARO1 mechanism leads to slightly slower consumption of toluene and early formation of O_3 in the initial stage of the experiments. In addition, the O_3 concentrations at the end of the experiments are slightly lower than those based on the original mechanism, which agrees the O_3 responses to the toluene-OH branching ratios modification. Figure S2-15 shows a representative O_3 time series in one chamber experiments, and the comparison of the final O_3 concentrations in these simulations against observations is shown in Figure S2-16. The lower O_3 formation with higher yield for the cresol pathway has also been verified in MCM box model simulations.

Table S2-7 Selected smog chamber experiments used to test the modified SAPRC-11 toluene and ARO1 mechanisms.

Run ID	Initial Reactant Summary			Light Type	J_{NO2}	Avg. Temp	Running time	PM Wall Loss
	Test. VOC	VOC	NO_x					
		(ppm)	(ppb)		(min^{-1})	(K)	(min)	
EPA210A	TOLUENE	0.26	42	Arc light solar simulator	0.260	305	377	6.3E-03
EPA210B	TOLUENE	0.26	93	Arc light solar simulator	0.260	305	377	5.4E-03
EPA443A	TOLUENE	0.17	31	Arc light solar simulator	0.260	304	364	3.3E-03
EPA443B	TOLUENE	0.36	99	Arc light solar simulator	0.260	304	364	3.3E-03
EPA289B	TOLUENE	0.22	25	Blacklights	0.165	301	492	2.9E-03
EPA1098A	TOLUENE	0.08	16	Blacklights	0.131	298	583	5.0E-03
EPA1098B	TOLUENE	0.08	30	Blacklights	0.131	298	360	3.5E-03
EPA1099B	TOLUENE	0.04	10	Blacklights	0.131	298	487	3.9E-03
EPA1101A	TOLUENE	0.08	19	Blacklights	0.401	300	491	5.5E-03
EPA1101B	TOLUENE	0.08	9	Blacklights	0.401	300	360	4.5E-03
EPA1102A	TOLUENE	0.08	43	Blacklights	0.401	300	474	6.0E-03
EPA1102B	TOLUENE	0.08	32	Blacklights	0.401	300	474	4.5E-03
EPA1106A	TOLUENE	0.03	20	Blacklights	0.401	300	630	6.2E-03
EPA1106B	TOLUENE	0.03	11	Blacklights	0.401	300	630	6.0E-03
EPA1107A	TOLUENE	0.04	40	Blacklights	0.401	300	595	5.7E-03
EPA1107B	TOLUENE	0.04	30	Blacklights	0.401	300	595	6.4E-03

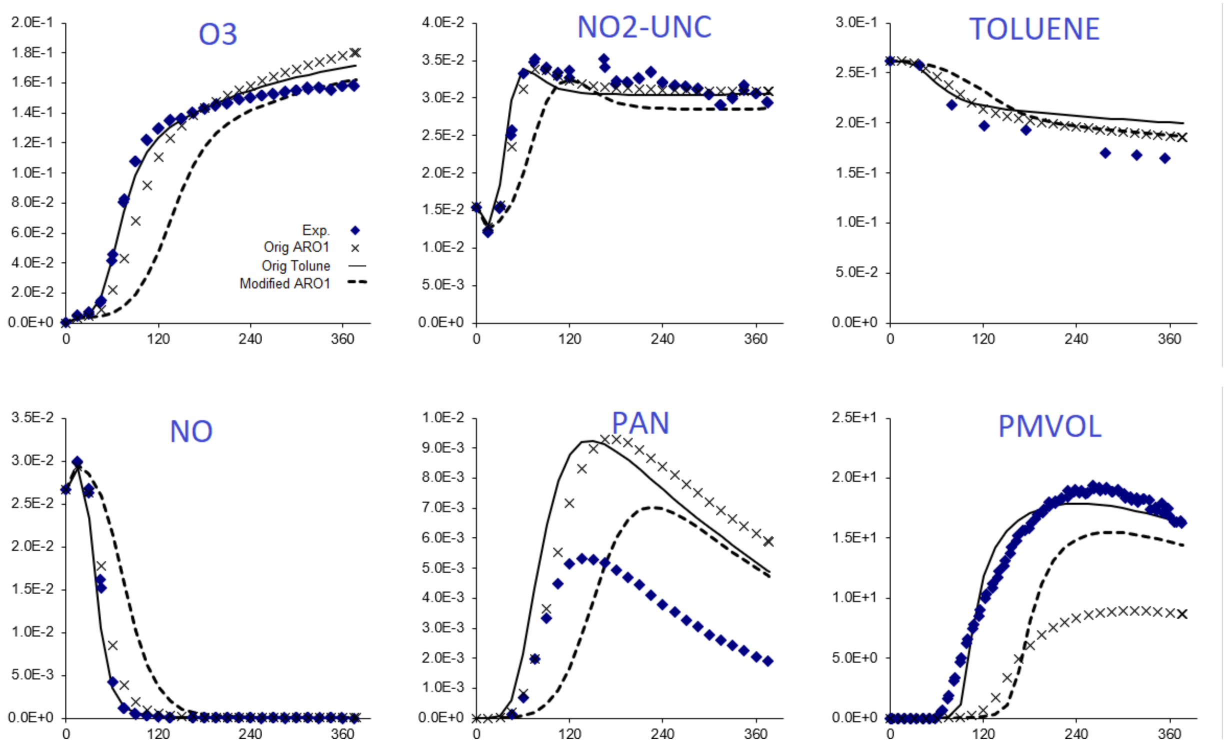


Figure S2-15 Time series of O_3, NO_2 (NO2-UNC), toluene (TOLUENE), NO, PAN and PM volume in the smog chamber experiments using the original lumped ARO1 mechanism, the original toluene mechanism and the modified ARO1 mechanism for the chamber experiment EPA210A.

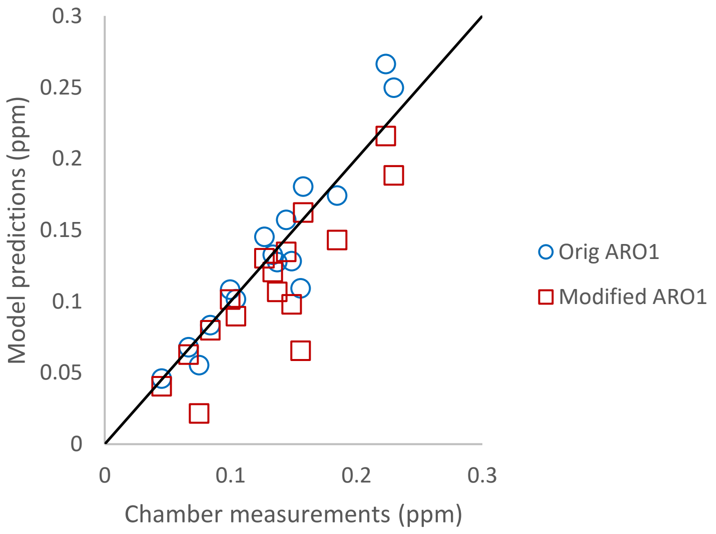


Figure S2-16 Comparison of the final O_3 concentrations in the chamber experiments simulated using the original ARO1 and modified ARO1 mechanisms with observations.

Section S2.4 Evaluation of the MCM-based box model in simulating other chamber experiments

We obtained three sets of chamber data reported by Ng et al.[30], including the time series of toluene, and wall-loss corrected and particle and vapor wall-loss corrected SOA, from Dr. Shantanu Jathar of Colorado State (with consent from Sally Ng). Two of the datasets (Cases 1 and 3) are for high-NO_x conditions, and one dataset (Case 2) is for low-NOx conditions. HONO was used as the OH source for Cases 1 and 3, and NO and NO_2 were added to ensure an initial NOx concentration of 1 ppm. H_2O_2 was used as the OH source for Case 2 with an initial concentration of 5 ppm. The OH concentrations in the chamber experiments were not directly measured. We estimated OH concentrations by assuming decay of toluene is caused by OH only and using $k_{tol+OH}=5.63\times10^{-12}$ cm^3 $molecules^{-1}$ s^{-1} at 25 °C[30]. The photolysis rate of NO_2 was estimated to be 0.45 min^{-1} for all three cases, which correctly predicts the decay of toluene and the OH concentration in the chamber (see Figure S2-17).

Under low NO_x conditions, the model predicted SOA yields with both the original (low cresol pathway) and new (high cresol pathway) mechanisms were lower than the chamber measurements. This is consistent with the results obtained when simulating the low-NOx chamber experiments from Hildebrandt et al.[158] Under high NO_x conditions, the measured SOA yields in Ng et al. were lower (~0.1 at OM = 20 μg m^{-3}) than those in Hildebrandt et al.[158] (0.15-0.45 at OM = 20 μg m^{-3}). The differences in the predicted SOA yields from the original and the modified mechanism are small. The new mechanism still predicted higher SOA yields, as reported in the original manuscript when simulating the Hildebrandt et al.'s data, and has a slightly closer agreement with the vapor wall-loss corrected SOA yields.

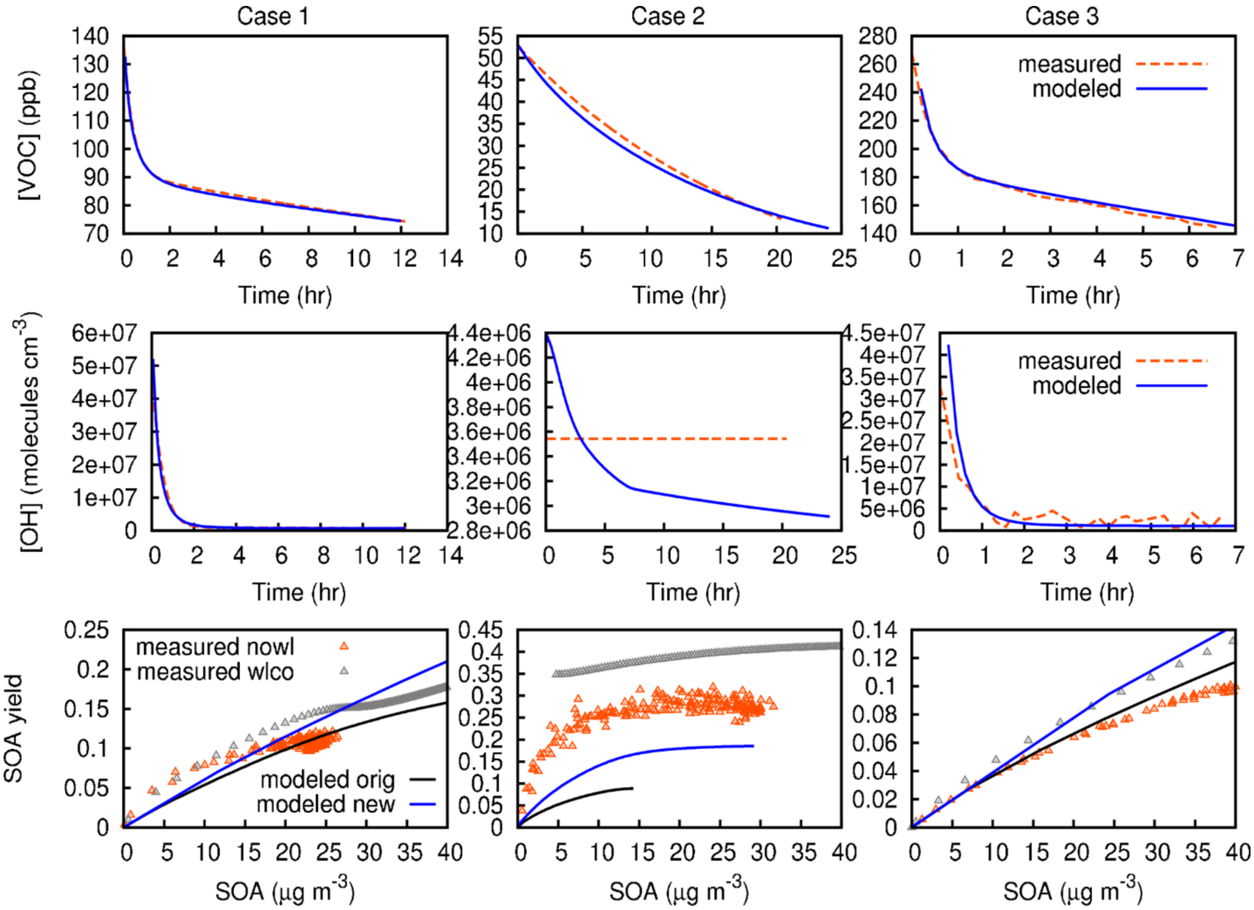


Figure S2-17 Comparison predicted and MCM modeled time series of toluene and OH concentrations, and SOA yields. “measured nowl” is the SOA yields without vapor wall-loss correction. “measured wlco” is SOA yields calculated based on vapor wall-loss corrected SOA concentrations.

APPENDIX B

Table S3-1 Isoprene source emission rate from the grid of SAES site location based on MEIC inventory (unit: kg day^{-1})

Source	Weekday	Weekend
Industrial	346.65	225.27
Residential	8.28	5.38
Transportation	107.92	73.52
MEIC	462.84	304.17
REAS3	58.04	37.72
Biogenic	50.24	50.24

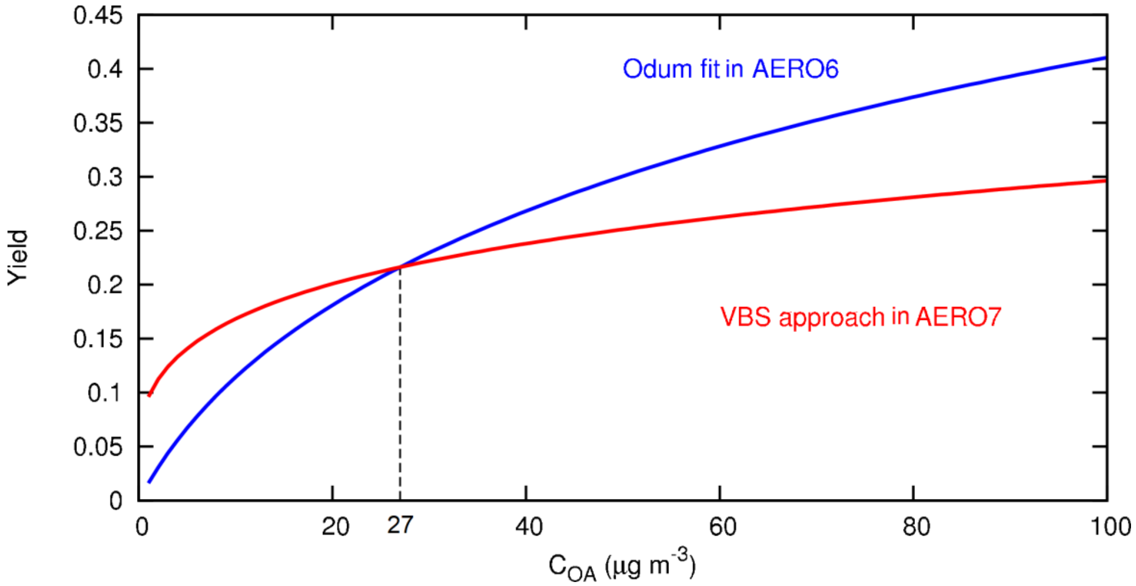


Figure S3-1 Monoterpene-SOA yield based on Odum fit in the AERO6 module and VBS approach in the AERO7 module.

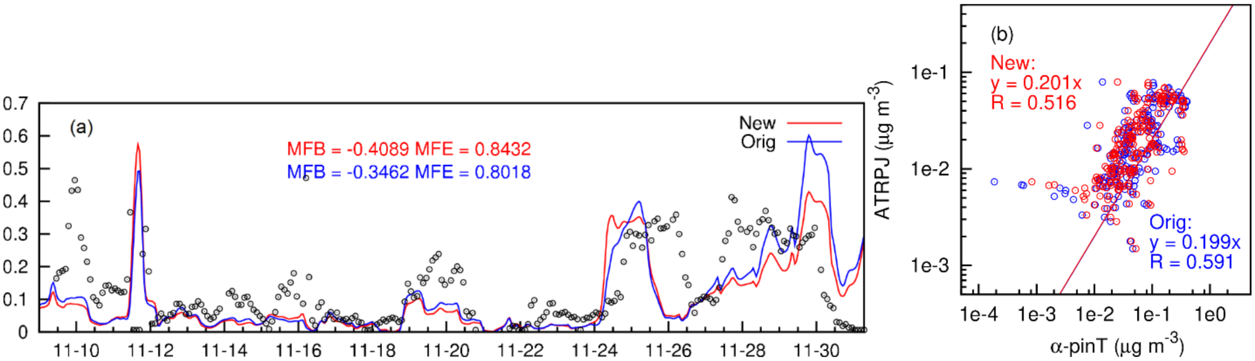


Figure S3-2 (a) Timeseries model predicted monoterpene SOA with new and original yields. Black circles represent tracer-derived monoterpene SOA mass concentrations. (b) The linear correlation between the model predicted monoterpene SOA and observed α-pinene-derived SOA tracer (α-pinT).

APPENDIX C

Section S4.1 Summertime DHOPA and f_{SOA}

To further strengthen the conclusions in this study, the modified CMAQ model was applied to simulate DHOPA in China from June to August 2012. This episode was chosen because it represents a different season and DHOPA data are available at 15 sites across China[136, 166] for model evaluation. A summary of this additional modeling exercise is provided below.

Emission and meteorology inputs for the summer 2012 simulation were generated using the same approaches as those used for the winter 2018 simulation. Since there was no ARO1 and ARO2 data for summer 2012, we compared the predicted precursor concentrations with August 2013 measurements in Nanjing, located in the PRD region. The modeled average ARO1 (5.6 ppb) and ARO2 (2.3 ppb) concentrations in August 2012 are close to the observed concentrations (4.4 ppb for toluene + ethylbenzene; 1.9 ppb for xylenes).

The predicted organic carbon (OC) concentrations were compared with reported concentrations at four Pearl River Delta (PRD) sites.[213] Overall, the predicted OC was lower than observations by a factor of 2 (mean fractional bias (MFB) = -0.6), with larger underpredictions of primary OC (POC, MFB = -1.0) and smaller underpredictions of secondary OC (SOC, MFB = -0.4). In addition, the observed average summertime OC concentration (25.2 μg m^{-3}) at Xianghe, a rural site in northern China, was also underpredicted by a factor of 3 (7.5 μg m^{-3}). Thus, OC concentrations in other locations may also be underestimated by a factor of 2-3, leading to errors in the gas-particle partitioning of DHOPA. Thus, primary organic aerosol (POA) emissions were increased uniformly in the domain by a factor of 3, and a new simulation was conducted to improve the OC predictions.

The model predicted DHOPA concentrations at the 15 sites show a moderate Pearson correlation with observations (R = 0.43), and the predictions are slightly lower than observations with a mean MFB of -0.11 and a mean fractional error (MFE) of 0.78 (Table S4-1; Figure S4-1). The model performance is slightly improved when the POA emission were increased by a factor of 3 (MFB = -0.03, MFE = 0.76, and R = 0.43). We also compared the predicted average DHOPA concentrations in Beijing and Fuzhou, which were made in summer and fall 2016, and the predictions are also in reasonable agreement with observations (Table S4-1)

Table S4-1 Model performance of DHOPA during the summertime in 2012

Site	Duration	Obs. (ng m^{-3})	Case	Pre. (ng m^{-3})	MFB	MFE	R
15 sites in China	June. - Aug. 2012[136, 166]	3.2	OA×1	3.59	-0.11	0.78	0.43
			OA×3	3.87	-0.03	0.76	0.43
Beijing	Summer 2008 [167] Summer 2016 [91]	13.3 (urban)	OA×1	13.5	-	-	-
		9.7 (rural), 11.0 (urban)	OA×3	13.7	-	-	-
Fuzhou	Fall 2016[168]	2.4	OA×1	2.67	-	-	-
			OA×3	3.05	-	-	-

We also looked at the modeled f_{SOA} values throughout the domain. As shown in Figure S4-2, the linear regression slopes between modeled aromatic SOA and DHOPA are similar to those for the winter episode. The f_{SOA} values are ~0.0043-0.0046 (excluding SOA from the ARO1 and ARO2 GLY and MGLY) and 0.0022-0.0023 (including all aromatic SOA components) are very close to the ratios from the winter data. Regional plots of f_{SOA} shown in Figure S4-3 also show rather small variations of f_{SOA} in polluted regions, further confirming the universal applicability of the f_{SOA} values in estimating aromatic SOA.

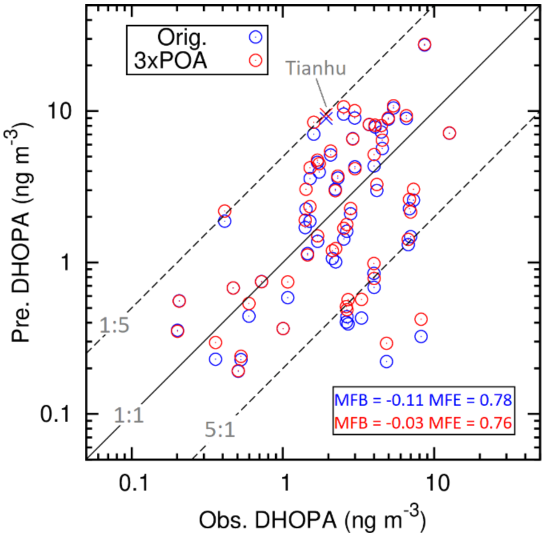


Figure S4-1 Model evaluation of DHOPA with 15 sites across China during the summertime in 2012. The blue points are model predictions with the original POA emissions. The red points are model predictions with the POA emissions increased by a factor of 3. The circles are 48-hour average based on observations reported by Ding et al.[136] The crosses are summertime average at the Tianhu site.[166]

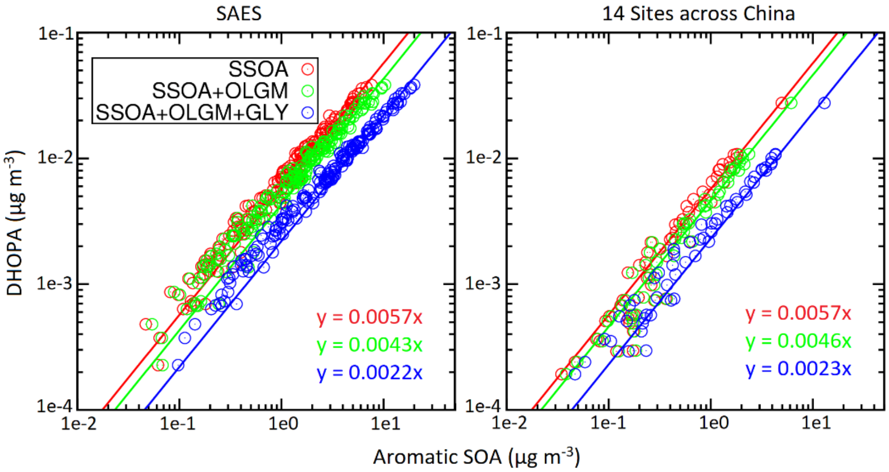


Figure S4-2 Average DHOPA/aromatic SOA ratio based on linear regression with a zero intercept, using data from the 2012 summer simulation. Red, green and blue dots represent semivolatile SOA (semi-SOA, SSOA), SSOA + oligomers (OLGM), and SSOA+ OLGM + glyoxal and methylglyoxal SOA products (GLY), respectively. The left panel shows daily data at the Shanghai Academy of Environmental Sciences (SAES) location, and the right panel shows 48-hour average data (2 per month) at 14 sites across China.

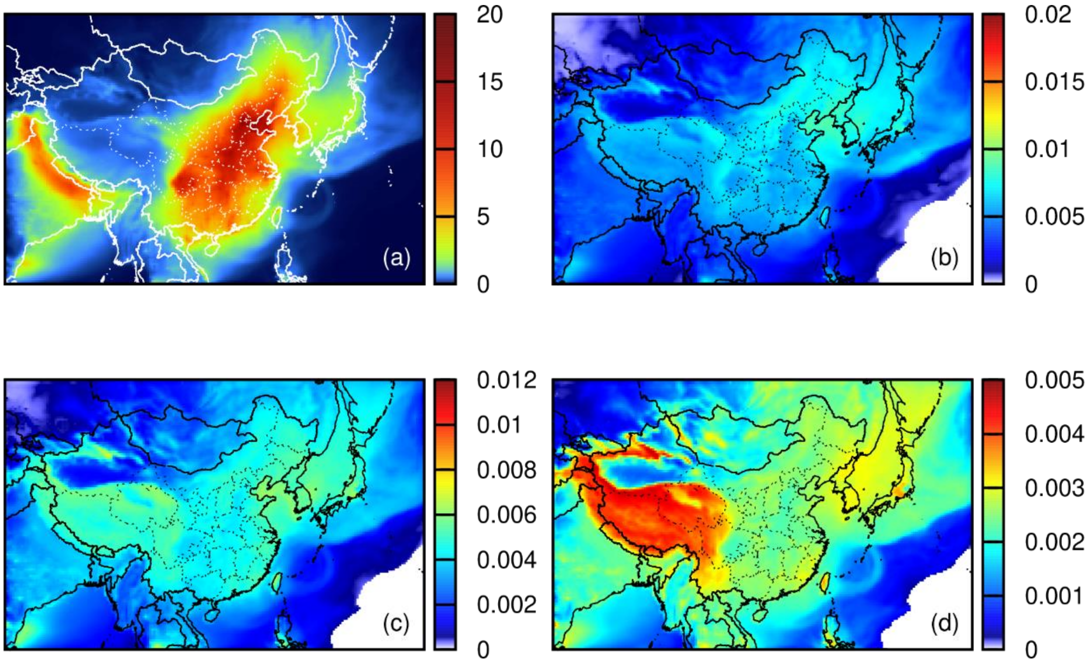


Figure S4-3 Regional distribution of 2012 summer average (a) DHOPA (ng m^{-3}) and $\boldsymbol{f_{SOA}}$ (b-d). The $\boldsymbol{f_{SOA}}$ represents the mass fraction of DHOPA in the aromatic (a) semi-SOA; (b) semi-SOA and oligomers; (c) semi-SOA, oligomers, and glyoxal and methylglyoxal products with a threshold of SOA larger than 0.01 μg m^{-3}.

Figure S4-4 shows the regional distribution of modeled DHOPA from each formation pathway in summer 2012. During the summertime, ARO1 photooxidation under high-NO$_x$ conditions is still the dominant formation pathway of DHOPA, but its relative contributions decrease. On the other hand, contributions from the ARO1 low-NO$_x$ pathway increase because daytime NO$_x$ concentrations in the summertime are lower than in wintertime.

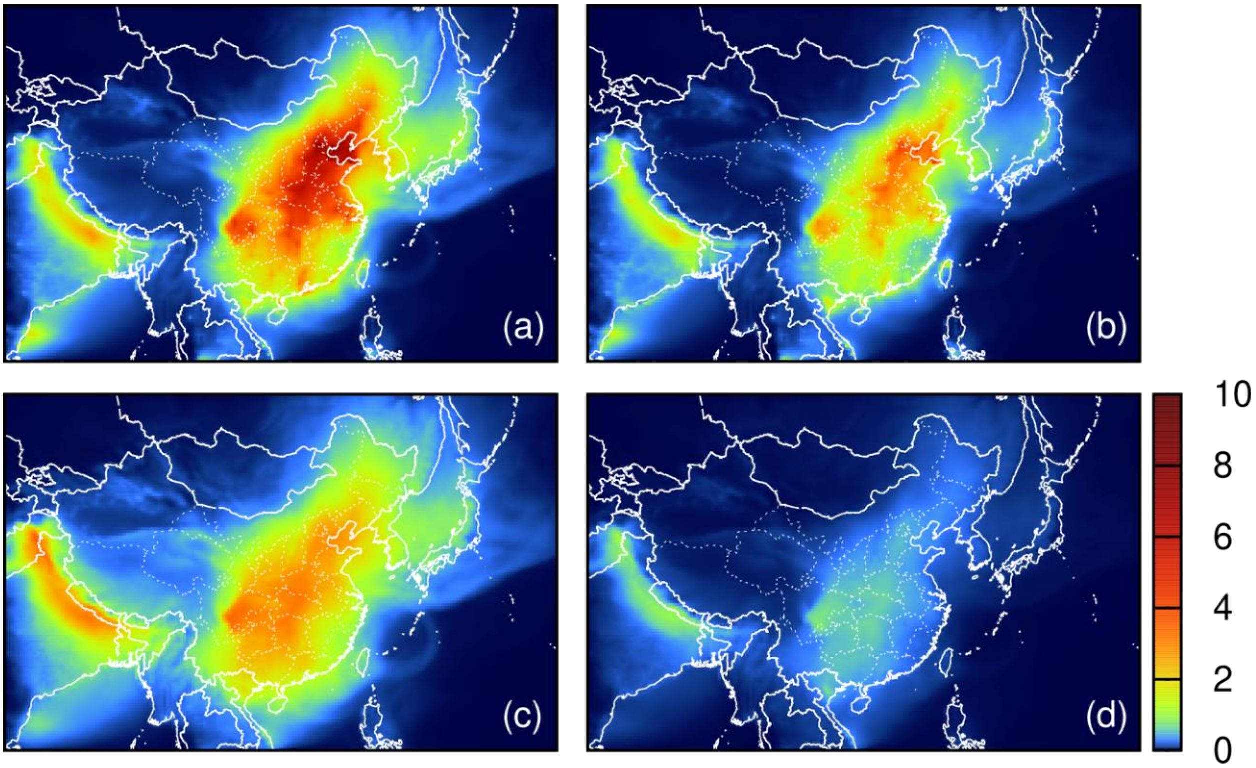


Figure S4-4 Regional distribution of modeled average DHOPA in summer (June – August) 2012 from (a) ARO1 high-NO_x pathway, (b) ARO2 high-NO_x pathway, (c) ARO1 low-NO_x pathway, and (d) ARO2 low-NO_x pathway. Units are ng m^{-3}

Section S4.2 Other species in ARO1 and ARO2 and their impacts on DHOPA yields and f_{SOA}

On average, the major components of ARO1 include toluene (88%, molar fraction) and ethylbenzene (10%), and the major components of ARO2 include xylene (88%), naphthalene (1%), and methylnaphthalene (2%). Other species account for 2% of the total ARO1 and 9% of total ARO2. In the case simulation, all species included in ARO1 use the same DHOPA yield, and so does the species in ARO2.

The f_{SOA} for ethylbenzene and trimethylbenzene can be found from Al-Naiema et al.'s study.[39] The f_{SOA} and DHOPA mass yield (α) is listed in Table S4-2. The average yields of DHOPA calculated based on the fractions of detailed speciated ARO1 and ARO2 are lower than the yields of toluene and xylene by approximately 10%, so they do not significantly affect the overall DHOPA and f_{SOA} predictions.

Table S4-2 The f_{SOA} and mass yield (α) of DHOPA from aromatics photooxidation reaction under high-NOx.

Species	Fraction	f_{SOA}	α
Toluene	88%	0.004	0.00189
Ethylbenzene	10%	0.00015	9.75E-05
Other ARO1	2%	n.a.[#]	n.a.
		Average:	0.00165
Xylenes	88%	0.0033	0.00096
Naphthalene	1%	n.a.	n.a.
Methylnaphthalene	2%	n.a.	n.a.
Other ARO2[^]	9%	0.0013	0.00034
		Average:	0.00087

[#] n.a.: there is no experimental data available.

[^] assume to have the same f_{SOA} as trimethylbenzene.

In this study, the SOA yield parameters for ARO1 were based on the experimental data of toluene, and the parameters for ARO2 were based on the experimental data of xylene. Based on the mass yields for ethylbenzene, trimethylbenzene[230], and naphthalene[231], and the molar fractions reported in Table S4-2, we calculated more accurate SOA yields for ARO1 and ARO2, as shown in Figure S4-5. The estimated SOA yield for lumped ARO1 species is slightly lower than that for toluene alone by less than 5%, and the estimated SOA yield for lumped ARO2 is pretty close to that for xylene alone. Thus, the toluene and xylene SOA yields are good enough to represent the yields for lumped ARO1 and ARO2 species.

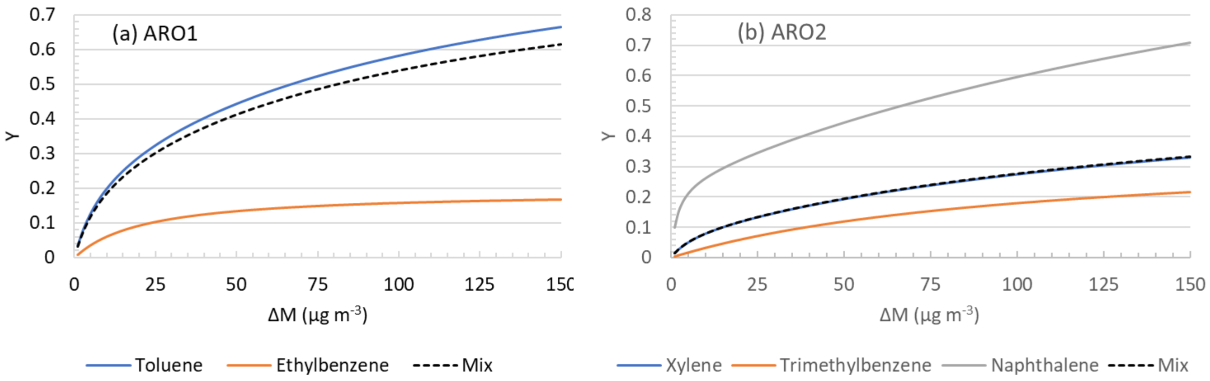


Figure S4-5 SOA yield curves for detailed ARO1 (a) and ARO2 (b) species, and mixture yield for lumped ARO1 and ARO2.

Section S4.3 Other supporting tables and figures

Table S4-3 SOA two-product model parameters used in the base case CMAQ modeling

	β_1[$]	C_1^* µg m^{-3}	β_2	C_2^* µg m^{-3}	ΔH_{vap} kJ/mol	β_3
ARO1[^]	0.2545	8.02	0.7623	119.3	18.0	0.57
ARO2[#]	0.0883	7.43	0.5647	192.7	32.0	0.612

[$] The mass yields ($\beta_1$, β_2) and saturation concentrations (C_1^*, C_2^*) are parameters for the high-NO$_x$ pathways. The enthalpy of evaporation is applied to both products. β_3 is based on the mass yield for the low-NO$_x$ pathway. The products are considered non-volatile.

[^] High-NO$_x$ pathway based on the 4-bin VBS parameters reported by Hildebrandt et al. [158], fitted to the Odum two-product model by Ying et al. [83] and then slightly increased in Hu et al.[88]

[#] High-NO$_x$ pathway Based on the 5-bin VBS parameters reported by Ahlberg et al. [145], fitted to the Odum two-product model in this study.

Table S4-4 List of base case and sensitivity CMAQ runs performed in this study.

		Base case	Sensitivity Cases				
			Case 1	Case 2	Case 3	Case 4	Case 5
p_L^0 and ΔH_{vap}		E-AIM Method 1	E-AIM Method 2	E-AIM Method 1	E-AIM Method 2	E-AIM Method 1	E-AIM Method 1
Xylene SOA yields		Ahlberg et al.		Same as toluene		Ahlberg et al.	
DHOPA mass yields (α)	Toluene high-NO$_x$	0.00189	0.00205	0.00189	0.00189	0.00189	0.00189
	Toluene low-NO$_x$	0.00391	0.00408	0.00391	0.00391	0.00391	0.00391
	Xylene high-NO$_x$	0.0009	0.00101	0.00304	0.0009	0.0009	0.0009
	Xylene low-NO$_x$	0.0009	0.00101	0.00304	0.0009	0.0009	0.0009
Temperature		unchanged	unchanged	unchanged	unchanged	+5K	+10K

Table S4-5 Meteorology model performance and comparison with proposed benchmarks.

			Statistics	Benchmarks[#]
TEMP	(K)	MB	0.95	≤ ±0.5
		GE	2.09	< 2.0
RH	(%)	MB	-7.04	
		GE	11.98	
WSPD	(m s^{-1})	MB	0.45	≤ ±0.5
		RMSE	1.61	< 2
WDIR	(°)	MB	12.17	≤ ±10
		GE	34.91	< 30

[#] Benchmarks for temperature, wind speed, and wind direction were proposed by Emery et al.[124]

Table S4-6 Distribution of $\boldsymbol{f_{SOA}}$ values calculated using the monthly average SOA and DHOPA

DHOPA concentration	2 – 5 ng m^{-3}		5 – 10 ng m^{-3}		≥ 10 ng m^{-3}	
	Mean	Stdv.	Mean	Stdv.	Mean	Stdv.
Semi-Volatile SOA (SSOA)	0.00843	0.00297	0.00779	0.00164	0.00665	0.00068
SSOA + OLGM	0.00469	0.00109	0.00436	0.00071	0.00399	0.00023
SSOA + OLGM + GLY	0.00231	0.00025	0.00225	0.00021	0.00220	0.00012

Three ways of calculating SOA: (1) semivolatile SOA (semi-SOA, SSOA) only, (2) SSOA + oligomers (OLGM), and (3) SSOA+ OLGM + glyoxal and methylglyoxal SOA products (GLY).

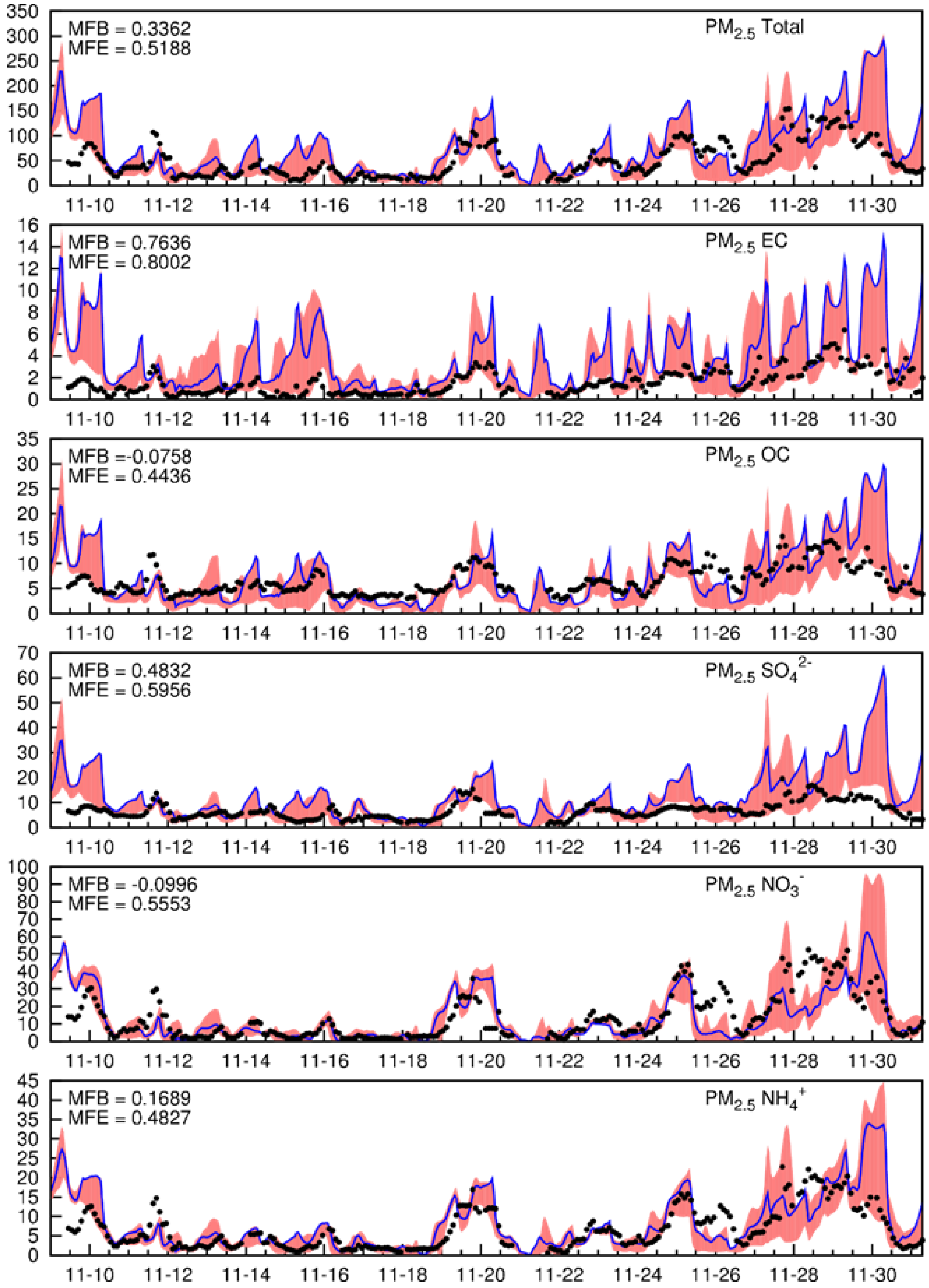


Figure S4-6 Time series and model performance statistics of PM$_{2.5}$ mass and major chemical components in Shanghai, based on the MEIC (left column) and REAS3 (right column) emission inventories from November 9 to December 1, 2018. Black dots are the observed concentrations; solid lines are predicted concentrations in the grid where the observation site is located; shaded areas represent the concentration ranges from 3×3 grids with the observation site in the center grid. Units are µg m^{-3}.

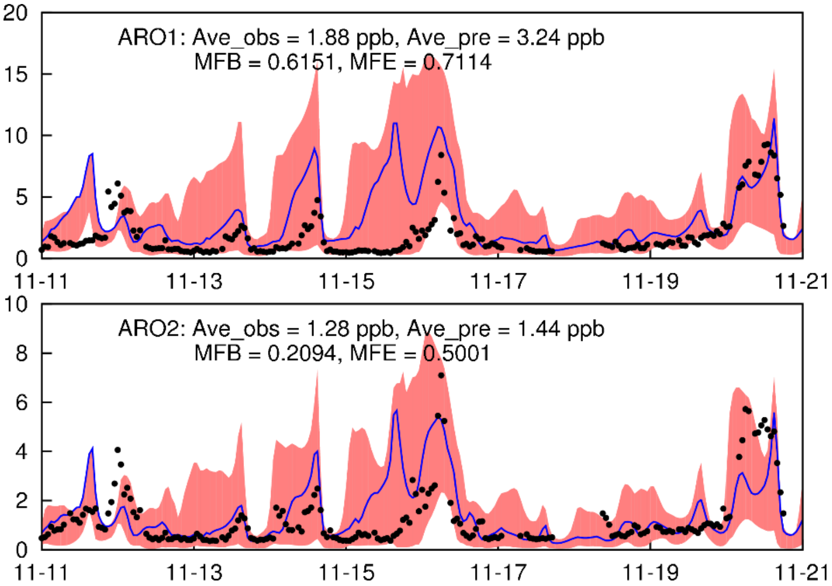


Figure S4-7 Time series of predicted VOCs using the MEIC and REAS3 emissions in Shanghai, from November 11 to 20, 2018. Observations are based on hourly individual VOCs lumped into SARPC11 model species. Black dots are the observed concentrations; solid lines are predicted concentrations in the grid where the observation site is located; shaded areas represent the concentration ranges from 3×3 grids with the observation site in the center grid. Units are ppb.

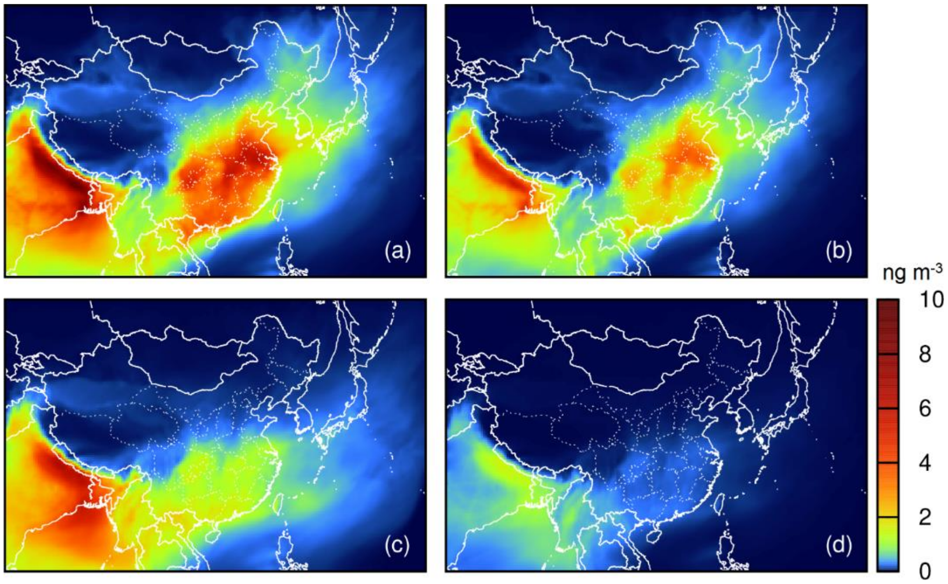


Figure S4-8 Regional distribution of modeled monthly average DHOPA from (a) ARO1 high-NOx pathway, (b) ARO2 high-NOx pathway, (c) ARO1 low-NOx pathway, and (d) ARO2 low-NOx pathway.

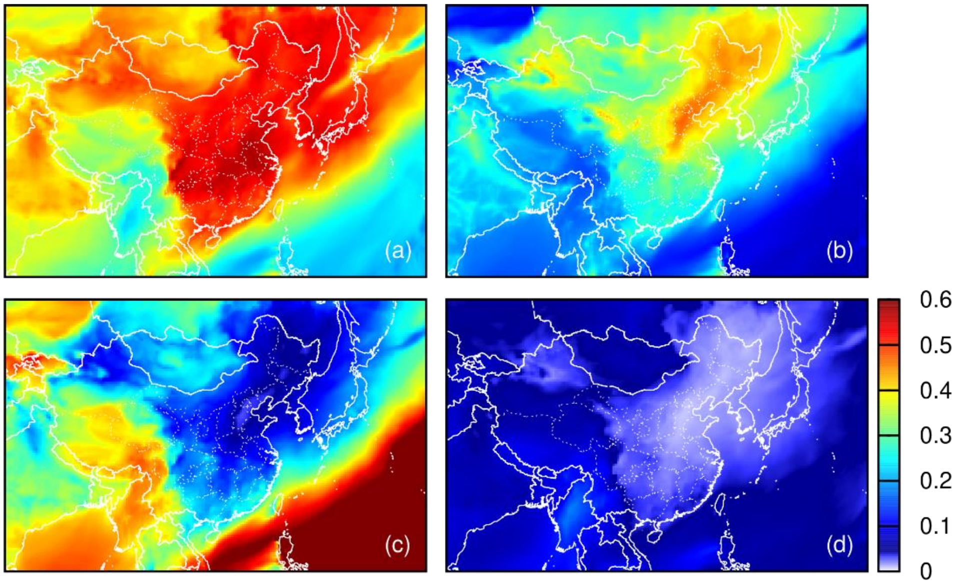


Figure S4-9 Fractional contributions to modeled monthly average DHOPA from (a) ARO1 high-NO_x pathway, (b) ARO2 high-NO_x pathway, (c) ARO1 low-NOx pathway, and (d) ARO2 low-NO_x pathway.

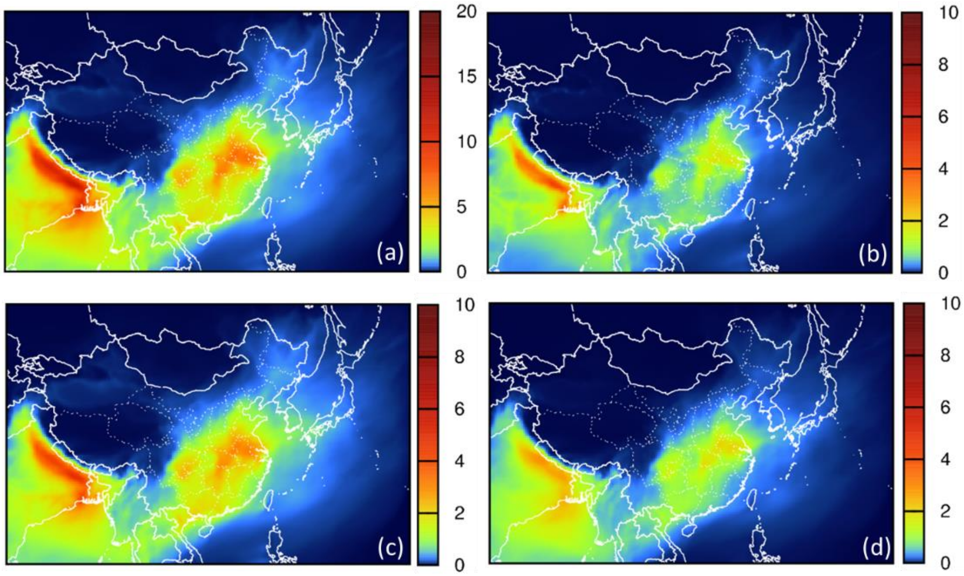


Figure S4-10 Monthly averaged concentration of (a) total aromatics SOA and contributions to total SOA from (b) semivolatile components, (c) surface uptake of glyoxal and methylglyoxal SOA, and (d) oligomers. Units are μg m^{-3}.

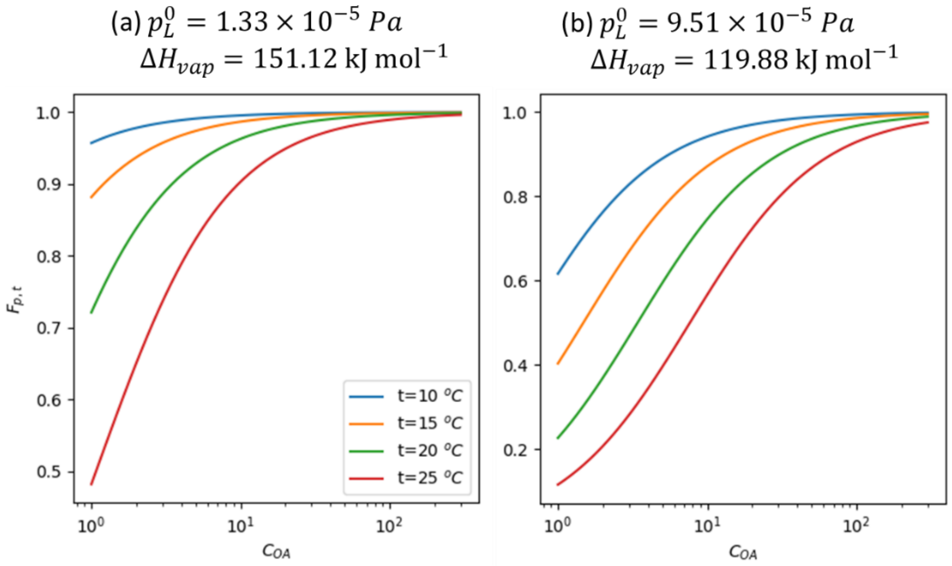


Figure S4-11 Fraction of DHOPA in the particle phase as a function of the OA loading estimated based on the thermodynamic data using (a) E-AIM method 1 and (b) E-AIM method 2 at different temperatures.

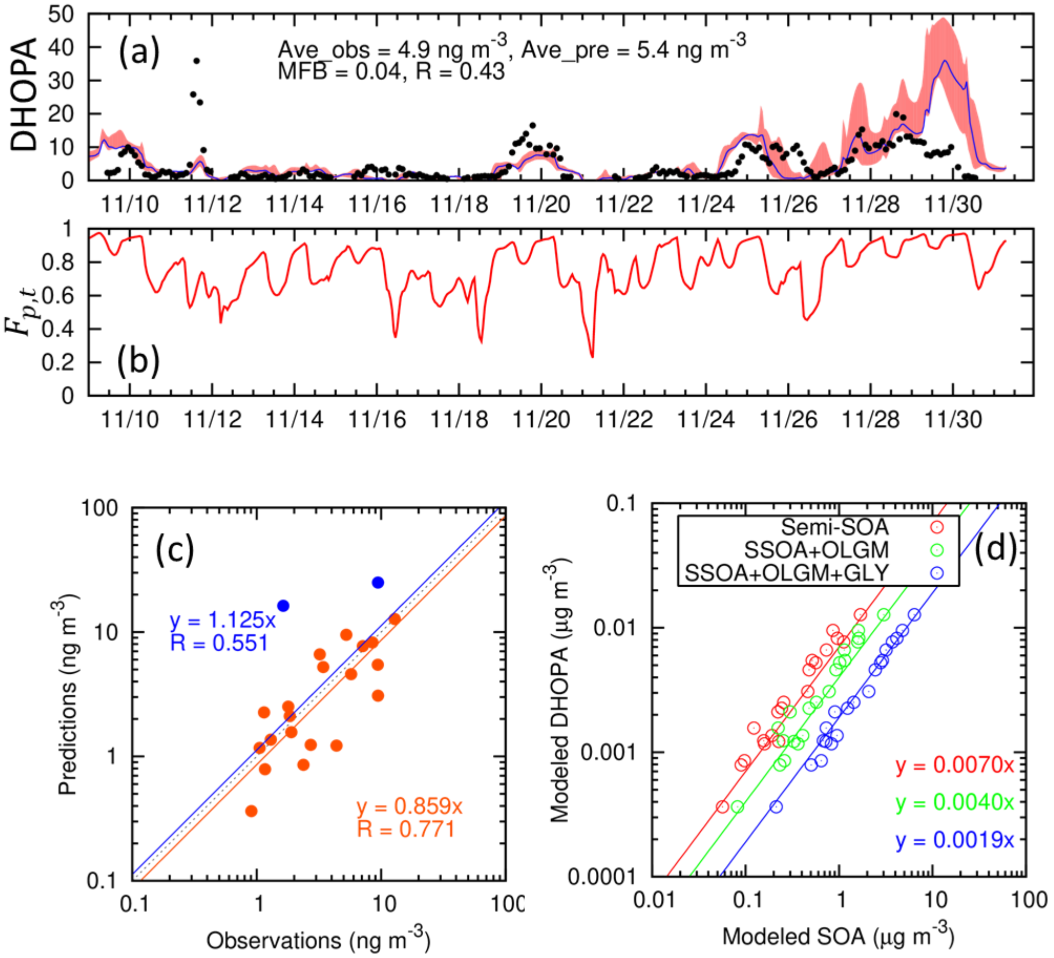


Figure S4-12 (a) Predicted and observed hourly DHOPA concentrations, (b) modeled fraction of DHOPA in the particle phase, (c) predicted vs. observed daily average DHOPA, and (d) modeled mass ratio of DHOPA to aromatic SOA with different SOA components. R is the Pearson correlation coefficient. Shaded areas in (a) represent the concentration ranges from the 3×3 grids with the observation site in the center grid. This set of results is different from the base case in that the DHOPA thermodynamic data are generated using E-AIM Method 2.

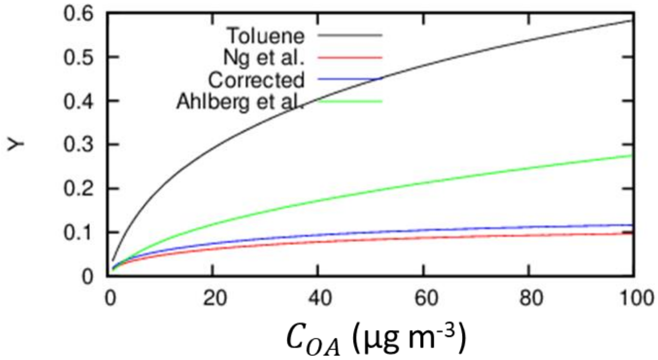


Figure S4-13 Comparison of xylene SOA yield based on Ng et al., Ng et al. with vapor wall-loss correction, Ahlberg et al., and the toluene SOA yield used in CMAQ.

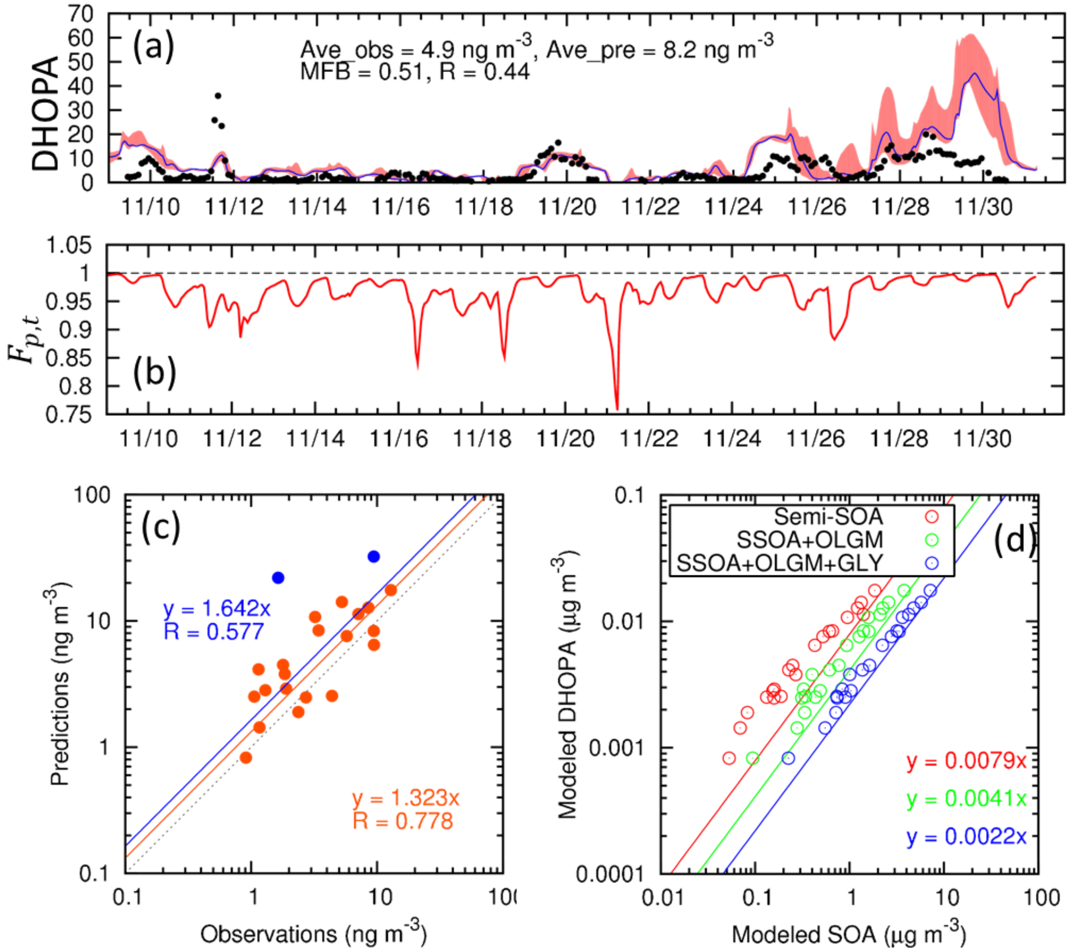


Figure S4-14 Same as Figure S4-12, but for sensitivity Case 2, in which DHOPA thermodynamic data are generated using E-AIM Method 1 and xylene SOA yields are taken to be the same as toluene SOA yields.

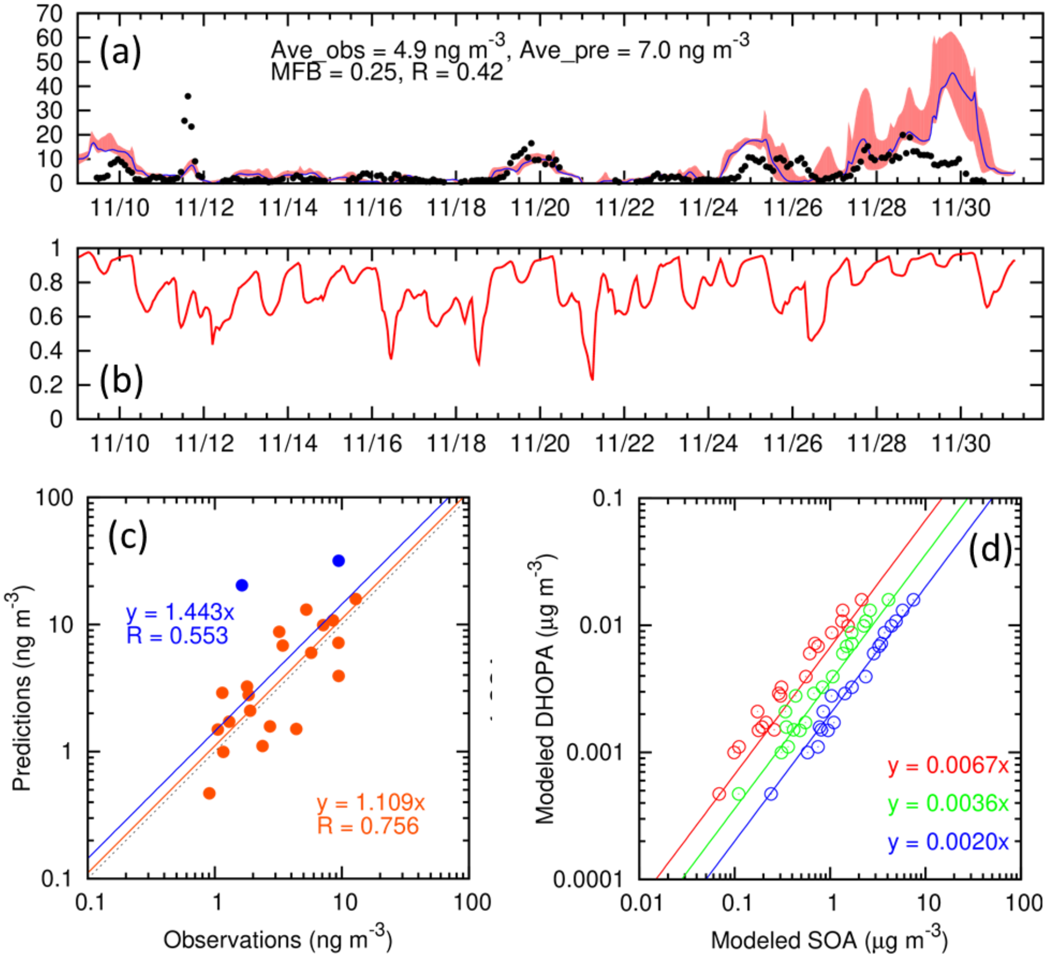


Figure S4-15 Same as Figure S4-12, but for sensitivity Case 2, in which DHOPA thermodynamic data are generated using the E-AIM Method 1 and xylene SOA yields are taken to be the same as toluene SOA yields.

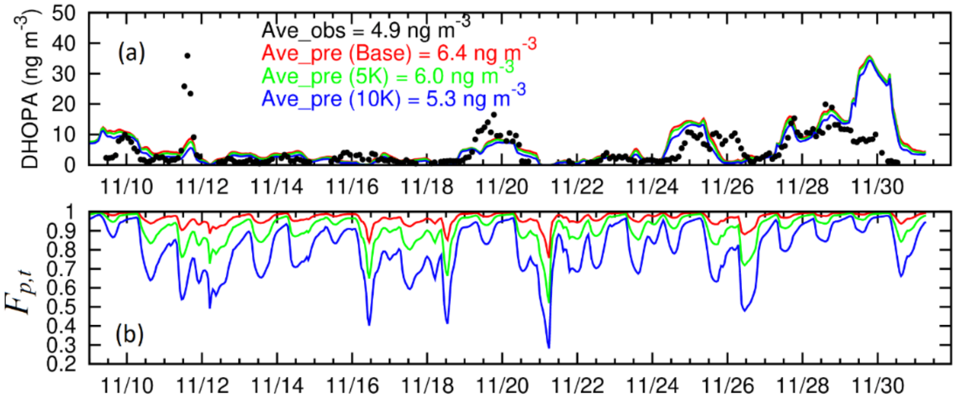


Figure S4-16 (a) Predicted hourly DHOPA with the base case (red), sensitivity cases with temperature increases of 5K (green) and 10K (blue), and observed hourly DHOPA concentrations (black dots). (b) Modeled fraction of DHOPA in the particle phase for three cases.

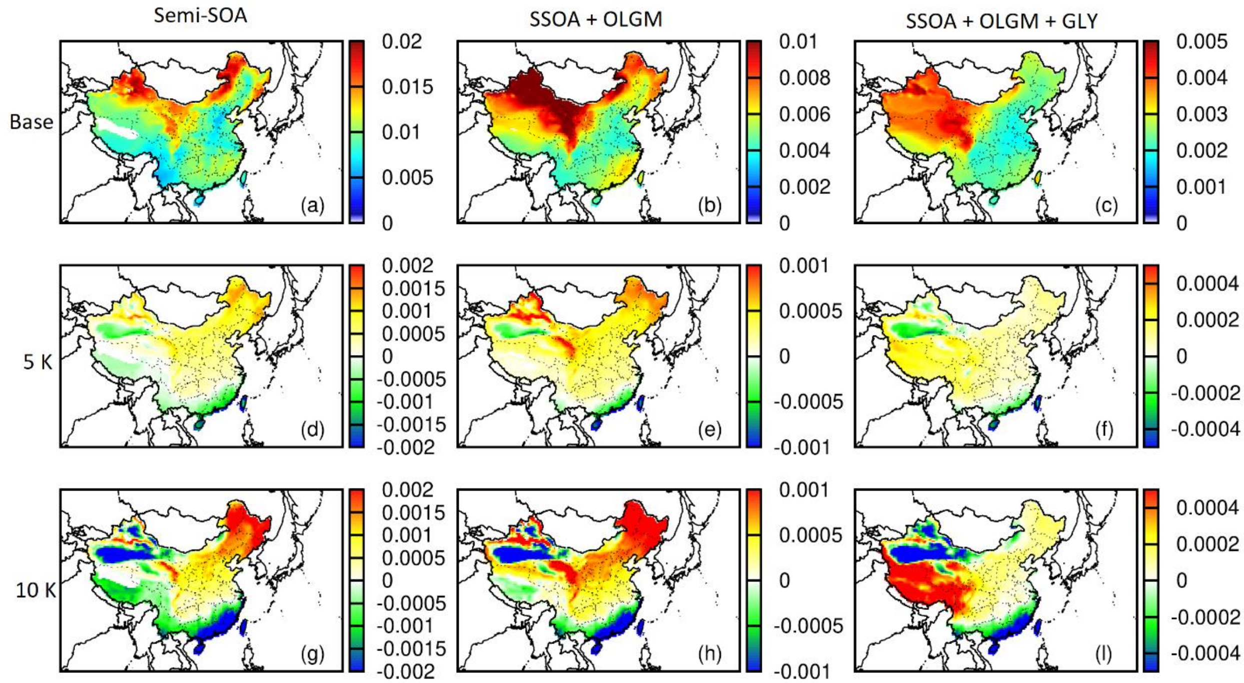


Figure S4-17 Regional distribution of monthly average $\boldsymbol{f_{SOA}}$ from (a-c) the base case; (b-f) the difference in $\boldsymbol{f_{SOA}}$ between Sen. Case 4 and the base case (Sen. Case 4 – base case); and (g-l) the difference in $\boldsymbol{f_{SOA}}$ between Sen. Case 5 and the base case (Sen. Case 5 – base case).

APPENDIX D

Table S5-1 Heterogeneous formation of isoprene SOA components through surface uptake and heterogeneous reactions in the aerosol aqueous phase, along with their parent hydrocarbons, nucleophiles, acids, and rate constants for the ring opening reactions.

Species	Parent hydrocarbon	Nucleophile	Acid	$k_{i,j}$ (M^{-2} s^{-1})
2-MT	IEPOX	Water	H^+	9.00×10^{-4}
	IEPOX	Water	HSO_4^-	1.31×10^{-5}
	IEPOX	Water	NH_4^+	3.10×10^{-7}
AIEOS	IEPOX	SO_4^{2-}	H^+	8.83×10^{-3}
	IEPOX	SO_4^{2-}	HSO_4^-	2.92×10^{-6}
2-MG	MAE	Water	H^+	9.00×10^{-4}
	MAE	Water	HSO_4^-	1.31×10^{-5}
AIMOS	MAE	SO_4^{2-}	H^+	2.00×10^{-4}
	MAE	SO_4^{2-}	HSO_4^-	2.92×10^{-6}

2-MT: 2-methyltetrols; 2-MG: 2-methylglyceric acid; AIEOS: isoprene-derived organosulfate; AIMOS: MAE-derived organosulfate.

Table S5-2 Summary of simulations in this study.

Case	Base case$^{\$}$	NV/Het. Chem.	Sens. 1	Sens. 2
Tracer Volatility	Semi-volatile	Non-volatile	Semi-volatile	Semi-volatile
Pathways	Het. Chem. + Additional	Het. Chem.	Het. Chem. + Additional	Het. Chem. + Additional
ΔH_{vap} (kJ mol^{-1}) (2-MG/2-MT)	130.0 / 117.7	-	43.2 / 38.4	130.0 / 117.7
Tracer Yield (2-MG/2-MT)	0.0026 / 0.084	-	0.0026 / 0.084	0.0078/ 0.168

Table S5-3 Model performance of major meteorological parameters in the PRD region [a].

Parameter	Month	Average observation	Average prediction	MB [e]	GE [f]	RMSE [g]
TEMP [b] (K)	Jun.	300.88	300.26	-0.62	1.67	1.98
	Jul.	301.80	300.55	-1.25	2.09	2.79
	Aug.	301.95	301.61	-0.34	1.84	2.16
RH (%)	Jun.	81.04	87.71	6.67	10.80	7.96
	Jul.	78.04	86.66	8.62	11.99	9.60
	Aug.	76.96	81.62	4.66	10.62	7.34
WSPD [c] (m s^{-1})	Jun.	3.81	4.75	0.94	1.67	3.19
	Jul.	3.70	4.74	1.04	1.65	2.98
	Aug.	3.27	4.10	0.82	1.62	3.79
WDIR [d] (°)	Jun.	156.10	158.55	12.49	35.33	22.08
	Jul.	163.39	170.70	11.19	37.46	23.24
	Aug.	189.02	186.96	11.63	48.35	26.03

[a] 40 surface monitoring stations from NOAA NCDC are available in the PRD region in 2012.

[b] WRF predicted temperature at height of 2 m.

[c] WRF predicted wind speed at height of 10 m.

[d] WRF predicted wind direction at height of 10 m.

[e] Mean Bias: $MB = \sum_i^N (pre_i - obs_i)/N.$

[f] Gross Error: $GE = \sum_i^N |pre_i - obs_i|/N.$

[g] Root mean square error: $RMSE = \sqrt{\sum_i^N (pre_i - obs_i)^2/N}.$

Table S5-4 Model performance of 2-MG and 2-MT in the PRD for each site.

Site	Average observation (ng m^{-3})	Case	Average prediction (ng m^{-3})	MFB	MFE	R
Dongguan	2-MG: 1.1	Base case	0.6	-1.012	1.081	0.690
		Repartitioned	1.6	0.055	0.594	0.852
		NV/Het. Chem.	6.7	0.783	1.118	0.392
		Sens. 1	3.6	0.644	0.789	0.838
		Sens. 2	4.1	0.881	0.882	0.904
Guangzhou	2-MG: 1.8	Base case	0.7	-1.032	1.089	0.859
		Repartitioned	2.0	0.069	0.494	0.837
		NV/Het. Chem.	6.3	0.745	0.962	0.439
		Sens. 1	4.6	0.697	0.784	0.867
		Sens. 2	5.1	0.906	0.943	0.916
Nanhai	2-MG: 2.3	Base case	1.1	-0.898	1.047	0.873
		Repartitioned	2.4	-0.135	0.319	0.983
		NV/Het. Chem.	12.5	0.816	1.011	0.746
		Sens. 1	5.7	0.587	0.717	0.982
		Sens. 2	6.1	0.722	0.722	0.976
Nansha	2-MG: 1.1	Base case	0.4	-0.819	0.943	0.653
		Repartitioned	0.8	-0.237	0.641	0.763
		NV/Het. Chem.	2.5	0.096	0.846	0.236
		Sens. 1	1.3	0.126	0.480	0.739
		Sens. 2	2.3	0.713	0.713	0.724
Dongguan	2-MT: 20.6	Base case	15.2	-0.815	1.064	0.660
		Repartitioned	35.8	0.168	0.744	0.807
		NV/Het. Chem.	44.0	0.267	0.869	0.364
		Sens. 1	74.6	0.696	0.969	0.806
		Sens. 2	82.0	0.809	1.008	0.833
Guangzhou	2-MT: 46.9	Base case	19.0	-0.988	1.162	0.537
		Repartitioned	48.3	-0.022	0.661	0.588
		NV/Het. Chem.	38.9	-0.142	0.689	0.455
		Sens. 1	101.2	0.518	0.745	0.603
		Sens. 2	112.7	0.669	0.832	0.589
Nanhai	2-MT: 31.4	Base case	20.9	-0.688	1.135	0.675
		Repartitioned	45.2	0.065	0.885	0.816
		NV/Het. Chem.	70.8	0.182	0.874	0.530
		Sens. 1	96.6	0.586	0.925	0.818
		Sens. 2	100.7	0.644	0.953	0.812
Nansha	2-MT: 16.5	Base case	18.0	-0.752	1.049	0.841
		Repartitioned	24.8	-0.218	0.679	0.888
		NV/Het. Chem.	20.4	0.025	0.756	0.886
		Sens. 1	40.0	0.162	0.625	0.886
		Sens. 2	56.2	0.497	0.688	0.895

Table S5-5 Model performance of 2-MG across the country for each site.

Site	Average observation (ng m^{-3})	Case	Average prediction (ng m^{-3})	MFB	MFE
Hailun (HL)	11.8	Base case	1.8	-1.501	1.501
		Repartitioned	8.1	-0.504	0.559
		NV/Het. Chem.	9.2	-0.974	1.424
		Sens. 1	8.2	-0.552	0.656
		Sens. 2	19.2	0.376	0.393
Tongyu (TYU)	25.0	Base case	1.2	-1.779	1.779
		Repartitioned	6.4	-1.157	1.157
		NV/Het. Chem.	8.7	-1.140	1.140
		Sens. 1	7.0	-1.107	1.107
		Sens. 2	14.8	-0.586	0.961
Beijing (BJ)	24.7	Base case	2.0	-1.694	1.694
		Repartitioned	9.8	-0.834	0.834
		NV/Het. Chem.	15.1	-0.548	0.950
		Sens. 1	12.4	-0.633	0.640
		Sens. 2	19.6	-0.221	0.346
Taiyuan (TY)	12.5	Base case	2.0	-1.391	1.391
		Repartitioned	8.7	-0.327	0.481
		NV/Het. Chem.	10.8	-0.280	0.683
		Sens. 1	9.0	-0.314	0.493
		Sens. 2	18.7	0.362	0.735
Dunhuang (DH)	12.5	Base case	0.04	-1.976	1.976
		Repartitioned	0.2	-1.875	1.875
		NV/Het. Chem.	0.5	-1.682	1.682
		Sens. 1	0.2	-1.877	1.877
		Sens. 2	0.7	-1.706	1.706
Shapotou (SPT)	12.5	Base case	0.7	-1.826	1.826
		Repartitioned	3.3	-1.292	1.292
		NV/Het. Chem.	10.0	-0.846	1.195
		Sens. 1	3.3	-1.282	1.282
		Sens. 2	5.8	-0.832	0.832
Hefei (HF)	19.3	Base case	1.4	-1.600	1.600
		Repartitioned	8.1	-0.665	0.918
		NV/Het. Chem.	8.5	-0.596	0.807
		Sens. 1	12.2	-0.335	0.704
		Sens. 2	20.6	0.097	0.546

Table S5-5 Continued.

Site	Average observation (ng m^{-3})	Case	Average prediction (ng m^{-3})	MFB	MFE
Wuxi (WX)	11.9	Base case	1.0	-1.605	1.605
		Repartitioned	5.8	-0.713	0.995
		NV/Het. Chem.	7.3	-0.450	0.988
		Sens. 1	9.6	-0.396	1.018
		Sens. 2	15.3	-0.115	1.113
Qianyunzhou (QYZ)	11.7	Base case	2.1	-1.544	1.544
		Repartitioned	10.2	-0.672	1.040
		NV/Het. Chem.	14.5	-0.565	1.099
		Sens. 1	12.5	-0.480	0.919
		Sens. 2	22.4	0.013	0.887
Kunming (KM)	5.1	Base case	3.8	-0.526	1.081
		Repartitioned	12.4	0.306	1.192
		NV/Het. Chem.	23.1	0.693	1.431
		Sens. 1	8.2	0.008	1.032
		Sens. 2	16.3	0.526	1.326
Xishuangbanna (BN)	11.5	Base case	0.1	-1.964	1.964
		Repartitioned	0.6	-1.802	1.802
		NV/Het. Chem.	2.3	-1.365	1.365
		Sens. 1	0.5	-1.831	1.831
		Sens. 2	1.3	-1.579	1.579
Linzhi (LZ)	5.0	Base case	0.2	-1.592	1.592
		Repartitioned	0.5	-1.396	1.396
		NV/Het. Chem.	0.4	-1.502	1.502
		Sens. 1	0.1	-1.801	1.801
		Sens. 2	1.1	-1.002	1.260
Namco (mCo)	3.6	Base case	0.1	-1.945	1.945
		Repartitioned	0.2	-1.776	1.776
		NV/Het. Chem.	0.4	-1.566	1.566
		Sens. 1	0.04	-1.963	1.963
		Sens. 2	0.5	-1.579	1.579
Sanya (SY)	6.8	Base case	0.1	-1.853	1.853
		Repartitioned	0.9	-1.338	1.338
		NV/Het. Chem.	0.6	-1.546	1.546
		Sens. 1	1.2	-1.238	1.305
		Sens. 2	2.6	-0.821	1.269

Table S5-6 Model performance of 2-MT across the country for each site.

Site	Average observation (ng m^{-3})	Case	Average prediction (ng m^{-3})	MFB	MFE
Hailun (HL)	178.3	Base case	43.4	-1.221	1.221
		Repartitioned	156.4	-0.158	0.294
		NV/Het. Chem.	55.7	-1.025	1.025
		Sens. 1	154.5	-0.206	0.378
		Sens. 2	286.4	0.424	0.339
Tongyu (TYU)	118.6	Base case	25.2	-1.316	1.316
		Repartitioned	121.2	-0.104	0.498
		NV/Het. Chem.	57.6	-0.785	0.785
		Sens. 1	124.9	-0.107	0.511
		Sens. 2	217.3	0.428	1.143
Beijing (BJ)	98.7	Base case	15.5	-1.316	1.316
		Repartitioned	66.7	-0.197	0.465
		NV/Het. Chem.	38.4	-0.664	0.664
		Sens. 1	77.2	-0.069	0.478
		Sens. 2	114.3	0.272	0.925
Taiyuan (TY)	50.3	Base case	18.8	-0.701	0.849
		Repartitioned	77.3	0.461	0.769
		NV/Het. Chem.	37.5	-0.146	0.694
		Sens. 1	77.5	0.458	0.768
		Sens. 2	134.8	0.867	0.589
Dunhuang (DH)	100.2	Base case	3.5	-1.863	1.863
		Repartitioned	23.0	-1.327	1.327
		NV/Het. Chem.	11.9	-1.608	1.608
		Sens. 1	23.8	-1.323	1.323
		Sens. 2	44.4	-0.969	1.886
Shapotou (SPT)	48.4	Base case	13.8	-1.240	1.240
		Repartitioned	70.3	0.115	0.512
		NV/Het. Chem.	67.3	-0.508	1.075
		Sens. 1	70.9	0.094	0.593
		Sens. 2	114.2	0.616	1.584
Hefei (HF)	300.8	Base case	30.7	-1.388	1.388
		Repartitioned	142.0	-0.558	1.093
		NV/Het. Chem.	45.8	-1.198	1.289
		Sens. 1	183.9	-0.360	0.951
		Sens. 2	264.6	-0.092	1.240

Table S5-6 Continued.

Site	Average observation (ng m^{-3})	Case	Average prediction (ng m^{-3})	MFB	MFE
Wuxi (WX)	174.7	Base case	24.8	-1.165	1.427
		Repartitioned	119.4	-0.339	1.114
		NV/Het. Chem.	49.7	-0.594	1.352
		Sens. 1	160.0	-0.148	0.976
		Sens. 2	219.7	0.036	1.579
Qianyunzhou (QYZ)	152.2	Base case	38.9	-1.340	1.340
		Repartitioned	164.5	-0.314	0.809
		NV/Het. Chem.	89.0	-0.936	1.252
		Sens. 1	188.1	-0.153	0.679
		Sens. 2	288.4	0.207	1.319
Kunming (KM)	86.6	Base case	42.4	-0.651	1.125
		Repartitioned	125.4	0.200	0.938
		NV/Het. Chem.	125.9	0.175	0.988
		Sens. 1	95.0	-0.039	0.960
		Sens. 2	176.7	0.467	0.658
Xishuangbanna (BN)	47.9	Base case	6.8	-1.478	1.478
		Repartitioned	41.4	-0.221	0.521
		NV/Het. Chem.	28.4	-0.523	0.577
		Sens. 1	36.3	-0.337	0.575
		Sens. 2	75.4	0.334	1.736
Linzhi (LZ)	23.9	Base case	14.4	-0.874	0.874
		Repartitioned	31.2	-0.118	0.544
		NV/Het. Chem.	5.1	-1.420	1.420
		Sens. 1	10.5	-1.094	1.094
		Sens. 2	60.0	0.475	0.466
Namco (mCo)	55.3	Base case	1.6	-1.880	1.880
		Repartitioned	7.2	-1.528	1.528
		NV/Het. Chem.	2.4	-1.787	1.787
		Sens. 1	1.7	-1.889	1.889
		Sens. 2	13.5	-1.214	1.477
Sanya (SY)	43.3	Base case	9.5	-1.107	1.107
		Repartitioned	61.2	0.090	1.124
		NV/Het. Chem.	16.3	-0.772	0.940
		Sens. 1	72.7	0.200	1.113
		Sens. 2	118.7	0.503	1.779

Table S5-7 f_{SOA} values at 14 sites across China

Site	Ave_tracer (ng m^{-3})	$f_{SOA,1}$	$f_{SOA,2}$	$f_{SOA,3}$	$f_{SOA,4}$
Beijing	0.023	0.075	0.069	0.029	0.009
Dunhuang	0.003	0.112	0.065	0.013	0.005
Hailun	0.041	0.140	0.123	0.047	0.016
Hefei	0.038	0.064	0.059	0.029	0.009
Kunming	0.114	0.730	0.473	0.149	0.059
Linzhi	0.031	0.862	0.676	0.174	0.076
mCo	0.006	0.760	0.518	0.085	0.032
QYZ	0.044	0.072	0.067	0.035	0.012
Sanya	0.014	0.068	0.063	0.026	0.009
Shapotou	0.045	0.357	0.231	0.067	0.024
Taiyuan	0.038	0.139	0.123	0.045	0.014
Tongyu	0.021	0.119	0.101	0.032	0.010
Wuxi	0.021	0.058	0.053	0.026	0.008
XSBN	0.049	0.326	0.213	0.109	0.050

$f_{SOA,1}$ for isoprene semivolatile SOA only.

$f_{SOA,2}$ for isoprene SOA including semivolatile products and species formed from IEPOX and MAE in the aerosol liquid water.

$f_{SOA,2}$ for isoprene SOA including semivolatile products, corresponding oligomers, and species formed from IEPOX and MAE in the aerosol liquid water.

$f_{SOA,2}$ for isoprene SOA including semivolatile products, corresponding oligomers, species formed from IEPOX and MAE in the aerosol liquid water, and dicarbonyl products (GLY/MGLY).

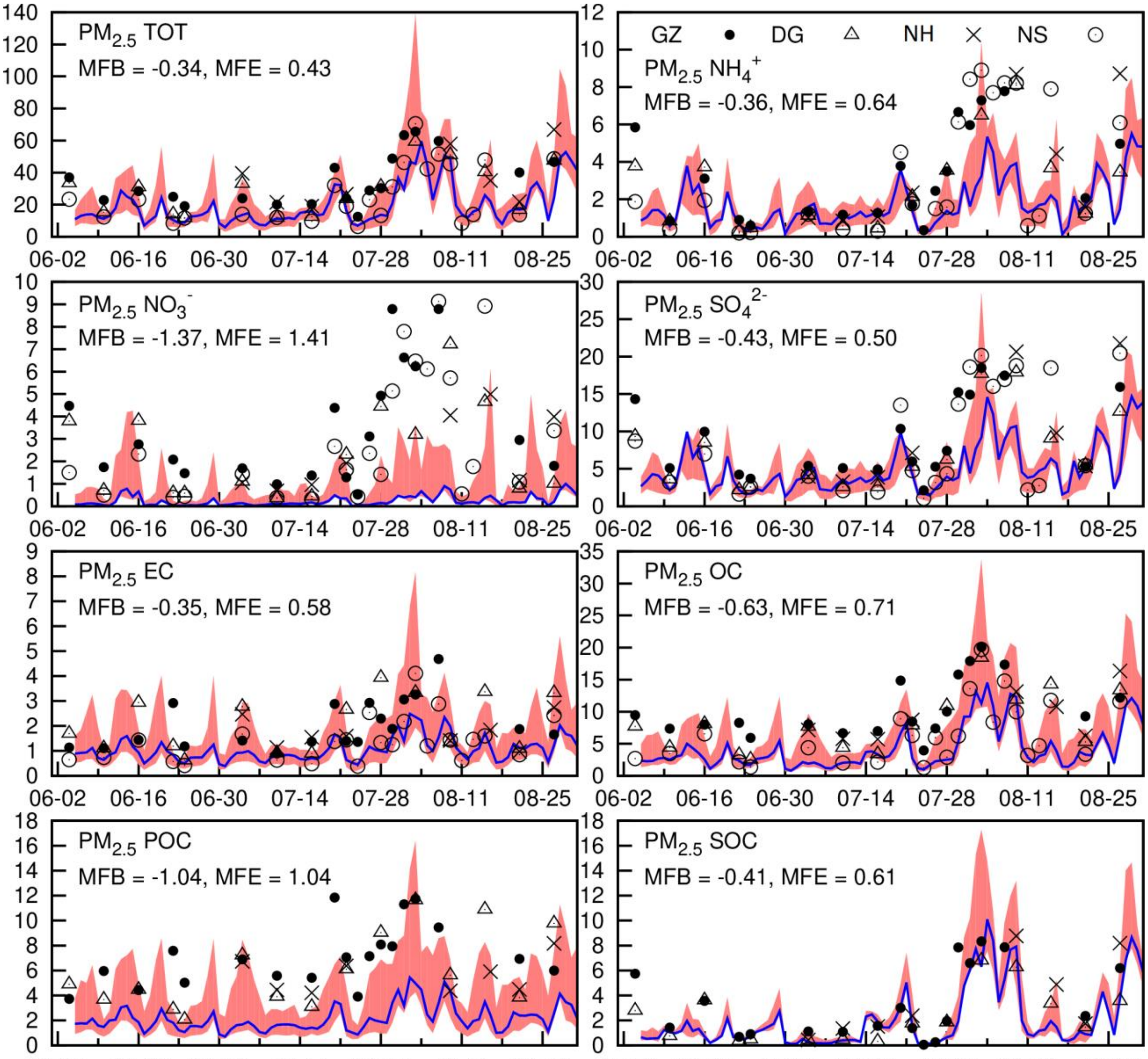


Figure S5-1 Model performance of $PM_{2.5}$ mass concentrations and major chemical components in the PRD region. The blue line represents the predictions from the grid cell of the Guangzhou site. The red shaded area represents the range of concentrations from the 3×3 grid cells around the Guangzhou (GZ) site, where the other three sites, Dongguan (DG), Nanhai (NH), and Nansha (NS), are located.

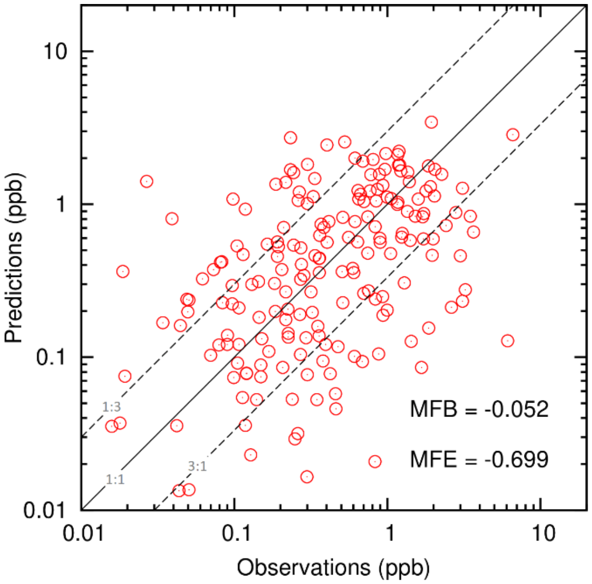


Figure S5-2 Predicted and observed daily isoprene concentrations at 20 sites across China from June to August 2012. The locations of the monitoring sites are shown in Figure S5-5(b).

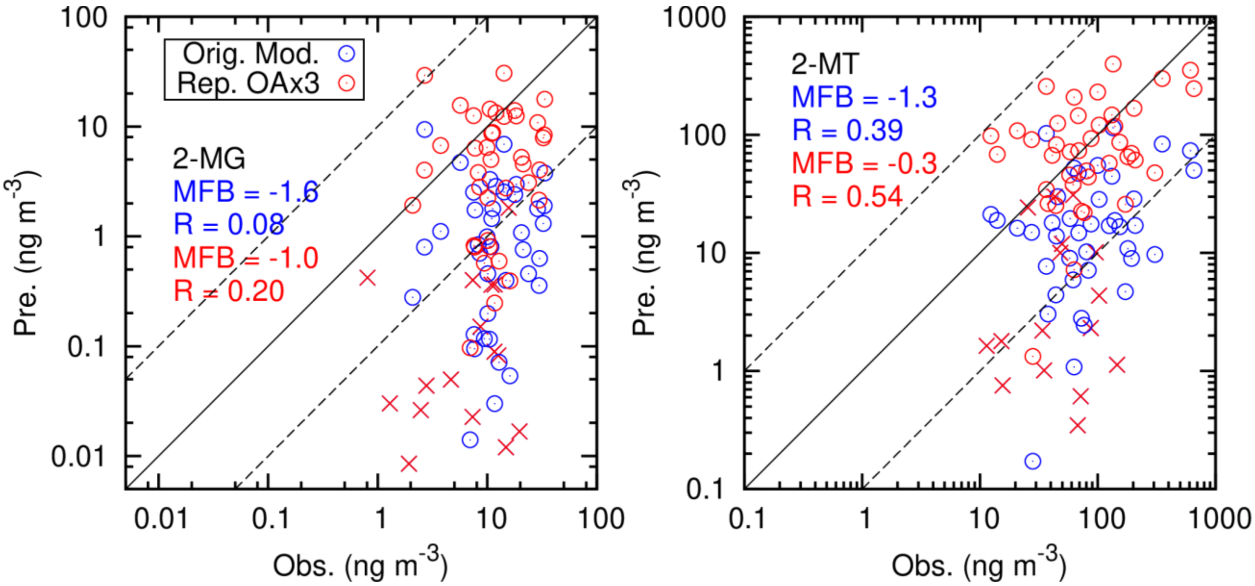


Figure S5-3 Predicted and observed daily average 2-MT and 2-MG at 14 sites. The blue markers show original model predictions, and the red markers show repartitioned predictions by increasing the predicted organic aerosol (OA) by a factor of 3. The cross symbols indicate data at three remote sites, Linzhi (LZ) and Namco (mCo) in Tibet and Dunhuang (DH) in Gansu.

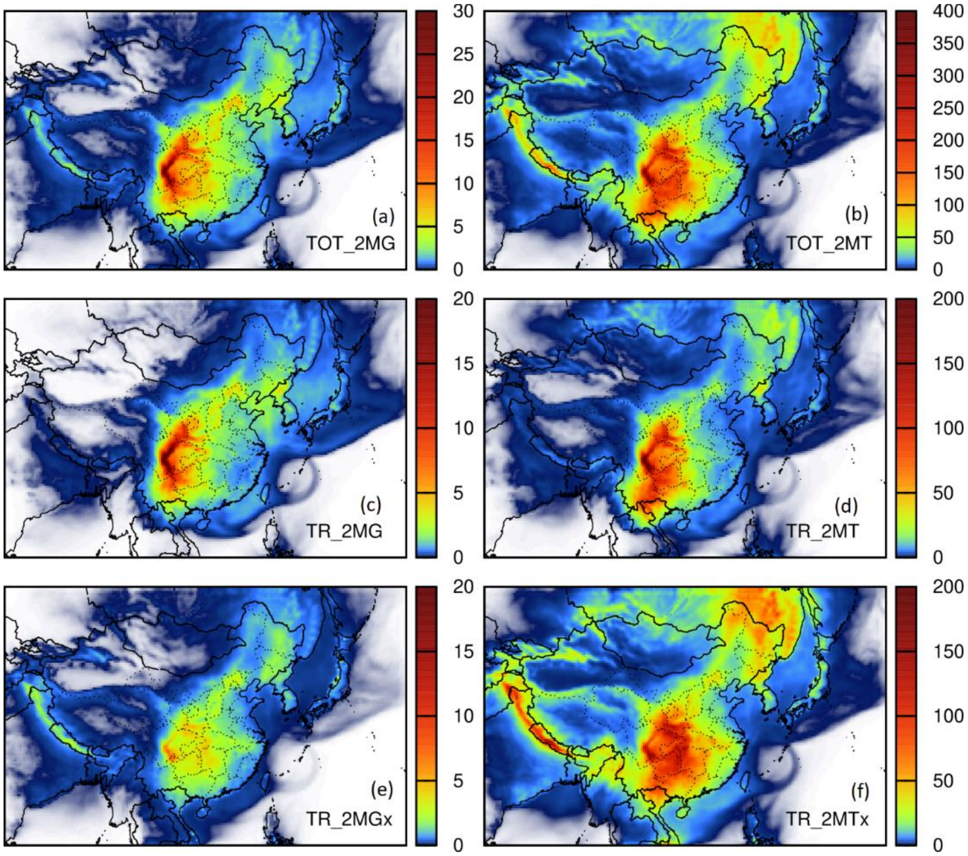


Figure S5-4 The average model predicted total 2-MG and 2-MT (a, b), their formation in the aqueous phase following the multiphase chemistry (c, d), and formation from additional pathway (e, f) in June – August 2012.

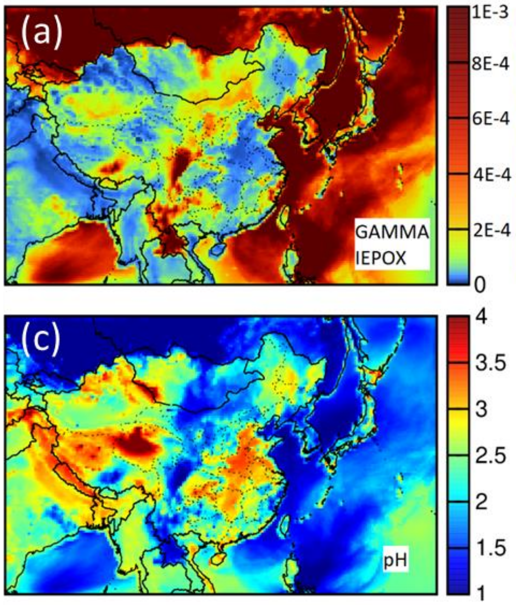


Figure S5-5 Predicted average uptake coefficient of (a) IEPOX and (b) MAE, and (c) pH of aerosol water based on ISORROPIA for June – August 2012.

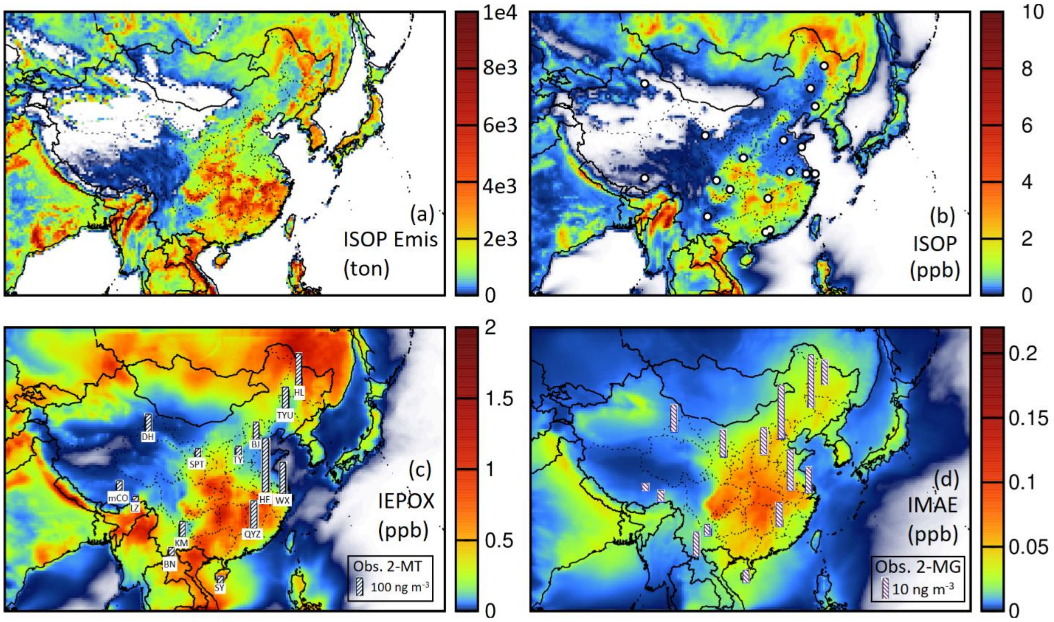


Figure S5-6 The average modeled (a) isoprene emissions, (b) isoprene concentrations, (c) gas-phase IEPOX, and (d) gas phase MAE, in June – August 2012. Circles on panel (b) show the location of the isoprene monitors. The 14 sites where 2-MT and 2-MG were measured are: Hailun (HL), Tongyu (TYU), Beijing (BJ), Taiyuan (TY), Hefei (HF), Wuxi (WX), Qianyunzhou (QYZ), Sanya (SY), Kunming (KM), Xishuangbanna (BN), Sapotou (SPT), Dunhuang (DH), Linzhi (LZ), and Namco (mCO).

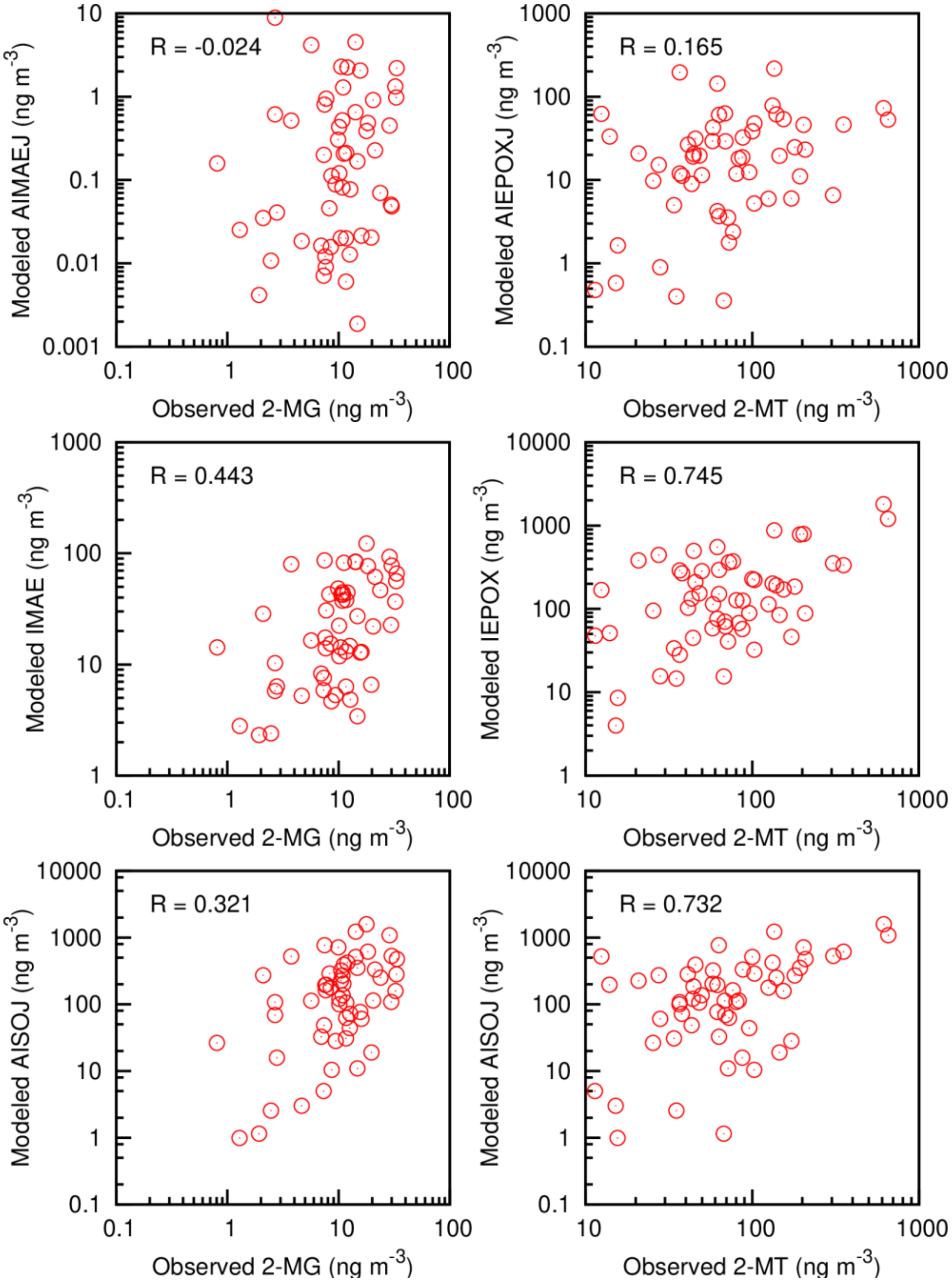


Figure S5-7 Correlation between observed daily 2-MG and 2-MT from Ding et al. and the corresponding model predictions of the corresponding heterogenous chemistry products (AIMAJ=TR_2MG+AIMOS, AIEPOXJ=TR_2MT+AIEOS), MAE and IEPOX in the gas, and the isoprene semi-volatile SOA (AISOJ) in the aerosol phase. R is the Pearson correlation coefficient.

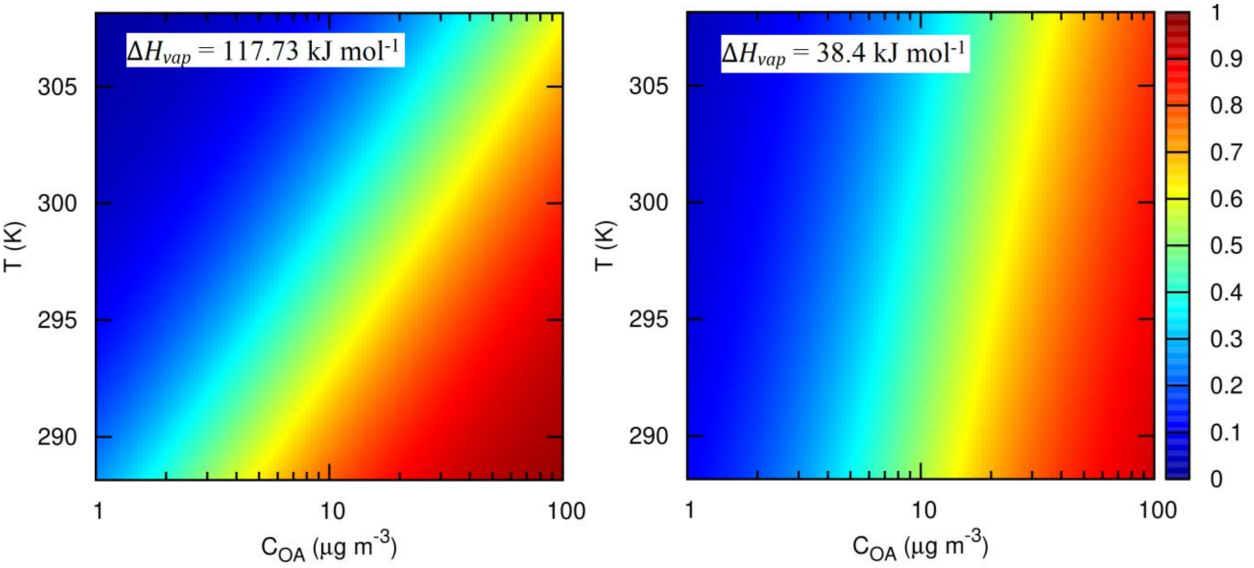


Figure S5-8 Theoretically calculated fraction ($F_{p,t}$) of 2-MT in the particle phase with OA loadings and temperature change.

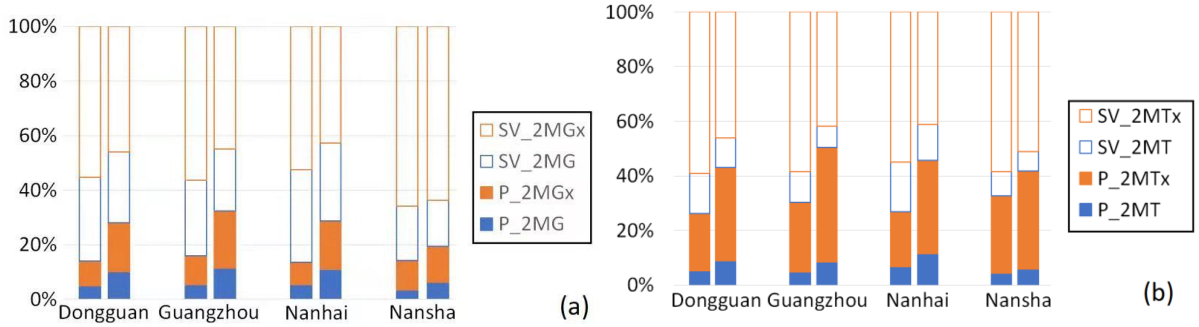


Figure S5-9 Gas-particle distribution of 2-MG (a) and 2-MT (b) in the base case simulation (left bars) and the sensitivity simulation with lower ΔH_{vap} estimations (Sens. 1; right bars), based on repartitioned concentrations.

APPENDIX E

Section S6.1 Two-product SOA model parameters

The traditional semi-volatile SOA yields in the original AERO6 module of the CMAQ model are mostly based on the data from historical chamber experiments,[30, 138] which generally lead to lower SOA yields due to the lack of vapor wall-loss correction.[161] In this study, the semivolatile product yields (β) for long-chain alkanes, toluene, benzene, monoterpenes, isoprene, and sesquiterpenes are adjusted by considering vapor wall-loss, which were used in the previous study.[134] The ethylbenzene (S11D model species C2B) and other lumped less reactive aromatics (S11D model species ARO1) yields are kept the same as toluene. Yields of xylenes under high-NO_x conditions are based on the 2-product fitting of the 5-bin volatility basis set (VBS) parameters reported by Ahlberg et al.,[145] which included vapor wall-loss correction, and the yields of other lumped more reactive aromatics (ARO2) are the same as xylenes. Naphthalene (S11D model species NAPH) and methylnaphthalene (S11D model species MNAP) yields are determined based on the 2-product model parameters reported by Chan et al.[231] and vapor wall-loss bias correction for naphthalene.[161]

The yields from lumped other monoterpene species (S11D model species TERP) photooxidation with OH radical are obtained from Carlton et al.[212] and corrected for vapor wall-loss[161]. α-Pinene and d-limonene SOA yields with OH and NO_3 radicals are based on the most recent experimental data from Mutzel et al.[206] β-Pinene, Δ^3-carene, and sabinene yields with OH and NO_3 radicals are derived from Griffin et al.'s experiments [138], and the reported β values are corrected for the SOA density[212] and vapor wall-loss.[161] The yields of the monoterpene species with O_3 are based on Takekawa and Takasu.[205] The yields for monoterpene species with $O(^3P)$ are taken to be the same as OH reactions, as no experimental data is available. Table S1 lists all the yields and the saturation concentration (C^*) values used in this study. The enthalpy of vaporization (ΔH_{vap}) for semivolatile products generated from ozonolysis are also obtained from Takekawa and Takasu[205] (α-pinene 37 kJ mol^{-1}, β-pinene 33 kJ mol^{-1}, d-limonene 49 kJ mol^{-1}, Δ^3-carene 40 kJ mol^{-1}, sabinene 40 kJ mol^{-1}, and other terpenes 41 kJ mol^{-1}). In addition, ΔH_{vap} values for semivolatile products from the OH and NO_3 pathways are kept the same as the lumped monoterpene SOA in the original CMAQ model (40 kJ mol^{-1}).

Section S6.2 Meteorology and emission inputs

The meteorological inputs are generated using WRFv4.2 with initial and boundary conditions from the ERA5 Reanalysis 0.25° gridded data (available at https://rda.ucar.edu/datasets/ds633.0/). The land use/land cover and topographical data are based on the 30 s resolution default WRF input dataset. Reanalysis nudging is enabled to improve the agreement between the predicted and observed meteorological parameters.[119] Other major physics options for the WRF simulations are described by Zhang et al.[120]

Anthropogenic emissions are based on the Regional Emission inventory in Asia v3.1 (REAS3)[121] using an in-house emission processor. The selected detailed speciation profiles from the SPECIATE database developed by the US EPA are used to estimate emissions of CMAQ-ready VOCs. Windblown dust emissions in the entire domain are generated by the inline module.[86] Biogenic emissions are produced by the Model for Emissions of Gaseous and Aerosols from Nature (MEGAN) v2.10.[123] The MEGAN model is modified to output emissions of the individual monoterpene species and other explicitly represented model species for detailed S11D mechanism. In addition, emissions of all monoterpenes (MTs) and sesquiterpenes (SQTs) are increased by 75% to match those reported by Qin et al., which used China-specific plant functional type data to improve biogenic emission estimations.[176] Figure S2 shows the average emission rates of total MTs and SQTs, and emission rates for individual MT species in the model are shown in Figure S3.

Table S6-1 Two-product model parameters to model terpene SOA from different precursors and oxidation pathways

SOA name	Precursor	SVOC1		SVOC2 $^{\#}$		NVOC §
		β_1	C_1^*	β_2	C_2^*	β
ALK	Alkane	0.0865	0.020	-	-	-
XYL	Xylene	0.0883	7.430	0.5647	192.7	0.612
NAPH	Naphthalene	0.252	1.695	1.284	270.3	0.876
MNAP	1-Methylnaphthalene	0.600	9.091	-	-	0.816
ARO2	Other more reactive aromatics	0.0883	7.430	0.5647	192.7	0.612
TOL	Toluene	0.2545	8.024	0.7623	119.3	0.570
C2B	Ethylbenzene	0.2545	8.024	0.7623	119.3	0.570
ARO1	Other less reactive aromatics	0.2545	8.024	0.7623	119.3	0.570
BNZ	Benzene	0.090	0.302	1.110	111.1	0.666
ISOP	Isoprene	0.063	0.617	0.510	116.0	-
SQT	Sesquiterpene	1.537	24.98	-	-	-
oTERP_1	Other Terpenes + OH	0.1807	14.79	0.5905	133.7	-
oTERP_2	Other Terpenes + O_3	0.0533	0.46	0.1430	11.63	-
oTERP_3	Other Terpenes + NO_3	0.1807	14.79	0.5905	133.7	-
oTERP_4	Other Terpenes + O(3P)	0.1807	14.79	0.5905	133.7	-
APIN_1	α-Pinene + OH	0.0680	3.72	-	-	-
APIN_2	α-Pinene + O_3	0.0624	4.72	0.0845	4.76	-
APIN_3	α-Pinene + NO_3	0.0980	3.60	-	-	-
APIN_4	α-Pinene + O(3P)	0.0680	3.72	-	-	-
BPIN_1	β-Pinene + OH	0.2315	31.62	0.6965	284.9	-
BPIN_2	β-Pinene + O_3	0.0377	0.55	0.3107	204.1	-
BPIN_3	β-Pinene + NO_3	1.6900	61.35	-	-	-
BPIN_4	β-Pinene + O(3P)	0.2315	31.62	0.6965	284.9	-
DLIM_1	d-Limonene + OH	0.2910	22.73	-	-	-
DLIM_2	d-Limonene + O_3	0.1690	0.10	0.6747	29.41	-
DLIM_3	d-Limonene + NO_3	0.1890	5.38	-	-	-
DLIM_4	d-Limonene + O(3P)	0.2910	22.73	-	-	-
CARN_1	Δ^3-Carene + OH	0.2662	67.57	0.9888	689.7	-
CARN_2	Δ^3-Carene + O_3	0.1976	2.18	0.0845	158.73	-
CARN_3	Δ^3-Carene + NO_3	0.4343	109.9	1.2557	113.6	-
CARN_4	Δ^3-Carene + O(3P)	0.2662	67.57	0.9888	689.7	-
SABI_1	Sabinene + OH	0.1015	4.298	0.6478	291.5	-

Table S6-1 Continued

SOA name	Precursor	SVOC1		SVOC2 [#]		NVOC [§]
		β_1	C_1^*	β_2	C_2^*	β
SABI_2	Sabinene + O_3	0.0598	0.94	0.0052	71.43	-
SABI_3	Sabinene + NO_3	1.6900	86.96	-	-	-
SABI_4	Sabinene + O(3P)	0.1015	4.298	0.6478	291.5	-

[#] The SVOC1 and SVOC2 represent the semi-volatile products formed under high-NO_x conditions.

[§] The NVOC represent the non-volatile products formed under low-NO_x conditions

Table S6-2 Terpene SOA tracer vapor pressure estimation and saturation concentration estimation.

Tracer	p_L^0 (298K) [Pa]	C^* [μg m^{-3}]	Method	Reference
Pinic acid ($C_9H_{14}O_4$)	**1.98×10^{-4}**	**16.0**	E-AIM Method 1	-
	7.02×10^{-4}	56.7	E-AIM Method 2	-
	4.00×10^{-4}	32.4	Group contribution	Capouet and Müller [232]
	1.20×10^{-4}	9.7	UNIFAC	Pankow et al. [55]
	1.01×10^{-4}	8.2	-	Zhang et al. [233]
	4.27×10^{-5}	3.5	Experimental	Bilde and Pandis [208]
	1.24×10^{-4}	10.0	SIMPOL	Müller et al. [209]
	3.96×10^{-5}	3.2	Optimized fitting of chamber data	Jenkin [201]
Pinonic acid ($C_{10}H_{16}O_3$)	**7.19×10^{-5}**	**5.8**	UNIFAC	Pankow et al.[55]
	7×10^{-5} (296K)	5.6 (296K)	-	Bilde and Pandis [208]
	9.67×10^{-3}	780.6	E-AIM Method 1	-
	3.68×10^{-4}	29.4	Optimized fitting of chamber data	Jenkin [201]
3-methyl-1,2,3-butanetricarboxylic acid (MBTCA) ($C_8H_{12}O_6$)	**2.23×10^{-8}**	**1.8×10^{-3}**	Experimental	Kostenidou et al.[211]
	1.24×10^{-7}	0.01	SIMPOL	Müller et al. [209]
β-Caryophyllinic acid ($C_{14}H_{22}O_4$)	**2.14×10^{-7}**	**0.02**	E-AIM Method 1	-
	1.3×10^{-8}	1.3×10^{-3}	SPARC on-line calculator	Li et al. [234]
	6.6×10^{-6}	0.55	-	Alfarra, et al. [235]

Bold fonts are the C^* values used in the base case simulations.

Table S6-3 Model performance of the mass concentration of $PM_{2.5}$ mass and major components

Species	Mean Obs.	Mean Pred.	MFB[#]	MFE[^]	R[*]
Total $PM_{2.5}$	30.4	22.7	-0.34	0.43	0.70
$PM_{2.5}$ NH_4^+	3.2	1.7	-0.36	0.64	0.68
$PM_{2.5}$ NO_3^-	2.9	0.5	-1.37	1.41	0.53
$PM_{2.5}$ SO_4^{2-}	8.5	5.0	-0.43	0.50	0.67
$PM_{2.5}$ EC	1.8	1.3	-0.35	0.58	0.37
$PM_{2.5}$ OC	8.5	5.4	-0.57	0.68	0.70
$PM_{2.5}$ POC	6.5	2.2	-1.00	1.00	0.62
$PM_{2.5}$ SOC	2.9	2.3	-0.34	0.61	0.83

[#] Mean Fractional Bias: $MFB = \frac{1}{N}\sum_{i=1}^{N}\frac{2(P_i-O_i)}{P_i+O_i}$, N is the number of data points, P is observation, and O is observation.

[^] Mean Fractional Error: $MFE = \frac{1}{N}\sum_{i=1}^{N}\frac{2|P_i-O_i|}{P_i+O_i}$

[*] R: Pearson correlation coefficient

Table S6-4 Average mixing ratios of monoterpenes

Site	Site Type	Sampling Period	Monoterpenes (ppb_v)	Source
Beijing		**Jun. - Aug. 2012**	**0.04 – 0.16**[#]	**This work**
			0.08 – 0.31[^]	
	Urban	Nov. - Dec. 2016	0.57 ± 0.20	Zhang, et al.[215]
	Urban	Jun. 2017	0.16 ± 0.10	Zhang, et al.[215]
	Suburban	Dec. 2014	0.04 ± 0.04	Li, et al.[217]
	Urban	Feb. 2015	0.06 ± 0.06	Li, et al.[217]
	Suburban	Nov. 2014	0.05 – 0.09	Li, et al.[218]
Guangzhou		**Jun. – Aug. 2012**	**0.05 – 0.43**	**This work**
			0.08 – 0.79	
	Urban	Sep. 2005	0.16 ± 0.03	Barletta, et al.[216]
Dongguan		**Jun. – Aug. 2012**	**0.01 – 0.34**	**This work**
			0.02 – 0.62	
	Urban	Sep. 2005	0.15 ± 0.02	Barletta, et al.[216]

[#] Total mixing ratio range of α-pinene and β-pinene within 3×3 grid cells with the center of cities.

[^] Total mixing ratio range of monoterpenes within 3×3 grid cells with the center of cities.

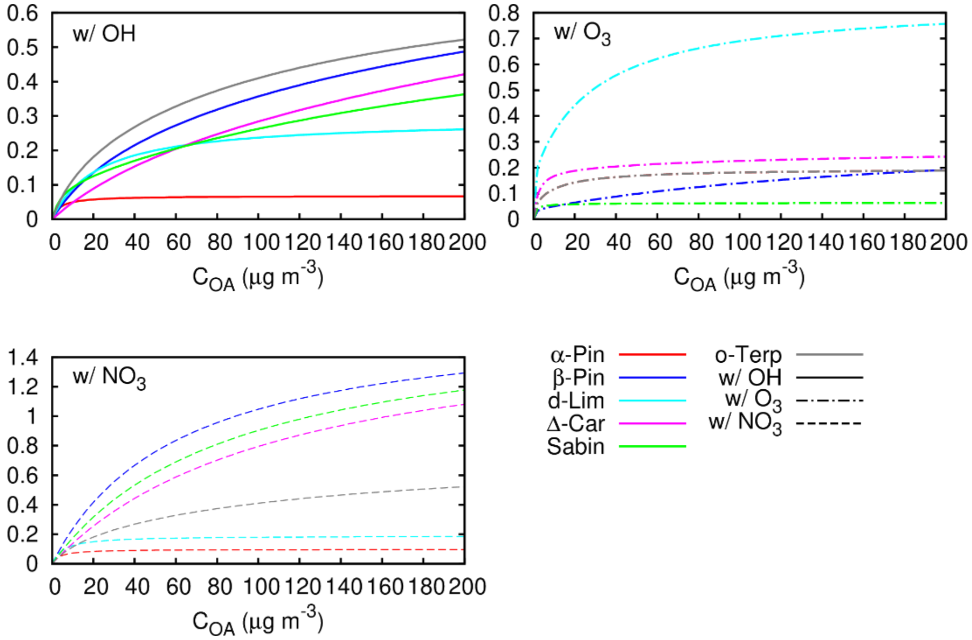


Figure S6-1 SOA yields of individual monoterpene species used in this model.

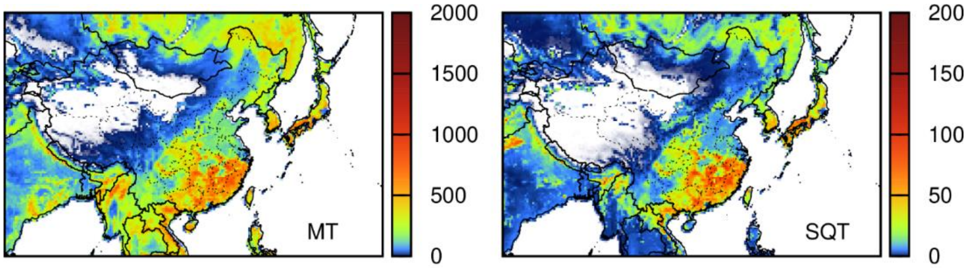


Figure S6-2 Average summertime (June – August 2012) emission rates of total MTs and SQTs. Units: gC km^{-2} h^{-1}.

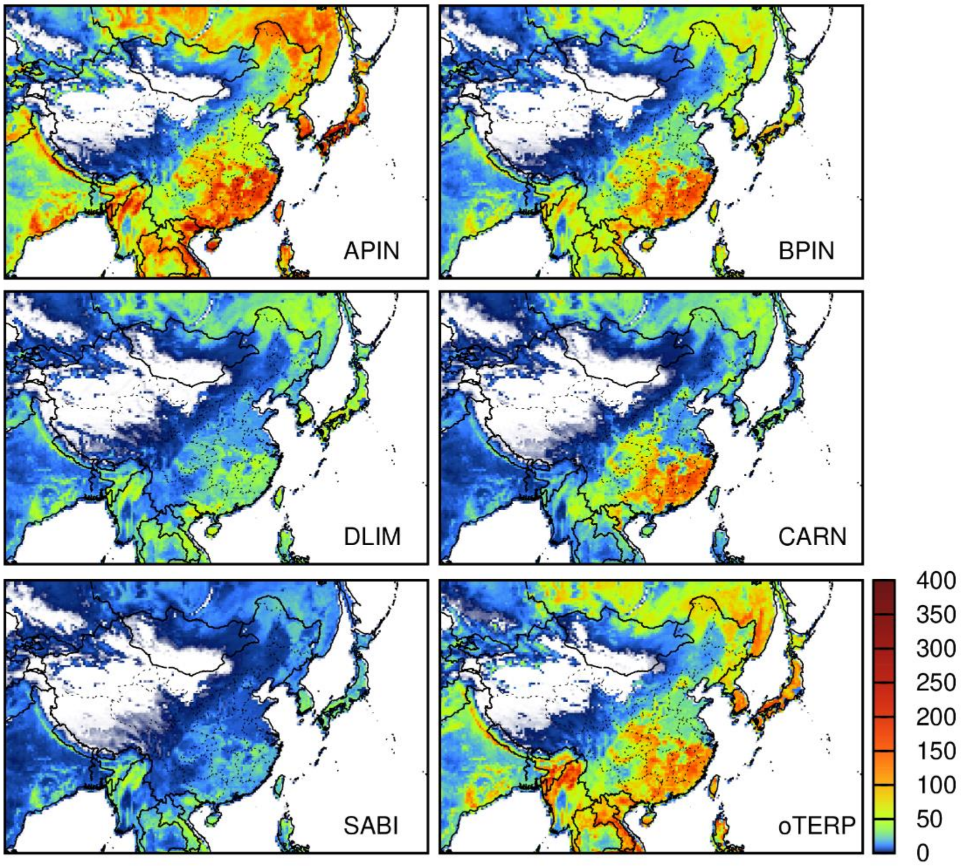


Figure S6-3 Average emission rates for α-pinene (APIN), β-pinene (BPIN), d-limonene (DILM), Δ^3-carene (CARN), sabinene (SABI), and the lumped other terpene (oTERP) species. Units: gC km^{-2} h^{-1}.

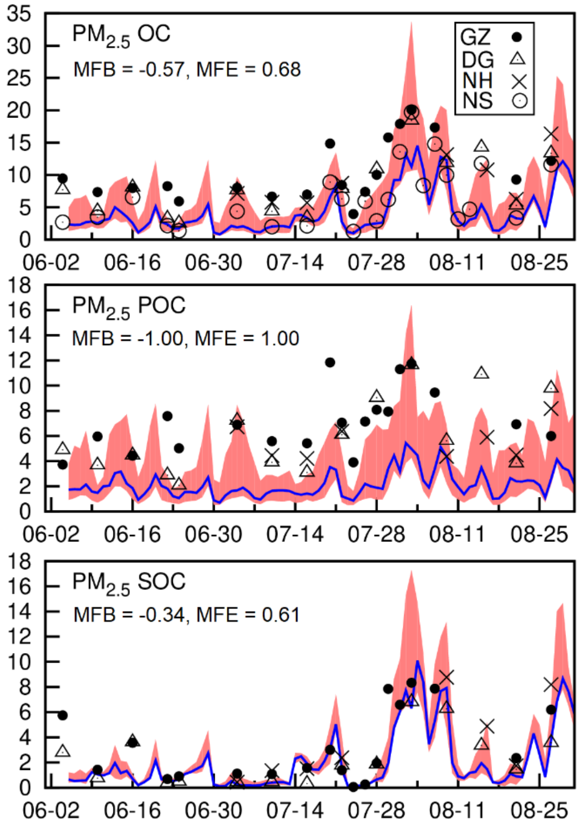


Figure S6-4 Predicted and observed organic carbon (OC), primary OC (POC), and secondary OC (SOC) in the PRD region. The blue line represents the predictions from the grid cell of the Guangzhou site. The red shaded area represents the range of concentrations from the 3×3 grid cells around the Guangzhou (GZ) site, where the other three sites, Dongguan (DG), Nanhai (NH), and Nansha (NS), are located.

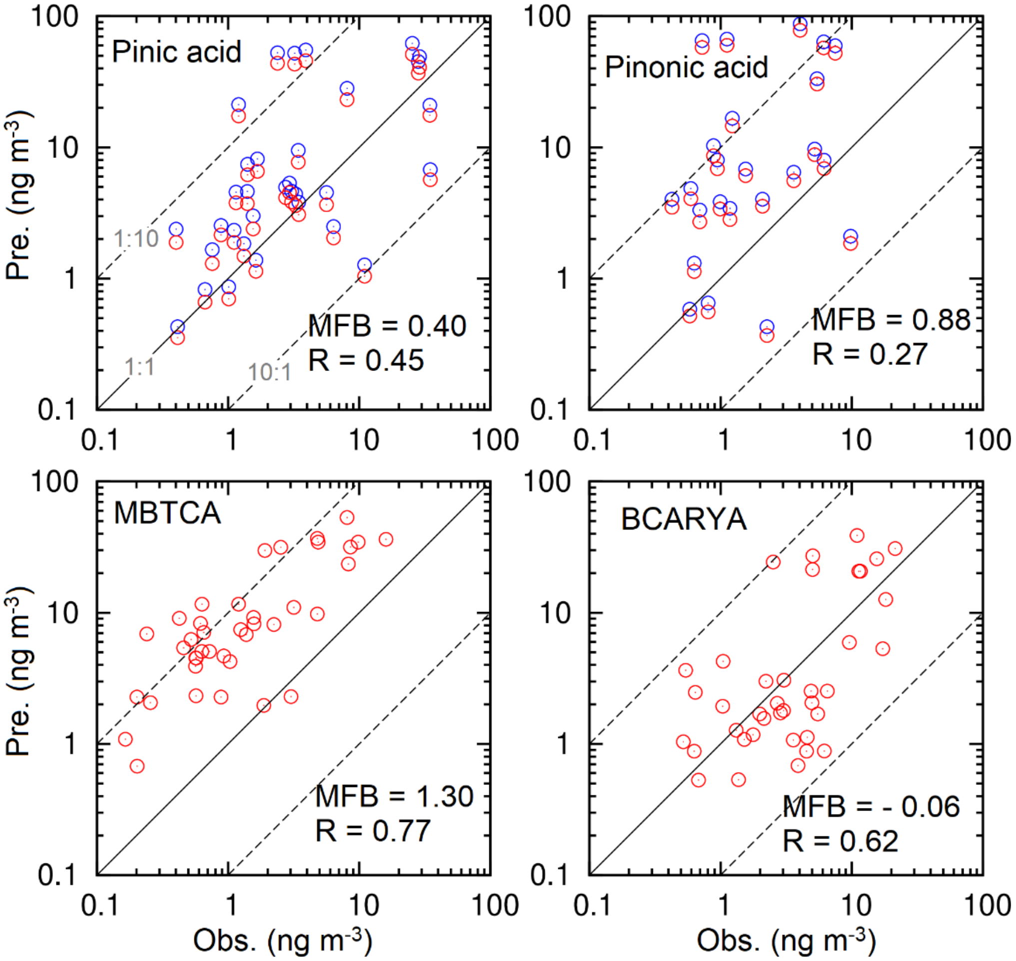


Figure S6-5 Predicted and observed pinic acid, pinonic acid, MBTCA and BCARYA at four PRD sites. Pinic acid and pinonic acids predictions are repartitioned using the observed OC and an OA/OC ratio of 1.6. The blue dots represent repartitioned predictions using an OA/OC ratio of 2.1. Repartitioning does not affect MBTCA and BCARYA because they are almost non-volatile.

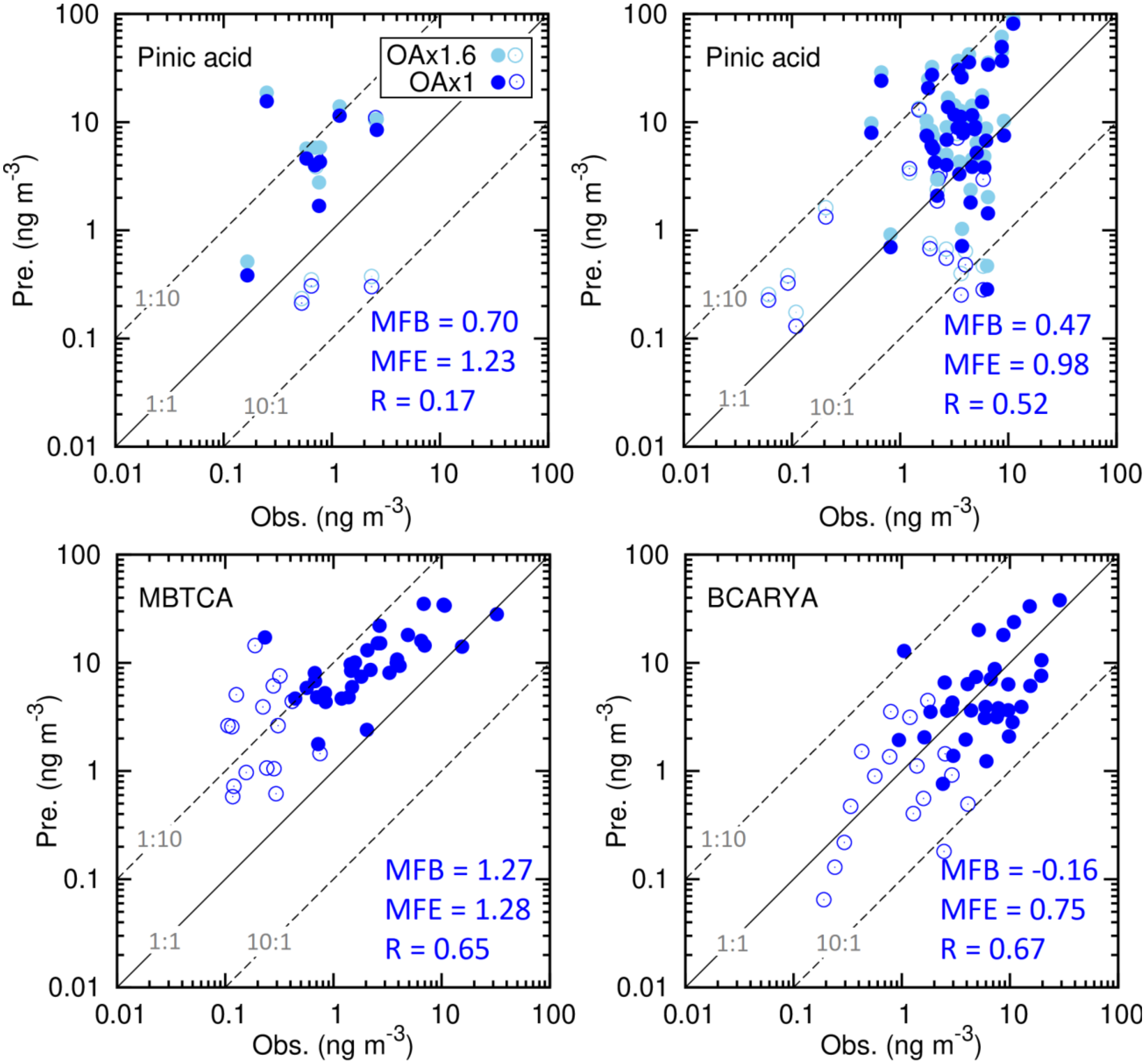


Figure S6-6 Model evaluation of organic tracers (a) pinic acid, (b) pinonic acid, (c) MBTCA, and (d) BCARYA from the 14 sites across China. The open circles represent the data from sites in the clean areas of west and southwest China (Linzhi, Namco, Dunhuang, Shapotou, Xishuangbanna, and Sanya).

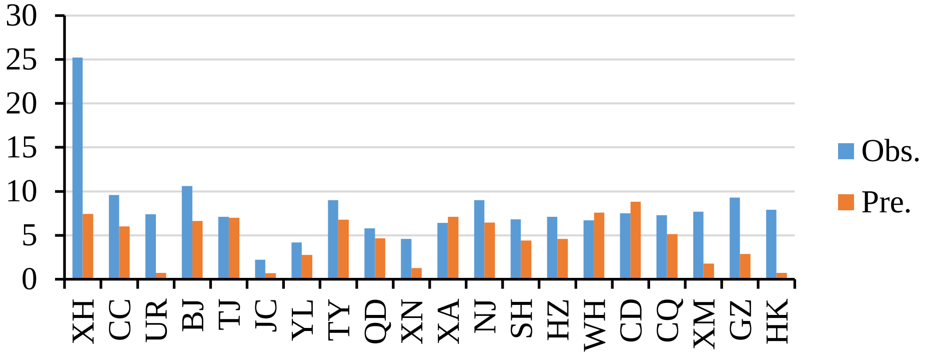
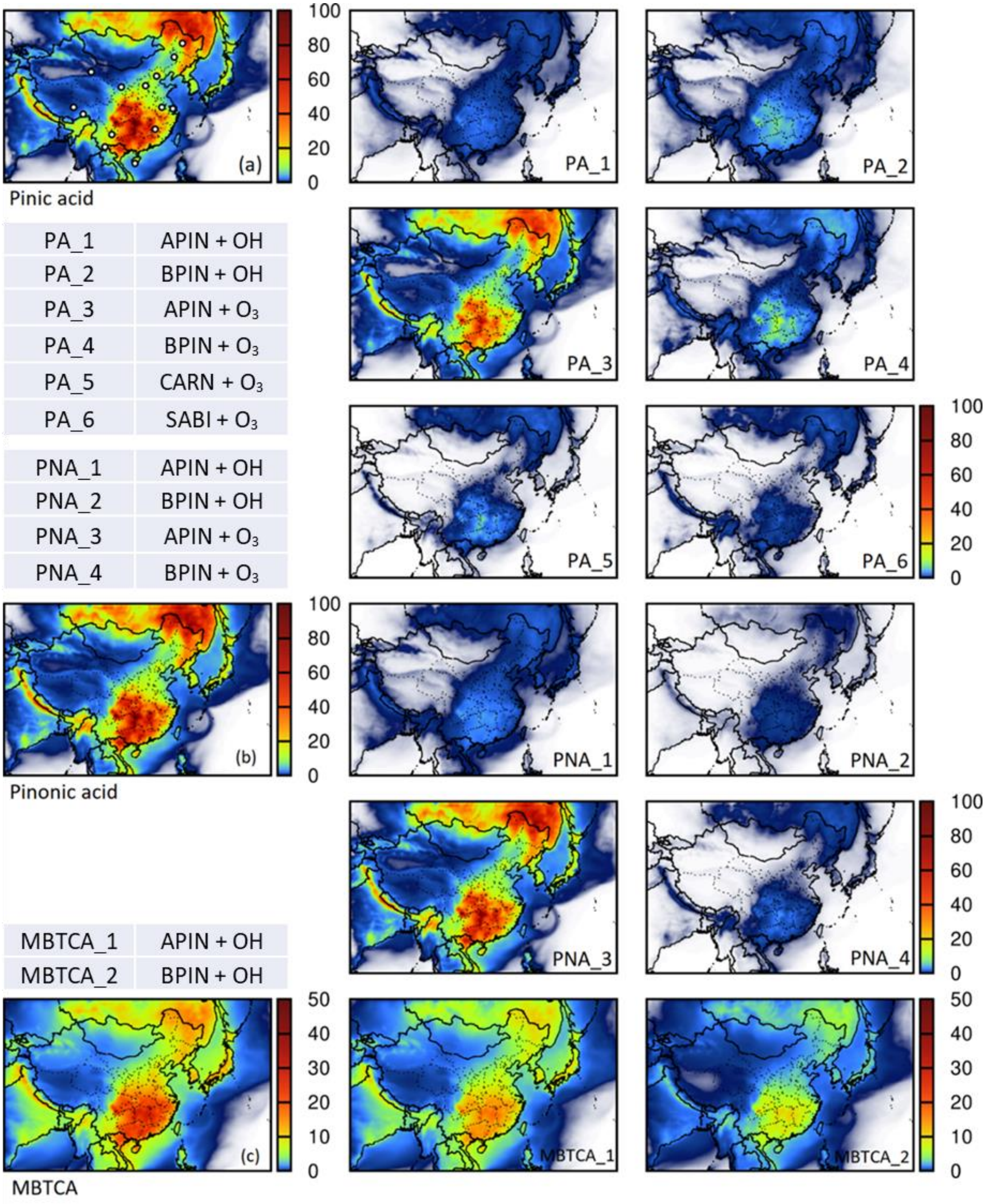


Figure S6-7 Model evaluation of OC by comparing with historical data.[190, 236] Units are μg m^{-3}. (The city abbreviations are XH: Xianghe, CC: Changchun, UR: Urumqi, BJ: Beijing, TJ: Tianjin, JC: Jinchang, YL: Yulin, TY: Taiyuan, QD: Qingdao, XN: Xining, XA: Xi'an, NJ: Nanjing, SH: Shanghai, HZ: Hangzhou, WH: Wuhan, CD: Chengdu, CQ: Chongqing, XM: Xiamen, GZ: Guangzhou, HK: Hongkong.)



PA_1	APIN + OH
PA_2	BPIN + OH
PA_3	APIN + O_3
PA_4	BPIN + O_3
PA_5	CARN + O_3
PA_6	SABI + O_3

PNA_1	APIN + OH
PNA_2	BPIN + OH
PNA_3	APIN + O_3
PNA_4	BPIN + O_3

MBTCA_1	APIN + OH
MBTCA_2	BPIN + OH

Figure S6-8 Regional distributions the monoterpene SOA tracers (pinic acid, pinonic acid, and MBTCA) from individual formation pathways. Units are ng m^{-3}. Circles show the location of the 14 sites.

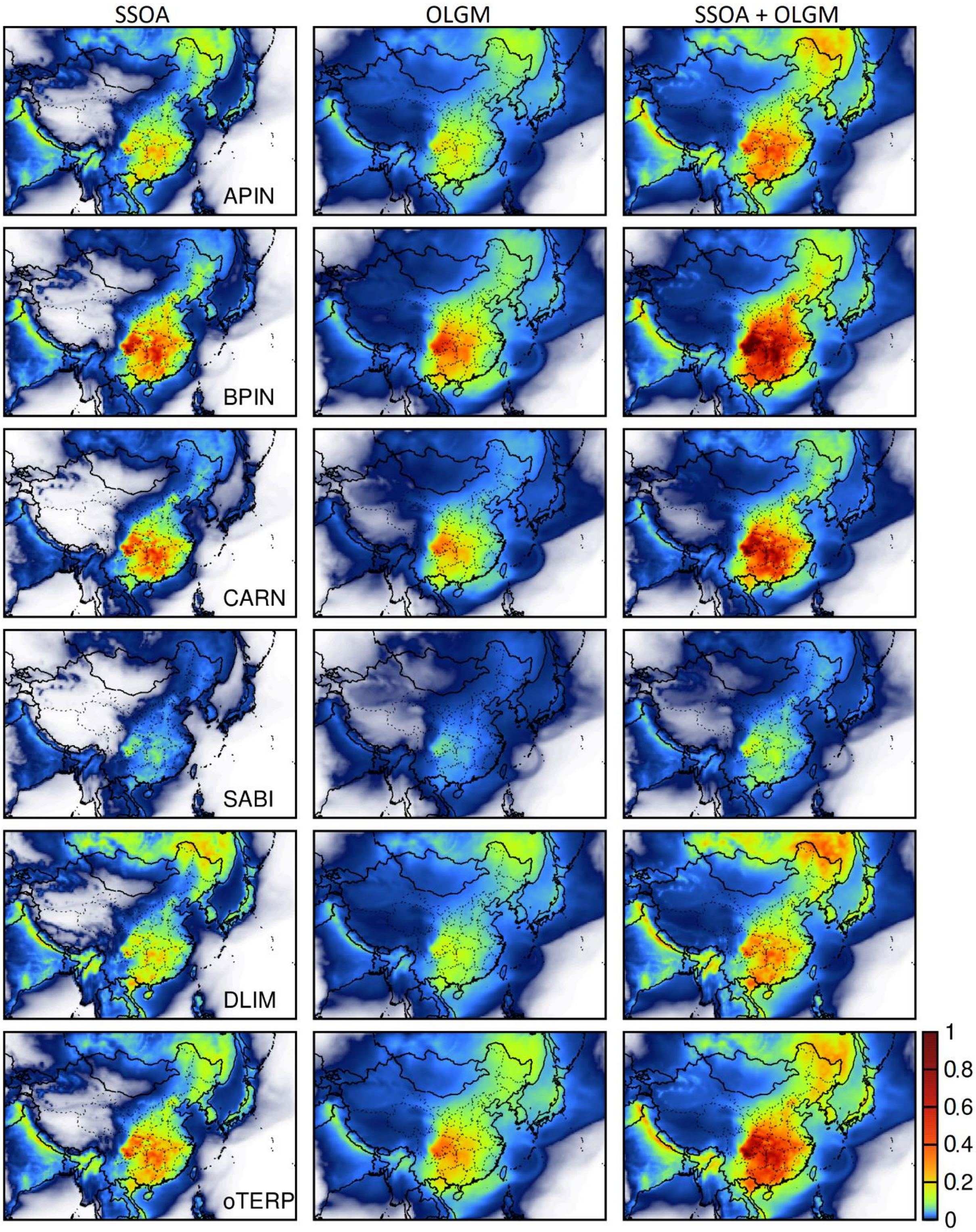


Figure S6-9 Average concentration of semi-volatile SOA (SSOA; left column), corresponding oligomers (OLGM; middle column), and total SOA (SSOA+OLGM; right column). Units: μg m^{-3}.

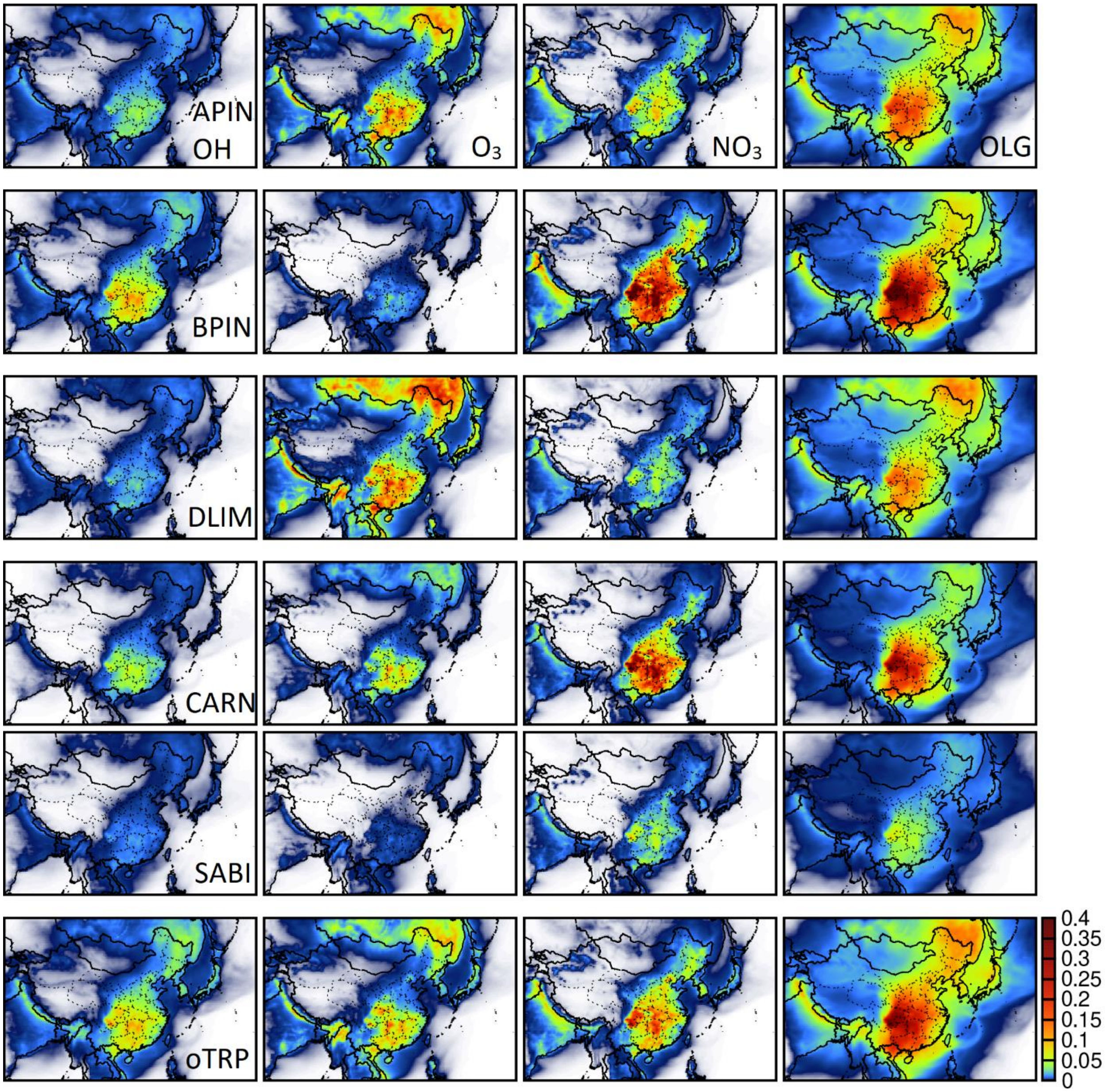


Figure S6-10 Average contributions to the monoterpene SOA from individual species and formation pathways. Units are µg m^{-3}.

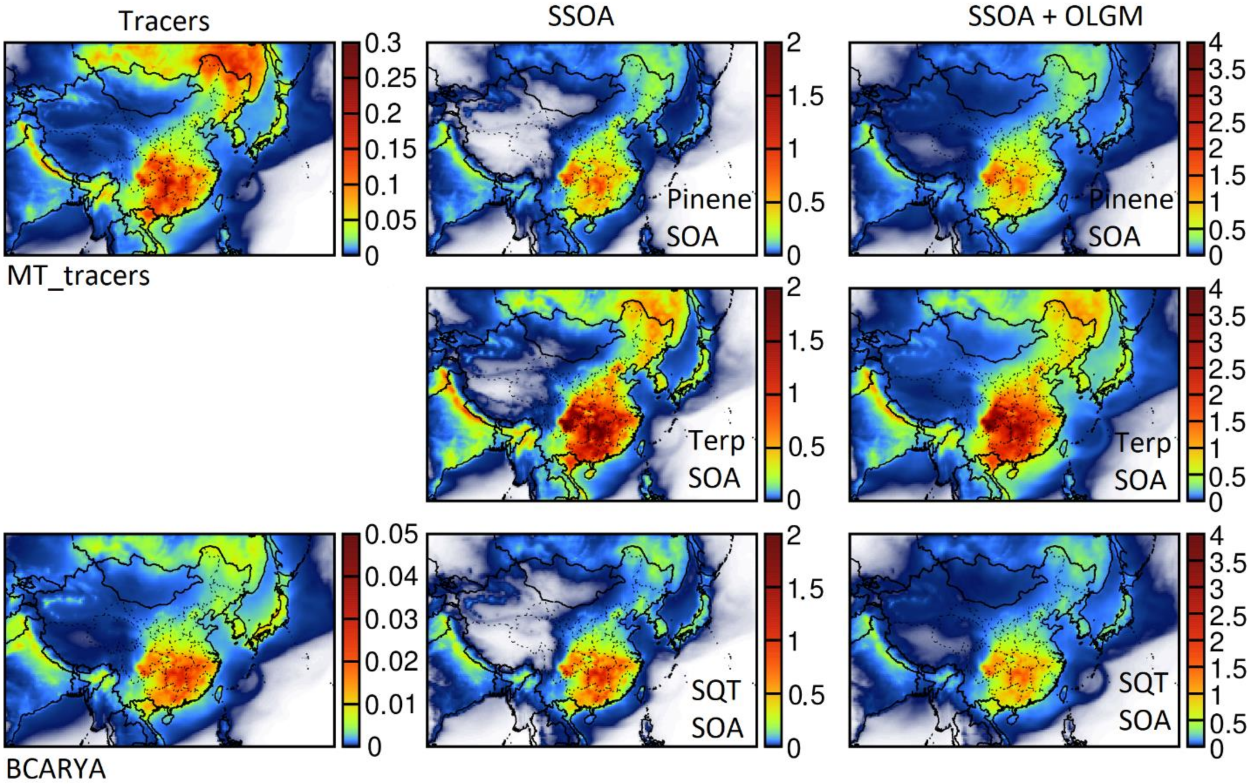


Figure S6-11 Average concentration of monoterpene tracers (MT_tracers, sum of PA, PNA and MBTCA) and sesquiterpene tracer BCARYA (left column). Semi-volatile SOA (SSOA) and total SOA (SSOA + oligomers (OLGM)) from α-pinene + β-pinene (pinene SOA), all monoterpenes (Terp SOA), and sesquiterpenes (SQT SOA). Units are μg m^{-3}.

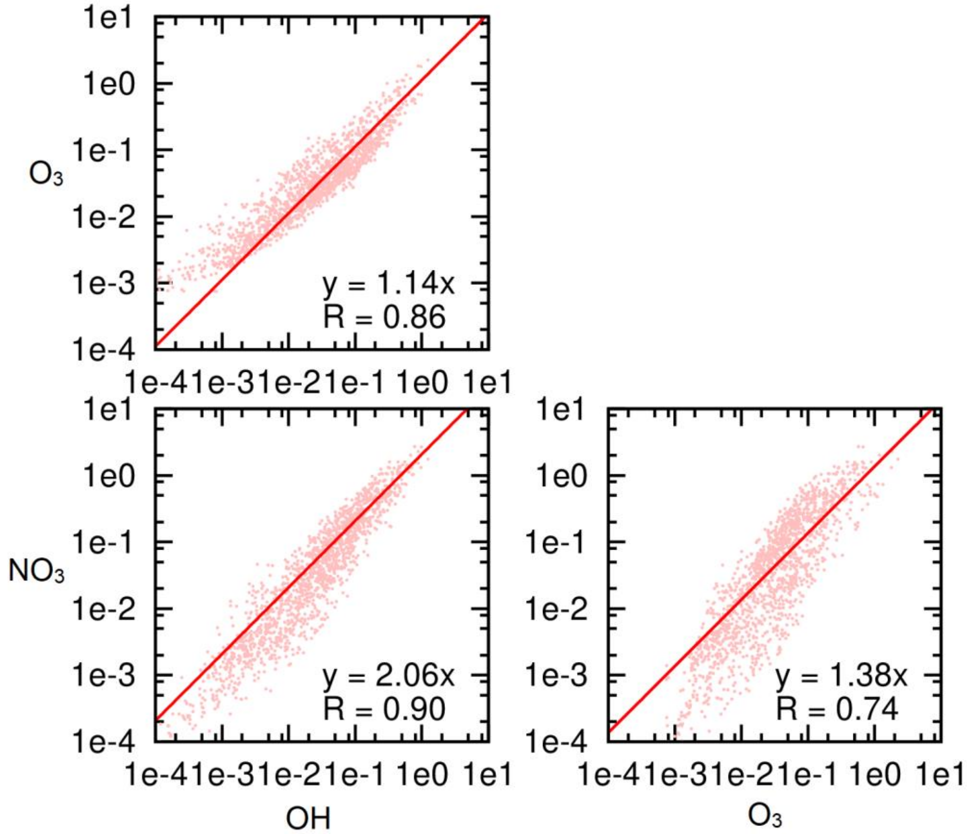


Figure S6-12 Correlation of daily MT SOA from three major formation pathways (OH, NO_3, and O_3).

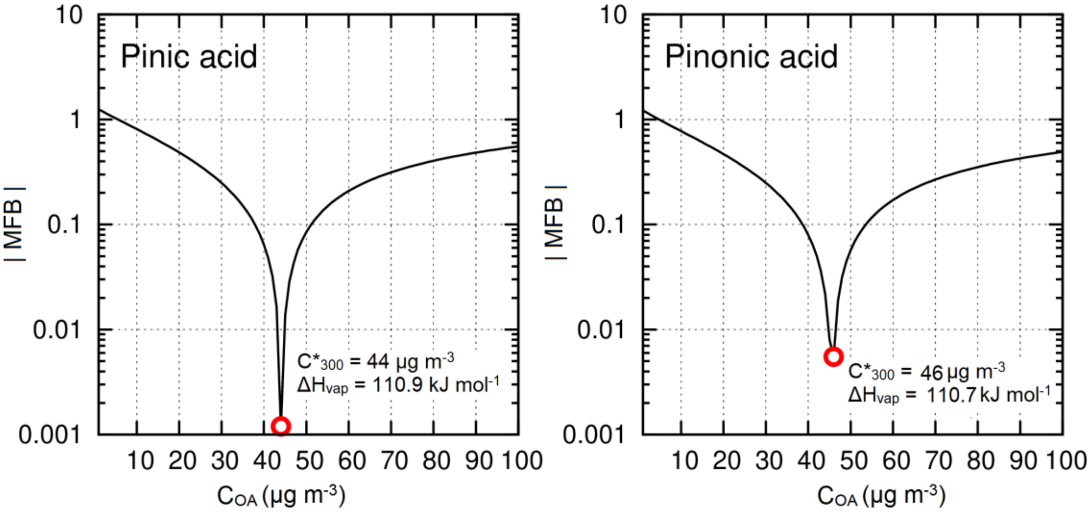


Figure S6-13 Saturation mass concentration optimization for PA and PNA.

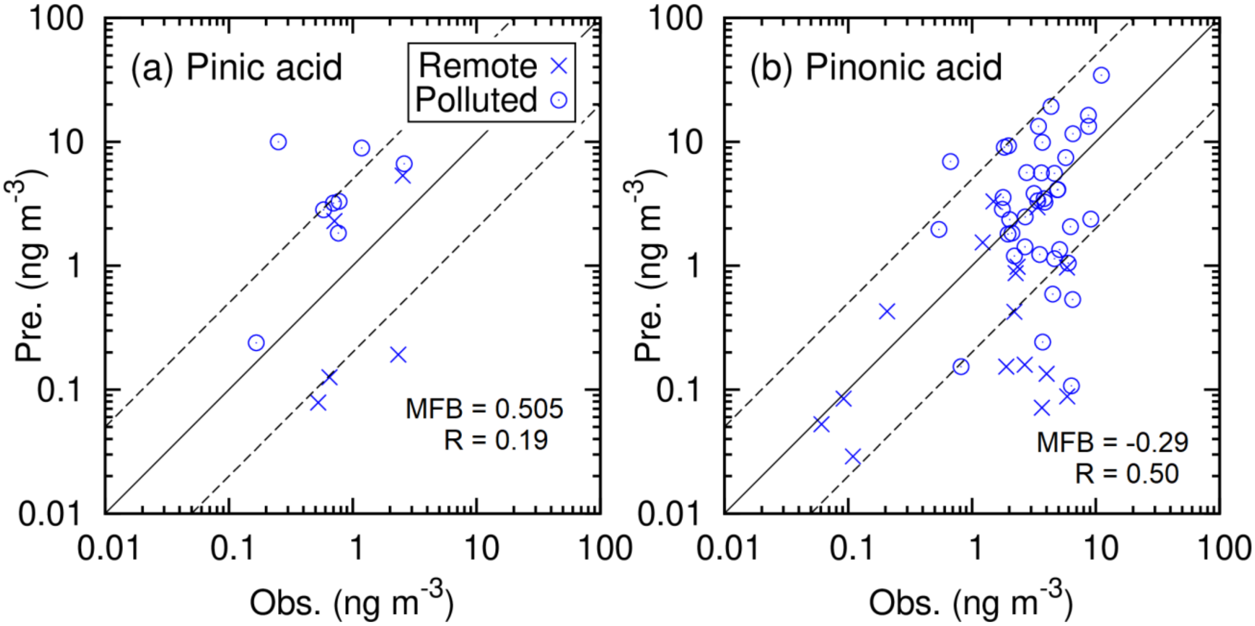


Figure S6-14 Model evaluation of PA and PNA predictions with optimized saturation mass concentrations against measurements at the 14 sites across China. Predictions are based on repartitioning with OA increased by a factor of 1.6. The cross markers represent data from the sites in the clean areas of west and southwest China (Linzhi, Namco, Dunhuang, Shapotou, Xishuangbanna, and Sanya).

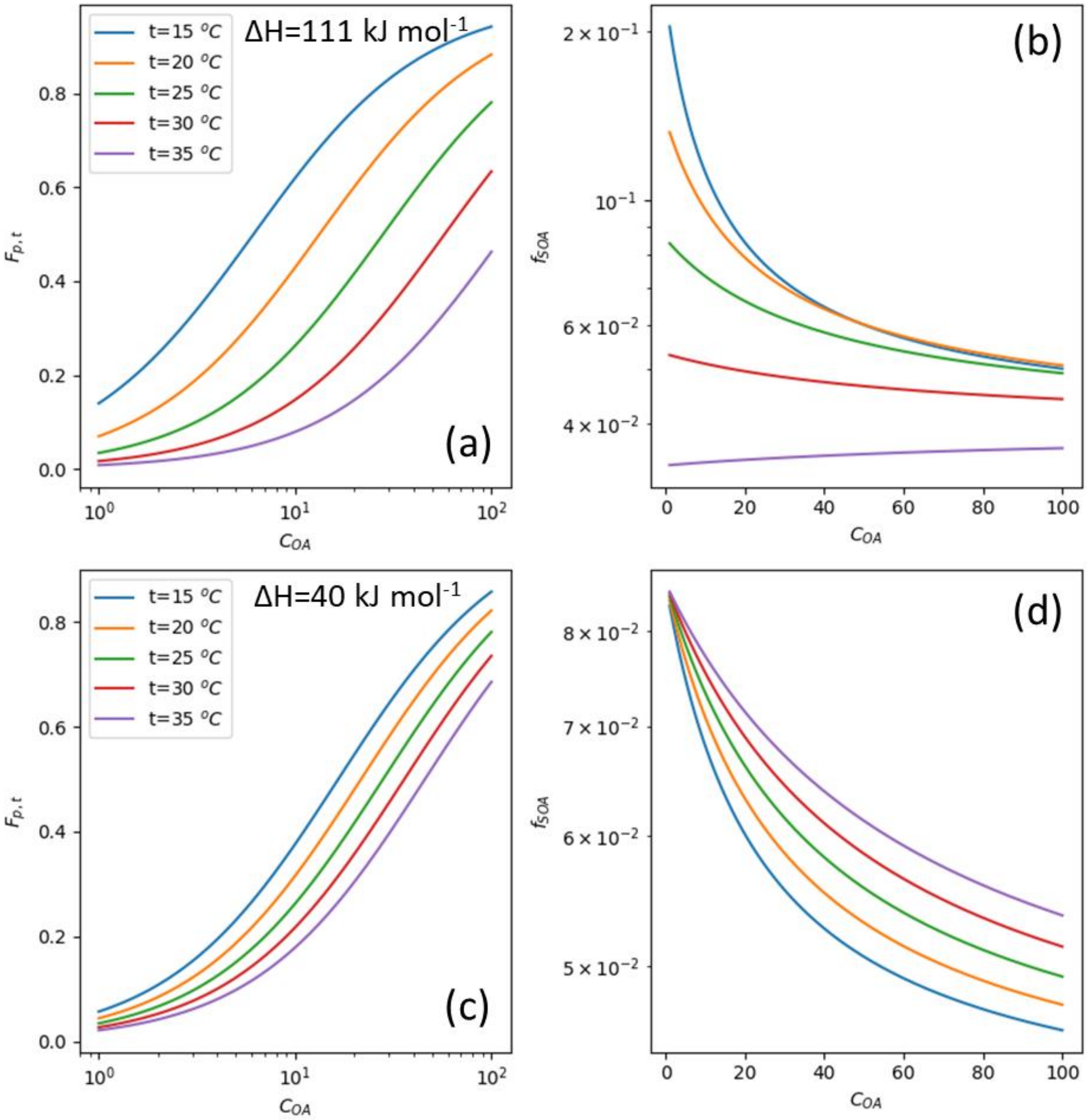


Figure S6-15 Fraction of pinic acid (PA) in the aerosol phase ($F_{p,t}$) (panels a and c) and Tracer-to-SOA ratio ($\boldsymbol{f_{SOA}}$) (panels b and d) using enthalpy of vaporization ($\Delta \boldsymbol{H_{vap}}$) of 111 kJ mol^{-1} (a, b) and 40 kJ mol^{-1} (c, d). The calculations are done assuming that PA is from α-pinene ozonolysis and monoterpene SOA is due to β-pinene reaction with NO_3, and the reacted α-pinene/β-pinene molar ratio is 1:1. Optimized C^* of PA (33.4 μg m^{-3}) is used in the calculation. Mass yield of PA is 0.0661, from Table 6-1. C^* and semivolatile product mass yields (β) for β-pinene + O_3 is from Table S6-1.

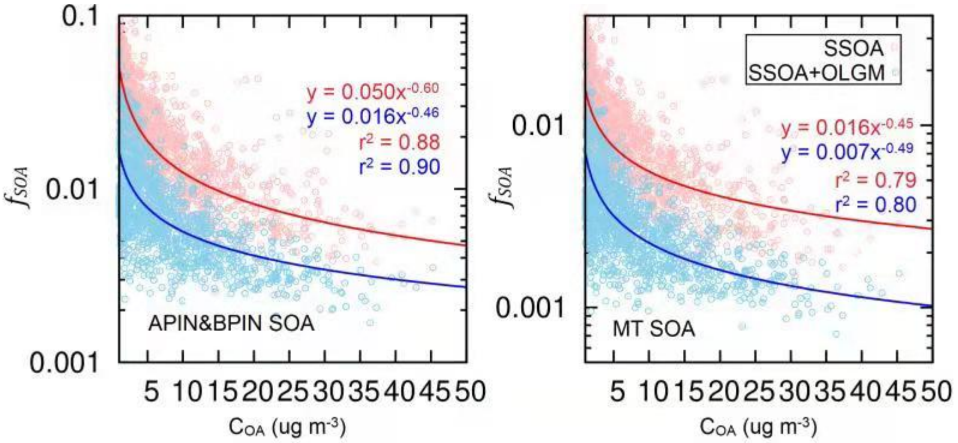


Figure S6-16 Variation of daily average MBTCA-to-SOA ratio ($\boldsymbol{f_{SOA}}$) with organic aerosol concentration (C$_{OA}$). The $\boldsymbol{f_{SOA}}$ values are calculated using (a) α-pinene and β-pinene SOA and (b) monoterpene SOA.

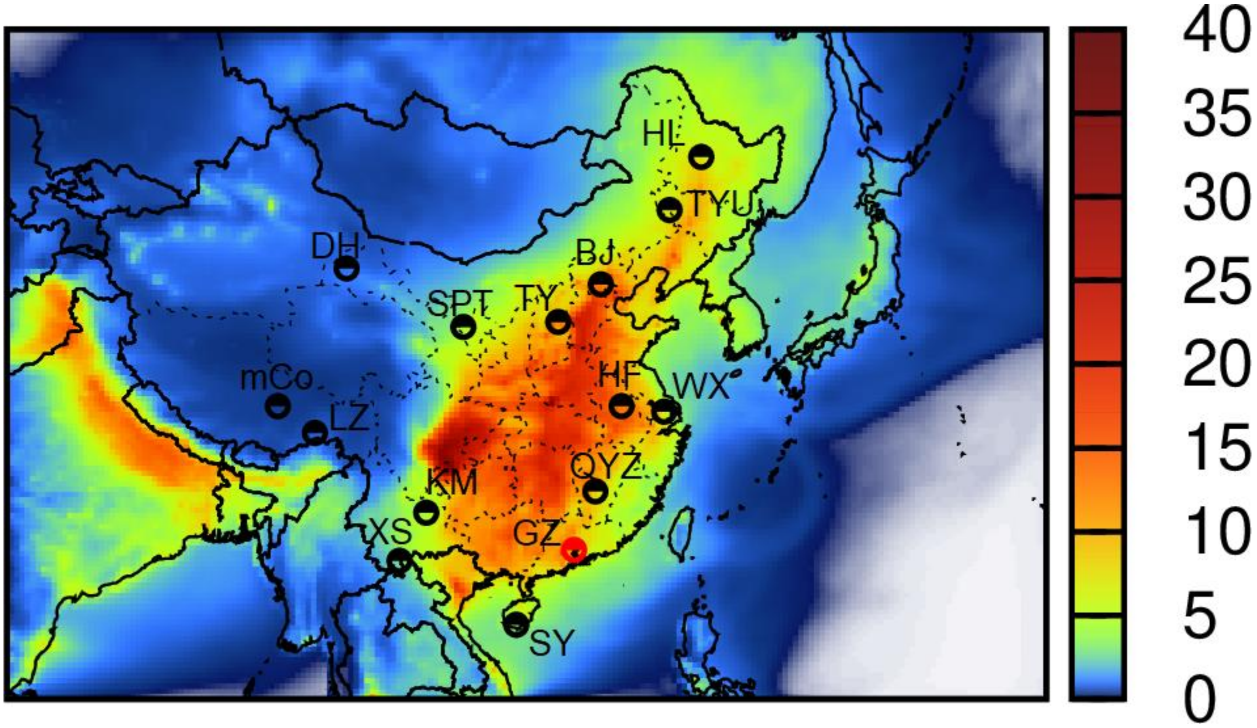


Figure S6-17 Average OA loadings during the summertime. Units: μg m^{-3}. The 14 sites marked in black open circles are: Hailun (HL), Tongyu (TYU), Beijing (BJ), Taiyuan (TY), Hefei (HF), Wuxi (WX), Qianyunzhou (QYZ), Sanya (SY), Kunming (KM), Xishuangbanna (BN), Shapotou (SPT), Dunhuang (DH), Linzhi (LZ), and Namco (mCO). Guangzhou (GZ) in the PRD region is marker with a red circle. Three other PRD sites, Dongguan, Nanhai, and Nansha are very close to GZ and are not shown on the map.

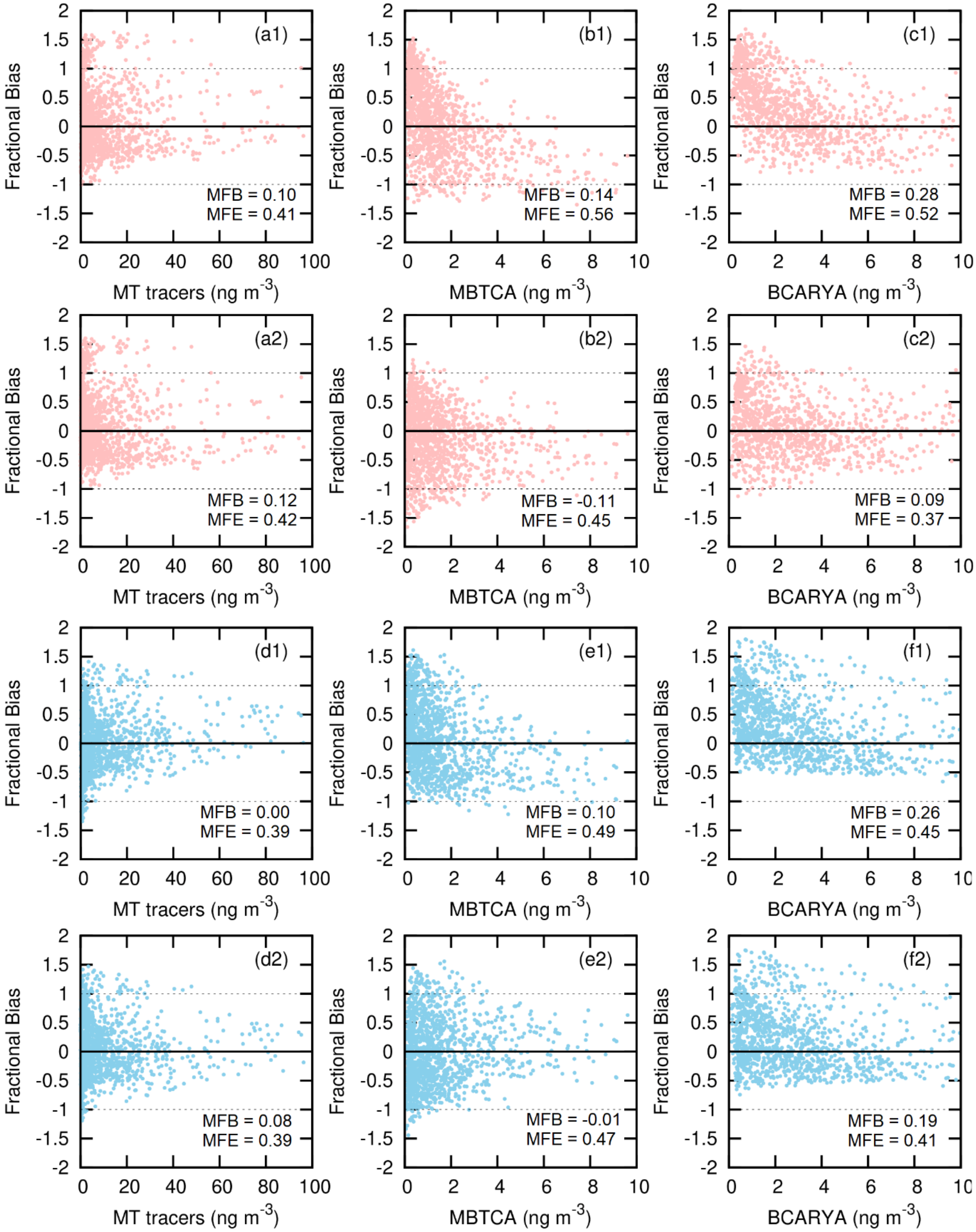


Figure S6-18 Fractional bias of predicted semivolatile SOA (SSOA) (a1-c2), and total SOA (SSOA + oligomers) (d1-f2) using the linear equations (a1-c1, d1-f1) and power-law equations (a2-c2, d2-f2). MT tracers (sum of PA, PNA and MBTCA) (first column) and MBTCA (second column) are used to estimate MT SOA. BCARYA is used to estimate SQT SOA.

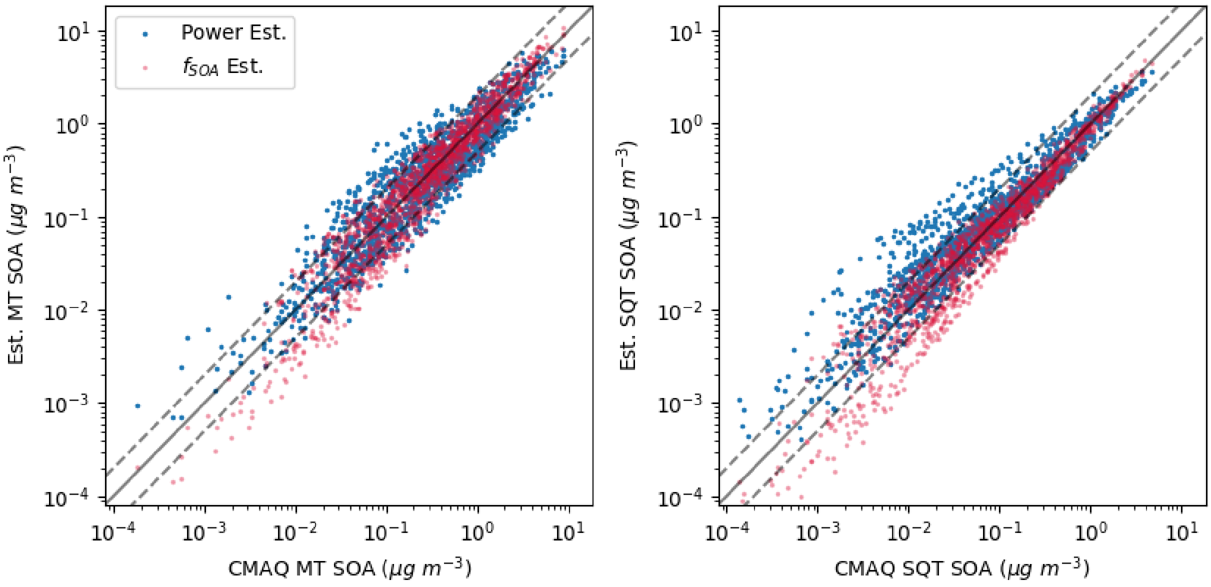


Figure S6-19 CMAQ-predicted and tracer estimated daily (a) MT SOA and (b) SQT SOA at monitor sites. The dash lines show 1:2 and 2:1 ratio. The blue dots are estimations using the power equations derived in Section 6.4.3, and the red dots are estimations using the f_{SOA} equations **Error! Reference source not found.** and (6.9) in Section 6.4.2.

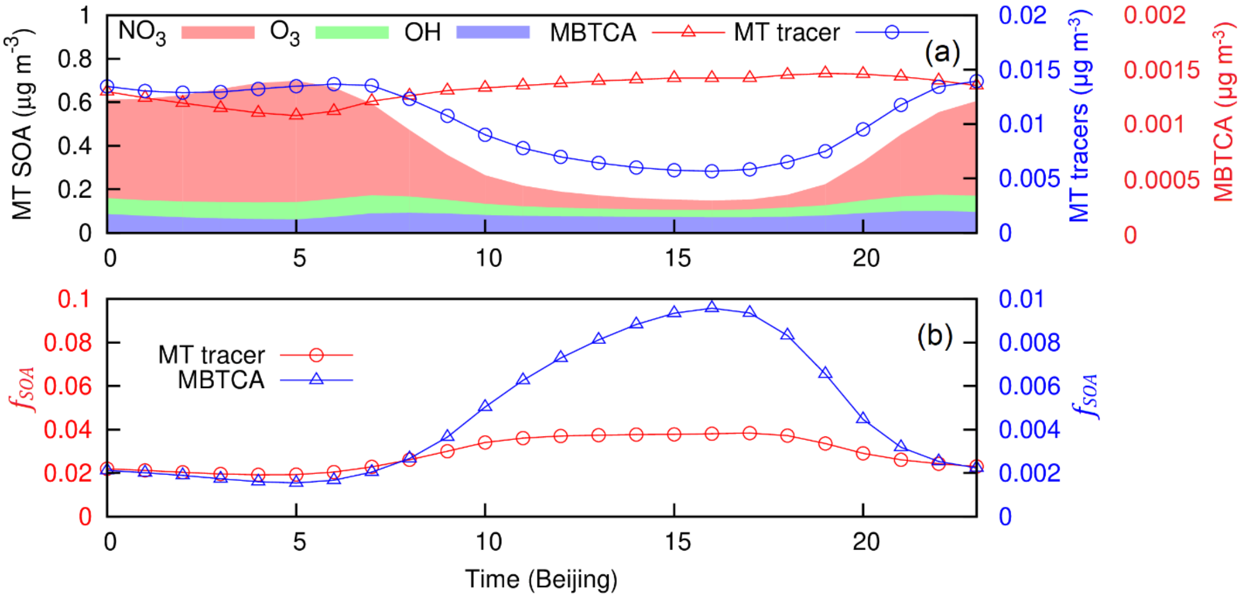


Figure S6-20 Diurnal variation of average (a) MT SSOA formation from each oxidation pathway and tracer concentrations, and (b) the f_{SOA} to estimate SSOA using MT tracers (sum of PA, PNA, and MBTCA) and MBTCA.