We thank the two reviewers for their constructive and detailed comments. In response, we have added additional analyses and re-written most of the main text to improve clarity throughout. We respond to each specific comment below. The reviewers' original comments are shown in red. Our replies are shown in black. The corresponding changes in the manuscript are shown in blue.

Reviewer 1:

General Description of manuscript:

The authors use satellite observations of glyoxal and formaldehyde to estimate a range in emissions of non-methane volatile organic compounds (NMVOCs) in China for 2007 using an adjoint inversion. Results from their inversion are discussed in the context of other top-down estimates for China and the a posteriori NMVOCs emissions are used to simulate surface ozone. The updated ozone concentrations increase consistency between the model and observed surface ozone concentrations in winter (December) and summer (July).

General Comments:

R1.1 What are the implications of the updated NMVOCs emissions on organic aerosol (and hence PM2.5) over China?

Thank you for the suggestion. We added in Section 6 an assessment of the impacts of our average top-down NMVOC emission estimates on simulated Chinese surface SOC (Figure S8), as well as comparison to surface SOC measurements (Table S10). We found that, by driving the model with our average top-down NMVOC emissions, the simulated surface SOC concentrations in June increased by 0.1 to 0.8 μgC m^{-3} over eastern China relative to the simulation using the *a priori* NMVOC emissions. This increase in simulated SOC concentrations brought the model to closer to the surface measurements, but the model still severely underestimated observed SOC concentrations.

[Main text, lines 765 to 782]: Figure S8 compares the simulated monthly mean surface SOC concentrations using our averaged top-down NMVOCs emissions against those simulated using the *a priori* NMVOC emissions for January and June in 2007. Also shown are the SOC measurements at 12 surface sites in June of 2006 and 2007 from Zhang et al. (2012) (Table S10). By driving the model with our average top-down NMVOC emissions, the simulated surface SOC

concentrations in June increased by 0.1 to 0.8 μgC m^{-3} over eastern China relative to the simulation using the *a priori* NMVOC emissions. This increase in simulated SOC concentrations brought the model to closer to the surface measurements, but the model still severely underestimated observed SOC concentrations. We note our version of the GEOS-Chem model only included two pathways for secondary organic aerosol formation: (1) the reversible partitioning of semi-volatile products from the oxidation of isoprene, monoterpenes, and aromatics formation pathways (Liao et al., 2007; Henze et al., 2008), and (2) the irreversible uptake of dicarbonyl by aqueous aerosols and cloud drops (Fu et al., 2008). Other pathways, such as the atmospheric aging semi-volatile and intermediate volatility organic compounds (S/IVOC), has been shown to be an important source of secondary organic aerosols (Robinson et al., 2007; Pye and Seinfeld, 2010) but they were not included in our version of GEOS-Chem. In any case, the precursors and formation pathways of secondary organic aerosols in China are still poorly understood (Fu et al., 2012), such that no quantitative conclusions can be drawn regarding the impacts of our top-down NMVOC emission estimates on regional secondary organic aerosol formation.

[Supplementary information, Figure S8]:

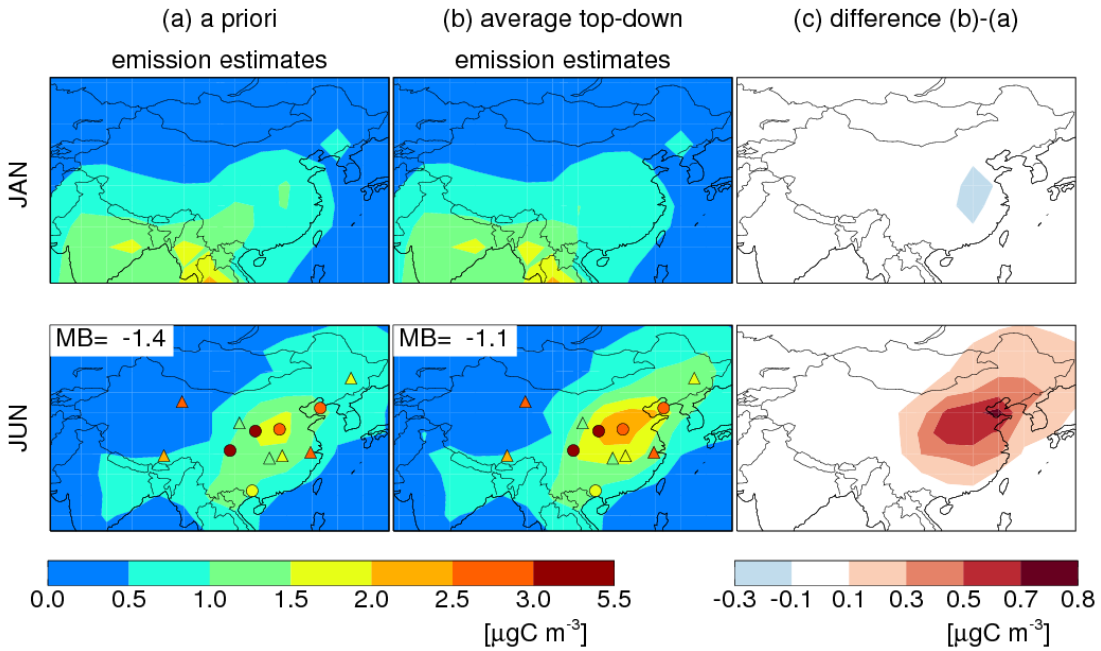


Figure S8. Simulated monthly mean surface secondary organic carbon (SOC) concentrations in June and December 2007 driven by (a) the *a priori* emissions and (b) our average top-down emissions, respectively, as well as (c) the differences. Overlaid symbols show the SOC measurements at 12 urban (circles) and regional (triangles) sites in China in June (Table S10). Mean biases (MB) of the simulated concentrations relative to surface measurements in June are shown inset.

[Supplementary information, Table S10]: Surface measurements of SOC concentrations in June during 2006 and 2007 (Zhang et al., 2012)[a] and comparison to simulated SOC concentrations

Site	Site type	SOC concentration (μg C m^{-3})			Bias (model - measurement)	
		measurement	*a priori* simulation	average top-down emission estimates simulation	*a priori* simulation	average top-down emission estimates simulation
Chengdu (30.65°N, 104.03°E)	urban	3.79	1.31	1.61	-2.49	-2.18
Dalian (38.9°N, 121.63°E)	urban	2.64	1.32	2.09	-1.32	-0.55
Dunhuang (40.15°N, 94.68°E)	regional	2.51	0.38	0.41	-2.13	-2.11
Gaolanshan (36.0°N, 105.85°E)	regional	1.29	0.73	0.97	-0.56	-0.32
Jinsha (29.63°N, 114.2°E)	regional	1.81	1.40	1.85	-0.42	0.03
Lhasa (29.67°N, 91.13°E)	regional	2.34	0.47	0.48	-1.88	-1.86
LinAn (30.3°N, 119.73°E)	regional	2.51	0.95	1.29	-1.55	-1.22
Longfengshan (44.73°N, 127.6°E)	regional	1.89	0.85	1.09	-1.04	-0.79
Nanning (22.82°N, 108.35°E)	urban	1.70	0.72	0.74	-0.98	-0.96
Taiyangshan (29.17°N, 111.71°E)	regional	1.11	1.38	1.72	0.27	0.61
XiAn (34.43°N, 108.97°E)	urban	5.41	1.70	2.39	-3.71	-3.02
Zhengzhou	urban	2.78	1.59	2.17	-1.19	-0.62

(34.78°N, 113.68°E)						
Average		2.48	1.07	1.40	-1.42	-1.08

[a] SOC concentrations were computed using organic carbon measurements (μgC m^{-3}) and the EC-tracer approach (Zhang et al., 2012).

R1.2 Why use such a coarse resolution version of GEOS-Chem (5x4), when higher resolution versions of GEOS-Chem are available for the globe (2.5x2) and nested over China (0.667x0.5 for GEOS-5 meteorology)?

We agree with the Reviewer that our methodology is applicable to inversions at higher-resolutions and that it would be worthwhile to do so. However, the computation cost would be overwhelming for our analyses, which involved 48 inversion experiments (4 sets of satellite observations × 12 months, each inversion needed 10 to 50 calculations of forward and backward integrations) at higher resolutions (which would also require shorter time steps). We do plan to do higher resolution inversions focusing on a shorter periods of time, which would be more computationally feasible. We added a comment on this in the main text:

[Main text, lines 824 to 828]: The monthly inversions presented in this work, conducted at 5° longitude × 4° latitude resolutions due to limited computation resources, quantified the Chinese NMVOC emissions on regional/sub-regional scales. Future inversions and sensitivity studies targeting shorter periods of time may be conducted on finer resolutions to quantify Chinese NMVOC emissions and to evaluate their impacts on photochemistry at city cluster scales.

R1.3 What is the effect of the updates to the model (Section 2.1) on simulated column concentrations of HCHO and CHOCHO?

Thank you for pointing out this lack of clarity. We added more detailed description of the updated chemical mechanisms, as well as a summary of the yields of formaldehyde and glyoxal from the oxidation of individual NMVOC precursors in our updated mechanisms (Table S1).

[Main text, lines 172 to 213]: We updated the dicarbonyl chemical mechanism in GEOS-Chem developed by Fu et al. (2008), which in turn was originally adapted from the Master Chemical Mechanism (MCM) version 3.1 (Jenkin et al., 1997; Saunders et al.,

2003). Table S1 lists the yields of formaldehyde and glyoxal from the OH-oxidation of NMVOC precursors in our updated chemical mechanism. The lumped NMVOC precursors of formaldehyde in our mechanism included ethane, propane, ≥C4 alkanes, ethene, ≥C3 alkenes, benzene, toluene, xylenes, isoprene, monoterpenes, acetone, hydroxyacetone, methygloxal, glycolaldehyde, acetaldehyde, 2-methyl-3-bute-nol, methy ethyl ketone, methanol, and ethanol (lumped into ≥C4 alkanes). The lumped NMVOC precursors of glyoxal in our mechanism included ethene, ethyne, benzene, toluene, xylenes, isoprene, monoterpenes, glycolaldehyde, and 2-methyl-3-bute-2-nol (MBO). Hereinafter we focused our discussion on these NMVOC precursors only, as their emissions may be constrained by formaldehyde and glyoxal observations.

The OH-oxidation of isoprene is a major source of both formaldehyde and glyoxal over China (Fu et al., 2007, 2008; Myriokefalitakis et al., 2008). We replaced the isoprene photochemical scheme with that used in GEOS-Chem v10.1, which included updates from Paulot et al. (2009a,b) and Mao et al. (2013). In this updated scheme, oxidation of isoprene by OH under high-NO_x conditions produces formaldehyde and glyoxal at yields of 0.436 molecules per C and 0.0255 molecules per C, respectively (Table S1), mainly via the RO_2+NO pathways. Under low-NOx conditions, oxidation of isoprene by OH produces formaldehyde and glyoxal at yields of 0.38 molecules per C and 0.073 molecules per C, respectively (Table S1), via both RO_2+HO_2 and RO_2-isomerization reactions. Li et al. (2016) implemented this same isoprene photochemical scheme into a box model and compared the productions of formaldehyde and glyoxal from isoprene oxidation with those in the MCM version 3.3.1 (Jenkin et al., 2015). They showed that the production pathways and yields of formaldehyde and glyoxal were similar in the two schemes under the high-NOx conditions typical of eastern China.

We updated the molar yields of glyoxal from the OH oxidations of benzene (33.3%), toluene (26.2%), and xylenes (21.0%) following the latest literature (Arey et al., 2009; Nishino et al., 2010). These new molar yields were higher than those used in Fu et al. (2008) (based on averaged yields in the literature: 25.2% for benzene, 16.2% for toluene, and 15.6% for xylenes) but still lower than those used by Chan Miller et al. (2016) (75% for benzene, 70% for toluene, and 36% for xylenes), which were taken from the aromatic chemical scheme in MCM version 3.2 (Jenkin et al., 2003; Bloss et al., 2005). In MCM version 3.2, more than half of the glyoxal from aromatics oxidation were produced during second- and later-generation photochemistry, but such productions are with

limited experimental support and uncertain (Bloss et al., 2005).

Formaldehyde and glyoxal in the GEOS-Chem model were both removed by photolysis, as well as dry and wet deposition (Fu et al., 2008). We updated the Henry's law constant for glyoxal from $3.6 \times 10^5 \times \exp[7.2 \times 10^3 \times (1/T\text{-}1/298)]$ (Fu et al., 2008) to $4.19 \times 10^5 \times \exp[(62.2 \times 10^3/R) \times (1/T\text{-}1/298)]$ (Ip et al., 2009) and added the dry deposition of formaldehyde, glyoxal, methyglyoxal and glycolaldehyde on leaves (Mao et al., 2013). In addition, we assumed that glyoxal was reactively uptaken by wet aerosols and cloud droplets with an uptake coefficient = 2.9×10^{-3} (Liggio et al., 2005; Fu et al., 2008). All other physical and chemical processes in our forward model were as described in Fu et al. (2008).

[Supplementary information, Table S1]:Ultimate yields of formaldehyde and glyoxal from the oxidation of NMVOC precursors by OH in our model under high-NOx and low-NOx conditions

NMVOCs	Formaldehyde (molecules per C)		Glyoxal (molecules per C)	
	High-NO_x [a]	Low-NO_x [b]	High-NO_x [a]	Low-NO_x [b]
Ethene	0.995	0.366	0.0665	0.067
Glycolaldehyde	0.366	0.366	0.067	0.067
Isoprene	0.436	0.38	0.0255	0.073
2-methyl-3-bute-nol (MBO)	0.092	0.092	0.0168	0.0168
Benzene	0.001	0.001	0.0555	0.0555
Toluene	0.198	0.18	0.037	0.037
Xylenes	0.269	0.155	0.026	0.026
Monoterpenes (lumped)	0.006	0.006	0.005 [c]	0.005 [c]
Ethyne	-	-	0.318	0.318
Methanol	1.0	1.0	-	-
Ethane	0.5	0.5	-	-
Acetaldehyde (lumped)	0.5	0.5	-	-
Propane	0.49	0.317	-	-
$\geq C_3$ alkenes (lumped)	0.657	0.333	-	-
Acetone	0.64	0.383	-	-
Hydroxyacetone	0.333	0.333	-	-
Methyglyoxal	0.333	0.333	-	-
$\geq C_4$ alkanes (lumped)	0.578	0.187	-	-
Methy ethyl ketone (lumped)	0.465	0.25	-	-

[a] Yields under high-NO_x conditions were calculated assuming that all RO_2 radicals from the oxidation of the NMVOC precursor reacted with NO.

[b] Yields under low-NO_x conditions were calculated assuming RO_2:HO_2 concentration ratio of 1:1.

[c] Glyoxal produced from the oxidation of monoterpenes by ozone

Specific Comments:

R1.4 Lines 74-75: Include Millet et al. (2006) as a reference for the high yields of formaldehyde from NMVOCs.

Added as suggested. Thank you.

> [Main text, lines 76 to 78]: Formaldehyde is produced at high yields during the oxidation of many NMVOC species (Millet et al., 2006) and also emitted directly from anthropogenic and biomass burning activities (Akagi et al., 2011; Li et al., 2017).

R1.5 Lines 77-80: Biogenic emissions doesn't always dominate HCHO columns over the Amazon and Africa. Both locations include a large and often dominant contribution from biomass burning to HCHO.

> We revised this sentence as follows to avoid misunderstanding:
>
> [Main text, lines 78 to 83]: Early inversions of satellite-observed formaldehyde columns mostly focused on areas where the local NMVOC fluxes were dominated by biogenic sources during the growing season and in the absence of substantial biomass burning, such as the southeast U.S. (Palmer et al., 2003, 2006; Millet et al., 2006, 2008), Europe (Dufour et al., 2009; Curci et al., 2010), the Amazon (Barkley et al., 2008, 2009, 2013), and Africa (Marais et al., 2012, 2014a).

R1.6 Lines 81-82: "linearly proportional to the local biogenic isoprene flux during the growing season" seems odd, in particular when the HCHO columns are used to estimate isoprene emissions. Do you mean that the vegetation distribution and HCHO are spatially correlated?

We revised this sentence as follows to improve clarity:

[Main text, lines 83 to 84]: These studies showed that the observed local enhancements of formaldehyde column concentrations can be used to quantitatively constrain the local biogenic NMVOC fluxes.

R1.7 Line 80: Marais et al. (2012; 2014) obtained isoprene emissions for all of Africa, not just the tropical portion.

We re-wrote this sentence to correct for this error:

[Main text, lines 78 to 83]: Early inversions of satellite-observed formaldehyde columns mostly focused on areas where the local NMVOC fluxes were dominated by biogenic sources during the growing season and in the absence of substantial biomass burning, such as the southeast U.S. (Palmer et al., 2003, 2006; Millet et al., 2006, 2008), Europe (Dufour et al., 2009; Curci et al., 2010), the Amazon (Barkley et al., 2008, 2009, 2013), and Africa (Marais et al., 2012, 2014a).

R1.8 Line 84: The chronology is odd. The line starts with "Later studies", but many of these studies precede the studies in the previous paragraph.

Thank you for pointing this out. We re-wrote this sentences to improve clarity:

[Main text, lines 86 to 89]: In other areas, the NMVOC emissions from various sources may be comparable in magnitudes. Several studies constrained the NMVOC emissions from multiple sources over such areas by analyzing the spatiotemporal variability of the observed formaldehyde columns (Shim et al., 2005; Fu et al., 2007; Stavrakou et al., 2009b; Curci et al., 2010; Gonzi et al., 2011; Marais et al., 2014b; Zhu et al., 2014).

R1.9 Line 102: "diffused" should be "diffuse".

Corrected. Thank you.

R1.10 Line 117: Is "anonymous" a typo?

Yes, thank you for pointing out this typo. It should be "anomalous". We re-wrote the

sentence to avoid confusion:

[Main text line 120 to 123]: They suggested that the missing glyoxal source over eastern China was anthropogenic, on the basis that the anomalous glyoxal columns observed by SCIAMACHY (relative to the glyoxal columns simulated by their model) were spatially correlated with anthropogenic NOx emissions.

R1.11 Lines 163-177: This paragraph needs more context for readers not familiar with the array of GEOS-Chem model versions and chemistry mechanisms. Is this a separate branch of the model that includes detailed carbonyl chemistry not included in the standard version? What exactly are the updates that are applied to GEOS-Chem in this work? Has this branch of the model fallen behind the other model versions and so is being updated in this work to include the isoprene chemistry that is currently in the standard version of the model?

Thank you for pointing out this lack of clarity. We added the following description on the GEOS-Chem model version used in this work.

[Main text, lines 158 to 164]: We updated the GEOS-Chem global 3D chemical transport model (version 8.2.1) to simulate the emission, transport, chemistry, and deposition of NMVOCs, as well as the resulting formaldehyde and glyoxal column concentrations for the year 2007. The use of an older version of the GEOS-Chem forward model was necessary because, at the time of our study, the GEOS-Chem adjoint (version 34) was based on this older version. However, we updated the NMVOC chemical schemes (described below) and corrected several model errors in both our forward model and its adjoint by following the progress of the forward model up to version 10.1.

R1.12 Line 171: Is v10-01 correct? The isoprene chemistry of Paulot et al. (2009a; b) was added to v9-02.

Yes. The isoprene photochemical scheme in v10.1 included updates from Paulot et al. (2009a,b) and Mao et al (2013). We re-wrote this paragraph to clarify this point, as well as provide additional details on the updated isoprene photochemical scheme:

[Main text, lines 185 to 191]: We replaced the isoprene photochemical scheme with that used in GEOS-Chem v10.1, which included updates from Paulot et al. (2009a,b) and Mao et al. (2013). In this updated scheme, oxidation of isoprene by OH under high-NO_x

conditions produces formaldehyde and glyoxal at yields of 0.436 molecules per C and 0.0255 molecules per C, respectively (Table S1), mainly via the RO_2+NO pathways. Under low-NO_x conditions, oxidation of isoprene by OH produces formaldehyde and glyoxal at yields of 0.38 molecules per C and 0.073 molecules per C, respectively (Table S1), via both RO_2+HO_2 and RO_2-isomerization reactions.

R1.13 Line 181: Provide the yield values for Fu et al. (2008).

Thank you for the suggestion. We re-wrote this sentence to include the glyoxal yields from aromatics in Fu et al. (2008), those in our updated model, as well as those used by Chan Miller et al. (2016):

[Main text line 197 to 202]: We updated the molar yields of glyoxal from the OH oxidations of benzene (33.3%), toluene (26.2%), and xylenes (21.0%) following the latest literature (Arey et al., 2009; Nishino et al., 2010). These new molar yields were higher than those used in Fu et al. (2008) (based on averaged yields in the literature: 25.2% for benzene, 16.2% for toluene, and 15.6% for xylenes) but still lower than those used by Chan Miller et al. (2016) (75% for benzene, 70% for toluene, and 36% for xylenes), which were taken from the aromatic chemical scheme in MCM version 3.2 (Jenkin et al., 2003; Bloss et al., 2005).

R1.14 Line 183: Bloss et al. (2005) was used above as the reference for MCM v3.1. What is the appropriate reference for MCM v3.2?

Thank you for pointing out this error. We have updated the references for different updates to MCM:

[Main text, lines 172 to 174]: We updated the dicarbonyl chemical mechanism in GEOS-Chem developed by Fu et al. (2008), which in turn was originally adapted from the Master Chemical Mechanism (MCM) version 3.1 (Jenkin et al., 1997; Saunders et al., 2003).

[Main text, lines 198 to 205]: These new molar yields were higher than those used in Fu et al. (2008) (based on averaged yields in the literature: 25.2% for benzene, 16.2% for toluene, and 15.6% for xylenes) but still lower than those used by Chan Miller et al. (2016) (75% for benzene, 70% for toluene, and 36% for xylenes), which were taken from the aromatic chemical scheme in MCM version 3.2 (Jenkin et al., 2003; Bloss et al., 2005). In MCM version 3.2, more than half of the glyoxal from aromatics oxidation were produced during second- and later-generation photochemistry, but such productions are with limited

experimental support and uncertain (Bloss et al., 2005).

R1.15 Line 187: Does "our model" refer to GEOS-Chem?

Yes. Corrected to improve clarity:

[Main text, lines 207 to 208]: Formaldehyde and glyoxal in the GEOS-Chem model were both removed by photolysis, as well as dry and wet deposition (Fu et al., 2008).

R1.16 Line 188: What was the Henry's law constant updated from and to?

We added details about the updated Henry's law constant:

[Main text, lines 208 to 211]: We updated the Henry's law constant for glyoxal from $3.6 \times 10^5 \times \exp[7.2 \times 10^3 \times (1/T\text{-}1/298)]$ (Fu et al., 2008) to $4.19 \times 10^5 \times \exp[(62.2 \times 10^3/R) \times (1/T\text{-}1/298)]$ (Ip et al., 2009) and added the dry deposition of formaldehyde, glyoxal, methyglyoxal and glycolaldehyde on leaves (Mao et al., 2013).

R1.17 Line 220: Specify which version of MEGAN is used in GEOS-Chem.

We used MEGAN v2.0 (Guenther et al., 2006). This sentence was re-written as follows:

[Main text, lines 240 to 242]: The *a priori* biogenic NMVOC emissions from China and from the rest of the world were calculated with the MEGAN v2.0 algorithm (Guenther et al., 2006) and dependent on temperature, shortwave radiation, and monthly mean leaf area index.

R1.18 Liner 245: Was MEIC also scaled to 2007? As written this isn't clear.

No, we did not scale the MEIC emission estimates to the year 2007, because the uncertainty of the anthropogenic NMVOC emission estimates were much larger than the differences in emissions between the years 2007 and 2010. We added the following comment:

[Main text, lines 268 to 271]: As such, we did not scale the MEIC Chinese NMVOC emissions to the year 2007, because the uncertainty in the emission estimates were much larger than the differences in emissions between the years 2007 and 2010 (Chinese anthropogenic NMVOC emissions increased 14% from 2006 to 2010 according to Li et al, 2017).

R1.19 Lines 250, 252, 643: "burnt" should be "burned".

Corrected. Thank you.

R1.20 Lines 253-254: Is the CO flux scaled or is CO used to estimate (or perhaps scale) NMVOC emissions?

We used the CO emissions from crop residue burning estimated by Huang et al. (2012) and NMVOC-to-CO emission ratios for crop residue burning (Hays et al., 2002; Akagi et al., 2011) to estimate NMVOC emissions from crop residue burning. We rewrote the following sentences to make our treatment clear.

[Main text, lines 288 to 295]: Huang et al. (2012) estimated the Chinese CO emission from crop residue burning to be 4.0 Tg y^{-1}, based on MODIS daily thermal anomalies, Chinese provincial burned biomass data, and emission factors from Akagi et al. (2011). We scaled this CO flux using speciated NMVOC emission factors from crop residue burning from the literature (Hays et al., 2002; Akagi et al., 2011) and then multiplied the resulting NMVOC flux estimate by two. The reason for doubling the scaled NMVOC flux was that the emission factors for many NMVOC species were not measured, such that the sum of the speciated NMVOC emission factors was only half of the total NMVOC emission factor (Akagi et al., 2011).

R1.21 Line 308: Should "IMAGE" by "IMAGES"?

Corrected. Thank you.

R1.22 Lines 445-447: The sentence beginning "As biogenic emissions…" is challenging to follow. Seems there's a logical step missing.

We rewrote this sentence to improve clarity:

[Main text, lines 490 to 493]: During winter (particularly in January), the GOME-2A glyoxal VCDs show an enhancement over eastern China, which was not apparent in the GOME-2A formaldehyde VCDs. This indicated that the glyoxal VCDs were more reflective of anthropogenic source than formaldehyde VCDs.

R1.23 Line 464: "OMI formaldehyde VCDs were higher" than what? The *a priori*?

We rewrote this sentence to improve clarity:

[Main text, lines 513 to 515]: The spatial patterns and seasonal variations of the formaldehyde VCDs observed by OMI were similar to those observed by GOME-2A, with high formaldehyde over eastern China and during the warmer months.

R1.24 Lines 574-575: What does "strong traction" mean?

We rewrote this sentence to improve clarity:

[Main text, lines 590 to 593]: For precursors that produced large amounts of both formaldehyde and glyoxal (most importantly biogenic isoprene), the inversion reduced the top-down emissions as the formaldehyde observations had more weight in the cost function than the glyoxal observations, due to the lower observational errors in the formaldehyde VCDs.

R1.25 Figures 5 and 6: Are ground-based observations sampled at the same time as the satellite overpass?

Yes. We added clarification on this point in the main text, in the captions of Figures 3 to 10, and in the title of Table S3.

[Main text, lines 461 to 464]: A few ground-based measurements of tropospheric formaldehyde VCDs have been made in China using the Multi-Axis Differential Optical Absorption Spectrometry (MAX-DOAS) technique (Li et al., 2013; Vlemmix et al., 2015; Wang et al., 2017); these measurements (sampled at GOME-2A overpass time) are shown in Figure 3, Figure 4, and Table S3.

R1.26 Figures 4,6: "Monthly mean formaldehyde" in the figure caption is deceptive if seasonal means are shown for the ground-based observations.

Thank you for the suggestion. We now show all comparisons between satellite observations, model simulations, and ground-based MAX-DOAS measurements on a monthly basis, with the exception of measurements at Wuxi, which were only available as bi-monthly means. We added Table S3 to show the details of the MAX-DOAS measurements.

[Supplementary information, Table S3]: Ground-based MAX-DOAS measurements of formaldehyde and glyoxal vertical column densities in China at GOME-2A and OMI overpass times

Reference	Location	Time of measurement		Vertical column densities	
				9-10 local time	13-14 local time
Formaldehyde [10^{16} molecules cm^{-2}]					
Vlemmix et al. (2015)	Xianghe, Heibei (39.75N, 116.96E)	2011	JAN	0.24	0.54
			FEB	0.78	0.99
			MAR	0.77	0.95
			APR	0.99	0.98
			MAY	1.08	1.53
			JUN	2.06	2.67
			JUL	1.49	2.10
			AUG	1.47	2.03
			SEP	1.05	1.36
			OCT	1.11	1.64
			NOV	0.85	1.18

		2010	DEC	0.49	0.79
Lee et al. (2015)	Beijing (39.59°N, 116.18°E)	August 16 to September 11, 2006		-	1.79
Wang et al. (2017)	Wuxi, Jiangsu (31.57°N,120.31°E)	2011 - 2014	JF	0.7 [a]	0.8 [a]
			MA	0.9 ±0.15 [a]	1.1 ±0.26 [a]
			MJ	1.5 ±0.12 [a]	1.9 ±0.15 [a]
			JA	1.7 ±0.10 [a]	2.2 ±0.26 [a]
			SO	1.2 ±0.12 [a]	1.7 ±0.12 [a]
			ND	0.8 ±0.30 [a]	1.4 ±0.32 [a]
Li et al. (2013)	Back Garden, Guangdong (23.50°N, 113.03°E)	July 2006		1.3 ±1.0 [b]	1.3 ±0.7 [b]
Glyoxal [10^{14} molecules cm^{-2}]					
Li et al. (2013)	Back Garden, Guangdong (23.50°N, 113.03°E)	July 2006		6.8 ±5.2 [c]	11.4 ±6.8 [c]

[a] From Figure 12 of Wang et al. (2017)

[b] From Figure 4 of Li et al. (2013)

[c] From Figure 5 of Li et al. (2013)

R.1.27 Figures 4-7, 10: Increase the size of the points showing the ground-based measurements.

The symbols in Figures 3 to 10, Figure 13, and Figure S8 have been enlarged as recommended. Thank you.

Reviewer 2:

This study reports top-down estimates of non-methane volatime organic compound emissions over China based on formaldehyde and glyoxal column observations from two sounders, OMI and GOME-2 for 2007. Based on model simulations with the adjoint of the GEOS-Chem model, Cao et al. analyze the impacts of the different satellite datasets on the top-down emission estimates. They find that the annual total top-down VOC emission amounts to 30 Tg C, by 10% higher than the *a priori* inventory. In addition, using glyoxal retrievals from OMI, the authors estimate the annual aromatics Chinese source from 5 to 7.3 Tg C, also higher than in the bottom-up inventory. This study addresses an interesting subject for Atmospheric Chemistry and Physics journal. However, there are several weaknesses in the current work.

For example, the figures are not informative enough and cannot properly feed the discussion, the tables appear in an illogical order, some key statements appear without citation, references are missing. In addition, I see contradictions in the top-down emission estimates mentioned in the abstract and not enough details (and possibly errors) in the chemical scheme. Therefore, I have doubts regarding the validity of the conclusions and think that the manuscript will need a major revision before it becomes suitable for publication.

General comments :

R2.1 The chemical mechanism described very briefly in Section 2.1 is the core ingredient of the top-down VOC studies.

In l.164-165, several NMVOC precursors of formaldehyde are mentioned, but key precursors like methanol, acetaldehyde, ethanol, acetone, etc. do not show in the list. Why are these compounds omitted? Provide also more details on C4 alkanes (l.165).

Thank you for pointing out this lack of clarity. We made major revisions to our mechanism and now included methanol as an independent tracer. Anthropogenic ethanol was lumped into $\geqslant C_4$ alkanes. Chinese biogenic ethanol was not included due to its small source. We rewrote the description of our NMVOC precursors to formaldehyde and glyoxal to improve clarity:

[Main text, lines 175 to 182]: The lumped NMVOC precursors of formaldehyde in our mechanism included ethane, propane, $\geqslant$C4 alkanes, ethene, $\geqslant$C3 alkenes, benzene, toluene, xylenes, isoprene, monoterpenes, acetone, hydroxyacetone, methygloxal, glycolaldehyde, acetaldehyde, 2-methyl-3-bute-nol, methy ethyl ketone, methanol, and ethanol (lumped into $\geqslant$C4 alkanes). The lumped NMVOC precursors of glyoxal in our mechanism included ethene, ethyne, benzene, toluene, xylenes, isoprene, monoterpenes, glycolaldehyde, and 2-methyl-3-bute-2-nol (MBO). Hereinafter we focused our discussion on these NMVOC precursors only, as their emissions may be constrained by formaldehyde and glyoxal observations.

R2.2 In l.166 propane and (higher) alkanes are mentioned as glyoxal precursors. I have serious doubts on this. Please elaborate on the degradation scheme leading to glyoxal in your model.

Thank you for pointing this out. Propane and high alkanes were not glyoxal precursors in our model. The original statement was a typo on our part, which has been removed. We

rewrote the description of our NMVOC precursors to formaldehyde and glyoxal to improve clarity:

[Main text, lines 175 to 182]: The lumped NMVOC precursors of formaldehyde in our mechanism included ethane, propane, ⩾C4 alkanes, ethene, ⩾C3 alkenes, benzene, toluene, xylenes, isoprene, monoterpenes, acetone, hydroxyacetone, methygloxal, glycolaldehyde, acetaldehyde, 2-methyl-3-bute-nol, methy ethyl ketone, methanol, and ethanol (lumped into ⩾C4 alkanes). The lumped NMVOC precursors of glyoxal in our mechanism included ethene, ethyne, benzene, toluene, xylenes, isoprene, monoterpenes, glycolaldehyde, and 2-methyl-3-bute-2-nol (MBO). Hereinafter we focused our discussion on these NMVOC precursors only, as their emissions may be constrained by formaldehyde and glyoxal observations.

R2.3 l.170-172 : provide more details on how glyoxal is formed at both high- and low-NOx levels.

Thank you for the suggestion. We added more details on the formation of glyoxal from isoprene oxidation in the main text, as well as a summary of the yields of formaldehyde and glyoxal from individual NMVOC precursors (Table S1).

[Main text, lines 187 to 195]: In this updated scheme, oxidation of isoprene by OH under high-NOx conditions produces formaldehyde and glyoxal at yields of 0.436 molecules per C and 0.0255 molecules per C, respectively (Table S1), mainly via the RO_2+NO pathways. Under low-NO_x conditions, oxidation of isoprene by OH produces formaldehyde and glyoxal at yields of 0.38 molecules per C and 0.073 molecules per C, respectively (Table S1), via both RO_2+HO_2 and RO_2-isomerization reactions. Li et al. (2016) implemented this same isoprene photochemical scheme into a box model and compared the productions of formaldehyde and glyoxal from isoprene oxidation with those in the MCM version 3.3.1 (Jenkin et al., 2015). They showed that the production pathways and yields of formaldehyde and glyoxal were similar in the two schemes under the high-NOx conditions typical of eastern China.

R2.4 l.172-176 : I don't get this. Li et al. (2016) discusses the AM3 mechanism, not the GEOS-Chem mechanism. Furthermore, the statement that the updated scheme matches the MCM yields is not correct, the NOx-dependence of the yield is completely different in the two schemes.

Thank you for pointing out this lack of clarity. The isoprene photochemistry mechanism which Li et al. (2016) implemented into the AM3 model was from GEOS-Chem v10.1 and

identical to the one we used. The Reviewer was correct in that the NOx dependence of the glyoxal yield from isoprene oxidation were different between the GEOS-Chem v10.1 mechanism (which we used) and the MCM v3.3.1. Our point was that, under the high-NOx conditions typical of eastern China, the production pathways and yields of formaldehyde and glyoxal were similar in these two mechanisms.

We rewrote this paragraph to improve clarity:

[Main text, lines 184 to 195]: The OH-oxidation of isoprene is a major source of both formaldehyde and glyoxal over China (Fu et al., 2007, 2008; Myriokefalitakis et al., 2008). We replaced the isoprene photochemical scheme with that used in GEOS-Chem v10.1, which included updates from Paulot et al. (2009a,b) and Mao et al. (2013). In this updated scheme, oxidation of isoprene by OH under high-NO_x conditions produces formaldehyde and glyoxal at yields of 0.436 molecules per C and 0.0255 molecules per C, respectively (Table S1), mainly via the RO_2+NO pathways. Under low-NO_x conditions, oxidation of isoprene by OH produces formaldehyde and glyoxal at yields of 0.38 molecules per C and 0.073 molecules per C, respectively (Table S1), via both RO_2+HO_2 and RO_2-isomerization reactions. Li et al. (2016) implemented this same isoprene photochemical scheme into a box model and compared the productions of formaldehyde and glyoxal from isoprene oxidation with those in the MCM version 3.3.1 (Jenkin et al., 2015). They showed that the production pathways and yields of formaldehyde and glyoxal were similar in the two schemes under the high-NO_x conditions typical of eastern China.

R2.5 Provide a table with formation yields at high and low NOx conditions for formaldehyde and glyoxal from their respective precursors.

Thank you for this suggestion. We added Table S1 to summarize the yields of formaldehyde and glyoxal from the oxidation of individual NMVOC precursors under high- and low-NO_x conditions.

[Supplementary information, Table S1]: Ultimate yields of formaldehyde and glyoxal from the oxidation of NMVOC precursors by OH in our model under high-NOx and low-NOx conditions

NMVOCs	Formaldehyde (molecules per C)		Glyoxal (molecules per C)	
	High-NO_x [a]	Low-NO_x [b]	High-NO_x [a]	Low-NO_x [b]

Ethene	0.995	0.366	0.0665	0.067
Glycolaldehyde	0.366	0.366	0.067	0.067
Isoprene	0.436	0.38	0.0255	0.073
2-methyl-3-bute-nol (MBO)	0.092	0.092	0.0168	0.0168
Benzene	0.001	0.001	0.0555	0.0555
Toluene	0.198	0.18	0.037	0.037
Xylenes	0.269	0.155	0.026	0.026
Monoterpenes (lumped)	0.006	0.006	0.005 [c]	0.005 [c]
Ethyne	-	-	0.318	0.318
Methanol	1.0	1.0	-	-
Ethane	0.5	0.5	-	-
Acetaldehyde (lumped)	0.5	0.5	-	-
Propane	0.49	0.317	-	-
$\geq C_3$ alkenes (lumped)	0.657	0.333	-	-
Acetone	0.64	0.383	-	-
Hydroxyacetone	0.333	0.333	-	-
Methyglyoxal	0.333	0.333	-	-
$\geq C_4$ alkanes (lumped)	0.578	0.187	-	-
Methy ethyl ketone (lumped)	0.465	0.25	-	-

[a] Yields under high-NO_x conditions were calculated assuming that all RO_2 radicals from the oxidation of the NMVOC precursor reacted with NO.

[b] Yields under low-NO_x conditions were calculated assuming RO_2:HO_2 concentration ratio of 1:1.

[c] Glyoxal produced from the oxidation of monoterpenes by ozone

R2.6 The comparisons between emission estimates shown in Table 1 relies heavily on conversion factors of 0.84, 0.57 and 0.85 for anthropogenic, biomass burning C2 and biogenic VOC, respectively. There is no reference on how these numbers are calculated. In particular, for isoprene and monoterpenes the factor of 0.85 is wrong. For methanol the real factor is also much lower.

Thank you for pointing out this lack of clarity. In response, we have changed the unit for NMVOC emissions from Tg C y^{-1} to Tg y^{-1} to avoid the use of NMVOC carbon conversion factors.

R2.7 Table 1 misses emission estimates from widely used recent bottom-up and topdown inventories, e.g. GFED4 (van der Werf et al. 2017) on biomass burning emissions, HTAPV2 (Janssens-Maenhout et al. 2015) and EDGARv4.3.2 (Huang et al. 2017) on anthropogenic emissions, MEGAN-MACC (Sindelarova et al. 2014) and MEGAN-MOHYCAN (Stavrakou

et al. 2014) on biogenic VOC, MACCity (Granier et al. 2011) on global anthropogenic and fire inventories. Especially for China, top-down estimates from Fu et al. (2008), Bauwens et al. (2016), Stavrakou et al. (2017), Granier et al. (2017) are missing.

Thank you for this suggestion. We added most of these additional emission estimates to Table 2. The HTAPv2 emission estimates (Janssens-Maenhout et al., 2015) was not included in Table 2, as they were actually the MEIC emission estimates from Li et al. (2017).

[Main text, Table 2]: Comparison of Chinese annual NMVOC emission estimates for the years 2000 to 2014

Literature	Target year	NMVOC [Tg y^{-1}]				
		Anthropogenic		Biogenic		Biomass burning
		Total	Aromatics	Total	Isoprene	
Bottom-up estimates						
Bo et al. (2008) [a]	2005	12.7				3.8 [d]
Zhang et al. (2009) [a]	2006	23.2 (±68%)	2.4			
Cao et al. (2011) [a]	2007	35.46				
Huang et al. (2017) [a]	2007	24.6				
Granier et al. (2017) [a]	2007	29.0				
Kurokawa et al. (2013) [a]	2008	27.1 (±46%)				
Li et al. (2017) [a]	2010	23.6	5.4			
Wu et al. (2016) [a]	2008	18.62				3.83 [d]
	2009	21.8				3.32 [d]
	2010	23.83				3.75 [d]
	2011	24.78				3.76 [d]
	2012	25.65				4.20 [d]
Huang et al. (2012) [a]	2006					2.2 (1.08 to 3.46)
van der Werf et al. (2010)	2007					0.47
van der Werf et al. (2017) [a]	2007					0.91
Sindelarova et al. (2014)	2005				9.9	
Guenther et al.(2006)	2007			17.3 [e]	7.5 [e]	
Stavrakou et al. (2014)	2007				7.6	
Top-down estimates						
Fu et al. (2007)	2000	4.27 [g]		12.7		5.1
Liu et al. (2012) [b]	2007	34.2	13.4			
Stavrakou et al. (2014)	2007				8.6	
Stavrakou et al. (2015) [c]	2010	20.6 to 24.6			5.9 to 6.5	2.0 to 2.7
Stavrakou et al. (2017) [c]	2005	24.4			5.8	

	2006	24.0			(average of emissions from 2005 to 2011)	
	2007	26.7				
	2008	25.9				
	2009	26.5				
	2010	26.1				
	2011	25.5				
	2012	25.6				
	2013	27.7				
	2014	27.8				
This work	2007	20.2 [f] (16.4 - 23.6)	6.5 [f] (5.5 - 7.9)	19.2 [f] (12.2 - 22.8)	9.6 [f] (5.4 - 11.7)	2.48 [f] (2.08 – 3.13)

[a] These emission estimates included some NMVOC species which were not precursors to formaldehyde or glyoxal and therefore not included in this work. See color keys in Figure 2 for NMVOC species whose emissions were included in this work.

[b] Used SCIAMACHY-observed glyoxal VCDs as constraints.

[c] Used GOME-2A-observed and OMI-observed formaldehyde VCDs as constraints.

[d] Consisted of emissions from open burning of crop residues and from biofuel burning.

[e] Calculated by the GEOS-Chem model using GEOS-5 meteorological data.

[f] Average of top-down estimates from four inversion experiments.

[g] Only anthropogenic emissions of reactive alkenes, formaldehyde, and xylenes from northeastern, northern, central and southern China were included

R2.8 The Table ordering is illogical. Table 3 should rather become Table 1 or move to the supplement. Table 2 describes the simulations, so it should come first. Table 3 shows results and comparisons to previous studies so it should be called in the result section.

Thank you for the suggestion. We re-ordered the presentation of the tables in the main text as follows:

[Main text, Table 1]: Inversion experiments to constrain Chinese NMVOC emissions

Inversion experiments	Observational constraints from satellites [±uncertainties]	Annual Chinese NMVOC emission estimates [Tg y^{-1}]			
		Anthropogenic	Biogenic	Biomass burning	Total
		A priori emission estimates [uncertainty]			
		18.8 (5.4 for aromatics) [a]	17.3 (7.5 for isoprene) [b]	2.27 [factor of	38.3

		[factor of two uncertainty]	[±55% uncertaitny	three uncertainty] [c]	
		A posteriori emission estimates [range of estimates]			
IE-1	GOME-2A formaldehyde [±90%] and glyoxal [±150%]	17.8 (5.8 for aromatics)	20.0 (9.8 for isoprene)	2.27	40.1
IE-2	OMI formaldehyde [±90%] and glyoxal [±150%]	16.4 (5.5 for aromatics)	12.2 (5.4 for isoprene)	2.08	30.7
IE-3	GOME-2A formaldehyde × 170% [±90%]	23.6 (6.6 for aromatics)	22.8 (11.3 for isoprene)	3.13	49.5
IE-4	OMI glyoxal [±150%]	23.0 (7.9 for aromatics)	21.6 (11.7 for isoprene)	2.43	47.0
Our top-down estimates		20.2 [d] [16.4 - 23.6] (6.5 [d] [5.5 - 7.9] for aromatics)	19.2 [d] [12.2 – 22.8] (9.6 [d] [5.4 – 11.7] for isoprene)	2.48 [d] [2.08 – 3.13]	41.9 [d] [30.7 – 49.5]

[a] From Li et al. (2017)

[b] From Guenther et al. (2006).

[c] Compiled from the emission estimated by van der Werf et al. (2010) plus a scaling of the emission estimated by Huang et al. (2012). See text (section 2.2) for details.

[d] Average of top-down estimates from the four inversion experiments.

[Main text, Table 2]: Comparison of Chinese annual NMVOC emission estimates for the years 2000 to 2014

Literature	Target year	NMVOC [Tg y^{-1}]				
		Anthropogenic		Biogenic		Biomass burning
		Total	Aromatics	Total	Isoprene	
Bottom-up estimates						
Bo et al. (2008) [a]	2005	12.7				3.8 [d]
Zhang et al. (2009) [a]	2006	23.2 (±68%)	2.4			
Cao et al. (2011) [a]	2007	35.46				
Huang et al. (2017) [a]	2007	24.6				
Granier et al. (2017) [a]	2007	29.0				
Kurokawa et al. (2013) [a]	2008	27.1 (±46%)				
Li et al. (2017) [a]	2010	23.6	5.4			
Wu et al. (2016) [a]	2008	18.62				3.83 [d]
	2009	21.8				3.32 [d]

	2010	23.83				3.75 [d]
	2011	24.78				3.76 [d]
	2012	25.65				4.20 [d]
Huang et al. (2012) [a]	2006					2.2 (1.08 to 3.46)
van der Werf et al. (2010)	2007					0.47
van der Werf et al. (2017) [a]	2007					0.91
Sindelarova et al. (2014)	2005				9.9	
Guenther et al.(2006)	2007			17.3 [e]	7.5 [e]	
Stavrakou et al. (2014)	2007				7.6	
Top-down estimates						
Fu et al. (2007)	2000	4.27 [g]		12.7		5.1
Liu et al. (2012) [b]	2007	34.2	13.4			
Stavrakou et al. (2014)	2007				8.6	
Stavrakou et al. (2015) [c]	2010	20.6 to 24.6			5.9 to 6.5	2.0 to 2.7
Stavrakou et al. (2017) [c]	2005	24.4			5.8 (average of emissions from 2005 to 2011)	
	2006	24.0				
	2007	26.7				
	2008	25.9				
	2009	26.5				
	2010	26.1				
	2011	25.5				
	2012	25.6				
	2013	27.7				
	2014	27.8				
This work	2007	20.2 [f] (16.4 - 23.6)	6.5 [f] (5.5 - 7.9)	19.2 [f] (12.2 - 22.8)	9.6 [f] (5.4 - 11.7)	2.48 [f] (2.08 – 3.13)

[a] These emission estimates included some NMVOC species which were not precursors to formaldehyde or glyoxal and therefore not included in this work. See color keys in Figure 2 for NMVOC species whose emissions were included in this work.

[b] Used SCIAMACHY-observed glyoxal VCDs as constraints.

[c] Used GOME-2A-observed and OMI-observed formaldehyde VCDs as constraints.

[d] Consisted of emissions from open burning of crop residues and from biofuel burning.

[e] Calculated by the GEOS-Chem model using GEOS-5 meteorological data.

[f] Average of top-down estimates from four inversion experiments.

[g] Only anthropogenic emissions of reactive alkenes, formaldehyde, and xylenes from northeastern, northern, central and southern China were included

R2.9 In the abstract you mention that the annual total NMVOC emissions ranges from 23.5 to 35.4 Tg C (mean of 30.8). This does not match the sum of individual categories given in lines 27-29 of the abstract (23.5-36 Tg C). This brings confusion to the reader already from the first lines. Which one is correct? Change accordingly throughtout the paper and the Tables. In l.29-30 provide a name for the "most widely used bootom-up inventory".

Thank you for point out this lack of clarity, which was originally due to the expression of NMVOC emissions in units of Tg C y^{-1}. We have changed the unit for NMVOC emissions from Tg C y^{-1} to Tg y^{-1} to make the numbers consistent. We also re-wrote the description of the *a priori* emission inventories.

[Main text, Abstract, line 27 to 30]: Our top-down estimates for Chinese annual total NMVOC emission were 30.7 to 49.5 (average 41.9) Tg y^{-1}, including 16.4 to 23.6 (average 20.2) Tg y^{-1} from anthropogenic sources, 12.2 to 22.8 (average 19.2) Tg y^{-1} from biogenic sources, and 2.08 to 3.13 (average 2.48) Tg y^{-1} from biomass burning.

[Main text, lines 798 to 806]: Our top-down estimates of total annual Chinese NMVOC emission from the four inversion experiments ranged from 30.7 to 49.5 Tg y^{-1}. Our top-down estimates of Chinese anthropogenic NMVOC emission was 16.4 to 23.6 Tg y^{-1}. In particular, our top-down estimates for Chinese anthropogenic aromatic emissions ranged from 5.5 to 7.9 Tg y^{-1}, much smaller than the top-down estimate of 13.4 Tg y^{-1} by Liu et al. (2012). Our top-down estimate of Chinese biogenic NMVOC emission ranged from 12.2 to 22.8 Tg y^{-1}, with 5.4 to 11.7 Tg y^{-1} attributed to isoprene. Our top-down estimate for Chinese biomass burning NMVOC emission range from 2.08 to 3.13 Tg y^{-1} and was mostly associated with seasonal open burning of crop residue after local harvests, such as those over the NCP in June.

[Main text, lines 788 to 791]: The a priori NMVOC emission estimates from biogenic, anthropogenic, and biomass burning sources were taken from the inventories developed by Guenther et al. (2006), Li et al (2014, 2017), and Huang et al. (2012), as well as van der Werf et al. (2010), respectively.

R2.10 l.231 : Do you mean 19.8 Tg C from Table or am I missing something? I have several doubts about the reported numbers. Check again before you resubmit.

Please see the response to the previous comment.

R2.11 In l. 239, the uncertainty of *a priori* emissions is given, ±200%. Is this what is really meant here? It would correspond to a range of -20 to 60 Tg C. This makes no sense given the reported numbers for the anthropogenic flux from different inventories. Same for l. 224, 267.

Thank you for pointing out this lack of clarity. We re-wrote the statements on the uncertainty of the *a priori* emission estimates to avoid confusion and to maintain consistency with the original descriptions in the paper by Li et al. (2017).

[Main text line 266 to line 268]: We therefore estimated the uncertainty for the a priori Chinese anthropogenic NMVOC emission estimates to be a factor of two.

[Main text line 305 to line 307]: We therefore estimated the uncertainty of the a priori Chinese biomass burning NMVOC flux to be a factor of three.

R2.12 l.249 : Liu et al. (2015) is based fire radiative power, not burnt area.

Thank you for pointing out this lack of clarity. The original sentence meant that Liu et al. (2015) pointed out the underestimation of emissions in inventories based on satellite burned area observations. We re-wrote this paragraph to improve clarity:

[Main text, lines 281 to 288]: Post-harvest, in-field burning of crop residue has been recognized as a large seasonal source of NMVOCs in China (Fu et al., 2007; Huang et al., 2012; Liu et al., 2015; Stavrakou et al., 2016). These emissions from crop residue fires have been severely underestimated in inventories based on burned area observations from satellites, such as the Global Fire Emissions Database version 3 (GFED3, van der Werf et al., 2010). The recent Global Fire Emissions Database version 4 (GFED4, van der Werf et al., 2017) included small fires by scaling burned area with satellite fire pixel observations, but the resulting Chinese NMVOC emission estimate from biomass burning (0.91 Tg y^{-1}) was still much lower than the bottom-up inventory by Huang et al. (2012).

R2.13 l. 265 : GFED4 (van der Werf et al. 2017) acoounts for agricultural fire burning, which was not the case in GFED3. You should compare with GFED4 for this emission category.

Thank you for the suggestion. However, the NMVOC emissions from small fires in GFED4 were still much lower than both the estimates by Huang et al. (2012) and our top-down estimates. We added these comparisons in the main text and in Table 2.

[Main text, lines 285 to 288]: The recent Global Fire Emissions Database version 4 (GFED4, van der Werf et al., 2017) included small fires by scaling burned area with satellite fire pixel observations, but the resulting Chinese NMVOC emission estimate from biomass burning (0.91 Tg y^{-1}) was still much lower than the bottom-up inventory by Huang et al. (2012).

[Main text, lines 715 to 718]: The updated GFED4 (van der Werf et al., 2017) partially accounted for emissions for small fires, but its estimate for Chinese biomass burning NMVOC emissions was still lower than our top-down estimates by at least a factor of two.

[Main text, Table 2]: Comparison of Chinese annual NMVOC emission estimates for the years 2000 to 2014

Literature	Target year	NMVOC [Tg y^{-1}]				
		Anthropogenic		Biogenic		Biomass burning
		Total	Aromatics	Total	Isoprene	
Bottom-up estimates						
Bo et al. (2008) [a]	2005	12.7				3.8 [d]
Zhang et al. (2009) [a]	2006	23.2 (±68%)	2.4			
Cao et al. (2011) [a]	2007	35.46				
Huang et al. (2017) [a]	2007	24.6				
Granier et al. (2017) [a]	2007	29.0				
Kurokawa et al. (2013) [a]	2008	27.1 (±46%)				
Li et al. (2017) [a]	2010	23.6	5.4			
Wu et al. (2016) [a]	2008	18.62				3.83 [d]
	2009	21.8				3.32 [d]
	2010	23.83				3.75 [d]
	2011	24.78				3.76 [d]
	2012	25.65				4.20 [d]
Huang et al. (2012) [a]	2006					2.2 (1.08 to 3.46)
van der Werf et al. (2010)	2007					0.47
van der Werf et al. (2017) [a]	2007					0.91
Sindelarova et al. (2014)	2005				9.9	
Guenther et al.(2006)	2007			17.3 [e]	7.5 [e]	
Stavrakou et al. (2014)	2007				7.6	
Top-down estimates						
Fu et al. (2007)	2000	4.27 [g]		12.7		5.1
Liu et al. (2012) [b]	2007	34.2	13.4			
Stavrakou et al. (2014)	2007				8.6	
Stavrakou et al. (2015) [c]	2010	20.6 to 24.6			5.9 to 6.5	2.0 to 2.7
Stavrakou et al. (2017) [c]	2005	24.4			5.8	

	2006	24.0			(average of emissions from 2005 to 2011)	
	2007	26.7				
	2008	25.9				
	2009	26.5				
	2010	26.1				
	2011	25.5				
	2012	25.6				
	2013	27.7				
	2014	27.8				
This work	2007	20.2 [f] (16.4 - 23.6)	6.5 [f] (5.5 - 7.9)	19.2 [f] (12.2 - 22.8)	9.6 [f] (5.4 - 11.7)	2.48 [f] (2.08 – 3.13)

[a] These emission estimates included some NMVOC species which were not precursors to formaldehyde or glyoxal and therefore not included in this work. See color keys in Figure 2 for NMVOC species whose emissions were included in this work.

[b] Used SCIAMACHY-observed glyoxal VCDs as constraints.

[c] Used GOME-2A-observed and OMI-observed formaldehyde VCDs as constraints.

[d] Consisted of emissions from open burning of crop residues and from biofuel burning.

[e] Calculated by the GEOS-Chem model using GEOS-5 meteorological data.

[f] Average of top-down estimates from four inversion experiments.

[g] Only anthropogenic emissions of reactive alkenes, formaldehyde, and xylenes from northeastern, northern, central and southern China were included

R2.14 There are many language errors in the manuscript. This decreases its readability. I strongly recommend that the manuscript is corrected by a native speaker among the co-authors and thoroughly re-read.

Thank you for the suggestion. We have carefully rewritten most of the manuscript. The revised manuscript have been proof read by one of the native-English-speaking coauthors.

R2.15 The discussion in Sections 3, 4 is not quantitative. The reader does not get enough information about absolute differences. This should be improved in the revised version.

Thank you for the suggestion. We added Tables S4 to S8 to summarize the statistics of the comparison between satellite-observed and model-simulated formaldehyde and glyoxal VCDs over eastern China. We also added quantitative comparisons in the main text:

[Main text, lines 451 to 459]: The *a priori* simulated formaldehyde VCDs generally reproduced the observed seasonal contrast and spatial patterns over eastern China, with correlation coefficients (R) between 0.74 and 0.94 year-round, except in December (R = 0.51). The *a priori* simulated formaldehyde VCDs were significantly higher than the GOME-2A observations over eastern China between late fall and winter (November, December, January, and February), with normalized mean biases (NMB) of 13% to 67%, implying an overestimate of the anthropogenic formaldehyde precursors in the *a priori* emission estimates. The *a priori* simulated formaldehyde VCDs were lower than the GOME-2A observations over eastern China during May to July (NMB between -11% to -6.4%), implying an underestimation of the emissions of formaldehyde precursors in the *a priori* during May to July.

[Main text, lines 493 to 495]: The *a priori* simulated glyoxal VCDs were generally lower than the GOME-2A glyoxal VCDs over eastern China year-round, especially during the warmer months (NMB between -52% and -59% during May to September, Table S6).

[Main text, lines 519 to 521]: The *a priori* simulated formaldehyde VCDs (at OMI overpass time) were higher than the OMI observations over eastern China year-round (NMB between 22% and 70%, Table S7), suggesting an overestimation of NMVOC emissions year-round.

[Main text, lines 536 to 538]: The *a priori* simulated glyoxal VCDs were lower than the OMI observations throughout the year (NMB between -32% to -66%, Table S8) and especially from March to October, indicating an underestimation of NMVOC sources in the *a priori* year-round.

[Main text, lines 565 to 568]: The optimization was especially effective in optimizing the spatial pattern of the a posteriori formaldehyde VCDs, such that the a posteriori R against the GOME-2A formaldehyde VCDs exceeded 0.85 over eastern China for all twelve months (Table S4).

R2.16 l. 466 : Can you specify what are the differences between the retrievals algorithms? I wonder why you didn't use retrievals from GOME-2 and OMI based on the same retrieval algorithm. These products are available. This should remove undesirable biases due to the different retrieval methodologies.

The Review is correct in pointing out that there are MAX-DOAS retrievals of both GOME-2A and OMI formaldehyde and glyoxal VCDs (De Smedt et al., 2012, 2015; Lerot et al., 2010). However, (1) the GOME-2A and OMI formaldehyde and glyoxal products retrieved using different algorithms provided disparate information on seasonal NMVOC emissions (as shown in Section 3), (2) none of these products have been sufficiently validated over China, and (3) several studies have used these different satellite product to derive top-down emission estimates (e.g., Chan Miller et al., 2016; Stavrakou et al., 2015, 2016). Therefore, the uncertainty associated with the use of different satellite retrievals in top-down Chinese NMVOC emission estimates should be explored in a consistent way.

We emphasized this point in the main text:

[Main text, lines 150 to 153]: In this study, we used satellite retrievals of both formaldehyde and glyoxal, along with a chemical transport model and its adjoint, to constrain NMVOC emissions from China for the year 2007. We conducted sensitivity experiments to evaluate the impacts on the top-down estimates due to different satellite observations, with the goal of bracketing a probable range of top-down estimates.

[Main text, lines 550 to 553]: The qualitative analyses in Section 3 showed that the GOME-2A and OMI retrievals of formaldehyde and glyoxal VCDs provided disparate information on seasonal Chinese NMVOC emissions. Therefore, our four inversion experiments using different satellite observations as constraints represented the range of probable top-down estimates given current satellite observations.

We also provided additional details on the GOME-2A and OMI observations of formaldehyde and glyoxal in Table S2.

[Supplementary information, Table S2]: Technical details for the GOME-2A and OMI formaldehyde and glyoxal observations used in this study

Technical details	GOME-2A		OMI	
	Formaldehyde	Glyoxal	Formaldehyde	Glyoxal
Product reference	De Smedt et al. (2012)	Lerot et al. (2010)	González Abad et al. (2015)	Chan Miller et al. (2014)
Platform	European MetOp-A satellite		NASA Aura satellite	
Operation time	October 2006 – present		July 2004 – present	
Overpass time	9:30 local time		13:30 local time	
Global coverage	Every 1.5 days before June 2013; every 3 days after June 2013		Every 1 day	
Spatial resolution	80 km × 40 km		13 km × 24 km	
Spectral window	240-790 nm		270-500 nm	
Spectral resolution	0.26-0.5 nm		0.42 nm and 0.63 nm	

Selected absorption band		328.5 - 346 nm	435 - 460 nm	328.5 - 356.5 nm	435 - 461 nm
Retrieval algorithm		Differential Optical Absorption Spectroscopy (DOAS) fitting		Direct fitting	
Cloud parameter data		FRESCO+ (Wang et al., 2008)		OMCLDO2 (Acarreta et al., 2004)	
Surface albedo data		Kleipool et al. (2008)		Kleipool et al. (2008)	
Air mass factor calculation	Radiative transfer model	LIDORT (Spurr, 2008)		VLIDORT (Spurr, 2006)	
	Tracer gas profiles	IMAGES model outputs (Stavrakou et al., 2009b)		GEOS-Chem model outputs (González Abad et al., 2015)	
Extinction by aerosols		Considered implicitly via cloud correction (Boersma et al., 2004)		Considered implicitly in the cloud retrieval (Acarreta et al., 2004)	
Discarded pixels		Pixels with cloud fraction >40% or zenith angles >60^{o} were discarded		Pixels with cloud fraction > 40% were discarded	Pixels flagged as impacted by random telegraph signals were discarded [a]

[a] Some pixels were flagged as impacted by random telegraph signals in the level 1-B product (Kleipool, 2005).

R2.17 All figures are based on model/data comparisons only for January, April, June, October. By doing that, we miss important information for other months, especially for July and August (maximum of biogenic emissions). The figures are also hard to read. More synthetic figures should be added, for instance showing the monthly variation of the satellite/model columns over large regions.

Thank you for the suggestion. All comparisons (Figures 3 to 10, Tables S4 to S8) between satellite observations, model simulations, and ground-based MAX-DOAS measurements are now presented on a monthly basis, with the exception of measurements at Wuxi, which were only available as bi-monthly means.

We have also increased the figure resolutions and enlarged the symbols in the Figures.

R2.18 Detailed comparisons with ground-based measurements are missing. The ground-based measurements shown in Figures 4-7 leave a lot to be desired. No concrete conclusion can be drawn from these plots with regards to the observed monthly variation and how well the model can reproduce it.

Thank you for pointing out the issue. In response, we added Table S3 to show the details of the MAX-DOAS measurements. We also added the monthly MAX-DOAS measurements of formaldehyde VCDs at Xianghe (a site in the NCP) at GOME-2A and OMI overpass

time, as well as comparisons against satellite observations and model simulations (Figures S4 and S5). We added discussions in the main text.

[Supplementary information, Table S3]: Ground-based MAX-DOAS measurements of formaldehyde and glyoxal vertical column densities in China at GOME-2A and OMI overpass times

Reference	Location	Time of measurement		Vertical column densities	
				9-10 local time	13-14 local time
Formaldehyde [10^{16} molecules cm^{-2}]					
Vlemmix et al. (2015)	Xianghe, Heibei (39.75N, 116.96E)	2011	JAN	0.24	0.54
			FEB	0.78	0.99
			MAR	0.77	0.95
			APR	0.99	0.98
			MAY	1.08	1.53
			JUN	2.06	2.67
			JUL	1.49	2.10
			AUG	1.47	2.03
			SEP	1.05	1.36
			OCT	1.11	1.64
			NOV	0.85	1.18
		2010	DEC	0.49	0.79
Lee et al. (2015)	Beijing (39.59^{o}N, 116.18^{o}E)	August 16 to September 11, 2006		-	1.79
Wang et al. (2017)	Wuxi, Jiangsu (31.57^{o}N,120.31^{o}E)	2011 - 2014	JF	0.7 [a]	0.8 [a]
			MA	0.9 ±0.15 [a]	1.1 ±0.26 [a]
			MJ	1.5 ±0.12 [a]	1.9 ±0.15 [a]
			JA	1.7 ±0.10 [a]	2.2 ±0.26 [a]
			SO	1.2 ±0.12 [a]	1.7 ±0.12 [a]
			ND	0.8 ±0.30 [a]	1.4 ±0.32 [a]
Li et al. (2013)	Back Garden, Guangdong (23.50^{o}N, 113.03^{o}E)	July 2006		1.3 ±1.0 [b]	1.3 ±0.7 [b]
Glyoxal [10^{14} molecules cm^{-2}]					
Li et al. (2013)	Back Garden, Guangdong	July 2006		6.8 ±5.2 [c]	11.4 ±6.8 [c]

	(23.50°N, 113.03°E)			

[a] From Figure 12 of Wang et al. (2017)

[b] From Figure 4 of Li et al. (2013)

[c] From Figure 5 of Li et al. (2013)

[Supplementary information, Figure S4]:

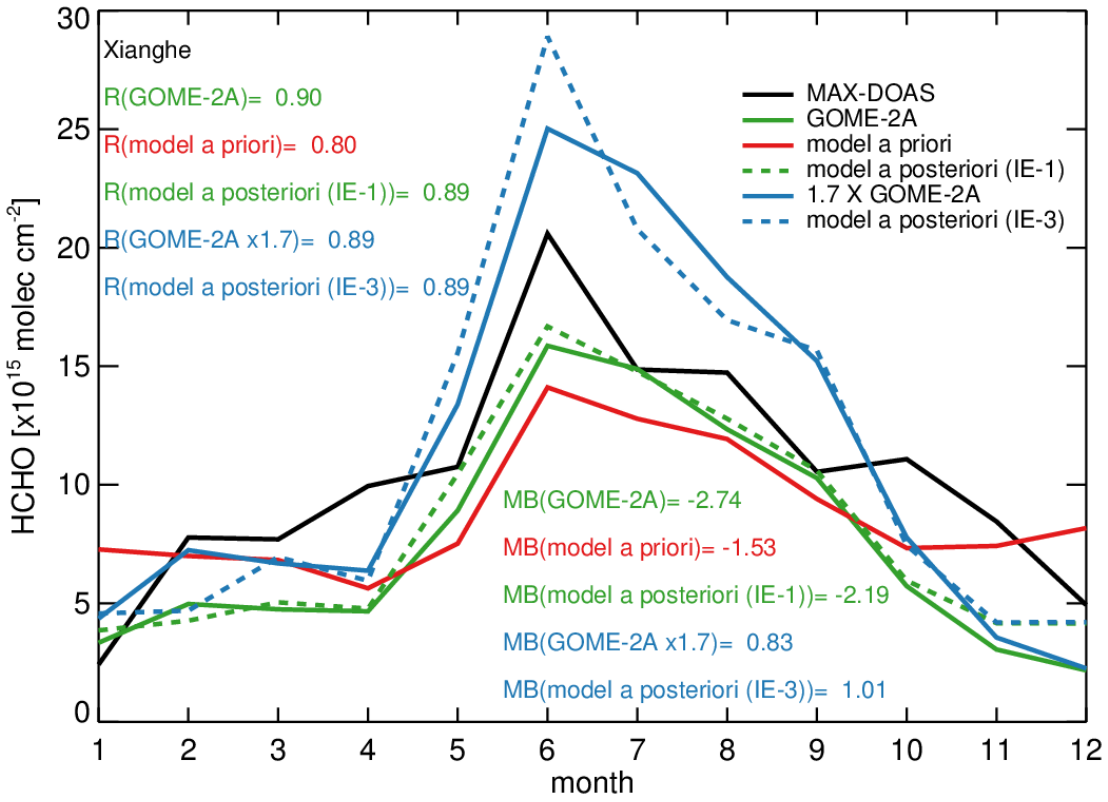


Figure S4. Measured and simulated monthly mean formaldehyde VCDs at Xianghe at GOME-2A overpass time: MAX-DOAS measurements (black line, Vlemmix et al., 2015), GOME-2A measurements (green solid line), GOME-2A measurements multiplied by 1.7 (blue solid line), monthly mean formaldehyde VCDs from the *a priori* simulation (red line), the IE-1 *a posteriori* simulation (green dashed line), and the IE-3 *a posteriori* simulation (blue dashed line). Pearson correlation coefficients (R) of the satellite-observed and simulated formaldehyde VCDs against the MAX-DOAS measurements are shown in the top left. Annual mean bias (MB, in units of 10^{15} molecules cm^{-2}) of the satellite-observed and simulated formaldehyde VCDs against the MAX-DOAS measurements are shown in the bottom right.

[Supplementary information, Figure S5]:

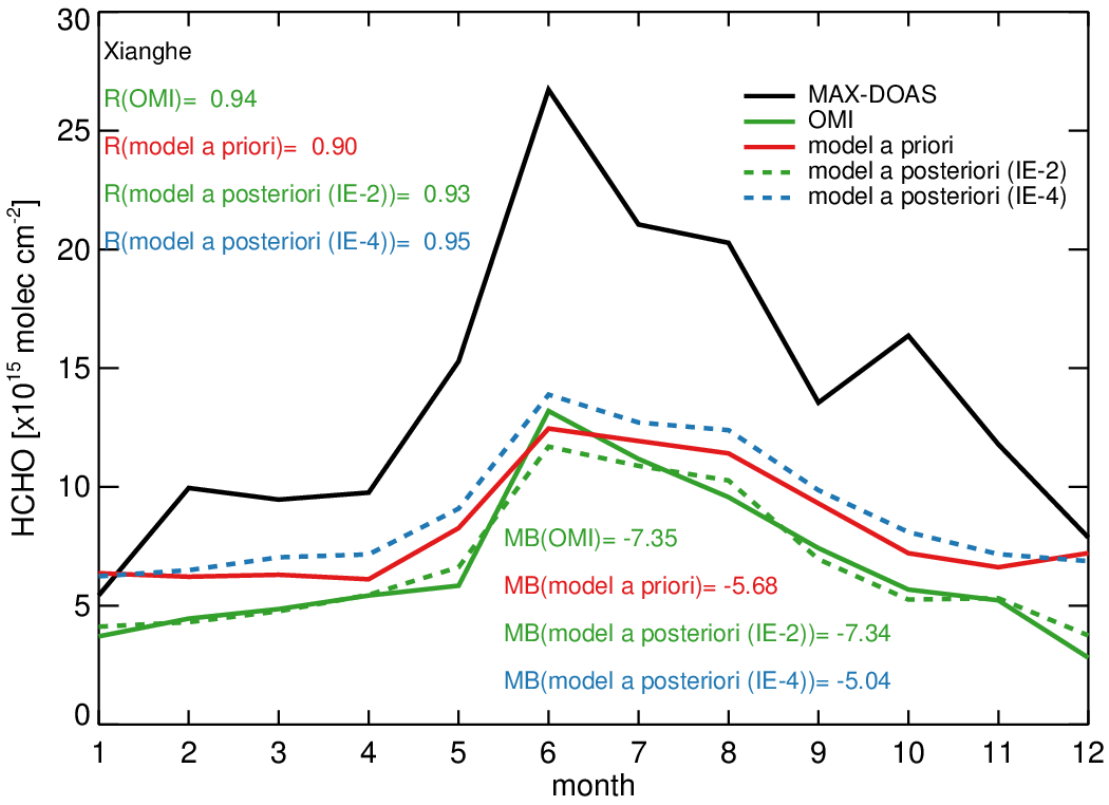


Figure S5 Measured and simulated monthly mean formaldehyde VCDs at Xianghe at OMI overpass time: MAX-DOAS measurements (black line, Vlemmix et al., 2015), OMI measurements (green solid line), monthly mean formaldehyde VCDs from the *a priori* simulation (red line), the IE-2 *a posteriori* simulation (green dashed line), and the IE-4 *a posteriori* simulation (blue dashed line). Pearson correlation coefficients (R) of the satellite-observed and simulated formaldehyde VCDs against the MAX-DOAS measurements are shown in the top left. Annual mean bias (MB, in units of 10^{15} molecules cm^{-2}) of the satellite-observed and simulated formaldehyde VCDs against the MAX-DOAS measurements are shown in the bottom right.

[Main text, lines 461 to 473.]: A few ground-based measurements of tropospheric formaldehyde VCDs have been made in China using the Multi-Axis Differential Optical Absorption Spectrometry (MAX-DOAS) technique (Li et al., 2013; Vlemmix et al., 2015; Wang et al., 2017); these measurements (sampled at GOME-2A overpass time) are shown in Figure 3, Figure 4, and Table S3. In principle, these ground-based measurements are not directly comparable to the satellite-observed and model-simulated formaldehyde VCDs, due to the coarse spatial resolution of our analyses. Nevertheless, these ground-based measurements showed that (1) formaldehyde VCDs were higher during the warmer months relative to the colder months; (2) formaldehyde VCDs over Wuxi (in central eastern China) were higher than those over Xianghe (in northern China) and Back Garden (in southern China) for most months; (3) with the exception of June, when the formaldehyde VCDs over Xianghe were the highest among the three MAX-DOAS sites, reflecting the

strong emissions from biomass burning in the NCP. Thus, the seasonal patterns shown in these few ground-based measurements were consistent with both the GOME-2A-observed and model-simulated formaldehyde VCDs.

[Main text, lines 475 to 484]: Figure S4 compares the GOME-2A-observed and model-simulated formaldehyde VCDs against the monthly MAX-DOAS measurements at Xianghe (Vlemmix et al., 2015). The GOME-2A formaldehyde VCDs were consistent with the MAX-DOAS measurements in terms of the seasonal variation (R = 0.9) but showed an annual mean bias of -2.74×10^{15} molecules cm^{-2}. In comparison, by multiplying the GOME-2A formaldehyde VCD observations by 1.7, the annual mean bias against the MAX-DOAS measurements at Xianghe was reduced to 0.83×10^{15} molecules cm^{-2}. Figures 3 and 4 show that the bias between the satellite and MAX-DOAS measurements was also reduced at Wuxi when the GOME-2A formaldehyde VCDs were scaled up by 1.7. These findings offered some support for using the GOME-2A formaldehyde VCDs scaled by 1.7 as an upper-bound constraints for Chinese NMVOC emissions.

Reference

Akagi, S. K., Yokelson, R. J., Wiedinmyer, C., Alvarado, M. J., Reid, J. S., Karl, T., Crounse, J. D., and Wennberg, P. O.: Emission factors for open and domestic biomass burning for use in atmospheric models, Atmos. Chem. Phys., 11, 4039-4072, doi: 10.5194/acp-11-4039-2011, 2011.

Arey, J., Obermeyer, G., Aschmann, S. M., Chattopadhyay, S., Cusick, R. D., and Atkinson, R.: Dicarbonyl Products of the OH Radical-Initiated Reaction of a Series of Aromatic Hydrocarbons, Environ. Sci. Technol., 43, 683-689, doi: 10.1021/es8019098, 2009.

Barkley, M. P., Palmer, P. I., Kuhn, U., Kesselmeier, J., Chance, K., Kurosu, T. P., Martin, R. V., Helmig, D., and Guenther, A.: Net ecosystem fluxes of isoprene over tropical South America inferred from Global Ozone Monitoring Experiment (GOME) observations of HCHO columns, J. Geophys. Res., 113, doi: 10.1029/2008jd009863, 2008.

Barkley, M. P., Palmer, P. I., De Smedt, I., Karl, T., Guenther, A., and Van Roozendael, M.: Regulated large-scale annual shutdown of Amazonian isoprene emissions?, Geophys. Res. Lett., 36, doi:10.1029/2008gl036843, 2009.

Barkley, M. P., Smedt, I. D., Van Roozendael, M., Kurosu, T. P., Chance, K., Arneth, A., Hagberg, D., Guenther, A., Paulot, F., Marais, E., and Mao, J.: Top-down isoprene emissions over tropical South America inferred from SCIAMACHY and OMI formaldehyde columns, J. Geophys. Res. Atmos., 118, 6849-6868, doi:10.1002/jgrd.50552, 2013.

Bloss, C., Wagner, V., Jenkin, M. E., and Volkamer, R.: Development of a detailed chemical mechanism (MCMv3.1) for the atmospheric oxidation of aromatic hydrocarbons, Atmos. Chem. Phys., 5, 641-664, doi:10.5194/acp-5-641-2005, 2005.

Chan Miller, C., Jacob, D. J., Abad, G. G., and Chance, K.: Hotspot of glyoxal over the Pearl River delta seen from the OMI satellite instrument: implications for emissions of aromatic hydrocarbons, Atmos. Chem. Phys., 16, 4631-4639, doi: 10.5194/acp-16-4631-2016, 2016.

Curci, G., Palmer, P. I., Kurosu, T. P., Chance, K., and Visconti, G.: Estimating European volatile organic compound emissions using satellite observations of formaldehyde from the Ozone Monitoring Instrument, Atmos. Chem. Phys., 10, 11501-11517, doi: 10.5194/acp-10-11501-2010, 2010.

De Smedt, I., Van Roozendael, M., Stavrakou, T., Müller, J. F., Lerot, C., Theys, N., Valks, P., Hao, N., and van der A, R.: Improved retrieval of global tropospheric formaldehyde columns from GOME-2/MetOp-A addressing noise reduction and instrumental degradation issues, Atmos. Meas. Tech., 5, 2933-2949, doi:10.5194/amt-5-2933-2012, 2012.

De Smedt, I., Stavrakou, T., Hendrick, F., Danckaert, T., Vlemmix, T., Pinardi, G., Theys, N., Lerot, C., Gielen, C., Vigouroux, C., Hermans, C., Fayt, C., Veefkind, P., Müller, J. F., and Van Roozendael, M.: Diurnal, seasonal and long-term variations of global formaldehyde columns inferred from combined OMI and GOME-2 observations, Atmos. Chem. Phys., 15, 12519-12545, doi: 10.5194/acp-15-12519-2015, 2015.

Dufour, G., Wittrock, F., Camredon, M., Beekmann, M., Richter, A., Aumont, B., and Burrows, J. P.: SCIAMACHY formaldehyde observations: constraint for isoprene emission estimates over Europe?, Atmos. Chem. Phys., 9, 1647-1664, doi:10.5194/acp-9-1647-2009, 2009.

Fu, T. M., Cao, J. J., Zhang, X. Y., Lee, S. C., Zhang, Q., Han, Y. M., Qu, W. J., Han, Z., Zhang, R., Wang, Y. X., Chen, D., and Henze, D. K.: Carbonaceous aerosols in China: top-down constraints on primary sources and estimation of secondary contribution, Atmos. Chem. Phys., 12, 2725-2746, doi: 10.5194/acp-12-2725-2012, 2012.

Fu, T.-M., Jacob, D. J., Palmer, P. I., Chance, K., Wang, Y. X., Barletta, B., Blake, D. R., Stanton, J. C., and Pilling, M. J.: Space-based formaldehyde measurements as constraints on volatile organic compound emissions in east and south Asia and implications for ozone, J. Geophys. Res., 112, doi: 10.1029/2006jd007853, 2007.

Fu, T.-M., Jacob, D. J., Wittrock, F., Burrows, J. P., Vrekoussis, M., and Henze, D. K.: Global budgets of atmospheric glyoxal and methylglyoxal, and implications for formation of secondary organic aerosols, J. Geophys. Res., 113, doi:10.1029/2007jd009505, 2008.

Gonzi, S., Palmer, P. I., Barkley, M. P., De Smedt, I., and Van Roozendael, M.: Biomass burning emission estimates inferred from satellite column measurements of HCHO: Sensitivity to co-emitted aerosol and injection height, Geophys. Res. Lett., 38, doi: 10.1029/2011gl047890, 2011.

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

Atmos. Chem. Phys., 6, 3181-3210, doi: 10.5194/acp-6-3181-2006, 2006.

Hays, M. D., Geron, C. D., Linna, K. J., Smith, N. D., and Schauer, J. J.: Speciation of gas-phase and fine particle emissions from burning of foliar fuels, Environ. Sci. Technol., 36, 2281-2295, doi:10.1021/es0111683, 2002.

Henze, D. K., Seinfeld, J. H., Ng, N. L., Kroll, J. H., Fu, T. M., Jacob, D. J., and Heald, C. L.: Global modeling of secondary organic aerosol formation from aromatic hydrocarbons: high- vs. low-yield pathways, Atmos. Chem. Phys., 8, 2405-2420, doi:10.5194/acp-8-2405-2008, 2008.

Huang, X., Li, M., Li, J., and Song, Y.: A high-resolution emission inventory of crop burning in fields in China based on MODIS Thermal Anomalies/Fire products, Atmos. Environ., 50, 9-15, doi: 10.1016/j.atmosenv.2012.01.017, 2012.

Ip, H. S. S., Huang, X. H. H., and Yu, J. Z.: Effective Henry's law constants of glyoxal, glyoxylic acid, and glycolic acid, Geophys. Res. Lett., 36, doi: 10.1029/2008GL036212, 2009.

Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Dentener, F., Muntean, M., Pouliot, G., Keating, T., Zhang, Q., Kurokawa, J., Wankmuller, R., van der Gon, H. D., Kuenen, J. J. P., Klimont, Z., Frost, G., Darras, S., Koffi, B., and Li, M.: HTAP_v2.2: a mosaic of regional and global emission grid maps for 2008 and 2010 to study hemispheric transport of air pollution, Atmos. Chem. Phys., 15, 11411-11432, doi: 10.5194/acp-15-11411-2015, 2015.

Jenkin, M. E., Saunders, S. M. Wagner, V., and Pilling, M. J.: The tropospheric degradation of volatile organic compounds: a protocol for mechanism development, Atmos. Environ., 31, 81-104, doi:10.1016/S1352-2310(96)00105-7, 1997.

Jenkin, M. E., Saunders, S. M., Wagner, V., and 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. and Phys., 3, 181-193, doi:10.5194/acp-3-181-2003, 2003.

Jenkin, M. E., Young, J. C., and Rickard, A. R.: The MCM v3.3.1 degradation scheme for isoprene, Atmos. Chem. Phys., 15, 11433-11459, doi: 10.5194/acp-15-11433-2015, 2015.

Kleipool, Q. L., Dobber, M. R., de Haan, J. F., and Levelt, P. F.: Earth surface reflectance climatology from 3 years of OMI data, Journal of Geophysical Research, 113, doi: 10.1029/2008jd010290, 2008.

Kleipool, Q. L.: Transient signal flagging algorithm definition for radiance data, Tech. Rep. TN-OMIE-KNMI-717 TN-OMIEKNMI-717 TN-OMIE-KNMI-717 TN-OMIE-KNMI-717 TNOMIE-KNMI-717, Royal Netherlands Meteorological Institute, De Bilt, the Netherlands, 2005.

Lerot, C., Stavrakou, T., De Smedt, I., Muller, J. F., and Van Roozendael, M.: Glyoxal vertical columns from GOME-2 backscattered light measurements and comparisons with a global model, Atmos. Chem. Phys., 10, 12059-12072, doi: 10.5194/acp-10-12059-2010, 2010.

Li, M., Zhang, Q., Streets, D. G., He, K. B., Cheng, Y. F., Emmons, L. K., Huo, H., Kang, S. C., Lu, Z.,

Shao, M., Su, H., Yu, X., and Zhang, Y.: Mapping Asian anthropogenic emissions of non-methane volatile organic compounds to multiple chemical mechanisms, Atmos. Chem. Phys., 14, 5617-5638, doi:10.5194/acp-14-5617-2014, 2014.

Li, M., Zhang, Q., Kurokawa, J. I., Woo, J. H., He, K., Lu, Z., Ohara, T., Song, Y., Streets, D. G., Carmichael, G. R., Cheng, Y., Hong, C., Huo, H., Jiang, X., Kang, S., Liu, F., Su, H., and Zheng, B.: MIX: a mosaic Asian anthropogenic emission inventory under the international collaboration framework of the MICS-Asia and HTAP, Atmos. Chem. Phys., 17, 935-963, doi:10.5194/acp-17-935-2017, 2017.

Li, X., Brauers, T., Hofzumahaus, A., Lu, K., Li, Y. P., Shao, M., Wagner, T., and Wahner, A.: MAX-DOAS measurements of NO2, HCHO and CHOCHO at a rural site in Southern China, Atmos. Chem. Phys., 13, 2133-2151, doi: 10.5194/acp-13-2133-2013, 2013.

Liao, H., Henze, D. K., Seinfeld, J. H., Wu, S. L., and Mickley, L. J.: Biogenic secondary organic aerosol over the United States: Comparison of climatological simulations with observations, Journal of Geophysical Research, 112, doi:10.1029/2006JD007813, 2007.

Li, J. Y., Mao, J. Q., Min, K. E., Washenfelder, R. A., Brown, S. S., Kaiser, J., Keutsch, F. N., Volkamer, R., Wolfe, G. M., Hanisco, T. F., Pollack, I. B., Ryerson, T. B., Graus, M., Gilman, J. B., Lerner, B. M., Warneke, C., de Gouw, J. A., Middlebrook, A. M., Liao, J., Welti, A., Henderson, B. H., McNeill, V. F., Hall, S. R., Ullmann, K., Donner, L. J., Paulot, F., and Horowitz, L. W.: Observational constraints on glyoxal production from isoprene oxidation and its contribution to organic aerosol over the Southeast United States, J. Geophys. Res. Atmos., 121, 9849-9861, doi: 10.1002/2016JD025331, 2016.

Liggio, J., Li, S. M., and McLaren, R.: Reactive uptake of glyoxal by particulate matter, J. Geophys. Res. Atmos., 110, doi: 10.1029/2004JD005113, 2005.

Liu, M., Song, Y., Yao, H., Kang, Y., Li, M., Huang, X., and Hu, M.: Estimating emissions from agricultural fires in the North China Plain based on MODIS fire radiative power, Atmos. Environ., 112, 326-334, doi: 10.1016/j.atmosenv.2015.04.058, 2015.

Mao, J. Q., Paulot, F., Jacob, D. J., Cohen, R. C., Crounse, J. D., Wennberg, P. O., Keller, C. A., Hudman, R. C., Barkley, M. P., and Horowitz, L. W.: Ozone and organic nitrates over the eastern United States: Sensitivity to isoprene chemistry, J. Geophys. Res. Atmos., 118, 11256-11268, doi: 10.1002/jgrd.50817, 2013.

Marais, E. A., Jacob, D. J., Kurosu, T. P., Chance, K., Murphy, J. G., Reeves, C., Mills, G., Casadio, S., Millet, D. B., Barkley, M. P., Paulot, F., and Mao, J.: Isoprene emissions in Africa inferred from OMI observations of formaldehyde columns, Atmos. Chem. Phys., 12, 6219-6235, doi: 10.5194/acp-12-6219-2012, 2012.

Marais, E. A., Jacob, D. J., Guenther, A., Chance, K., Kurosu, T. P., Murphy, J. G., Reeves, C. E., and Pye, H. O. T.: Improved model of isoprene emissions in Africa using Ozone Monitoring Instrument (OMI) satellite observations of formaldehyde: implications for oxidants and particulate matter, Atmos. Chem. Phys., 14, 7693-7703, doi: 10.5194/acp-14-7693-2014, 2014a.

Marais, E. A., Jacob, D. J., Wecht, K., Lerot, C., Zhang, L., Yu, K., Kurosu, T. P., Chance, K., and Sauvage, B.: Anthropogenic emissions in Nigeria and implications for atmospheric ozone pollution: A view from space, Atmos. Environ., 99, 32-40, doi: 10.1016/j.atmosenv.2014.09.055, 2014b.

Millet, D. B., Jacob, D. J., Turquety, S., Hudman, R. C., Wu, S., Fried, A., Walega, J., Heikes, B. G., Blake, D. R., Singh, H. B., Anderson, B. E., and Clarke, A. D.: Formaldehyde distribution over North America: Implications for satellite retrievals of formaldehyde columns and isoprene emission, J. Geophys. Res., 111, doi: 10.1029/2005jd006853, 2006.

Millet, D. B., Jacob, D. J., Boersma, K. F., Fu, T.-M., Kurosu, T. P., Chance, K., Heald, C. L., and Guenther, A.: Spatial distribution of isoprene emissions from North America derived from formaldehyde column measurements by the OMI satellite sensor, J. Geophys. Res., 113, doi: 10.1029/2007jd008950, 2008.

Myriokefalitakis, S., Vrekoussis, M., Tsigaridis, K., Wittrock, F., Richter, A., Bruehl, C., Volkamer, R., Burrows, J. P., and Kanakidou, M.: The influence of natural and anthropogenic secondary sources on the glyoxal global distribution, Atmos. Chem. Phys., 8, 4965-4981, doi: 10.5194/acp-8-4965-2008, 2008.

Nishino, N., Arey, J., and Atkinson, R.: Formation Yields of Glyoxal and Methylglyoxal from the Gas-Phase OH Radical-Initiated Reactions of Toluene, Xylenes, and Trimethylbenzenes as a Function of NO2 Concentration, J. Phys. Chem. A, 114, 10140, doi: 10.1021/jp105112h, 2010.

Palmer, P. I., Jacob, D. J., Fiore, A. M., Martin, R. V., Chance, K., and Kurosu, T. P.: Mapping isoprene emissions over North America using formaldehyde column observations from space, J. Geophys. Res. Atmos., 108, doi: 10.1029/2002jd002153, 2003.

Palmer, P. I., Abbot, D. S., Fu, T.-M., Jacob, D. J., Chance, K., Kurosu, T. P., Guenther, A., Wiedinmyer, C., Stanton, J. C., Pilling, M. J., Pressley, S. N., Lamb, B., and Sumner, A. L.: Quantifying the seasonal and interannual variability of North American isoprene emissions using satellite observations of the formaldehyde column, J. Geophys. Res., 111, doi: 10.1029/2005jd006689, 2006.

Paulot, F., Crounse, J. D., Kjaergaard, H. G., Kroll, J. H., Seinfeld, J. H., and Wennberg, P. O.: Isoprene photooxidation: new insights into the production of acids and organic nitrates, Atmos. Chem. Phys., 9, 1479-1501, doi: 10.5194/acp-9-1479-2009, 2009a.

Paulot, F., Crounse, J. D., Kjaergaard, H. G., Kurten, A., St Clair, J. M., Seinfeld, J. H., and Wennberg, P. O.: Unexpected Epoxide Formation in the Gas-Phase Photooxidation of Isoprene, Science, 325, 730-733, doi: 10.1126/science.1172910, 2009b.

Pye, H. O. T., and Seinfeld, J. H.: A global perspective on aerosol from low-volatility organic compounds, Atmos. Chem. Phys., 10, 4377-4401, doi:10.5194/acp-10-4377-2010.

Robinson, A. L., Donahue, N. M., Shrivastava, M. K., Weitkamp, E. A., Sage, A. M., Grieshop, A. P., Lane, T. E., Pierce, J. R., and Pandis, S. N.: Rethinking organic aerosols: Semivolatile emissions and photochemical aging, Science, 315, 1259-1262, doi: 10.1126/science.1133061 2007.

Saunders, S. M., Jenkin, M. E., Derwent, R. G., and Pilling, M. J.: Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part A): tropospheric degradation of non-aromatic volatile organic compounds, Atmos. Chem. Phys., 3, 181-193, doi: 10.5194/acp-3-161-2003, 2003.

Shim, C., Wang, Y., Choi, Y., Palmer, P. I., Abbot, D. S., and Chance, K.: Constraining global isoprene emissions with Global Ozone Monitoring Experiment (GOME) formaldehyde column measurements, J. Geophys. Res., 110, doi: 10.1029/2004jd005629, 2005.

Spurr, R.: LIDORT and VLIDORT: Linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems, in: Light Scattering Reviews, edited by: Kokhanovsky, A., Springer, 3, 229–275, 2008.

Spurr, R. J. D.: VLIDORT: A linearized pseudo-spherical vector discrete ordinate radiative transfer code for forward model and retrieval studies in multilayer multiple scattering media, J. Quant. Spectrosc. Radiat. Transf., 102, 316-342, doi: 10.1016/j.jqsrt.2006.05.005, 2006.

Stavrakou, T., Muller, J. F., Bauwens, M., De Smedt, I., Lerot, C., Van Roozendael, M., Coheur, P. F., Clerbaux, C., Boersma, K. F., van der, A. R., and Song, Y.: Substantial Underestimation of Post-Harvest Burning Emissions in the North China Plain Revealed by Multi-Species Space Observations, Sci. Rep., 6, 32307, doi: 10.1038/srep32307, 2016.

Stavrakou, T., Müller, J. F., Bauwens, M., De Smedt, I., Van Roozendael, M., De Mazière, M., Vigouroux, C., Hendrick, F., George, M., Clerbaux, C., Coheur, P. F., and Guenther, A.: How consistent are top-down hydrocarbon emissions based on formaldehyde observations from GOME-2 and OMI?, Atmos. Chem. Phys., 15, 11861-11884, doi: 10.5194/acp-15-11861-2015, 2015.

Stavrakou, T., Muller, J. F., De Smedt, I., Van Roozendael, M., van der Werf, G. R., Giglio, L., and Guenther, A.: Global emissions of non-methane hydrocarbons deduced from SCIAMACHY formaldehyde columns through 2003-2006, Atmos. Chem. Phys., 9, 3663-3679, doi:10.5194/acp-9-3663-2009, 2009b.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y., and van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), Atmos. Chem. Phys., 10, 11707-11735, doi:10.5194/acp-10-11707-2010, 2010.

van der Werf, G. R., Randerson, J. T., Giglio, L., van Leeuwen, T. T., Chen, Y., Rogers, B. M., Mu, M. Q., van Marle, M. J. E., Morton, D. C., Collatz, G. J., Yokelson, R. J., and Kasibhatla, P. S.: Global fire emissions estimates during 1997-2016, Earth Syst Sci Data, 9, 697-720, doi: 10.5194/essd-9-697-2017, 2017.

Vlemmix, T., Hendrick, F., Pinardi, G., Smedt, I., De Fayt, C., Hermans, C., Piters, A., Wang, P., and Levelt, P.: MAX-DOAS observations of aerosols, formaldehyde and nitrogen dioxide in the Beijing area: comparison of two profile retrieval, Atmos. Meas. Tech., 2, 941–963, doi:10.5194/amt-8-941-2015, 2015.

Wang, Y., Beirle, S., Lampel, J., Koukouli, M., De Smedt, I., Theys, N., Li, A., Wu, D. X., Xie, P. H., Liu, C., Van Roozendael, M., Stavrakou, T., Muller, J. F., and Wagner, T.: Validation of OMI, GOME-2A and GOME-2B tropospheric NO2, SO2 and HCHO products using MAX-DOAS observations from 2011 to 2014 in Wuxi, China: investigation of the effects of priori profiles and aerosols on the satellite products, Atmos. Chem. Phys., 17, 5007-5033, doi: 10.5194/acp-17-5007-2017, 2017.

Wang, P., Stammes, P., R., v. d. A., Pinardi, G., and Roozendael, M. V.: FRESCO+: an improved O2 A-band cloud retrieval algorithm for tropospheric trace gas retrievals, Atmos. Chem. Phys., 8, 6565-6576, doi: 10.5194/acp-8-6565-2008, 2008.

Zhang, X. Y., Wang, Y. Q., Niu, T., Zhang, X. C., Gong, S. L., Zhang, Y. M., and Sun, J. Y.: Atmospheric aerosol compositions in China: spatial/temporal variability, chemical signature, regional haze distribution and comparisons with global aerosols, Atmospheric Chemistry and Physics, 12, 779-799, doi:10.5194/acp-12-779-2012, 2012.

Zhu, L., Jacob, D. J., Mickley, L. J., Marais, E. A., Cohan, D. S., Yoshida, Y., Duncan, B. N., González Abad, G., and Chance, K. V.: Anthropogenic emissions of highly reactive volatile organic compounds in eastern Texas inferred from oversampling of satellite (OMI) measurements of HCHO columns, Environ. Res. Lett., 9, 114004, doi: 10.1088/1748-9326/9/11/114004, 2014.

Adjoint inversion of Chinese non-methane volatile organic compound emissions using space-based observations of formaldehyde and glyoxal

Hansen Cao[1], Tzung-May Fu[1,*], Lin Zhang[1], Daven K. Henze[2], Christopher Chan Miller[3], Christophe Lerot[4], Gonzalo González Abad[3], Isabelle De Smedt[4], Qiang Zhang[5], Michel van Roozendael[4], Francoois Hendrick[4], Kelly Chance[3], Jie Li[6], Junyu Zheng[7], Yuanhong Zhao[1]

[1]Department of Atmospheric and Oceanic Sciences and Laboratory for Climate and Ocean-Atmosphere Studies, School of Physics, Peking University, Beijing,100871, China
[2]Department of Mechanical Engineering, University of Colorado, Boulder, USA
[3]Atomic and Molecular Physics Division, Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts, USA
[4]Belgian Institute for Space Aeronomy ~~(BIRA-IASB),~~ Brussels, Belgium
[5]Center for Earth System Science, Tsinghua University, Beijing, China
[6]Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
[7]College of Environmental Science and Engineering, South China University of Technology, Guangzhou, China

Correspondence to: Tzung-May Fu (tmfu@pku.edu.cn)

Formatted: Tab stops: Not at 10.73 cm

Formatted: Not Superscript/

Abstract. We used the GEOS-Chem model and its adjoint to quantify Chinese non-methane volatile organic compound (NMVOC) emissions for the year 2007, using the ~~vertical~~tropospheric column concentrations of formaldehyde and glyoxal observed by the Global Ozone Monitoring Experiment-2A (GOME-2A) instrument and the Ozone Monitoring Instrument (OMI) as quantitative constraints. We conducted a series of inversion experiments using different combinations of satellite observations to explore ~~the~~ their impacts on the top-down emission estimates ~~due to different satellite retrievals.~~. Our top-down estimates for Chinese annual total NMVOC emission ~~was 23.4~~were 30.7 to ~~35.4~~ 49.5 (average ~~30.8~~41.9) Tg ~~C~~ y^{-1}, including ~~13.5~~16.4 to ~~19.7~~ 23.6 (average ~~17.0~~20.2) Tg ~~C~~ y^{-1} from anthropogenic sources, ~~8.9~~12.2 to ~~14~~22.8 (average ~~12.6~~19.2) Tg ~~C~~ y^{-1} from biogenic sources, and ~~1.1~~2.08 to ~~1.5~~3.13 (average ~~1.~~2.48) Tg ~~C~~ y^{-1} from biomass burning. In comparison, the ~~most widely used bottom-up estimate~~*a priori* estimate for Chinese annual total NMVOC emission was ~~27.4~~38.3 Tg ~~C~~ y^{-1}, including ~~15.5~~18.8 Tg ~~C~~ y^{-1} from anthropogenic sources, ~~10.8~~17.3 Tg ~~C~~ y^{-1} from

biogenic sources, and ~~1.1~~2.27 Tg ~~C~~ y^{-1} from biomass burning. The simultaneous use of glyoxal and formaldehyde observations helped distinguish the NMVOC species from different sources and was essential in constraining anthropogenic emissions. Our four ~~inversions~~inversion experiments consistently showed that the ~~emissions of~~ Chinese anthropogenic emissions of NMVOC precursors of glyoxal were larger than the *a priori* estimates. Our top-down estimates for ~~the~~ Chinese annual emission of anthropogenic aromatics (benzene, toluene, and xylene) ranged from 5.~~0~~5 to 7.~~3~~9 Tg ~~C~~ y^{-1}, 2% to ~~49~~46% larger than the estimate of the ~~bottom-up~~ *a priori* emission inventory (5.~~4~~.~~9~~ Tg ~~C~~ y^{-1}). Three out of our four inversion experiments indicated that the seasonal variation of Chinese NMVOC emissions was significantly stronger than indicated in the *a priori* inventory. Model simulations ~~using~~driven by the average of our top-down NMVOC emission estimates (which had a stronger seasonal variation~~cycle~~ than the *a priori*) showed that surface afternoon ozone concentrations over ~~northern and central~~eastern China increased ~~5-12~~by 1-8 ppb in June and decreased ~~5-13~~by 1-10 ppb in December relative to the simulations using the *a priori* emissions and were in better agreement with measurements. We concluded that the satellite observations of ~~glyoxal and~~ formaldehyde and glyoxal together provided quantitative constraints on the emissions and source types of NMVOCs over China and improved our understanding on regional chemistry.

Formatted: Font: Not Bold

1 Introduction

Non-methane volatile organic compounds (NMVOCs) are emitted into the atmosphere from surface anthropogenic, biogenic, and biomass burning sources. NMVOCs are precursors to tropospheric ozone and secondary organic aerosols, both of which are climate forcers and major air pollutants. NMVOC also affect the oxidation capacity of the atmosphere, which in turn changes the lifetimes of greenhouse gases and other pollutants ~~(Monks, 2005; Lelieveld et al., 2008).~~ It is thus crucial to quantify NMVOC emissions in order to understand their impacts on atmospheric chemistry and climate on both global and regional scales. Here we used satellite observations and a chemical transport model to constrain NMVOC emissions from China and assessed their impacts on seasonal surface ozone.

Emissions of trace species are traditionally estimated in a “bottom-up” manner using activity data and emission factors, but these bottom-up estimates are sometimes susceptible to large uncertainties. This

is especially true for NMVOC emissions in developing countries such as China, because (1) a wide range of species, source activities, and technologies are involved (Q. Zhang et al., 2009; Kurokawa et al., 2013; Li et al., 2014; Qiu et al., 2014), (2) locally-representative emission factors are often not measured (Wei et al., 2008; Zhao et al., 2011), and (3) reliable activity data are often incomplete, particularly for small-scale industries, residential activities, and agricultural waste burning (Q. Zhang et al., 2009). ~~Table 1 shows bottom~~Bottom-up estimates for Chinese total annual NMVOC emissions for the years 2005 to 2012~~, which~~ ranged from ~~21.6~~31 to ~~51.7~~57 Tg y^{-1} (Guenther et al., 2006; ~~Bolscher et al., 2007;~~ Bo et al., 2008; Q. Zhang et al., 2009; van der Werf et al., 2010, 2017; Cao et al., 2011; Huang et al., 2012; Kurokawa et al., 2013; Li et al., 2014; Stavrakou et al., 2014; Sindelarova et al., 2014; Wu et al., 2016~~). The~~; Huang et al., 2017; Granier et al., 2017). Such large uncertainties in ~~these Chinese NMVOC~~the emission estimates of Chinese NMVOCs have led to great difficulty in evaluating their impacts on regional chemistry (Han et al., 2013; Wang et al., 2014).

Formatted: Not Highlight

A complementary, "top-down" approach ~~of~~for quantifying emissions uses observations of the targeted species or its chemical derivatives, combined with a chemical transport model acting as a transfer function, to invert for the fluxes of the targeted species. In particular, tropospheric column concentrations of formaldehyde ~~(HCHO),~~ retrieved from satellite UV-backscatter ~~observations~~measurements, have been used to constrain NMVOC emissions. Formaldehyde is produced at high yields during the oxidation of many NMVOC species~~, as well as~~ (Millet et al., 2006) and also emitted directly from anthropogenic and biomass burning activities (Akagi et al., 2011; Li et al., 2017). Early ~~applications~~inversions of satellite-observed formaldehyde columns ~~mainly~~mostly focused on areas where the local NMVOC fluxes were dominated by biogenic ~~emissions~~sources during the growing season and in the absence of substantial biomass burning, such as the southeast U.S. (Palmer et al., 2003, 2006; Millet et al., 2006, 2008), Europe (Dufour et al., 2009; Curci et al., 2010), the Amazon (Barkley et al., 2008, 2009, 2013), and ~~the tropical central~~ Africa (Marais et al., 2012, 2014a). These studies showed that the observed ~~high~~local enhancements of formaldehyde column concentrations ~~of formaldehyde over densely vegetated areas were linearly proportional~~can be used to quantitatively constrain ~~the~~ local biogenic ~~isoprene flux during the growing season~~NMVOC fluxes.

~~Later~~In other areas, the NMVOC emissions from various sources may be comparable in magnitudes.

Several studies constrained the NMVOC emissions from multiple sources over such areas by analyzing the spatiotemporal variability of the observed formaldehyde columns (Shim et al., 2005; Fu et al., 2007; Stavrakou et al., 2009b; Curci et al., 2010; Gonzi et al., 2011; Marais et al., 2014b; Zhu et al., 2014). Fu et al. (2007~~), a forerunner of this study,~~) analyzed the spatial and seasonal variation of the formaldehyde column observations from the Global Ozone Monitoring Experiment (GOME) over East and South Asia. They showed that, during the early 2000s, Chinese reactive NMVOC fluxes from biogenic, anthropogenic, and biomass burning sources were 3, 1.2, and 8.8 times their respective bottom-up estimates~~.~~ at that time. In particular, Fu et al. (2007) found a large, annually-recurring NMVOC source over the North China Plain (NCP) in June, which they attributed to crop residue burning after the local harvest of winter wheat. However, these top-down studies using only formaldehyde as constraints relied exclusively on bottom-up activity statistics to differentiate between NMVOC source types.

More recently, satellite measurements of tropospheric glyoxal columns emerged as an additional constraint on NMVOC emissions (Stavrakou et al., 2009a). Like formaldehyde, glyoxal is produced during the oxidation of many NMVOCs (including most importantly isoprene), as well as emitted directly from biomass burning (Fu et al., 2008; Myriokefalitakis et al., 2008). In ~~addition~~particular, glyoxal is produced at high yields at the initial ring-cleaving stage during the oxidation of aromatics (Volkamer, 2001; Nishino et al., 2010), which are mainly anthropogenic. In contrast, the production of formaldehyde ~~production~~ from the oxidation of aromatics is further downstream and thus spatially ~~diffused~~diffuse (Volkamer, 2001). As such, simultaneous analyses of formaldehyde and glyoxal observations can help differentiate between biogenic and anthropogenic NMVOC emissions. Stavrakou et al. (2009a) pioneered a two-compound inversion using tropospheric glyoxal and formaldehyde column observations from the ~~SCIAMACHY~~Scanning Imaging Absorption spectrometer for Atmospheric CHartographY (SCIAMACHY) satellite instrument to constrain the global sources of glyoxal. They estimated that the anthropogenic NMVOC fluxes over East Asia for the year 2005 were a factor of 2~~-~~ to 3 larger than the bottom-up estimates of the Emission Database for Global Atmospheric Research (EDGAR, v3.3) inventory (Olivier et al., 2001, 2002) and the REanalysis TROpospheric (RETRO) emission inventory (Schultz et al., 2007). In addition, they inferred a large missing source of glyoxal over the global continents, which they attributed to production from an unknown biogenic

precursor.

Over eastern China, Liu et al. (2012) showed that the glyoxal column concentrations observed by SCIAMACHY in August 2007 was more than twice the simulated glyoxal columns using the bottom-up emission inventory developed by Q. Zhang et al. (2009). Over the Pearl River Delta area (PRD) in southern China, the discrepancy was at least a factor of three. They suggested that the missing glyoxal source over eastern China was anthropogenic, on the basis that the ~~anonymous~~anomalous glyoxal columns observed by SCIAMACHY (relative to the glyoxal columns simulated by their model) were spatially correlated with anthropogenic NO_x emissions. ~~Their~~ They estimated the Chinese anthropogenic aromatics emission ~~was~~to be 13.4 Tg y^{-1}, which was six times the 2.4 Tg y^{-1} anthropogenic aromatic flux estimated by Q. Zhang et al. (2009). In contrast, Chan Miller et al. (2016) simulated the formaldehyde and glyoxal column concentrations over the Pearl River Delta area (PRD) in southern China for the years 2006 and 2007 using the same inventory developed by Q. Zhang et al. (2009). They found that their simulated formaldehyde columns were consistent with the OMI formaldehyde observations, while their simulated glyoxal columns were lower than OMI observations by only 40%. They attributed the high anthropogenic aromatics emission estimate by Liu et al. (2012) in part to a regional high-bias in the SCIAMACHY data, ~~as well as underestimated yields of~~and in part to the lower glyoxal yields from ~~the~~ aromatics oxidation ~~of aromatics.~~used in Liu et al. (2012).

One limitation in the use of satellite observations of formaldehyde and glyoxal for constraining NMVOC sources is their inherent uncertainty. Several studies have compared GOME-2A and OMI formaldehyde column observations against aircraft or ground-based measurements at a few locations around the world (De Smedt et al., 2015; Lee et al., 2015; Wang et al., 2017; Zhu et al., 2016). Zhu et al. (2016) compared the GOME-2A-observed formaldehyde column concentrations over the Southeast U.S. in summer 2013 against aircraft measurements and found the satellite measurements to be too low by a factor of approximately 1.7. Chan Miller et al. (2017) found that glyoxal column concentrations observed by OMI were lower than the aircraft measurements over the Southeast U.S. in summer 2013 by a factor of 1.5. Wang et al. (2017) compared the bi-monthly mean GOME-2A and OMI formaldehyde column concentrations retrieved by De Smedt et al. (2012, 2015) against

ground-based multi-axis differential optical absorption spectroscopy (MAX-DOAS) measurements at a rural site in eastern China. They found that both satellite retrievals were systematically lower than the ground-based measurements by approximately 20%. These studies ~~highlight~~highlighted the potential impacts on top-down NMVOC emission estimates due to uncertainty associated with satellite retrievals.

In this study, we used satellite retrievals of both formaldehyde and glyoxal, along with ~~an updated~~a chemical transport model and its adjoint, to constrain NMVOC emissions from China for the year 2007. We conducted sensitivity experiments to evaluate the impacts on the top-down estimates due to different satellite ~~retrieval constraints~~observations, with the goal of ~~obtaining~~bracketing a ~~most~~ probable range of top-down estimates. Finally, we examined the impacts of our top-down NMVOC emission estimates on surface ~~ozone concentrations~~air quality over China.

2 Model and data

2.1 The GEOS-Chem model and its adjoint

We ~~used~~updated the GEOS-Chem global 3D chemical transport model (version 8.2.1) to simulate the emission, transport, chemistry, and deposition of NMVOCs, as well as the resulting formaldehyde and glyoxal column concentrations for the year 2007. The use of an older version of the GEOS-Chem forward model was necessary because, at the time of our study, the GEOS-Chem adjoint (version 34) was based on this older version. However, we updated the NMVOC chemical schemes (described below) and corrected several model errors in both our forward model and its adjoint by following the progress of the forward model up to version 10.1. GEOS-Chem was driven by the assimilated meteorological data from the NASA Goddard Earth Observing System (GEOS-5) (Bey et al., 2001). To drive our simulations, the horizontal resolution of GEOS-5 data was downgraded from its native 2/3° longitude × 1/2° latitude to 5° longitude × 4° latitude. The number of vertical levels was reduced from 72 to 47 by merging layers in the stratosphere. The lower 2 km of the atmosphere was resolved by 14 levels. The temporal resolution of GEOS-5 data into GEOS-Chem is 3 h for atmospheric variables and 1 h for surface variables.

We updated the dicarbonyl chemical mechanism in GEOS-Chem developed by Fu et al. (2008), which in turn was originally adapted from the Master Chemical Mechanism (MCM) version 3.1 (~~Saunders et al., 2003; Bloss et al., 2005).~~Jenkin et al., 1997; Saunders et al., 2003). Table S1 lists the yields of formaldehyde and glyoxal from the OH-oxidation of NMVOC precursors in our updated chemical mechanism. The lumped NMVOC precursors of formaldehyde in our mechanism included ethane, propane, $\geqslant C_4$ alkanes, ethene, $\geqslant C_3$ alkenes, benzene, toluene, xylenes, isoprene, ~~and~~ monoterpenes~~.~~, acetone, hydroxyacetone, methygloxal, glycolaldehyde, acetaldehyde, 2-methyl-3-bute-nol, methy ethyl ketone, methanol, and ethanol (lumped into $\geqslant C_4$ alkanes). The lumped NMVOC precursors of glyoxal in our mechanism included ~~propane, alkanes,~~ ethene~~, $\geqslant C_3$ alkenes~~, ethyne, benzene, toluene, xylenes, isoprene, monoterpenes, glycolaldehyde, and 2-methyl-3-bute-2-nol~~.~~ (MBO). Hereinafter we focused our discussion on these NMVOC precursors only, as their emissions may be constrained by formaldehyde and glyoxal observations.

The OH-oxidation of isoprene is a major source of both formaldehyde and glyoxal over China (Fu et al., 2007, 2008; Myriokefalitakis et al., 2008). We replaced the isoprene photochemical scheme with that used in GEOS-Chem v10.1 ~~(~~, which included updates from Paulot et al.~~,~~. (2009a,b~~;~~) and Mao et al. (2013~~), where~~). In this updated scheme, oxidation of isoprene by OH under high-NO_x conditions produces formaldehyde and glyoxal ~~were produced from isoprene oxidation~~ at yields of 0.436 molecules per C and 0.0255 molecules per C, respectively (Table S1), mainly via the RO_2+NO ~~pathway under high-NO_x conditions and~~pathways. Under low-NO_x conditions, oxidation of isoprene by OH produces formaldehyde and glyoxal at yields of 0.38 molecules per C and 0.073 molecules per C, respectively (Table S1), via both RO_2+HO_2 and RO_2-isomerization ~~under low-NO_x conditions.~~reactions. Li et al. ~~(2016)~~(2016) implemented this same isoprene photochemical scheme in a box model and compared the productions of formaldehyde and glyoxal from isoprene oxidation ~~in this updated scheme~~ with those in the MCM version 3.3.1 (Jenkin et al., 2015). They showed that the production pathways and yields of formaldehyde and glyoxal were similar in the two schemes under the high-NO_x conditions typical of eastern China.

We updated the molar yields of glyoxal from the OH oxidations of benzene (33.3%), toluene (26.2%), and xylenes (21.0%) following the latest literature (Arey et al., 2009; Nishino et al., 2010). These new

Formatted: Subscript

molar yields were higher than those used in Fu et al. (2008) (based on averaged yields in the literature: 25.2% for benzene, 16.2% for toluene, and 15.6% for xylenes) but still lower than those used by Chan Miller et al. (2016) (75% for benzene, 70% for toluene, and 36% for xylenes), which were taken from the aromatic chemical scheme in MCM version 3.2 (Jenkin et al., 2003; Bloss et al., 2005). In MCM version 3.2, more than half of the glyoxal from aromatics oxidation were produced during second- and later-generation photochemistry, but such productions are ~~still uncertain,~~ with limited experimental support and uncertain (Bloss et al., 2005).

Formaldehyde and glyoxal in ~~our~~the GEOS-Chem model were both removed by photolysis, as well as dry and wet deposition (Fu et al., 2008). We updated the Henry's law constant for glyoxal from $3.6 \times 10^5 \times \exp[7.2 \times 10^3 \times (1/T\text{-}1/298)]$ (Fu et al., 2008) to $4.19 \times 10^5 \times \exp[(62.2 \times 10^3/R) \times (1/T\text{-}1/298)]$ (Ip et al., 2009) and added the dry deposition of formaldehyde, glyoxal, methyglyoxal and glycolaldehyde on leaves (Mao et al., 2013). In addition, we assumed that glyoxal was reactively uptaken by wet aerosols and cloud droplets with an uptake coefficient ~~γ=~~γ = 2.9×10^{-3} (Liggio et al., 2005; Fu et al., 2008). All other physical and chemical processes in our forward model were as described in Fu et al. (2008).

For the forward model described above, we developed the adjoint by modifying the standard GEOS-Chem adjoint (version 34) (Henze et al., 2007). We used the Kinetic PreProcessor (KPP) (Daescu et al., 2003; Sandu et al., 2003) to construct the adjoint of the updated photochemical mechanism. Adjoint algorithms were updated to include the emission and deposition processes of formaldehyde and glyoxal precursors. The aqueous uptake rate of glyoxal by wet aerosols was a function of the ambient glyoxal concentration and the total wet aerosol surface area (Fu et al., 2008). We linearized this uptake process in the backward integrations by ~~archiving~~using the archived wet aerosol surface areas ~~in~~from the forward simulations ~~for use in the backward integrations~~.

We verified the adjoint model mathematically in two ways. Firstly, we used the adjoint model to calculate the sensitivities of global glyoxal and formaldehyde burdens to biogenic isoprene and anthropogenic xylenes emissions, respectively, and found that the results reproduced the calculated sensitivities from the forward model (Figure S1 ~~in Supplementary Information).~~). Secondly, we used ~~an~~

a ~~*priori*~~set of bottom-up NMVOC emission ~~inventory~~inventories (Section 2.2) to drive the forward model and took the resulting global tropospheric formaldehyde and glyoxal column concentrations as pseudo observations. We then used the pseudo observations of formaldehyde and glyoxal~~, respectively,~~ to successfully optimize back to ~~close to~~ the ~~*a priori*~~bottom-up NMVOC emission estimates over high-emission areas from an initial ~~emission~~ guess that was five times larger (Figure S2~~ in Supplementary Information~~). These experiments demonstrated the usefulness of the adjoint model for the inversion of NMVOCs emissions.

Formatted: Font: Not Italic

2.2 *A priori* emission estimates of Chinese NMVOCs

As a starting point for our inversion, we ~~used~~compiled the ~~most widely used~~*a priori* Chinese NMVOC emission estimates ~~for China as the *a priori*.~~from recent bottom-up emission inventories. Table ~~2~~1 summarizes the annual total of these *a priori* emission estimates and their associated uncertainties.

The *a priori* biogenic NMVOC emissions from China and from the rest of the world were calculated with the MEGAN v2.0 algorithm (Guenther et al., 2006) and dependent on temperature, shortwave radiation, and monthly mean leaf area index. ~~The~~Previous top-down studies suggested that MEGAN overestimates global biogenic methanol by a factor of two to three (Stavrakou et al., 2011; Wells et al., 2012). We scaled our global biogenic methanol emissions to the value (100 Tg y^{-1}) reported by Stavrakou et al. (2011) to be the *a priori* in this study. The contributions of Chinese biogenic ethanol to formaldehyde are expected to be low due to its small emissions (Guenther et al., 2012); thus the Chinese biogenic ethanol emissions were neglected in this study. The resulting annual total biogenic NMVOC emissions over China for the year 2007 was ~~10.8~~17.3 Tg ~~C~~ y^{-1}, including ~~6.6~~7.5 Tg ~~C~~ y^{-1} of isoprene~~.~~, 4.6 Tg y^{-1} of methanol, and 5.2 Tg y^{-1} of other species (including monoterpenes, ethene, acetone, $\geqslant C_3$ alkenes, and MBO). Previous estimates of Chinese biogenic ~~NMVOC~~isoprene emissions ranged from 5.~~0~~8 to ~~11.0Tg C~~9.9 Tg y^{-1} (Guenther et al., 2006; ~~Fu~~Sindelarova et al., ~~2007~~2014; Stavrakou et al., 2014, 2015, 2017). Based on this range, we estimated the uncertainty of the *a priori* biogenic emissions over China to be ±55%.

The *a priori* emissions for Chinese anthropogenic NMVOCs were from the Multi-resolution Emission

Inventory for China ~~(MEIC)~~ inventory (MEIC, http://meicmodel.org) (Li et al., 2014~~),~~ , 2017), which was developed ~~for the year 2010~~ at 0.25° × 0.25° resolution~~.~~ for the year 2010. The MEIC inventory, including emissions from industry, transportation, power generation and residential activities, was compiled using monthly Chinese provincial activity data and a combination of Chinese and western emission factors. The estimated ~~annual~~ Chinese annual anthropogenic ~~NMVOC~~ emission of NMVOCs was ~~15~~18.8 Tg ~~C~~ y^{-1}, including ~~64~~63% from industries, ~~24~~26% from residential activities, 10% from transportation, and 1% from power generation. The estimated annual Chinese anthropogenic emission of aromatics was 5.~~4.9~~ Tg ~~C~~ y^{-1}, including 73% from industries, 15% from residential activities, 9% from transportation, and 3% from power generation. Previous estimates of Chinese anthropogenic NMVOC emissions for the years 2005 to 2012 ranged from ~~10~~12.7 to ~~29.8~~35.5 Tg ~~C~~ y^{-1}, with aromatics emissions ranging from 2.~~1~~4 to ~~11.3~~13.4 Tg ~~C~~ y^{-1} (Bo et al., 2008; Q. Zhang et al., 2009; Cao et al., 2011; Liu et al., 2012; Kurokawa et al., 2013; Li et al., 2014, 2017; Stavrakou et al., 2015; Wu et al., 2016; Huang et al., 2017; Granier et al., 2017). We therefore estimated the uncertainty for the *a priori* Chinese anthropogenic NMVOC emission estimates to be ~~±200%. Anthropogenic NMVOC emissions for the rest of the Asia were from the inventory compiled by Li et al. (2017) for the year 2010. Anthropogenic NMVOC~~a factor of two. As such, we did not scale the MEIC Chinese NMVOC emissions to the year 2007, because the uncertainty in the emission estimates were much larger than the differences in emissions between the years 2007 and 2010 (Chinese anthropogenic NMVOC emissions increased 14% from 2006 to 2010 according to Li et al, 2017). The spatial distribution of Chinese anthropogenic NO_x emissions were from the MEIC inventory for the year 2010 (Li et al., 2017) but scaled to the year 2007 levels using top-down constraints from the GOME-2A NO_2 observations (Mijling et al., 2013). Anthropogenic NMVOC emissions for the rest of the Asia were from Li et al. (2017) for the year 2010. Anthropogenic emissions for Europe, U.S., and the rest of the world were from the European Monitoring and Evaluation Programme ~~(EMEP)~~ inventory (Vestreng, 2003), the U.S. EPA 2005 National Emission Inventory (~~NEI05) (Brioude et al., 2011; Kim et al., 2011~~https://www.epa.gov/air-emissions-inventories/national-emissions-inventory-nei), and the ~~Emission Database for Global Atmospheric Research (~~EDGAR~~)~~ inventory (version 2.0) (Olivier et al., 1999), respectively, and scaled to the year 2007 using CO_2 emissions (van Donkelaar et al., 2008).

Post-harvest, in-field burning of crop residue has been recognized as a large seasonal source of

NMVOCs in China (Fu et al., 2007; Huang et al., 2012; Liu et al., 2015; Stavrakou et al., 2016~~), but this emission has~~). These emissions from crop residue fires have been severely underestimated in inventories based on ~~satellite burnt~~burned area observations ~~(Liu et al., 2015).~~from satellites, such as the Global Fire Emissions Database version 3 (GFED3, van der Werf et al., 2010). The recent Global Fire Emissions Database version 4 (GFED4, van der Werf et al., 2017) included small fires by scaling burned area with satellite fire pixel observations, but the resulting Chinese NMVOC emission estimate from biomass burning (0.91 Tg y^{-1}) was still much lower than the bottom-up inventory by Huang et al. (2012). Huang et al. (2012) estimated the Chinese CO emission from crop residue burning to be 4.0 Tg y^{-1}, based on MODIS daily thermal anomalies, Chinese provincial ~~burnt~~ burned biomass data, and emission factors from Akagi et al. (2011). We scaled this CO flux using speciated NMVOC emission factors from crop residue burning from the literature (Hays et al., 2002; Akagi et al., 2011) and then multiplied the resulting NMVOC flux estimate by two. The reason for doubling the scaled NMVOC flux was that the emission factors for many NMVOC species were not measured, such that the sum of the speciated NMVOC emission factors was only half of the ~~measured~~ total NMVOC emission factor (Akagi et al., 2011). This difference may partially explain why the ~~top-down~~formaldehyde inversion study by Stavrakou et al. (2016) ~~using satellite observations of formaldehyde~~ found that Huang et al. (2012) underestimated the NMVOC ~~flux~~fluxes from crop fires over the North China Plain (NCP) in June by at least a factor of two.

Our resulting *a priori* estimate for Chinese annual NMVOC emissions from biomass burning was ~~1.1~~2.27 Tg ~~C~~ y^{-1}, including ~~0.86~~1.80 Tg ~~C~~ y^{-1} from crop residue burning (obtained by scaling Huang et al., 2012 as described above) and 0.~~24~~47 Tg ~~C~~ y^{-1} from other types of biomass burning activities ~~(taken from the Global Fire Emissions Database version 3,~~ GFED3~~)~~ (van der Werf et al., 2010). Previous estimates of Chinese NMVOC emissions from biomass burning for the years ~~1996~~2000 to 2012 ranged widely from 0.~~24~~47 to ~~3.2~~5.1 Tg ~~C~~ y^{-1} (Fu et al., 2007; van der Werf et al., 2010, 2017; Wiedinmyer et al., 2011; Huang et al., 2012; Liu et al., 2015; Stavrakou et al., 2015, 2016). We therefore ~~assigned an~~ estimated the uncertainty of ~~±300% to~~ the *a priori* Chinese biomass burning NMVOC flux~~.~~ to be a factor of three. Biomass burning emissions from the rest of the world were ~~taken~~ from GFED3 (van der Werf et al., 2010).

Figure 1 (a)-(~~c~~d) show the spatial distribution of the *a priori* Chinese NMVOC emissions from biomass burning, anthropogenic, biogenic, and total sources, respectively. Biomass burning emissions were highest over the NCP and southwest China, reflecting the strong emissions from crop residue burning over the NCP in June and ~~from land-clearing burning~~ over southwest China during February to April, respectively. Chinese anthropogenic and biogenic NMVOC sources were both ~~showed a general~~stronger in the east than in the west ~~to east gradient, following~~, reflecting the co-location of dense population and vegetation ~~densities. Biogenic~~in the east. The highest biogenic NMVOC emissions ~~reflected~~were over southern China, due to the combined modulation by vegetation densities, temperature, and sunlight. Anthropogenic NMVOC fluxes exceeded 10^3 kg ~~C~~ km^{-2} y^{-1} throughout the industrialized and densely populated eastern China, with the highest fluxes over the NCP and around the Yantze River Delta area.

Figure 2 shows the seasonal variation of the *a priori* Chinese NMVOC emissions. The *a priori* anthropogenic NMVOC fluxes were ~~higher~~larger during the cold months and lower during the warm months, driven by the seasonal strengths of industrial and residential activities (Li et al., 2017). The *a priori* biogenic NMVOC fluxes showed the opposite seasonal pattern, with ~~more than half~~65% of the total annual flux emitted in summer~~.~~ (June to August). The *a priori* biomass burning NMVOC source was relatively small, except when it peaked ~~in June~~ due to ~~the burst of~~ post-harvest burning over the NCP in June and ~~in spring due to land-clearing~~ over southwest China in spring. As a result, the *a priori* Chinese NMVOC emissions were predominantly anthropogenic in ~~January~~winter but mainly biogenic in ~~June.~~summer. During the transition ~~months~~seasons of ~~April~~spring and ~~October~~fall, the anthropogenic ~~and~~ , biogenic, and biomass burning contributions ~~to the total NMVOC emissions~~ were comparable.

2.3 Formaldehyde and glyoxal column concentrations observed by GOME-2A and OMI

We used the monthly mean tropospheric formaldehyde and glyoxal column concentrations retrieved from the Global Ozone Monitoring Experiment-2A (GOME-2A) instrument and the Ozone Monitoring Instrument (OMI) for the year 2007 to constrain Chinese NMVOC sources. The ~~technical details of these~~ four sets of satellite retrievals used in this study are briefly described below; further technical details are summarized in Table ~~3.~~ S2.

The native GOME-2A pixel vertical column densities (VCDs) of formaldehyde and glyoxal were retrieved by De Smedt et al. (2012) and Lerot et al. (2010), respectively~~, using protocols briefly described below. First, pixel~~. Pixel slant column densities (SCDs) of formaldehyde and glyoxal were retrieved in the 328.5-346 nm and 435-460 nm windows, respectively, using the Differential Optical Absorption Spectroscopy (DOAS) technique (Platt et al., 1979). Previous glyoxal SCD retrievals often showed biases over remote tropical oceans due to absorption from liquid water (~~Vrekoussis et al., 2010;~~ Wittrock et al., 2006; ~~Lerot~~Vrekoussis et al., 2010). This bias was corrected in Lerot et al. (2010) by explicitly accounting for liquid water absorption during the DOAS fitting. ~~Second, pixel~~Pixel SCDs were then converted into VCDs using air mass factors (AMF), which was calculated using Linearized Discrete Ordinate Radiative Transfer model (LIDORT) (Spurr, 2008) and trace gas profiles simulated by the ~~IMAGE~~IMAGES v2 model (Stavrakou et al., 2009b). The native pixel VCDs were gridded to daily means at 0.25° × 0.25° resolution (De Smedt et al., 2012; Lerot et al., 2010). We further averaged the daily means to monthly means at 5° longitude × 4° latitude resolution. The retrieval errors of the spatially-and-temporally averaged VCDs were estimated to be 30%-40% for formaldehyde and 40% for glyoxal, due to a combination of errors associated with the SCD retrievals, the reference sector correction, the *a priori* profile, and the AMFs (De Smedt et al., 2012; Lerot et al., 2010).

The OMI native pixel VCDs of formaldehyde and glyoxal were retrieved by González Abad et al. (2015) and Chan Miller et al. (2014), respectively. ~~Briefly, formaldehyde and glyoxal pixel~~Pixel SCDs were retrieved by directly fitting the absorption spectra in the 328.5 – 356.5 nm (formaldehyde) and 435 – 461 nm (glyoxal) windows, respectively (Chance, 1998; ~~Kurosu et al., 2004, 2007;~~ Chan Miller et al., 2014~~), and~~). Pixel SCDs were then converted to~~ pixel~~ VCDs using AMF calculated with a linearized vector discrete ordinate radiative transfer model, VLIDORT (Spurr, 2006), and trace gas profiles simulated by the GEOS-Chem model (González Abad et al., 2015). Liquid water absorption was also explicitly calculated for the glyoxal retrieval (Chan Miller et al., 2014). The typical uncertainties of OMI-observed pixel VCDs over polluted areas were estimated to be 30% to 45% for formaldehyde and ~~104~~100% for glyoxal (González Abad et al., 2015; Chan Miller et al., 2014). The native pixel VCDs were averaged to monthly means at 5° longitude × 4° latitude resolution. For

glyoxal, we further removed VCDs with signal-to-uncertainty ratios less than 100%. We assumed the retrieval uncertainty of monthly mean OMI formaldehyde and glyoxal VCDs at $4^o \times 5^o$ resolution to be 40% and 100%, respectively.

To remove ~~globally~~ global systematic biases in the satellite observations, we ~~adjusted~~aligned the ~~global~~ observed monthly mean ~~VCDs by aligning the observed~~ VCDs over remote reference areas to those simulated by the GEOS-Chem model (sampled at satellite overpass time) using the *a priori* NMVOC emissions. The remote Pacific (140^o-$160W^o$, 90^oS-90^oN) was chosen as the reference area for formaldehyde (Palmer et al., 2003, 2006; Fu et al., 2007; González Abad et al., 2015). The Sahara desert (20-30^oN, 10^oW-30^oE), where the interference from liquid water absorption was minimal, was chosen as the reference area for glyoxal (Chan Miller et al., 2014). The justification for performing the alignment was two-fold: firstly, the formaldehyde and glyoxal ~~column concentrations~~VCDs over these remote reference areas were small and well simulated by the model (Fu et al., ~~2008~~2007; Chan Miller et al., 2014). The removed biases over the remote areas were less than 20% and 10% of the typical formaldehyde (>~~8~~4 × 10^{15} molecule cm^{-2}) and glyoxal (>~~4~~2 × 10^{14} molecule cm^{-2}) monthly mean VCDs observed over eastern China, respectively. More importantly, our inversion was performed over China only, assuming that the *a priori* NMVOC emissions for the rest of the world were unbiased. As will be seen in Sections 3 and 4, the optimization of NMVOC sources were predominantly driven by local formaldehyde and glyoxal enhancements produced by relatively short-lived NMVOCs.

2.4 Inversion experiments using the GEOS-Chem adjoint

We used the GEOS-Chem model to perform Bayesian inversions on Chinese NMVOC emissions, using satellite observations of formaldehyde and glyoxal over China and the *a priori* emission estimates as constraints. The inversion minimized ~~the~~a cost function, $J(\boldsymbol{x})$ ~~in Eq. (1)~~, over China (Rodgers, 2000~~, which we calculated over China:~~):

~~$J(\boldsymbol{x}) = \gamma(\boldsymbol{x} - \boldsymbol{x}_a)\boldsymbol{S}_a^{-1}(\boldsymbol{x} - \boldsymbol{x}_a) + (\boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{y})\boldsymbol{S}_o^{-1}(\boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{y})$~~

$$J(\boldsymbol{x}) = \gamma \bullet (\boldsymbol{x} - \boldsymbol{x}_a)\boldsymbol{S}_a^{-1}(\boldsymbol{x} - \boldsymbol{x}_a) + (\boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{y})\boldsymbol{S}_o^{-1}(\boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{y}) \quad \text{Eq. (1)}$$

Field Code Changed

The first and second terms on the right-hand-side of Eq. (1) represented the penalty error and the

prediction error, respectively. $\boldsymbol{x}$, which we sought to optimize, was the vector of scale factors (for each NMVOC species from each emission sector and for each grid) applied to the *a priori* emissions. $\boldsymbol{x_a}$ was a unit vector applied to the *a priori* NMVOC emission estimates. $\boldsymbol{y}$ was the vector of satellite-observed monthly mean VCDs of the targeted tracer (formaldehyde and/or glyoxal). $\boldsymbol{F}(\boldsymbol{x})$ was the vector of VCDs of the targeted tracer simulated by the forward model $\boldsymbol{F}$. $\boldsymbol{S_a}$ was the *a priori* emission error covariance matrix, which ~~was a diagonal matrix with the uncertainties estimated based on ranges of previous NMVOC estimates (Section 2.2 and Table 1).~~we assumed to be diagonal

The observation error covariance matrix in Eq. (1), $\boldsymbol{S_o}$, was difficult to quantify, as it included contributions not only from the satellite retrieval, but also from the model representation of chemistry and transport. Zhu et al. (2016) and Chan Miller et al. (2017) compared vertical profiles of GEOS-Chem-simulated formaldehyde and glyoxal over the Southeast U.S. in summer against aircraft measurements. They reported that the simulated formaldehyde mixing ratios showed only a small bias (-3%±2%) in the lower troposphere but were lower than the observations by 41% in the free troposphere, likely due to insufficient deep convection in the model (Zhu et al., 2016). The simulated glyoxal mixing ratios were within 20% of the observations in the mixed layer, but they were too low in the upper troposphere by more than a factor of two, also likely due to insufficient model vertical transport (Chan Miller et al., 2017). It should be noted that these errors assessed by Zhu et al. (2016) and Chan Miller et al. (2017) likely also included the errors associated with precursor emissions. Nevertheless, based on these assessments, we ~~roughly~~ estimated that the model errors for formaldehyde and glyoxal VCDs to be ±80%, ±100%, respectively. Adding these estimated model errors in quadrature to the satellite retrieval errors (Section 2.3), we estimated that the observation error ($\boldsymbol{S_o}$) of formaldehyde and glyoxal to be about ±90% and ±150%, respectively.

The optimization of Eq. (1) was dependent on the relative weighting of the penalty error ($\boldsymbol{S_a}$) and the prediction error ($\boldsymbol{S_o}$~~). However, the errors and error correlations within $\boldsymbol{S_a}$ and $\boldsymbol{S_o}$~~), which were often incompletely represented. In addition, we found that due to the mathematical formulation of Eq. (1), the cost function $J(\boldsymbol{x})$ was heavily weighted by grids where the *a priori* estimates were too high, such that the optimization was less effective at increasing emissions where the *a priori* emissions were too low. These issues were empirically addressed in inversion studies by the introduction of a

regularization factor, γ, ~~in Eq. (1)~~ to adjust the relative weight of the penalty error. Henze et al. (2009) used the L-curve method (Hansen, 1998) to find an ~~optimized~~optimal γ value, which minimized the total cost function ~~and balanced~~while balancing the prediction term and the penalty term. We followed that methodology and found a γ value of 0.01 for July, which we applied ~~for~~ to all ~~of the~~ warmer months (March to October). An optimized γ value of 0.1 was found for January~~,~~ and ~~we~~ applied ~~that value~~ to ~~the~~ colder months.

Table ~~2 shows the setup of~~1 lists our inversion experiments. Figure S3 illustrates our protocol for the inversion experiments. We experimented with four different sets of satellite retrievals as constraints, with the goal of bracketing the ~~uncertainties~~possible range of ~~the~~ top-down ~~estimate of~~estimates for Chinese NMVOC emissions. The first two experiments (IE-1 and IE-2) constrained emissions using the formaldehyde and glyoxal VCDs observations from GOME-2A and OMI, respectively. Several studies showed that GOME-2A formaldehyde VCDs may be low by a factor of 1.3 to 1.7 (Lee et al., 2015; Zhu et al., 2016; Wang et al., 2017). As an "upper bound" constraint, we conducted a third inversion experiment (IE-3~~), which was~~) constrained by 1.7 times the GOME-2A formaldehyde ~~alone.~~VCDs. We conducted a fourth inversion experiment (IE-4) constrained by OMI glyoxal VCDs alone to explore the impacts of glyoxal observations on the inversions ~~of anthropogenic emissions~~.

~~Figure 3 illustrates our protocol for the inversion experiments. For each month, we began by driving the GEOS-Chem forward model with the *a priori* emissions ($\boldsymbol{x}_{i=1} = \boldsymbol{x}_a = 1$) to simulate the monthly mean formaldehyde and glyoxal VCDs at satellite crossing time. The simulated and satellite observed VCDs were used to calculate the cost function, $J(\boldsymbol{x})$, and the forcing arrays ($\frac{\partial J(\boldsymbol{x})}{\partial \boldsymbol{F}(\boldsymbol{x})}$). The adjoint of GEOS-Chem was then used to compute the cost function gradient ($\frac{\partial J(\boldsymbol{x})}{\partial \boldsymbol{x}}$), and the next guess of the emission scale factor ($\boldsymbol{x}_{i+1}$) was calculated using the Quasi-Newton L-BFGS-B algorithm (Byrd et al., 1995; Zhu et al., 1997), subject to the bounds $0.32 \leq \boldsymbol{x} \leq 10$. These bounds were selected based on the largest uncertainties quoted in the literature on Chinese NMVOC emission estimates (Q. Zhang et al., 2009; Liu et al., 2012). The process was then iterated until the incremental relative reduction of the cost function ($\frac{|J(\boldsymbol{x})_{i+1} - J(\boldsymbol{x})_i|}{\max(J(\boldsymbol{x})_{i+1}, J(\boldsymbol{x})_i)}$) was less than 1% after at least six iterations. We took $\boldsymbol{x}_{i+1}$ from the last iteration as the optimized emission scale factor ($\boldsymbol{x}_p$) and applied it to calculate the top-down emission~~

~~estimate.~~

3 Comparison of simulations using the *a priori* emissions against satellite observations and ground-based measurements

We first ~~qualitatively~~ compared the formaldehyde and glyoxal VCDs simulated by the ~~GEOS-Chem~~ model (sampled at satellite overpass times) using the *a priori* emissions against those observed by GOME-2A and OMI, as well as against measurements at ground-based ~~measurements. Figure 4 (a)-(d)~~sites (Table S3). Figures 3 and 4 show the monthly mean formaldehyde VCDs observed by GOME-2A over China ~~for January, April, June, and October~~in 2007. Observed formaldehyde VCDs over China showed a distinct west-to-east gradient year-round, ~~which was~~ driven by the higher vegetation and population densities in eastern China. Observed formaldehyde VCDs were higher ~~during the warmer months, reflecting the~~ in summer than in winter, due to a combination of stronger biogenic emissions ~~during the growing seasons.~~and photochemistry during the warmer months. In spring, GOME-2A formaldehyde VCDs were high over Southwest China and Southeast Asia, reflecting the seasonal biomass burning emissions there. Highest formaldehyde VCDs were observed over the NCP in June, ~~in response to~~likely because of the large emissions from in-field crop residue burning. ~~In April, high concentrations of formaldehyde were also observed near the southwestern border, reflecting the seasonal biomass burning there~~.

~~Figure~~ Figures 3 and 4 ~~(e)-(h) show~~compare the simulated monthly mean formaldehyde VCDs using the *a priori* emission ~~estimates.~~ against the GOME-2A formaldehyde VCDs. Table S4 summarizes the statistics of the comparison over eastern China. The ~~model~~*a priori* simulated formaldehyde VCDs generally reproduced the observed seasonal contrast and ~~regional~~spatial patterns~~.~~ over eastern China, with correlation coefficients (R) between 0.74 and 0.94 year-round, except in December (R = 0.51). The *a priori* simulated formaldehyde ~~columns~~VCDs were significantly higher than the GOME-2A observations ~~in~~ over eastern China between late fall and winter (November, December, January, and February), with normalized mean biases (NMB) of 13% to 67%, implying an overestimate of the anthropogenic formaldehyde precursors in the *a priori* ~~in January.~~ emission estimates. The *a priori* simulated formaldehyde ~~columns~~VCDs were lower than the GOME-2A observations over eastern

China ~~in June,~~during May to July (NMB between -11% to -6.4%), implying an underestimation of the emissions of formaldehyde precursors in ~~June in~~ the *a priori*~~.~~ during May to July.

A few ground-based measurements of tropospheric formaldehyde VCDs have been made in China using the Multi-Axis Differential Optical Absorption Spectrometry (MAX-DOAS) technique (Li et al., 2013; ~~Stavrakou~~Vlemmix et al., 2015; Wang et al., 2017~~) (Table S1). Figure 4 also shows the seasonal mean of);~~ these ~~ground-based~~ measurements (sampled at GOME-2A overpass time) are shown in Figure 3, Figure 4, and Table S3. In principle, these ground-based measurements ~~were~~are not directly comparable to the satellite-observed and model-simulated formaldehyde ~~columns~~VCDs, due to the ~~different inherent uncertainties and the~~ coarse spatial resolution of our analyses. Nevertheless, these ground-based measurements showed that (1) formaldehyde VCDs were higher during the warmer months relative to the colder months; (2) formaldehyde VCDs over Wuxi (in central eastern China) were higher than those over Xianghe (in northern China) and Back Garden (in southern China) for most months; (3) with the ~~seasonal progression presented by~~exception of June, when the formaldehyde VCDs over Xianghe were the highest among the three MAX-DOAS sites, reflecting the strong emissions from biomass burning in the NCP. Thus, the seasonal patterns shown in these few ground-based measurements were consistent with both the GOME-2A-observed and model-simulated formaldehyde VCDs.

Figure S4 compares the GOME-2A-observed and model-simulated formaldehyde VCDs against the monthly MAX-DOAS measurements at Xianghe (Vlemmix et al., 2015). The GOME-2A formaldehyde VCDs were consistent with the MAX-DOAS measurements in terms of the seasonal variation (R = 0.9) but showed an annual mean bias of -2.74×10^{15} molecules cm^{-2}. In comparison, by multiplying the GOME-2A formaldehyde VCD observations by 1.7, the annual mean bias against the MAX-DOAS measurements at Xianghe was reduced to 0.83×10^{15} molecules cm^{-2}. Figures 3 and 4 show that the bias between the satellite and MAX-DOAS measurements was also reduced at Wuxi when the GOME-2A formaldehyde VCDs were scaled up by 1.7. These findings offered some support for using the GOME-2A formaldehyde VCDs scaled by 1.7 as an upper-bound constraint for Chinese NMVOC emissions.

Figures 5 ~~shows~~and 6 compare the monthly mean glyoxal VCDs observed by GOME-2A~~.~~ with those simulated by the model using *a priori* emission estimates. Similar to the case of formaldehyde, GOME-2A-observed glyoxal VCDs ~~were highest~~ over China were higher in the east than in the west and higher in summer than in winter. High glyoxal VCDs were observed over the NCP in June and over Southwest China in spring, reflecting the strong seasonal biomass burning emissions. During winter (particularly in January), the GOME-2A glyoxal VCDs show an enhancement over eastern China ~~in June, reflecting large emissions of NMVOC species that are precursors to both formaldehyde and glyoxal. During January the eastern China glyoxal enhancement was more evident than formaldehyde. As biogenic emissions were small in winter, this~~, which was not apparent in the GOME-2A formaldehyde VCDs. This indicated that the glyoxal VCDs were more reflective of anthropogenic source~~. Figure 5 (e) (h) shows that the~~ than formaldehyde VCDs. The *a priori* simulated glyoxal VCDs were ~~higher than the GOME-2A observations in January and~~ generally lower than the GOME-2A glyoxal VCDs over eastern China year-round, especially during the warmer months (NMB between -52% and -59% during May to September, Table S6). The only exception was over the NCP in winter, when the *a priori* simulated glyoxal VCDs were significantly higher than the GOME-2A observations ~~in June~~. This suggested an overestimation of ~~anthropogenic~~NMVOC sources over the NCP in ~~January~~winter and ~~an~~ a substantial underestimation of ~~the biogenic~~NMVOCs sources ~~in June, which was~~over eastern China in summer, consistent with the constraints implied by the GOME-2A formaldehyde observations. During the transition months of April and October, when the anthropogenic and biogenic contributions to carbonyl productions were presumably more comparable, the simulated glyoxal VCDs were lower than the GOME-2A observations, while the simulated formaldehyde VCDs were higher than the GOME-2A observations (~~Figure~~ Figures 3 and 4 ~~(e) (h)).~~). This likely indicated that the *a priori* inventory underestimated the emissions of NMVOC species that preferentially produced glyoxal (e.g. aromatics, ethyne, ethene, and glyoxal~~) that preferentially produced glyoxal,~~), while ~~it overestimated~~overestimating the emissions of species ~~(e.g. $\geq C_4$ alkanes and $\geq C_3$ alkenes from anthropogenic activities)~~ that preferentially produced formaldehyde (e.g. $\geq C_4$ alkanes, $\geq C_3$ alkenes, and formaldehyde) during the transition months. Ground-based MAX-DOAS glyoxal measurements at Back Garden (a rural site in southern China ~~site~~) in July 2006 averaged 6.8(±5.2) × 10^{14} molecules cm^{-2}~~,~~ (Li et al., 2013), higher than both the GOME-2A-observed and simulated glyoxal VCDs. No other

ground-based measurements were available to provide spatial and seasonal information.

~~Figure 6 (a) (d) shows~~Figures 7 and 8 compare the ~~monthly mean~~ formaldehyde VCDs observed by ~~the~~ OMI ~~instrument. Similar to the GOME-2A observed~~ with those simulated by the model using *a priori* emission estimates. The spatial patterns and seasonal variations of the formaldehyde VCDs~~,~~ observed by OMI were similar to those observed by GOME-2A, with high formaldehyde ~~VCDs were higher~~ over ~~Eastern~~eastern China and ~~enhanced~~ during the warmer months. However, the ~~formaldehyde VCDs observed by~~ OMI ~~were~~observations were approximately 30% lower than ~~those observed by~~ the GOME-2A ~~by approximately 30%, likely~~observations over eastern China year-round. This difference may be due in part to the different ~~retrieval algorithms~~satellite overpass times, or it may be due to the inherent biases between the GOME-2A and OMI formaldehyde VCD retrievals (De Smedt et al., 2012; González Abad et al., 2015). The *a priori* simulated formaldehyde VCDs (at OMI overpass time) were higher than the OMI observations over eastern China year-round (NMB between 22% and 70%, Table S7), suggesting an overestimation of NMVOC emissions year-round. The simulated formaldehyde VCDs at the OMI overpass time were also lower ~~at OMI overpass time~~ than those at GOME-2A overpass time by ~~less than~~approximately 20% in all seasons~~.~~, due to stronger photolysis of formaldehyde in the afternoon in the model. However, the ~~ground-based~~MAX-DOAS measurements at ~~the~~ three Chinese ~~surface~~ sites ~~did not consistently~~ all showed higher formaldehyde VCDs at the OMI overpass time than at the GOME-2A overpass time (Li et al., 2013; Vlemmix et al., 2015; Wang et al, 2017).

Figures 9 and 10 show ~~such a diurnal pattern.~~

~~Figure 7 (a) (d) shows~~ the monthly mean glyoxal VCDs observed by ~~the OMI instrument.~~OMI and those simulated by the model. Valid OMI glyoxal ~~VCDs~~ observations were relatively sparse over China, especially during ~~colder months.~~the cold seasons. The seasonal and spatial patterns of the glyoxal VCDs observed by OMI were generally consistent with those observed by GOME-2A~~.~~ over eastern China. However, the glyoxal VCDs observed by OMI were consistently higher than those observed by GOME-2A, except in January. MAX-DOAS measurements of glyoxal at ~~a rural southern China site~~Back Garden in July 2006 were also higher in the afternoon than in mid-morning~~.~~ (Li et al., 2013).

In contrast, the *a priori* simulated glyoxal VCDs at OMI overpass time ~~(Figure 7 (e)-(h))~~ were lower than those at GOME-2A overpass time. This discrepancy among the glyoxal diurnal cycles represented by the MAX-DOAS measurements and the model indicated an uncertainty in the simulated local glyoxal budget.

~~Figures 6 and 7 also compare the formaldehyde and glyoxal VCDs observed by OMI to those simulated by the model using the *a priori* emission estimates over China. Formaldehyde VCDs observed by OMI~~ The *a priori* simulated glyoxal VCDs were lower than ~~those simulated by the model in all seasons, with the exception of a local hotspot over the NCP in June. However, the glyoxal VCDs observed by OMI were higher than those simulated by the model in all seasons.~~ the OMI observations throughout the year (NMB between -32% to -66%, Table S8) and especially from March to October, indicating an underestimation of NMVOC sources in the *a priori* year-round.

It thus appeared that the constraints on Chinese NMVOC emissions indicated by the OMI formaldehyde and glyoxal observations were contradictory~~, even during January and June when the NMVOC emissions over Eastern China were dominated by anthropogenic and biogenic sources, respectively.~~. There ~~may be~~are two ~~explanations~~possible causes for this apparent contradiction ~~indicated by the OMI formaldehyde and glyoxal observations.~~. Firstly, the simulated photochemical budgets of formaldehyde and glyoxal during the local afternoon may be in error~~. Errors in the model photochemical budget~~, which would also explain why the MAX-DOAS measurements of formaldehyde and glyoxal VCDs were both higher in the afternoon than in the morning, while the model showed an opposite diurnal contrast. ~~It~~Secondly, it is also possible that there were different inherent biases in the OMI formaldehyde and glyoxal retrievals.

4 ~~Top-down estimates~~Inversion experiments of Chinese NMVOC emissions

4.1 *A posteriori* formaldehyde and glyoxal VCDs from inversion experiments

The qualitative analyses in Section 3 showed that the GOME-2A and OMI retrievals of formaldehyde and glyoxal VCDs ~~provide~~provided disparate information on seasonal Chinese NMVOC emissions. ~~Thus~~Therefore, our four inversion experiments ~~on monthly Chinese NMVOC emissions~~ using different

satellite observations as constraints ~~(Table 2) represent a~~represented the range of probable top-down estimates given current satellite observations. ~~Figure 2 shows the monthly top-down Chinese NMVOC emission estimates from~~Here we first examined how the ~~four~~ inversion experiments ~~for January, April, June, and October and compares them against~~optimized the *a ~~priori~~* ~~emission estimates. The top-down emission estimates for~~ *posteriori* formaldehyde and glyoxal VCDs and the ~~full twelve months are shown in Figure S3. Figure S4~~resulting effects on the top-down monthly Chinese NMVOC emission estimates. Figure S7 shows the changes in the normalized cost functions over China in the four inversion experiments. Relative to their respective initial cost function values, the optimized cost function values were reduced by ~~10%-60% for all four experiments~~8% to 75% for all four experiments. Figure 2 shows the top-down monthly Chinese NMVOC emission estimates from the four inversion experiments and compares them against the *a priori* emission estimates. Figure S6 compares the *a priori* and *a posteriori* emission estimates for anthropogenic glyoxal precursors.

~~Figure 4 (i-l) and Figure 5 (i-l)~~Figures 3 to 6 show the *a posteriori* simulated monthly mean ~~VCDs of~~ formaldehyde and glyoxal~~, respectively,~~ VCDs from the GOME-2A formaldehyde~~-~~ and glyoxal inversion experiment (IE-1~~.~~). Overall, IE-1 greatly improved the agreement between the *a posteriori* VCDs and the GOME-2A observations for both formaldehyde (Table S4) and glyoxal~~.~~ (Table S6) over eastern China for most months. The ~~*a posteriori*~~ optimization was especially effective in optimizing the spatial pattern of the *a posteriori* formaldehyde VCDs, such that the *a posteriori* R against the GOME-2A formaldehyde VCDs exceeded 0.85 over eastern China for all twelve months (Table S4). Relative to the *a priori* VCDs, the *a posteriori* VCDs of formaldehyde and glyoxal ~~over eastern China~~ both decreased over NCP in ~~January~~winter and increased ~~in June,~~over eastern China between May and September. During the transition months of April and October, ~~IE-1~~the *a posteriori* formaldehyde VCDs decreased relative the *a ~~posteriori~~* ~~formaldehyde VCDs~~*priori*, while ~~increasing~~ the *a posteriori* glyoxal VCDs~~.~~ increased relative to the *a priori*. Figures 2 and Figure ~~2 illustrates~~S6 illustrated how these changes in ~~VCDs were driven by~~*a posteriori* formaldehyde and glyoxal VCDs affected the top-down monthly NMVOC emission estimates. For IE-1, the estimated emissions of all NMVOC species were reduced in ~~January~~winter but enhanced ~~in June.~~between May and September. In April and October, however, IE-1 decreased the total NMVOC emissions while preferentially increasing the emissions of anthropogenic glyoxal precursors.

~~Figure 6 (i-l) and Figure~~Figures 7 ~~(i-l)~~to 10 show the *a posteriori* monthly mean ~~VCDs of~~ formaldehyde and glyoxal~~, respectively,~~ VCDs from the OMI formaldehyde-/glyoxal inversion experiment (IE-2~~.~~). IE-2 was effective in reducing the *a posteriori* formaldehyde VCDs over eastern China year-round to better agree with the OMI formaldehyde observations~~.~~ (Table S7). However, ~~the inversion~~IE-2 increased the *a posteriori* glyoxal VCDs only slightly and was less effective in bringing agreement with the OMI glyoxal observations~~.~~ (Table S8). Figure 2 shows that the *a posteriori* NMVOC emission estimates from IE-2 were lower than the *a priori* estimates for all months. This was due to a combination of factors at work in the inversion. The low formaldehyde observations from OMI in all months drove a large reduction in the emissions of NMVOCs that produced only formaldehyde ($\geqslant C_4$ alkanes and $\geq C_3$ alkenes from anthropogenic activities, as well as primary formaldehyde from biomass burning). At the same time, the relatively high glyoxal observations from OMI drove an increase in the emissions of NMVOCs that produced mainly glyoxal (ethene, ethyne, and aromatics from anthropogenic activities, as well as primary glyoxal from biomass burning). For precursors that produced large amounts of both formaldehyde and glyoxal (most importantly biogenic isoprene), the inversion reduced the top-down emissions~~. This was because~~ as the formaldehyde observations had more ~~leverage on~~weight in the ~~inversion~~ cost function than the glyoxal observations, due to ~~their~~the lower observational errors~~. This manifested~~ in the formaldehyde VCDs. These findings showed the importance of well-characterized retrievals with reliable error estimates in inversion studies.

Formatted: Font: Not Italic

~~Figure~~Figures 3 and 4 ~~(q-t) shows~~showed the *a posteriori* formaldehyde VCDs from the inversion experiment IE-3, which was constrained by the GOME-2A-observed formaldehyde VCDs scaled by a factor of 1.7. The *a posteriori* formaldehyde VCDs in IE-3 ~~were further~~ increased further over eastern China during the warmer months relative to IE-1, especially over the NCP and central China in ~~June.~~summer. In December and January, the scaled-up GOME-2A observations over eastern China were still lower than the simulated formaldehyde VCDs using the *a priori* emissions, leading to a ~~small~~ reduction in the *a posteriori* formaldehyde VCDs~~.~~ over East China (Table S5). Figure 2 shows that the top-down monthly emission estimates ~~of~~for all NMVOC species were lower than the *a priori* in November, December, January, and February, and higher than the *a priori* ~~in June.~~for the warmer months. Consequently, although no observations of glyoxal were used as constraints in IE-3, the *a*

posteriori glyoxal VCDs also decreased in ~~January~~winter and increased in ~~June (Figure 5(m) and (o)). This is~~ summer, which were in better agreement with the GOME-2A observations (Figures 5 and 6). This is consistent with our ~~findings~~analyses in Section 3~~, whereby:~~ the constraints exerted by the GOME-2A formaldehyde and glyoxal observations were consistent in ~~January~~winter and in ~~June~~summer, when the NMVOC emissions were dominated by anthropogenic and biogenic sources, respectively. However, IE-3 had almost no effects on the ~~simulated~~*a posteriori* glyoxal VCDs and the top-down emission estimates of anthropogenic glyoxal precursors in April and October ~~(Figure 5 (n) and (p)).~~. This demonstrated the necessity of glyoxal observations on constraining the emissions of NMVOC species that preferentially produced glyoxal, including most importantly the aromatics.

The impacts of satellite glyoxal observations on constraining Chinese ~~NMVOC~~ glyoxal precursors emission estimates was further demonstrated in IE-4. ~~Figure 7 (m-p) shows~~ Figures 9 and 10 show that the *a posteriori* glyoxal VCDs from IE-4~~, which used only~~ were in better agreement with the OMI glyoxal observations ~~as constraints. The *a posteriori* glyoxal VCDs~~ for all months ~~increased, to an extent greater than those in IE~~ (Table S8). Figure 2~~.~~ and Figure ~~2 shows~~S6 show that this increase in the *a posteriori* glyoxal VCDs in IE-4 was achieved ~~mainly~~ by substantially increasing the emission estimates of anthropogenic glyoxal precursors for all months. In ~~June~~summer, the emissions of biogenic isoprene (precursor to both glyoxal and formaldehyde) also increased. As a result, the *a posteriori* formaldehyde VCDs in IE-4 increased in ~~June~~summer but remained similar to the *a priori* simulation for the other months (~~Figure 6 (m-p)).~~Figures 7 and 8, Table S7).

4.2 Top-down estimates of Chinese NMVOC emissions from inversion experiments

Table ~~2~~1 and Figure ~~8 shows~~11 show the top-down estimates for Chinese annual ~~total~~ NMVOC emissions from the four inversion experiments and compare them against the *a priori*. Our top-down annual total estimates for Chinese NMVOCs ranged from ~~23.4~~30.7 to ~~35.4~~49.5 Tg ~~C~~ y^{-1}, compared to the ~~27.4~~38.3 Tg ~~C~~ y^{-1} of the *a priori*. The highest top-down estimate was from IE-3, constrained by 1.7 times the GOME-2A formaldehyde VCD observations. The lowest top-down estimate was from IE-2, ~~due to~~mainly driven by the relatively low formaldehyde observations from OMI.

Anthropogenic sources constituted 44%-53% 57% of the total top-down total NMVOC emissions. The highest top-down anthropogenic emissions estimate was from IE-4 (19.7 Tg C y^{-1}), which reflected the strong traction of the OMI glyoxal observations on constraining anthropogenic NMVOC emissions. The lowest top-down total anthropogenic emission estimate was from IE-2 (13.516.4 Tg C y^{-1}). All four inversion experiments consistently showed larger annual emissions of anthropogenic glyoxal precursors than the *a priori*. (Figure 11). In particular, our top-down estimates for anthropogenic aromatics ranged from 5.05 to 7.39 Tg C y^{-1}, consistently larger than the *a priori* of 5.4.9 Tg C y^{-1} (Li et al., 2014).2017). The highest top-down anthropogenic glyoxal precursors (including aromatics, ethyne, ethane, and glyoxal) emissions estimate was from IE-4 (12.3 Tg y^{-1}), which reflected the strong impacts of the OMI glyoxal observations on constraining anthropogenic NMVOC emissions.

The top-down estimates for biogenic NMVOCs emission emissions from IE-1, IE-3, and IE-4 ranged between 8.920.0 and 1422.8 Tg C y^{-1}. The (top-down estimates for biogenic isoprene were 4.9 to 10.5 Tg C y^{-1}.emission estimates between 9.8 and 11.7 Tg y^{-1}), which were significantly larger than the *a priori*. As a result, the contrast between the NMVOC emissions in summer and those in winter were greatly enhanced in the top-down estimates in these three inversion experiments, relative to the *a priori* (Figure 2). The exception was IE-2, which estimated the biogenic NMVOC emissions to be 12.2 Tg y^{-1} (including 5.4 Tg y^{-1} of isoprene). The top-down estimate for biomass burning NMVOC emissions from the four inversion experiments were between 1.062.08 to 1.473.13 Tg C y^{-1}, with the largest top-down estimate driven by the scaled-up GOME-2A formaldehyde VCDs (IE-3).

Figure 912 shows the spatial distribution of the scale factors for the Chinese annual NMVOC emissions from each of the four inversion experiments relative to the *a priori* emission estimates. The use of GOME-2A formaldehyde and glyoxal observations as constraints in IE-1 led to a domain-wide increase in biogenic NMVOC emissions, except in the northeast. IE-1 also found an increase in biomass burning emissions over the NCP in June. A similar spatial distribution was foundIn contrast, anthropogenic NMVOC emissions were slightly reduced over northeast, north, and southwest China. In IE-3, the annual NMVOC emissions over eastern China increased for the emission scale factors of IE-3. Again, this indicated a consistency between theall three sources, due to constraints exerted by the scaled-up GOME-2A formaldehyde and glyoxal observations from GOME-2A.VCDs. The optimized emission

scale factors from IE-2 and IE-4 were of opposite signs. Using only OMI glyoxal observations as constraints in IE-4 led to a domain-wide increase in NMVOC emissions from all sectors. However, when ~~further~~ constraints of the relatively low OMI formaldehyde observations were added in IE-2, the top-down NMVOC emission estimates decreased across the domain.

As discussed previously, our four inversion experiments using different satellite retrievals as constraints ~~represent~~represented the range of probable top-down estimates given currently ~~-~~available satellite observations. To represent the difference between these top-down estimates relative to the *a priori*, we averaged the top-down estimates from the four inversion experiments. Our averaged top-down estimate for Chinese total annual NMVOC emissions was ~~30.8~~41.9 Tg ~~C~~ y^{-1}, including ~~17.0~~20.2 Tg ~~C~~ y^{-1}, ~~12.6~~19.2 Tg ~~C~~ y^{-1}, and ~~1.~~2.48 Tg ~~C~~ y^{-1} from anthropogenic, biogenic, and biomass burning sources, respectively. Our average emission estimate for anthropogenic ~~aromatic flux~~aromatics was 6.~~1~~5 Tg ~~C~~ y^{-1}, which was ~~24~~20% larger than the *a priori* estimate of Li et al. (~~2014).~~2017).

Figure 1 ~~(e-l)~~ shows the spatial ~~distribution of annual Chinese NMVOC emission~~distributions of our averaged top-down ~~estimate~~ Chinese NMVOC emissions and the scale factors relative to the *a priori* ~~estimates~~. Our averaged top-down estimate ~~of~~for Chinese NMVOC emissions were spatially consistent with the *a priori*, but the total fluxes ~~increased by 10% to 40%~~were larger than the a priori throughout eastern China ~~relative~~by 10% to ~~the *a priori*.~~30%. In particular, we found a 40% increase in the biomass burning emissions over the NCP. We also found a ~~20% 40~~10%-30% increase in the anthropogenic NMVOC emissions in coastal eastern China. ~~Largest scale factors for~~Large increases in the biogenic ~~NMVOC~~emissions were found near the northwestern border of China and along the northeast-to-southwest division line of vegetation density. This potentially indicated an underestimation of biogenic NMVOC emission from semi-arid ecosystems in the MEGAN inventory.

5 Comparison with previous estimates of Chinese NMVOC emissions

Table ~~1~~2 compares our top-down estimates of Chinese ~~annual~~ NMVOC emissions for the year 2007 against estimates in the literature for the years between 2000~~2005~~ and ~~2012.~~2014. It should be noted that most bottom-up anthropogenic and biomass burning emission inventories ~~often estimated~~

Formatted: Font: 9 pt

thequantified total NMVOCs emitted from a source sectorNMVOC emissions using emission factors for total NMVOCs, then distributed the emissions using different species profile data. As a result, bottom-up estimates for anthropogenic and biomass burning sources often included additional NMVOC species not represented here in our study.

Our top-down estimate for biogenic NMVOC emissions range from 8.912.2 to 1422.8 Tg C y^{-1}, on average 1711% larger than the flux *a priori* estimate calculated fromby the MEGAN inventoryalgorithm (Guenther et al., 2006). Our top-down estimate for isoprene emission (the single most emitted NMVOC species) ranged from 5.4.9 to 10.511.7 Tg C y^{-1}, bracketing the *a priori* of 6.6 previous bottom-up estimates of 7.5 to 9.9 Tg C y^{-1}. (Guenther et al., 2006; Stavrakou et al., 2014; Sindelarova et al., 2014). Stavrakou et al. (2015) previously used GOME-2A and OMI formaldehyde observations VCDs in 2010 to derive top-down estimates of isoprene emissions over China of 5.09 Tg C y^{-1} (using GOME-2A observations as constraints) and 5.6.5 Tg C y^{-1} (using OMI observations as constraints), respectively. In comparison, our top-down isoprene emission estimates constrained by GOME-2A and OMI (observations of both formaldehyde and glyoxal) observations, were 9.8.2 Tg C y^{-1} (from IE-1) and 5.4.9 Tg C y^{-1} (from IE-2), respectively. Our top-down estimates constrained by GOME-2A observations was larger than that of Stavrakou et al. (2015) due to the additional glyoxal constraints. Our estimate isoprene emission constrained by OMI observations was lower than that of Stavrakou et al. (2015)). This was because the OMI formaldehyde VCDs over China retrieved by González Abad et al. (2015)we used were systematically lower than the OMI formaldehyde VCDs used in Stavrakou et al. (2015), which was retrieved by De Smedt et al. (2015). A later study by Stavrakou et al. (2017) using OMI formaldehyde VCDs estimated the average Chinese biogenic isoprene emissions for the years 2005 to 2014 to be 5.8 Tg y^{-1} , consistent with our top-down estimate from IE-2.

Our top-down estimates for Chinese annual biomass burning NMVOC emissions ranged from 1.062.08 to 1.473.13 Tg C y^{-1}. These numbers are, in good agreement with the estimate bottom-up estimates of Huang et al. (2012). Previous bottom-up biomass burning NMVOC emission estimates by) and the top-down estimates of Stavrakou et al. (2015) (Table 2). Similar to the findings in Fu et al. (2007) and Stavrakou et al. (2015, 2016), our study also highlighted the large emissions from crop residue over the NCP in June, which were severely underestimated in some previous studies. Bo et al. (2008) and Wu et

al. (2016) ~~ranged from 1.9~~estimated the Chinese biomass burning NMVOC emissions to ~~2.~~be 3.32 to 4.2 Tg ~~C~~ y^{-1}, but only 25% to 30% of these emissions were from open burning of crop residues; the rest were emitted from biofuel burning~~.~~, which we categorized as anthropogenic in this study. The GFED3 inventory (van der Werf et al., 2010), based on satellite ~~burnt~~burned area observations, severely underestimated biomass burning emissions over China, particularly those associated with crop residue burning. ~~Top down~~The updated GFED4 (van der Werf et al., 2017) partially accounted for emissions for small fires, but its estimate ~~of~~for Chinese biomass burning NMVOC emissions ~~by Stavrakou et al. (2015)~~ was ~~between 1.1-1.5 Tg C y^{-1}, very close to~~still lower than our top-down ~~estimate range (1.06-1.47 Tg C y^{-1}). Similar to Fu et al. (2007) and Stavrakou et al. (2016), our study also highlighted the large emissions from crop residue over the NCP in June (Figure 2).~~estimates by at least a factor of two.

Previous bottom-up estimates of Chinese anthropogenic NMVOC emissions ranged widely from ~~10~~12.7 to ~~29.8~~35.46 Tg ~~C~~ y^{-1} (Bo et al., 2008; Zhang et al., 2009; Cao et al., 2011; Kurokawa et al., 2013; Li et al., ~~2014~~2017; Wu et al., 2016; Granier et al., 2017; Huang et al., 2017) due to the use of different emission factors, activity data, and statistical models. Previous top-down estimates of Chinese anthropogenic NMVOC emissions for the years 2007 to 2014 ranged from ~~17.3-28.7~~20.6-34.2 Tg ~~C~~ y^{-1} (Liu et al., 2012; Stavrakou et al., 2015, 2017). Our top-down estimates had a smaller range between ~~13.5~~16.4 to ~~19.7~~23.6 Tg ~~C~~ y^{-1}. Our top-down estimates for anthropogenic aromatics (5.~~0~~5 to 7.~~3~~9 Tg ~~C~~ y^{-1}) were approximately middle-of-the-range relative to previous estimates of 2.~~1-11.3~~4-13.4 Tg ~~C~~ y^{-1}. The large difference between previous top-down estimates and our top-down estimates of anthropogenic NMVOCs were predominantly due to the choices of satellite observation constraints, and to a lesser extent due to the choices of chemical transport model, the NMVOC species modeled, and the *a priori* emission estimates. Specifically, the much higher estimate of anthropogenic aromatic emission by Liu et al. (2012) (~~11.3~~13.4 Tg ~~C~~ y^{-1}) compared to our top-down estimates (5.~~0~~5-7.~~3~~9 Tg ~~C~~ y^{-1}) was due to (1) the ~~high~~higher glyoxal VCDs observed by the SCIAMACHY instrument compared to those observed ~~the~~by GOME-2A and OMI ~~instruments~~ over China; (2) the assumption made by Liu et al. (2012) that all anomalous glyoxal was produced by aromatics ~~oxidation;~~; and (3) the lower yields of glyoxal from aromatics oxidation used in Liu et al. (2012) than those used in our model.

Formatted: Not Highlight

~~Our four inversion experiments all indicated stronger anthropogenic NMVOC emissions in summer than in winter. In contrast, the *a priori* estimates showed a slightly stronger NMVOC emission in winter than in summer, which was driven by stronger activity levels in winter in the bottom-up inventories along with seasonally invariant emission factors (Li et al., 2017). However, studies showed that the NMVOC emission factors, in particular those for transport and industrial sectors, were strongly and positively correlated with temperature (Rubin et al., 2006; Wei et al., 2016).~~

6 Impacts on simulated surface ozone and secondary organic aerosol levels over China

As discussed above, three out of our four inversion experiments showed a stronger seasonal contrast in the top-down NMVOC emission estimates between summer and winter (Figure 2). We evaluated the impacts of this stronger seasonal contrast in NMVOC emissions on surface ozone and secondary organic carbon (SOC) aerosol concentrations ~~due to our~~ by driving the GEOS-Chem model with the *a priori* NMVOC emission estimates and with the average top-down emission estimates ~~of NMVOCs.~~ from our four inversion experiments, respectively (Table 1).

Figure ~~10~~13 compares the monthly mean afternoon (13~~:00~~ to 17~~:00 LT~~ local time) surface ozone concentrations simulated using our averaged top-down emission estimates against those simulated using the *a priori* emissions for June and December 2007. Also shown ~~in Figure 10~~ are surface observations at representative regional sites (Li et al., 2007; Xu et al., 2008; J. M. Zhang et al., 2009; Zheng et al., 2010; Wang et al., 2012; Wang et al., 2015; Li and Bian, 2015; Sun et al., 2016; Xu et al., 2016~~).~~) (Table S9). Using the *a priori* emissions, the highest simulated afternoon surface ozone concentrations were between 90-100~~-110~~ ppb over the NCP in June. This was lower than the observations at two sites in the NCP, including at a rural site outside Beijing ~~(~~(100 to 120 ppb) and at Mt. Tai (108 ppb). In comparison, by using our averaged top-down NMVOC emission estimate, the simulated afternoon surface ozone increased by ~~5-10~~1 to 8 ppb over ~~the NCP~~eastern China in June and were in better agreement with the observations~~.~~ (reducing the bias from -3.4 ppb to -0.7 ppb, Table S9). In December, the simulated afternoon surface ozone using the *a priori* ~~emission~~emissions consistently overestimated the observed concentrations in eastern China. In comparison, by using our averaged top-down NMVOC emission estimates, the simulated afternoon surface ozone over eastern China

decreased by ~~5~~1 to ~~13~~10 ppb, again in better agreement with the observations~~.~~ (mean bias reduced from 12 ppb to 6.9 ppb, Table S9). It thus ~~appears~~appeared that our average top-down NMVOC emission estimates ~~for Chinese NMVOCs~~, with stronger seasonal contrast in NMVOC emissions compared to the *a priori*, improved the simulation of regional ozone when compared to surface measurements.

Figure S8 compares the simulated monthly mean surface SOC concentrations using our averaged top-down NMVOCs emissions against those simulated using the *a priori* NMVOC emissions for January and June in 2007. Also shown are the SOC measurements at 12 surface sites in June of 2006 and 2007 from Zhang et al. (2012) (Table S10). By driving the model with our average top-down NMVOC emissions, the simulated surface SOC concentrations in June increased by 0.1 to 0.8 μgC m^{-3} over eastern China relative to the simulation using the *a priori* NMVOC emissions. This increase in simulated SOC concentrations brought the model to be closer to the surface measurements, but the model still severely underestimated observed SOC concentrations. We note that our version of the GEOS-Chem model only included two pathways for secondary organic aerosol formation: (1) the reversible partitioning of semi-volatile products from the oxidation of isoprene, monoterpenes, and aromatics formation pathways (Liao et al., 2007; Henze et al., 2008), and (2) the irreversible uptake of dicarbonyl by aqueous aerosols and cloud drops (Fu et al., 2008). Other pathways, such as the atmospheric aging of semi-volatile and intermediate volatility organic compounds (S/IVOC), have been shown to be an important source of secondary organic aerosol (Robinson et al., 2007; Pye and Seinfeld, 2010) but they were not included in our version of GEOS-Chem. Regardless, the precursors and formation pathways of secondary organic aerosols in China are still poorly understood (Fu et al., 2012), such that no quantitative conclusions can be drawn regarding the impacts of our top-down NMVOC emission estimates on regional secondary organic aerosol formation.

7 Conclusions

We used the GEOS-Chem model and its adjoint, as well as satellite observations of tropospheric column concentrations of formaldehyde and glyoxal, to constrain monthly Chinese NMVOC emissions from anthropogenic, biogenic, and biomass burning sources for the year 2007. We updated the gas-phase chemistry in the GEOS-Chem model and constructed its adjoint. The *a priori* NMVOC

emission estimates from biogenic, anthropogenic, and biomass burning sources were taken from ~~widely-used bottom-up emission~~ the inventories~~.~~ developed by Guenther et al. (2006), Li et al (2014, 2017), and Huang et al. (2012), as well as van der Werf et al. (2010), respectively. We conducted four inversion experiments, which were constrained by ~~formaldehyde and glyoxal observations from the GOME-2A instrument (IE-1),~~ the formaldehyde and glyoxal observations from ~~the~~ GOME-2A (IE-1), the formaldehyde and glyoxal observations from OMI ~~instrument~~ (IE-2), ~~1.7 times~~ the formaldehyde observations from ~~the~~ GOME-2A ~~instrument~~ scaled by 1.7 (IE-3), and the glyoxal observations from ~~the~~ OMI ~~instrument~~ (IE-4), respectively. The results from these experiments represented the range of probable top-down NMVOC emission estimates for China given current satellite ~~observational~~observation constraints.

Our top-down estimates of total annual Chinese NMVOC emission from the four inversion experiments ranged from ~~23.4~~30.7 to ~~35.4~~49.5 Tg ~~C~~ y^{-1}. Our top-down estimates of Chinese anthropogenic NMVOC emission was ~~13.5~~16.4 to ~~19.7~~ 23.6 Tg ~~C~~ y^{-1}. In particular, ~~we~~our top-down ~~estimate of~~estimates for Chinese anthropogenic aromatic emissions ~~range~~ranged from 5.~~0~~ 5 to 7.~~39~~ Tg ~~C~~ y^{-1}, much smaller than the top-down estimate of ~~11.3~~13.4 Tg ~~C~~ y^{-1} by Liu et al. (2012). Our top-down estimate of Chinese biogenic NMVOC emission ranged from ~~8.9~~ 12.2 to ~~14~~22.8 Tg ~~C~~ y^{-1}, with 5.4~~.9~~ to ~~10.5~~ 11.7 Tg ~~C~~ y^{-1} attributed to isoprene. Our top-down estimate for Chinese biomass burning NMVOC emission range from ~~1.1~~2.08 to ~~1.5~~3.13 Tg ~~C~~ y^{-1} and was mostly associated with seasonal open burning of crop residue after local harvests, such as those over the NCP in June.

~~We evaluated the impacts on regional surface ozone concentrations from~~Three out of our ~~average~~ four inversion experiments showed a stronger seasonal contrast in the top-down ~~Chinese~~ NMVOC emission estimates~~. We found that the simulated monthly mean~~ between summer and winter, relative to the *a priori* emission estimates. By applying this stronger seasonal contrast in monthly NMVOC emissions in the model, the simulated afternoon surface ozone concentrations over eastern China increased by ~~5-12~~1 to 8 ppb ~~over the NCP~~ in June~~, compared to the *a priori* simulation. In December, the simulated monthly mean afternoon surface ozone concentrations~~ and decreased by ~~5-13~~1 to 10 ppb ~~over northern and central China, compared to the *a priori* simulation. For both seasons,~~in December relative to the simulation using ~~our averaged top-down emission estimates were in~~ the *a priori* emissions, and

the model was brought to better ~~general~~ agreement with regional surface ~~observations.~~ ozone measurements. Similarly, compared to the *a priori* simulation, the simulated monthly mean SOC concentrations driven by our top-down NMVOCs emissions increased by 0.1 to 0.8 μgC m^{-3} over Eastern China in June. This increase in simulated SOC concentrations reduced, but did not eliminate, the large low biases in the simulated SOC concentrations relative to the surface measurements in June.

We concluded that formaldehyde and glyoxal observations from GOME-2A and OMI provide quantitative constraints on the monthly emissions of Chinese NMVOCs. In particular, the simultaneous use of the observations of both species ~~helps~~helped distinguish NMVOC precursors and thus ~~provides~~provided better quantification of individual sources. However, better validation of these satellite data over China are urgently needed, particularly ~~in terms of discrepancies between different retrievals for the same species.~~to resolve the apparent discrepancies between different retrievals for the same species. The monthly inversions presented in this work, conducted at 5^o longitude × 4^o latitude resolution due to limited computation resources, quantified the Chinese NMVOC emissions on regional/sub-regional scales. Future inversions and sensitivity studies targeting shorter periods of time may be conducted on finer resolutions to quantify Chinese NMVOC emissions and to evaluate their impacts on photochemistry at city cluster scales.

Acknowledgements

This work was supported by the Ministry of Science and Technology of China (2014CB441303) and the National Natural Sciences Foundation of China (41461164007, 41222035). We thank the National Super Computer Center in Tianjin for ~~supporting this work.~~providing computational support and the QA4ECV campaign for providing the MAX-DOAS measurements at Xianghe. DKH was supported by the National Strategic Project-Fine particle of the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (MSIT), the Ministry of Environment (ME), and the Ministry of Health and Welfare (MOHW) (2017M3D8A1092052)~~D. K. Henze recognizes support from NASA NNX17AF63G~~.

References

~~Acarreta, J. R., De Haan, J. F., and Stammes, P.: Cloud pressure retrieval using the O2-O2absorption band at 477 nm, J. Geophys. Res., 109, doi: 10.1029/2003jd003915, 2004.~~

Akagi, S. K., Yokelson, R. J., Wiedinmyer, C., Alvarado, M. J., Reid, J. S., Karl, T., Crounse, J. D., and Wennberg, P. O.: Emission factors for open and domestic biomass burning for use in atmospheric models, Atmos. Chem. Phys., 11, 4039-4072, doi: 10.5194/acp-11-4039-2011, 2011.

Arey, J., Obermeyer, G., Aschmann, S. M., Chattopadhyay, S., Cusick, R. D., and Atkinson, R.: Dicarbonyl Products of the OH Radical-Initiated Reaction of a Series of Aromatic Hydrocarbons, Environ. Sci. Technol., 43, 683-689, doi: 10.1021/es8019098, 2009.

Barkley, M. P., Palmer, P. I., Kuhn, U., Kesselmeier, J., Chance, K., Kurosu, T. P., Martin, R. V., Helmig, D., and Guenther, A.: Net ecosystem fluxes of isoprene over tropical South America inferred from Global Ozone Monitoring Experiment (GOME) observations of HCHO columns, J. Geophys. Res., 113, doi: 10.1029/2008jd009863, 2008.

Barkley, M. P., Palmer, P. I., De Smedt, I., Karl, T., Guenther, A., and Van Roozendael, M.: Regulated large-scale annual shutdown of Amazonian isoprene emissions?, Geophys. Res. Lett., 36, doi:10.1029/2008gl036843, 2009.

Barkley, M. P., Smedt, I. D., Van Roozendael, M., Kurosu, T. P., Chance, K., Arneth, A., Hagberg, D., Guenther, A., Paulot, F., Marais, E., and Mao, J.: Top-down isoprene emissions over tropical South America inferred from SCIAMACHY and OMI formaldehyde columns, J. Geophys. Res. Atmos., 118, 6849-6868, doi:10.1002/jgrd.50552, 2013.

Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., Li, Q., Liu, H. Y., Mickley, L. J., and Schultz, M. G.: Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation, J. Geophys. Res., 106, 23073-23095, doi: 10.1029/2001JD000807, 2001.

Bloss, C., Wagner, V., Jenkin, M. E., and Volkamer, R.: Development of a detailed chemical mechanism (MCMv3.1) for the atmospheric oxidation of aromatic hydrocarbons, Atmos. Chem. Phys., 5, 641-664, doi:10.5194/acp-5-641-2005, 2005.

Bo, Y., Cai, H., and Xie, S. D.: Spatial and temporal variation of historical anthropogenic NMVOCs emission inventories in China, Atmos. Chem. Phys., 8, 7297-7316, doi: 10.5194/acp-8-7297-2008, 2008.

~~Bolscher, M., Pulles, T., Brand, R., Pereira, J., Mota, B., Spessa, A., Dalsoren, S., Nojic, T., and Szopa, S.: Emission data sets and methodologies for estimating emissions, RETRO Deliverable D1-6, 2007.~~

~~Brioude, J., Kim, S. W., Angevine, W. M., Frost, G. J., Lee, S. H., McKeen, S. A., Trainer, M., Fehsenfeld, F. C., Holloway, J. S., Ryerson, T. B., Williams, E. J., Petron, G., and Fast, J. D.: Top-down~~

~~estimate of anthropogenic emission inventories and their interannual variability in Houston using a mesoscale inverse modeling technique, J. Geophys. Res. Atmos., 116, doi:10.1029/2011JD016215, 2011.~~

~~Byrd, R. H., Lu, P. H., Nocedal, J., and Zhu, C. Y.: A Limited Memory Algorithm for Bound Constrained Optimization, Siam J. Sci. Comput., 16, 1190-1208, doi: 10.1137/0916069, 1995.~~

Cao, G., Zhang, X., Gong, S., An, X., and Wang, Y.: Emission inventories of primary particles and pollutant gases for China, Chin. Sci. Bull., 56, 781-788, doi:10.1007/s11434-011-4373-7, 2011.

Chan Miller, C., Gonzalez Abad, G., Wang, H., Liu, X., Kurosu, T., Jacob, D. J., and Chance, K.: Glyoxal retrieval from the Ozone Monitoring Instrument, Atmos. Meas. Tech., 7, 3891-3907, doi:10.5194/amt-7-3891-2014, 2014.

Chan Miller, C., Jacob, D. J., Abad, G. G., and Chance, K.: Hotspot of glyoxal over the Pearl River delta seen from the OMI satellite instrument: implications for emissions of aromatic hydrocarbons, Atmos. Chem. Phys., 16, 4631-4639, doi: 10.5194/acp-16-4631-2016, 2016.

Chan Miller, C., Jacob, D. J., Marais, E. A., Yu, K. R., Travis, K. R., Kim, P. S., Fisher, J. A., Zhu, L., Wolfe, G. M., Hanisco, T. F., Keutsch, F. N., Kaiser, J., Min, K. E., Brown, S. S., Washenfelder, R. A., Abad, G. G., and Chance, K.: Glyoxal yield from isoprene oxidation and relation to formaldehyde: chemical mechanism, constraints from SENEX aircraft observations, and interpretation of OMI satellite data, Atmos. Chem. Phys., 17, 8725-8738, doi: 10.5194/acp-17-8725-2017, 2017.

Chance, K.: Analysis of BrO measurements from the Global Ozone Monitoring Experiment, Geophys. Res. Lett., 25, 3335-3338, doi: 10.1029/98gl52359, 1998.

Curci, G., Palmer, P. I., Kurosu, T. P., Chance, K., and Visconti, G.: Estimating European volatile organic compound emissions using satellite observations of formaldehyde from the Ozone Monitoring Instrument, Atmos. Chem. Phys., 10, 11501-11517, doi: 10.5194/acp-10-11501-2010, 2010.

Daescu, D. N., Sandu, A., and Carmichael, G. R.: Direct and adjoint sensitivity analysis of chemical kinetic systems with KPP: II—numerical validation and applications, Atmos. Environ., 37, 5097-5114, doi:10.1016/j.atmosenv.2003.08.020, 2003.

De Smedt, I., Van Roozendael, M., Stavrakou, T., Müller, J. F., Lerot, C., Theys, N., Valks, P., Hao, N., and van der A, R.: Improved retrieval of global tropospheric formaldehyde columns from GOME-2/MetOp-A addressing noise reduction and instrumental degradation issues, Atmos. Meas. Tech., 5, 2933-2949, doi:10.5194/amt-5-2933-2012, 2012.

De Smedt, I., Stavrakou, T., Hendrick, F., Danckaert, T., Vlemmix, T., Pinardi, G., Theys, N., Lerot, C., Gielen, C., Vigouroux, C., Hermans, C., Fayt, C., Veefkind, P., Müller, J. F., and Van Roozendael, M.: Diurnal, seasonal and long-term variations of global formaldehyde columns inferred from combined OMI and GOME-2 observations, Atmos. Chem. Phys., 15, 12519-12545, doi: 10.5194/acp-15-12519-2015, 2015.

Dufour, G., Wittrock, F., Camredon, M., Beekmann, M., Richter, A., Aumont, B., and Burrows, J. P.:

SCIAMACHY formaldehyde observations: constraint for isoprene emission estimates over Europe?, Atmos. Chem. Phys., 9, 1647-1664, doi:10.5194/acp-9-1647-2009, 2009.

Fu, T. M., Cao, J. J., Zhang, X. Y., Lee, S. C., Zhang, Q., Han, Y. M., Qu, W. J., Han, Z., Zhang, R., Wang, Y. X., Chen, D., and Henze, D. K.: Carbonaceous aerosols in China: top-down constraints on primary sources and estimation of secondary contribution, Atmos. Chem. Phys., 12, 2725-2746, doi: 10.5194/acp-12-2725-2012, 2012.

Fu, T.-M., Jacob, D. J., Palmer, P. I., Chance, K., Wang, Y. X., Barletta, B., Blake, D. R., Stanton, J. C., and Pilling, M. J.: Space-based formaldehyde measurements as constraints on volatile organic compound emissions in east and south Asia and implications for ozone, J. Geophys. Res., 112, doi: 10.1029/2006jd007853, 2007.

Fu, T.-M., Jacob, D. J., Wittrock, F., Burrows, J. P., Vrekoussis, M., and Henze, D. K.: Global budgets of atmospheric glyoxal and methylglyoxal, and implications for formation of secondary organic aerosols, J. Geophys. Res., 113, doi:10.1029/2007jd009505, 2008.

González Abad, G., Liu, X., Chance, K., Wang, H., Kurosu, T. P., and Suleiman, R.: Updated Smithsonian Astrophysical Observatory Ozone Monitoring Instrument (SAO OMI) formaldehyde retrieval, Atmos. Meas. Tech., 8, 19-32, doi:10.5194/amt-8-19-2015, 2015.

Gonzi, S., Palmer, P. I., Barkley, M. P., De Smedt, I., and Van Roozendael, M.: Biomass burning emission estimates inferred from satellite column measurements of HCHO: Sensitivity to co-emitted aerosol and injection height, Geophys. Res. Lett., 38, doi: 10.1029/2011gl047890, 2011.

Granier, C., Doumbia, T., Granier, L., Sindelarova, K., Frost, G., Bouarar, I., Liousse, C., Darras, S. and Stavrakou, J.: Anthropogenic emissions in Asia, Air Pollution in Eastern Asia : an integrated perspective, eds. Bouarar, I., Wang, X., Brasseur, G., Springer international Publishing, doi:10.1007/978-3-319-59489-7-6, pp. 107-133, 2017.

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., and Wang, X.: The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions, Geosci. Model Dev., 5, 1471-1492, doi: 10.5194/gmd-5-1471-2012, 2012.

Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., and Geron, C.: Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), Atmos. Chem. Phys., 6, 3181-3210, doi: 10.5194/acp-6-3181-2006, 2006.

Hallquist, M., Wenger, J. C., Baltensperger, U., Rudich, Y., Simpson, D., Claeys, M., Dommen, J., Donahue, N. M., George, C., Goldstein, A. H., Hamilton, J. F., Herrmann, H., Hoffmann, T., Iinuma, Y., Jang, M., Jenkin, M. E., Jimenez, J. L., Kiendler-Scharr, A., Maenhaut, W., McFiggans, G., Mentel, T. F., Monod, A., Prevot, A. S. H., Seinfeld, J. H., Surratt, J. D., Szmigielski, R., and Wildt, J.: The formation, properties and impact of secondary organic aerosol: current and emerging issues, Atmospheric Chemistry and Physics, 9, 5155-5236, doi: 10.5194/acp-9-5155-2009, 2009.

Han, K. M., Park, R. S., Kim, H. K., Woo, J. H., Kim, J., and Song, C. H.: Uncertainty in biogenic

isoprene emissions and its impacts on tropospheric chemistry in East Asia, Sci. Total Environ., 463-464, 754-771, doi: 10.1016/j.scitotenv.2013.06.003, 2013.

Hansen, P. C.: Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion, SIAM, Philadelphia, 1998.

Hays, M. D., Geron, C. D., Linna, K. J., Smith, N. D., and Schauer, J. J.: Speciation of gas-phase and fine particle emissions from burning of foliar fuels, Environ. Sci. Technol., 36, 2281-2295, doi:10.1021/es0111683, 2002.

Henze, D. K., Hakami, A., and Seinfeld, J. H.: Development of the adjoint of GEOS-Chem, Atmos. Chem. Phys., 7, 2413-2433, doi: 10.5194/acp-7-2413-2007, 2007.

Henze, D. K., Seinfeld, J. H., and Shindell, D. T.: Inverse modeling and mapping US air quality influences of inorganic PM2.5 precursor emissions using the adjoint of GEOS-Chem, Atmos. Chem. Phys., 9, 5877-5903, doi:10.5194/acp-9-5877-2009, 2009.

Henze, D. K., Seinfeld, J. H., Ng, N. L., Kroll, J. H., Fu, T. M., Jacob, D. J., and Heald, C. L.: Global modeling of secondary organic aerosol formation from aromatic hydrocarbons: high- vs. low-yield pathways, Atmos. Chem. Phys., 8, 2405-2420, doi:10.5194/acp-8-2405-2008, 2008.

Huang, G. L., Brook, R., Crippa, M., Janssens-Maenhout, G., Schieberle, C., Dore, C., Guizzardi, D., Muntean, M., Schaaf, E., and Friedrich, R.: Speciation of anthropogenic emissions of non-methane volatile organic compounds: a global gridded data set for 1970-2012, Atmospheric Chemistry and Physics, 17, 7683-7701, 10.5194/acp-17-7683-2017, 2017.

Huang, X., Li, M., Li, J., and Song, Y.: A high-resolution emission inventory of crop burning in fields in China based on MODIS Thermal Anomalies/Fire products, Atmos. Environ., 50, 9-15, doi: 10.1016/j.atmosenv.2012.01.017, 2012.

Ip, H. S. S., Huang, X. H. H., and Yu, J. Z.: Effective Henry's law constants of glyoxal, glyoxylic acid, and glycolic acid, Geophys. Res. Lett., 36, doi: 10.1029/2008GL036212, 2009.

Jenkin, M. E., Saunders, S. M. Wagner, V., and Pilling, M. J.: The tropospheric degradation of volatile organic compounds: a protocol for mechanism development, Atmos. Environ., 31, 81-104, doi:10.1016/S1352-2310(96)00105-7, 1997.

Jenkin, M. E., Saunders, S. M., Wagner, V., and 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. and Phys., 3, 181-193, doi:10.5194/acp-3-181-2003, 2003.

Jenkin, M. E., Young, J. C., and Rickard, A. R.: The MCM v3.3.1 degradation scheme for isoprene, Atmos. Chem. Phys., 15, 11433-11459, doi: 10.5194/acp-15-11433-2015, 2015.

Kim, S. W., McKeen, S. A., Frost, G. J., Lee, S. H., Trainer, M., Richter, A., Angevine, W. M., Atlas, E., Bianco, L., Boersma, K. F., Brioude, J., Burrows, J. P., de Gouw, J., Fried, A., Gleason, J., Hilboll, A., Mellqvist, J., Peischl, J., Richter, D., Rivera, C., Ryerson, T., Hekkert, S. T. L., Walega, J., Warneke, C.,

Weibring, P., and Williams, E.: Evaluations of NOx and highly reactive VOC emission inventories in Texas and their implications for ozone plume simulations during the Texas Air Quality Study 2006, Atmos. Chem. Phys., 11, 11361-11386, doi: 10.5194/acp-11-11361-2011, 2011.

Kleipool, Q. L., Dobber, M. R., de Haan, J. F., and Levelt, P. F.: Earth surface reflectance climatology from 3 years of OMI data, J. Geophys. Res. Atmos., 113, doi: 10.1029/2008JD010290, 2008.

Kleipool, Q. L.: Transient signal flagging algorithm definition for radiance data, Tech. Rep. TN-OMIE-KNMI-717 TN-OMIEKNMI-717 TN-OMIE-KNMI-717 TN-OMIE-KNMI-717 TNOMIE-KNMI-717, Royal Netherlands Meteorological Institute, De Bilt, the Netherlands, 2005.

Kurokawa, J., Ohara, T., Morikawa, T., Hanayama, S., Janssens-Maenhout, G., Fukui, T., Kawashima, K., and Akimoto, H.: Emissions of air pollutants and greenhouse gases over Asian regions during 2000-2008: Regional Emission inventory in ASia (REAS) version 2, Atmos. Chem. Phys., 13, 11019-11058, doi:10.5194/acp-13-11019-2013, 2013.

Kurosu, T. P., Chance, K., and Sioris, C. E.: Preliminary results for HCHO and BrO from the EOS Aura Ozone Monitoring Instrument,Conference on Passive Optical Remote Sensing of the Atmosphere and Clouds IV, 5652, 116, doi: 10.1117/12.578606, 2004.

Kurosu, T. P., Chance, K., Liu, X., Volkamer, R., Fu, T. M., Millet, D., and Jacob, D. J.: Seasonally resolved global distributions of glyoxal and formaldehyde observed from the Ozone Monitoring Instrument on EOS Aura, Simpósio Brasileiro de Sensoriamento Remoto 13 (SBSR): 6461-6464, 2007.

Lee, H., Ryu, J., Irie, H., Jang, S.-H., Park, J., Choi, W., and Hong, H.: Investigations of the Diurnal Variation of Vertical HCHO Profiles Based on MAX-DOAS Measurements in Beijing: Comparisons with OMI Vertical Column Data, Atmosphere, 6, 1816-1832, doi: 10.3390/atmos6111816, 2015.

Lelieveld, J., Butler, T. M., Crowley, J. N., Dillon, T. J., Fischer, H., Ganzeveld, L., Harder, H., Lawrence, M. G., Martinez, M., Taraborrelli, D., and Williams, J.: Atmospheric oxidation capacity sustained by a tropical forest, Nature, 452, 737-740, doi: 10.1038/nature06870, 2008.

Lerot, C., Stavrakou, T., De Smedt, I., Muller, J. F., and Van Roozendael, M.: Glyoxal vertical columns from GOME-2 backscattered light measurements and comparisons with a global model, Atmos. Chem. Phys., 10, 12059-12072, doi: 10.5194/acp-10-12059-2010, 2010.

Li, D., and Bian, J. C.: Observation of a Summer Tropopause Fold by Ozonesonde at Changchun, China: Comparison with Reanalysis and Model Simulation, Adv. Atmos. Sci., 32, 1354-1364, doi: 10.1007/s00376-015-5022-x, 2015.

Li, J., Wang, Z. F., Akimoto, H., Gao, C., Pochanart, P., and Wang, X. Q.: Modeling study of ozone seasonal cycle in lower troposphere over east Asia, J. Geophys. Res.Atmos., 112, doi: 10.1029/2006JD008209, 2007.

Li, J. Y., Mao, J. Q., Min, K. E., Washenfelder, R. A., Brown, S. S., Kaiser, J., Keutsch, F. N., Volkamer, R., Wolfe, G. M., Hanisco, T. F., Pollack, I. B., Ryerson, T. B., Graus, M., Gilman, J. B., Lerner, B. M.,

Warneke, C., de Gouw, J. A., Middlebrook, A. M., Liao, J., Welti, A., Henderson, B. H., McNeill, V. F., Hall, S. R., Ullmann, K., Donner, L. J., Paulot, F., and Horowitz, L. W.: Observational constraints on glyoxal production from isoprene oxidation and its contribution to organic aerosol over the Southeast United States, J. Geophys. Res. Atmos., 121, 9849-9861, doi: 10.1002/2016JD025331, 2016.

Li, M., Zhang, Q., Streets, D. G., He, K. B., Cheng, Y. F., Emmons, L. K., Huo, H., Kang, S. C., Lu, Z., Shao, M., Su, H., Yu, X., and Zhang, Y.: Mapping Asian anthropogenic emissions of non-methane volatile organic compounds to multiple chemical mechanisms, Atmos. Chem. Phys., 14, 5617-5638, doi:10.5194/acp-14-5617-2014, 2014.

Li, M., Zhang, Q., Kurokawa, J. I., Woo, J. H., He, K., Lu, Z., Ohara, T., Song, Y., Streets, D. G., Carmichael, G. R., Cheng, Y., Hong, C., Huo, H., Jiang, X., Kang, S., Liu, F., Su, H., and Zheng, B.: MIX: a mosaic Asian anthropogenic emission inventory under the international collaboration framework of the MICS-Asia and HTAP, Atmos. Chem. Phys., 17, 935-963, doi:10.5194/acp-17-935-2017, 2017.

Li, X., Brauers, T., Hofzumahaus, A., Lu, K., Li, Y. P., Shao, M., Wagner, T., and Wahner, A.: MAX-DOAS measurements of NO2, HCHO and CHOCHO at a rural site in Southern China, Atmos. Chem. Phys., 13, 2133-2151, doi: 10.5194/acp-13-2133-2013, 2013.

Liao, H., Henze, D. K., Seinfeld, J. H., Wu, S. L., and Mickley, L. J.: Biogenic secondary organic aerosol over the United States: Comparison of climatological simulations with observations, Journal of Geophysical Research, 112, doi:10.1029/2006JD007813, 2007.

Liggio, J., Li, S. M., and McLaren, R.: Reactive uptake of glyoxal by particulate matter, J. Geophys. Res. Atmos., 110, doi: 10.1029/2004JD005113, 2005.

Liu, M., Song, Y., Yao, H., Kang, Y., Li, M., Huang, X., and Hu, M.: Estimating emissions from agricultural fires in the North China Plain based on MODIS fire radiative power, Atmos. Environ., 112, 326-334, doi: 10.1016/j.atmosenv.2015.04.058, 2015.

Liu, Z., Wang, Y., Vrekoussis, M., Richter, A., Wittrock, F., Burrows, J. P., Shao, M., Chang, C.-C., Liu, S.-C., Wang, H., and Chen, C.: Exploring the missing source of glyoxal (CHOCHO) over China, Geophys. Res. Lett., 39, doi:10.1029/2012gl051645, 2012.

Mao, J. Q., Paulot, F., Jacob, D. J., Cohen, R. C., Crounse, J. D., Wennberg, P. O., Keller, C. A., Hudman, R. C., Barkley, M. P., and Horowitz, L. W.: Ozone and organic nitrates over the eastern United States: Sensitivity to isoprene chemistry, J. Geophys. Res. Atmos., 118, 11256-11268, doi: 10.1002/jgrd.50817, 2013.

Marais, E. A., Jacob, D. J., Kurosu, T. P., Chance, K., Murphy, J. G., Reeves, C., Mills, G., Casadio, S., Millet, D. B., Barkley, M. P., Paulot, F., and Mao, J.: Isoprene emissions in Africa inferred from OMI observations of formaldehyde columns, Atmos. Chem. Phys., 12, 6219-6235, doi: 10.5194/acp-12-6219-2012, 2012.

Marais, E. A., Jacob, D. J., Guenther, A., Chance, K., Kurosu, T. P., Murphy, J. G., Reeves, C. E., and Pye, H. O. T.: Improved model of isoprene emissions in Africa using Ozone Monitoring Instrument

(OMI) satellite observations of formaldehyde: implications for oxidants and particulate matter, Atmos. Chem. Phys., 14, 7693-7703, doi: 10.5194/acp-14-7693-2014, 2014a.

Marais, E. A., Jacob, D. J., Wecht, K., Lerot, C., Zhang, L., Yu, K., Kurosu, T. P., Chance, K., and Sauvage, B.: Anthropogenic emissions in Nigeria and implications for atmospheric ozone pollution: A view from space, Atmos. Environ., 99, 32-40, doi: 10.1016/j.atmosenv.2014.09.055, 2014b.

Mijling, B., van der A, R. J., and Zhang, Q.: Regional nitrogen oxides emission trends in East Asia observed from space, Atmospheric Chemistry and Physics, 13, 12003-12012, doi: 10.5194/acp-13-12003-2013, 2013.

Millet, D. B., Jacob, D. J., Turquety, S., Hudman, R. C., Wu, S., Fried, A., Walega, J., Heikes, B. G., Blake, D. R., Singh, H. B., Anderson, B. E., and Clarke, A. D.: Formaldehyde distribution over North America: Implications for satellite retrievals of formaldehyde columns and isoprene emission, J. Geophys. Res., 111, doi: 10.1029/2005jd006853, 2006.

Millet, D. B., Jacob, D. J., Boersma, K. F., Fu, T.-M., Kurosu, T. P., Chance, K., Heald, C. L., and Guenther, A.: Spatial distribution of isoprene emissions from North America derived from formaldehyde column measurements by the OMI satellite sensor, J. Geophys. Res., 113, doi: 10.1029/2007jd008950, 2008.

~~Monks, P. S.: Gas-phase radical chemistry in the troposphere, Chem. Soc. Rev., 34, 376-395, doi:10.1039/b307982c, 2005.~~

Myriokefalitakis, S., Vrekoussis, M., Tsigaridis, K., Wittrock, F., Richter, A., Bruehl, C., Volkamer, R., Burrows, J. P., and Kanakidou, M.: The influence of natural and anthropogenic secondary sources on the glyoxal global distribution, Atmos. Chem. Phys., 8, 4965-4981, doi: 10.5194/acp-8-4965-2008, 2008.

Nishino, N., Arey, J., and Atkinson, R.: Formation Yields of Glyoxal and Methylglyoxal from the Gas-Phase OH Radical-Initiated Reactions of Toluene, Xylenes, and Trimethylbenzenes as a Function of NO2 Concentration, J. Phys. Chem. A, 114, 10140, doi: 10.1021/jp105112h, 2010.

Olivier, J. G. J., Bouwman, A. F., Berdowski, J. J. M., Veldt, C., Bloos, J. P. J., Visschedijk, A. J. H., van der Maas, C. W. M., and Zandveld, P. Y. J.: Sectoral emission inventories of greenhouse gases for 1990 on a per country basis as well as on 1°×1°, Environ. Sci. Policy, 2, 241-263, doi:10.1016/S1462-9011(99)00027-1, 1999.

Olivier, J. G. J., Berdowski, J. J. M., Peters, J. A. H. W., Bakker, J., Visschedijk, A. J. H., and Bloos, J.-P. J.: Applications of EDGAR, Including a description of EDGAR 3.0: reference database with trend data for 1970–1995, RIVM report no. 773301 001/ NOP report no. 410200 051, RIVM, Bilthoven, 2001.

Olivier, J. G. J.: Part III: Greenhouse gas emissions. 1. Shares and trends in greenhouse gas emissions; 2. Sources and methods: greenhouse gas emissions for 1990 and 1995 in “CO2 emissions from fuel combustion 1971–2000”, International Energy Agency, Paris, ISBN 92-64-09794-5, 1–31, 1–31, 2002.

Palmer, P. I., Jacob, D. J., Fiore, A. M., Martin, R. V., Chance, K., and Kurosu, T. P.: Mapping isoprene emissions over North America using formaldehyde column observations from space, J. Geophys. Res. Atmos., 108, doi: 10.1029/2002jd002153, 2003.

Palmer, P. I., Abbot, D. S., Fu, T.-M., Jacob, D. J., Chance, K., Kurosu, T. P., Guenther, A., Wiedinmyer, C., Stanton, J. C., Pilling, M. J., Pressley, S. N., Lamb, B., and Sumner, A. L.: Quantifying the seasonal and interannual variability of North American isoprene emissions using satellite observations of the formaldehyde column, J. Geophys. Res., 111, doi: 10.1029/2005jd006689, 2006.

Paulot, F., Crounse, J. D., Kjaergaard, H. G., Kroll, J. H., Seinfeld, J. H., and Wennberg, P. O.: Isoprene photooxidation: new insights into the production of acids and organic nitrates, Atmos. Chem. Phys., 9, 1479-1501, doi: 10.5194/acp-9-1479-2009, 2009a.

Paulot, F., Crounse, J. D., Kjaergaard, H. G., Kurten, A., St Clair, J. M., Seinfeld, J. H., and Wennberg, P. O.: Unexpected Epoxide Formation in the Gas-Phase Photooxidation of Isoprene, Science, 325, 730-733, doi: 10.1126/science.1172910, 2009b.

Platt, U., Perner, D., ~~auml,~~ and ~~tz~~Pätz, H. W.: Simultaneous measurement of atmospheric CH2O, O3, and NO2 by differential optical absorption, J. Geophys. Res. Oceans, 84, 6329-6335, doi:10.1029/JC084iC10p06329, 1979.

Pye, H. O. T., and Seinfeld, J. H.: A global perspective on aerosol from low-volatility organic compounds, Atmos. Chem. Phys., 10, 4377-4401, doi:10.5194/acp-10-4377-2010.

Qiu, K., Yang, L., Lin, J., Wang, P., Yang, Y., Ye, D., and Wang, L.: Historical industrial emissions of non-methane volatile organic compounds in China for the period of 1980–2010, Atmos. Environ., 86, 102-112, doi:10.1016/j.atmosenv.2013.12.026, 2014.

Qu, Z., Henze, D. K., Capps, S. L., Wang, Y., Xu, X. G., Wang, J., and Keller, M.: Monthly top-down NOx emissions for China (2005-2012): A hybrid inversion method and trend analysis, J. Geophy. Res. Atmos., 122, 4600-4625, doi: 10.1002/2016JD025852, 2017.

Robinson, A. L., Donahue, N. M., Shrivastava, M. K., Weitkamp, E. A., Sage, A. M., Grieshop, A. P., Lane, T. E., Pierce, J. R., and Pandis, S. N.: Rethinking organic aerosols: Semivolatile emissions and photochemical aging, Science, 315, 1259-1262, doi: 10.1126/science.1133061 2007.

Rodgers, C. D.: Inverse methods for atmospheric sounding: theory and practice, World Scientific, Singapore, 2000.

~~Rubin, J. I., Kean, A. J., Harley, R. A., Millet, D. B., and Goldstein, A. H.: Temperature dependence of volatile organic compound evaporative emissions from motor vehicles, J. Geophys. Res. Atmos., 111, doi:10.1029/2005JD006458, 2006.~~

Sandu, A., Daescu, D. N., and Carmichael, G. R.: Direct and adjoint sensitivity analysis of chemical kinetic systems with KPP: Part I - theory and software tools, Atmos. Environ., 37, 5083-5096, doi:10.1016/j.atmosenv.2003.08.019, 2003.

Saunders, S. M., Jenkin, M. E., Derwent, R. G., and Pilling, M. J.: Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part A): tropospheric degradation of non-aromatic volatile organic compounds, Atmos. Chem. Phys., 3, 181-193, doi: 10.5194/acp-3-161-2003, 2003.

Schultz, M. G., Backman, L., and Balkanski, Y.: REanalysis of the TROpospheric chemical composition over the past 40 years (RETRO): A long-term global modeling study of tropospheric chemistry, J¨ulich/Hamburg, Germany, 48/2007 report on Earth System Science of the Max Planck Institute for Meteorology, Hamburg, http://retro.enes.org, ISSN 1614-1199, 2007.

Shim, C., Wang, Y., Choi, Y., Palmer, P. I., Abbot, D. S., and Chance, K.: Constraining global isoprene emissions with Global Ozone Monitoring Experiment (GOME) formaldehyde column measurements, J. Geophys. Res., 110, doi: 10.1029/2004jd005629, 2005.

Sindelarova, K., Granier, C., Bouarar, I., Guenther, A., Tilmes, S., Stavrakou, T., Muller, J. F., Kuhn, U., Stefani, P., and Knorr, W.: Global data set of biogenic VOC emissions calculated by the MEGAN model over the last 30 years, Atmospheric Chemistry and Physics, 14, 9317-9341, doi:10.5194/acp-14-9317-2014, 2014.

Spurr, R.: LIDORT and VLIDORT: Linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems, in: Light Scattering Reviews, edited by: Kokhanovsky, A., Springer, 3, 229–275, 2008.

Spurr, R. J. D.: VLIDORT: A linearized pseudo-spherical vector discrete ordinate radiative transfer code for forward model and retrieval studies in multilayer multiple scattering media, J. Quant. Spectrosc. Radiat. Transf., 102, 316-342, doi: 10.1016/j.jqsrt.2006.05.005, 2006.

Stavrakou, T., Guenther, A., Razavi, A., Clarisse, L., Clerbaux, C., Coheur, P. F., Hurtmans, D., Karagulian, F., De Maziere, M., Vigouroux, C., Amelynck, C., Schoon, N., Laffineur, Q., Heinesch, B., Aubinet, M., Rinsland, C., and Muller, J. F.: First space-based derivation of the global atmospheric methanol emission fluxes, Atmos. Chem. Phys., 11, 4873-4898, doi: 10.5194/acp-11-4873-2011, 2011.

Stavrakou, T., Muller, J. F., ~~Muller, J. F.,~~ Bauwens, M., De Smedt, I.: Sources and long-term trends of ozone precursors to Asian Pollution, Air Pollution in Eastern Asia : an integrated perspective, eds. Bouarar, I.,Wang, X., Brasseur, G., Springer international Publishing, doi:10.1007/978-3-319-59489-7-8, pp. 167-189, 2017.

Stavrakou, T., Muller, J. F., Bauwens, M., De Smedt, I., Lerot, C., Van Roozendael, M., Coheur, P. F., Clerbaux, C., Boersma, K. F., van der, A. R., and Song, Y.: Substantial Underestimation of Post-Harvest Burning Emissions in the North China Plain Revealed by Multi-Species Space Observations, Sci. Rep., 6, 32307, doi: 10.1038/srep32307, 2016.

Stavrakou, T., Muller, J. F., Bauwens, M., De Smedt, I., Van Roozendael, M., Guenther, A., Wild, M., and Xia, X.: Isoprene emissions over Asia 1979-2012: impact of climate and land-use changes, Atmos. Chem. and Phys., 14, 4587-4605, doi: 10.5194/acp-14-4587-2014, 2014.

Stavrakou, T., Müller, J. F., Bauwens, M., De Smedt, I., Van Roozendael, M., De Mazière, M.,

Vigouroux, C., Hendrick, F., George, M., Clerbaux, C., Coheur, P. F., and Guenther, A.: How consistent are top-down hydrocarbon emissions based on formaldehyde observations from GOME-2 and OMI?, Atmos. Chem. Phys., 15, 11861-11884, doi: 10.5194/acp-15-11861-2015, 2015.

Stavrakou, T., Muller, J. F., De Smedt, I., Van Roozendael, M., Kanakidou, M., Vrekoussis, M., Wittrock, F., Richter, A., and Burrows, J. P.: The continental source of glyoxal estimated by the synergistic use of spaceborne measurements and inverse modelling, Atmos. Chem. Phys., 9, 8431-8446, doi: 10.5194/acp-9-8431-2009, 2009a.

Stavrakou, T., Muller, J. F., De Smedt, I., Van Roozendael, M., van der Werf, G. R., Giglio, L., and Guenther, A.: Global emissions of non-methane hydrocarbons deduced from SCIAMACHY formaldehyde columns through 2003-2006, Atmos. Chem. Phys., 9, 3663-3679, doi:10.5194/acp-9-3663-2009, 2009b.

~~Stavrakou, T., Müller, J. F., Bauwens, M., De Smedt, I., Van Roozendael, M., De Mazière, M., Vigouroux, C., Hendrick, F., George, M., Clerbaux, C., Coheur, P. F., and Guenther, A.: How consistent are top-down hydrocarbon emissions based on formaldehyde observations from GOME-2 and OMI?, Atmos. Chem. Phys., 15, 11861-11884, doi: 10.5194/acp-15-11861-2015, 2015.~~

~~Stavrakou, T., Muller, J. F., Bauwens, M., De Smedt, I., Lerot, C., Van Roozendael, M., Coheur, P. F., Clerbaux, C., Boersma, K. F., van der, A. R., and Song, Y.: Substantial Underestimation of Post-Harvest Burning Emissions in the North China Plain Revealed by Multi-Species Space Observations, Sci. Rep., 6, 32307, doi: 10.1038/srep32307, 2016.~~

Sun, L., Xue, L. K., Wang, T., Gao, J., Ding, A. J., Cooper, O. R., Lin, M. Y., Xu, P. J., Wang, Z., Wang, X. F., Wen, L., Zhu, Y. H., Chen, T. S., Yang, L. X., Wang, Y., Chen, J. M., and Wang, W. X.: Significant increase of summertime ozone at Mount Tai in Central Eastern China, Atmos. Chem. Phys., 16, 10637-10650, doi: 10.5194/acp-16-10637-2016, 2016.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y., and van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), Atmos. Chem. Phys., 10, 11707-11735, doi:10.5194/acp-10-11707-2010, 2010.

van der Werf, G. R., Randerson, J. T., Giglio, L., van Leeuwen, T. T., Chen, Y., Rogers, B. M., Mu, M. Q., van Marle, M. J. E., Morton, D. C., Collatz, G. J., Yokelson, R. J., and Kasibhatla, P. S.: Global fire emissions estimates during 1997-2016, Earth Syst Sci Data, 9, 697-720, doi: 10.5194/essd-9-697-2017, 2017.

van Donkelaar, A., Martin, R. V., Leaitch, W. R., Macdonald, A. M., Walker, T. W., Streets, D. G., Zhang, Q., Dunlea, E. J., Jimenez, J. L., Dibb, J. E., Huey, L. G., Weber, R., and Andreae, M. O.: Analysis of aircraft and satellite measurements from the Intercontinental Chemical Transport Experiment (INTEX-B) to quantify long-range transport of East Asian sulfur to Canada, Atmos. Chem. Phys., 8, 2999-3014, doi: 10.5194/acp-8-2999-2008, 2008.

Vestreng, V.: Review and revision. Emission data reported to CLRTAP, Tech. rep., EMEP MSC-W,

(available at: http://www.emep.int/mscw/mscw_publications.html#2003), 2003.

Vlemmix, T., Hendrick, F., Pinardi, G., Smedt, I., De Fayt, C., Hermans, C., Piters, A., Wang, P., and Levelt, P.: MAX-DOAS observations of aerosols, formaldehyde and nitrogen dioxide in the Beijing area: comparison of two profile retrieval, Atmos. Meas. Tech., 2, 941–963, doi:10.5194/amt-8-941-2015, 2015.

Volkamer, R.: Primary and Secondary Glyoxal Formation from Aromatics: Experimental Evidence for the Bicycloalkyl-Radical Pathway from Benzene, Toluene, and p-Xylene, J. Phys. Chem., 105, 7865, doi:10.1021/jp010152w, 2001.

Vrekoussis, M., Wittrock, F., Richter, A., and Burrows, J. P.: GOME-2 observations of oxygenated VOCs: what can we learn from the ratio glyoxal to formaldehyde on a global scale?, Atmos. Chem. Phys., 10, 10145-10160, doi: 10.5194/acp-10-10145-2010, 2010.

Wang, F., An, J., Li, Y., Tang, Y., Lin, J., Qu, Y., Chen, Y., Zhang, B., and Zhai, J.: Impacts of uncertainty in AVOC emissions on the summer ROx budget and ozone production rate in the three most rapidly-developing economic growth regions of China, Adv. Atmos. Sci., 31, 1331-1342, doi: 10.1007/s00376-014-3251-z, 2014.

Wang, H. Q., Ma, J. M., Shen, Y. J., and Wang, Y. A.: Assessment of Ozone Variations and Meteorological Influences at a Rural Site in Northern Xinjiang, Bull. Environ. Contam. Toxicol., 94, 240-246, doi: 10.1007/s00128-014-1451-y, 2015.

Wang, P., Stammes, P., R., v. d. A., Pinardi, G., and Roozendael, M. V.: FRESCO+: an improved O2 A-band cloud retrieval algorithm for tropospheric trace gas retrievals, Atmos. Chem. Phys., 8, 6565-6576, doi: 10.5194/acp-8-6565-2008, 2008.

Wang, Y., Konopka, P., Liu, Y., Chen, H., Muller, R., Ploger, F., Riese, M., Cai, Z., and Lu, D.: Tropospheric ozone trend over Beijing from 2002-2010: ozonesonde measurements and modeling analysis, Atmos. Chem. Phys., 12, 8389-8399, doi: 10.5194/acp-12-8389-2012, 2012.

Wang, Y., Beirle, S., Lampel, J., Koukouli, M., De Smedt, I., Theys, N., Li, A., Wu, D. X., Xie, P. H., Liu, C., Van Roozendael, M., Stavrakou, T., Muller, J. F., and Wagner, T.: Validation of OMI, GOME-2A and GOME-2B tropospheric NO2, SO2 and HCHO products using MAX-DOAS observations from 2011 to 2014 in Wuxi, China: investigation of the effects of priori profiles and aerosols on the satellite products, Atmos. Chem. Phys., 17, 5007-5033, doi: 10.5194/acp-17-5007-2017, 2017.

Wei, W., ~~Lv, Z., F., Yang, G., Cheng, S. Y., Li, Y., and Wang, L. T.: VOCs emission rate estimate for complicated industrial area source using an inverse-dispersion calculation method: A case study on a petroleum refinery in Northern China, Environ. Pollut.,doi: 10.1016/j.envpol.2016.07.062, 218, 681-688, 2016.~~

~~Wei, W.,~~ Wang, S., Chatani, S., Klimont, Z., Cofala, J., and Hao, J.: Emission and speciation of non-methane volatile organic compounds from anthropogenic sources in China, Atmos. Environ., 42, 4976-4988, doi:10.1016/j.atmosenv.2008.02.044, 2008.

Wells, K. C., Millet, D. B., Hu, L., Cady-Pereira, K. E., Xiao, Y., Shephard, M. W., Clerbaux, C. L., Clarisse, L., Coheur, P. F., Apel, E. C., de Gouw, J., Warneke, C., Singh, H. B., Goldstein, A. H., and Sive, B. C.: Tropospheric methanol observations from space: retrieval evaluation and constraints on the seasonality of biogenic emissions, Atmos. Chem. Phys., 12, 5897-5912, doi: 10.5194/acp-12-5897-2012, 2012.

Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J., and Soja, A. J.: The Fire INventory from NCAR (FINN): a high resolution global model to estimate the emissions from open burning, Geosci. Model Dev., 4, 625-641, doi: 10.5194/gmd-4-625-2011, 2011.

Wittrock, F., Richter, A., Oetjen, H., Burrows, J. P., Kanakidou, M., Myriokefalitakis, S., Volkamer, R., Beirle, S., Platt, U., and Wagner, T.: Simultaneous global observations of glyoxal and formaldehyde from space, Geophys. Res. Lett., 33, doi: 10.1029/2006gl026310, 2006.

Wu, R., Bo, Y., Li, J., Li, L., Li, Y., and Xie, S.: Method to establish the emission inventory of anthropogenic volatile organic compounds in China and its application in the period 2008–2012, Atmos. Environ., 127, 244-254, doi:10.1016/j.atmosenv.2015.12.015, 2016.

Xu, W. Y., Lin, W. L., Xu, X. B., Tang, J., Huang, J. Q., Wu, H., and Zhang, X. C.: Long-term trends of surface ozone and its influencing factors at the Mt Waliguan GAW station, China - Part 1: Overall trends and characteristics, Atmos. Chem. Phys., 16, 6191-6205, doi: 10.5194/acp-16-6191-2016, 2016.

Xu, X., Lin, W., Wang, T., Yan, P., Tang, J., Meng, Z., and Wang, Y.: Long-term trend of surface ozone at a regional background station in eastern China 1991-2006: enhanced variability, Atmos. Chem. Phys., 8, 2595-2607, doi: 10.5194/acp-8-2595-2008, 2008.

Zhang, J. M., Wang, T., Ding, A. J., Zhou, X. H., Xue, L. K., Poon, C. N., Wu, W. S., Gao, J., Zuo, H. C., Chen, J. M., Zhang, X. C., and Fan, S. J.: Continuous measurement of peroxyacetyl nitrate (PAN) in suburban and remote areas of western China, Atmos. Environ., 43, 228-237, doi: 10.1016/j.atmosenv.2008.09.070, 2009.

Zhang, Q., Streets, D. G., Carmichael, G. R., He, K. B., Huo, H., Kannari, A., Klimont, Z., Park, I. S., Reddy, S., Fu, J. S., Chen, D., Duan, L., Lei, Y., Wang, L. T., and Yao, Z. L.: Asian emissions in 2006 for the NASA INTEX-B mission, Atmos. Chem. Phys., 9, 5131-5153, 10.5194/acp-9-5131-2009, 2009.

Zhang, X. Y., Wang, Y. Q., Niu, T., Zhang, X. C., Gong, S. L., Zhang, Y. M., and Sun, J. Y.: Atmospheric aerosol compositions in China: spatial/temporal variability, chemical signature, regional haze distribution and comparisons with global aerosols, Atmospheric Chemistry and Physics, 12, 779-799, doi:10.5194/acp-12-779-2012, 2012.

Zhao, Y., Nielsen, C. P., Lei, Y., McElroy, M. B., and Hao, J.: Quantifying the uncertainties of a bottom-up emission inventory of anthropogenic atmospheric pollutants in China, Atmos. Chem. Phys., 11, 2295-2308, doi:10.5194/acp-11-2295-2011, 2011.

Zheng, J. Y., Zhong, L. J., Wang, T., Louie, P. K. K., and Li, Z. C.: Ground-level ozone in the Pearl River Delta region: Analysis of data from a recently established regional air quality monitoring network, Atmos. Environ., 44, 814-823, doi: 10.1016/j.atmosenv.2009.11.032, 2010.

Zhu, ~~C. Y., Byrd, R. H., Lu, P. H., and Nocedal, J.: Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization, ACM T. Math. Software, 23, 550-560, doi: 10.1145/279232.279236, 1997.~~

~~Zhu,~~ L., Jacob, D. J., Mickley, L. J., Marais, E. A., Cohan, D. S., Yoshida, Y., Duncan, B. N., González Abad, G., and Chance, K. V.: Anthropogenic emissions of highly reactive volatile organic compounds in eastern Texas inferred from oversampling of satellite (OMI) measurements of HCHO columns, Environ. Res. Lett., 9, 114004, doi: 10.1088/1748-9326/9/11/114004, 2014.

Zhu, L., Jacob, D. J., Kim, P. S., Fisher, J. A., Yu, K., Travis, K. R., Mickley, L. J., Yantosca, R. M., Sulprizio, M. P., De Smedt, I., González Abad, G., Chance, K., Li, C., Ferrare, R., Fried, A., Hair, J. W., Hanisco, T. F., Richter, D., Jo Scarino, A., Walega, J., Weibring, P., and Wolfe, G. M.: Observing atmospheric formaldehyde (HCHO) from space: validation and intercomparison of six retrievals from four satellites (OMI, GOME2A, GOME2B, OMPS) with SEAC^{4}RS aircraft observations over the southeast US, Atmos. Chem. Phys., 16, 13477-13490, doi: 10.5194/acp-16-13477-2016, 2016.

Formatted: Normal

Formatted: Font: 9 pt, Bold

Table 1 Inversion experiments to constrain Chinese NMVOC emissions

Inversion experiments	Observational constraints from satellites [±uncertainties]	Annual Chinese NMVOC emission estimates [Tg y^{-1}]			
		Anthropogenic	Biogenic	Biomass burning	Total
		A priori emission estimates [uncertainty]			
		18.8 (5.4 for aromatics) [a] [factor of two uncertainty]	17.3 (7.5 for isoprene) [b] [±55% uncertaitny	2.27 [factor of three uncertainty] [c]	38.3
		A posteriori emission estimates [range of estimates]			
IE-1	GOME-2A formaldehyde [±90%] and glyoxal [±150%]	17.8 (5.8 for aromatics)	20.0 (9.8 for isoprene)	2.27	40.1
IE-2	OMI formaldehyde [±90%] and glyoxal [±150%]	16.4 (5.5 for aromatics)	12.2 (5.4 for isoprene)	2.08	30.7
IE-3	GOME-2A formaldehyde × 170% [±90%]	23.6 (6.6 for aromatics)	22.8 (11.3 for isoprene)	3.13	49.5
IE-4	OMI glyoxal [±150%]	23.0 (7.9 for aromatics)	21.6 (11.7 for isoprene)	2.43	47.0
Our top-down estimates		20.2 [d] [16.4 - 23.6] (6.5 [d] [5.5 - 7.9] for aromatics)	19.2 [d] [12.2 – 22.8] (9.6 [d] [5.4 – 11.7] for isoprene)	2.48 [d] [2.08 – 3.13]	41.9 [d] [30.7 – 49.5]

[a] From Li et al. (2017)

[b] From Guenther et al. (2006).

[c] Compiled from the emission estimated by van der Werf et al. (2010) plus a scaling of the emission estimated by Huang et al. (2012). See text (section 2.2) for details.

[d] Average of top-down estimates from the four inversion experiments.

Formatted: Normal

1407 **Table 2 Comparison of Chinese annual NMVOC emission estimates for the years 2000 to 2014**

Literature	Target year	NMVOC [Tg y^{-1}]				
		Anthropogenic		Biogenic		Biomass burning
		Total	Aromatics	Total	Isoprene	
Bottom-up estimates						
Bo et al. (2008) [a]	2005	12.7				3.8 [d]
Zhang et al. (2009) [a]	2006	23.2 (± 68%)	2.4			
Cao et al. (2011) [a]	2007	35.46				
Huang et al. (2017) [a]	2007	24.6				
Granier et al. (2017) [a]	2007	29.0				
Kurokawa et al. (2013) [a]	2008	27.1 (± 46%)				
Li et al. (2017) [a]	2010	23.6	5.4			
Wu et al. (2016) [a]	2008	18.62				3.83 [d]
	2009	21.8				3.32 [d]
	2010	23.83				3.75 [d]
	2011	24.78				3.76 [d]
	2012	25.65				4.20 [d]
Huang et al. (2012) [a]	2006					2.2 (1.08 to 3.46)
van der Werf et al. (2010)	2007					0.47
van der Werf et al. (2017) [a]	2007					0.91
Sindelarova et al. (2014)	2005				9.9	
Guenther et al.(2006)	2007			17.3 [e]	7.5 [e]	
Stavrakou et al. (2014)	2007				7.6	
Top-down estimates						
Fu et al. (2007)	2000	4.27 [g]		12.7		5.1
Liu et al. (2012) [b]	2007	34.2	13.4			
Stavrakou et al. (2014)	2007				8.6	
Stavrakou et al. (2015) [c]	2010	20.6 to 24.6			5.9 to 6.5	2.0 to 2.7
Stavrakou et al. (2017) [c]	2005	24.4			5.8 (average of emissions from 2005 to 2014)	
	2006	24.0				
	2007	26.7				
	2008	25.9				
	2009	26.5				
	2010	26.1				
	2011	25.5				
	2012	25.6				
	2013	27.7				
	2014	27.8				
This work	2007	20.2 [f] (16.4 - 23.6)	6.5 [f] (5.5 - 7.9)	19.2 [f] (12.2 - 22.8)	9.6 [f] (5.4 - 11.7)	2.48 [f] (2.08 – 3.13)

[a] These emission estimates included some NMVOC species which were not precursors to formaldehyde or glyoxal and therefore not included in this work. See color keys in Figure 2 for NMVOC species whose emissions were included in this work.

[b] Used SCIAMACHY-observed glyoxal VCDs as constraints.

[c] Used GOME-2A-observed and OMI-observed formaldehyde VCDs as constraints.

[d] Consisted of emissions from open burning of crop residues and from biofuel burning.

[e] Calculated by the GEOS-Chem model using GEOS-5 meteorological data.

[f] Average of top-down estimates from four inversion experiments.

[g] Only anthropogenic emissions of reactive alkenes, formaldehyde, and xylenes from northeastern, northern, central and southern China were included

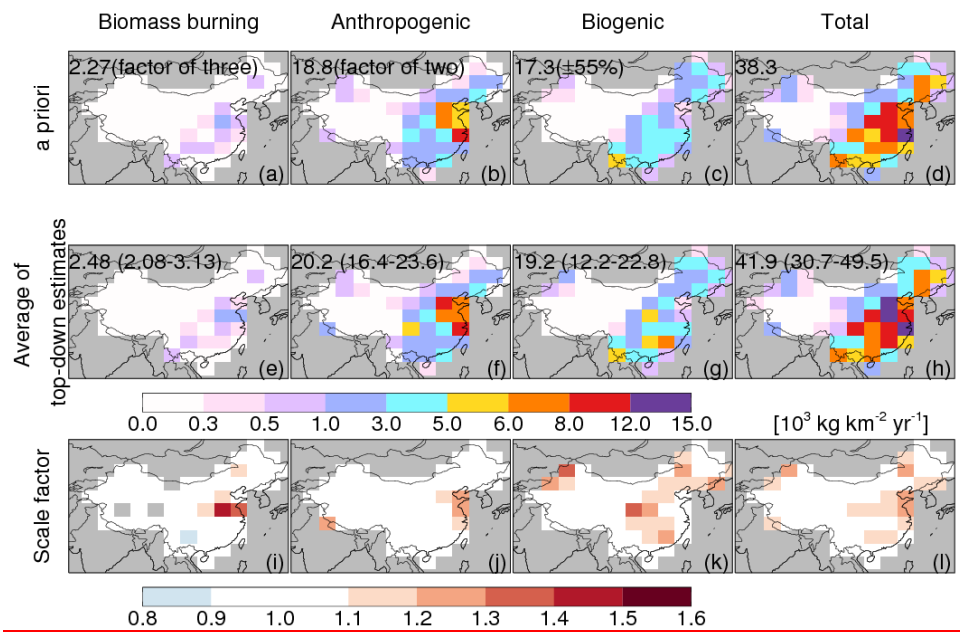


Figure 1. Spatial distributions of annual NMVOC emissions from China. (a)-(d): The *a priori* annual NMVOC emission estimates from (a) biomass burning, (b) anthropogenic, (c) biogenic, and (d) total sources. (e)-(h): The average of our four sets of top-down estimates of annual NMVOC emissions. Annual Chinese total emission estimates are shown inset in units of [Tg y^{-1}]. The uncertainties of the *a priori* emission estimates and the range of top-down emission estimates are shown in parentheses. (i)-(l): Scale factors for our averaged top-down estimates relative to the *a priori* estimates.

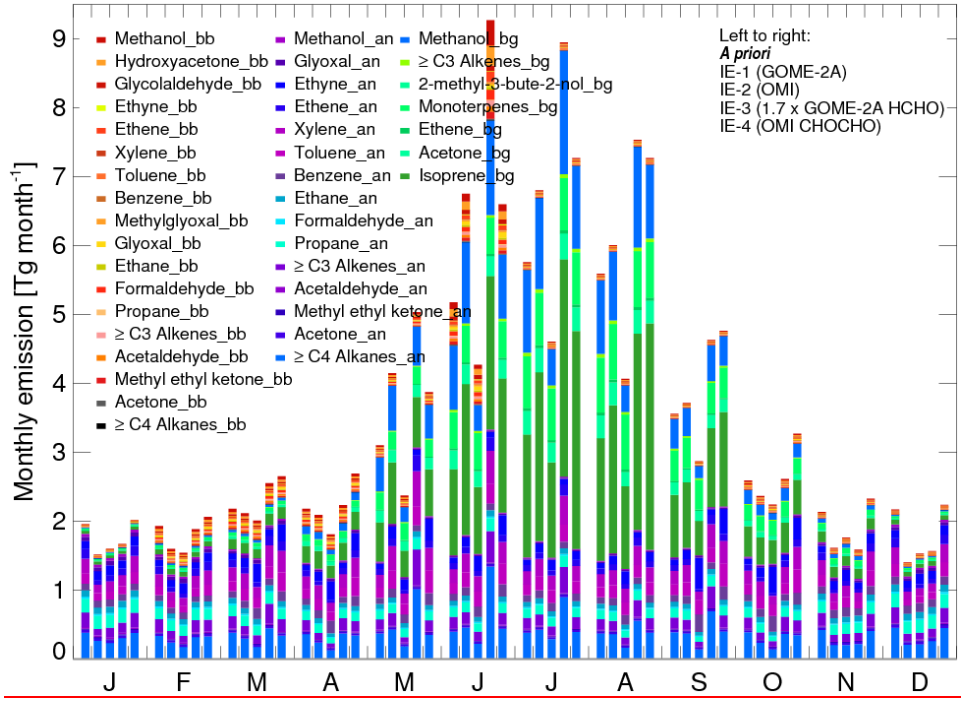


Figure 2. Estimates of monthly Chinese NMVOC emissions. For each month, the bars from left to right represent: the *a priori* emission estimates and the *a posteriori* emission estimates from IE-1, IE-2, IE-3, and IE-4, respectively. Color keys for NMVOC species are shown inset, with the suffixes of 'bb', 'an' and 'bg' indicating emissions from biomass burning, anthropogenic, and biogenic activities, respectively.

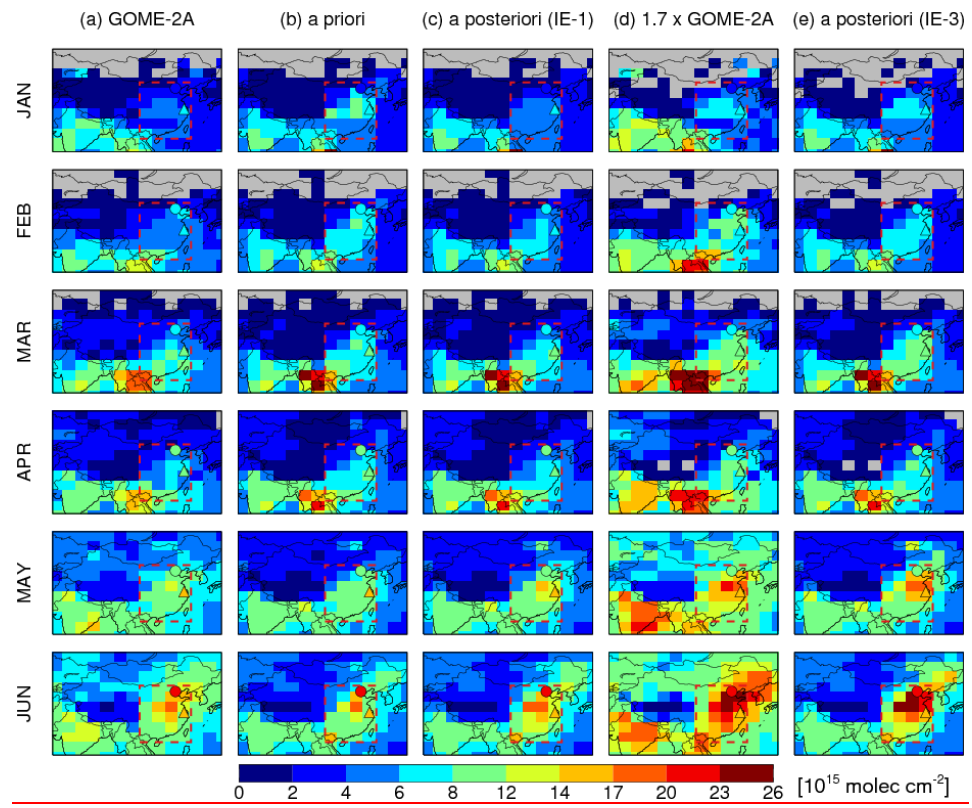


Figure 3. Monthly mean formaldehyde VCDs over China from January to June. For each month, the panels from left to right show: (a) formaldehyde VCDs observed by GOME-2A, (b) formaldehyde VCDs simulated by the model using *a priori* emission estimates, (c) the *a posteriori* formaldehyde VCDs from IE-1, (d) GOME-2A formaldehyde VCDs scaled by a factor of 1.7, and (e) the *a posteriori* formaldehyde VCDs from IE-3. All model results were sampled at GOME-2A overpass time. Also shown are ground-based MAX-DOAS measurements at 9:30 local time at Xianghe (monthly mean, circles) and Wuxi (bimonthly mean, upward triangles).

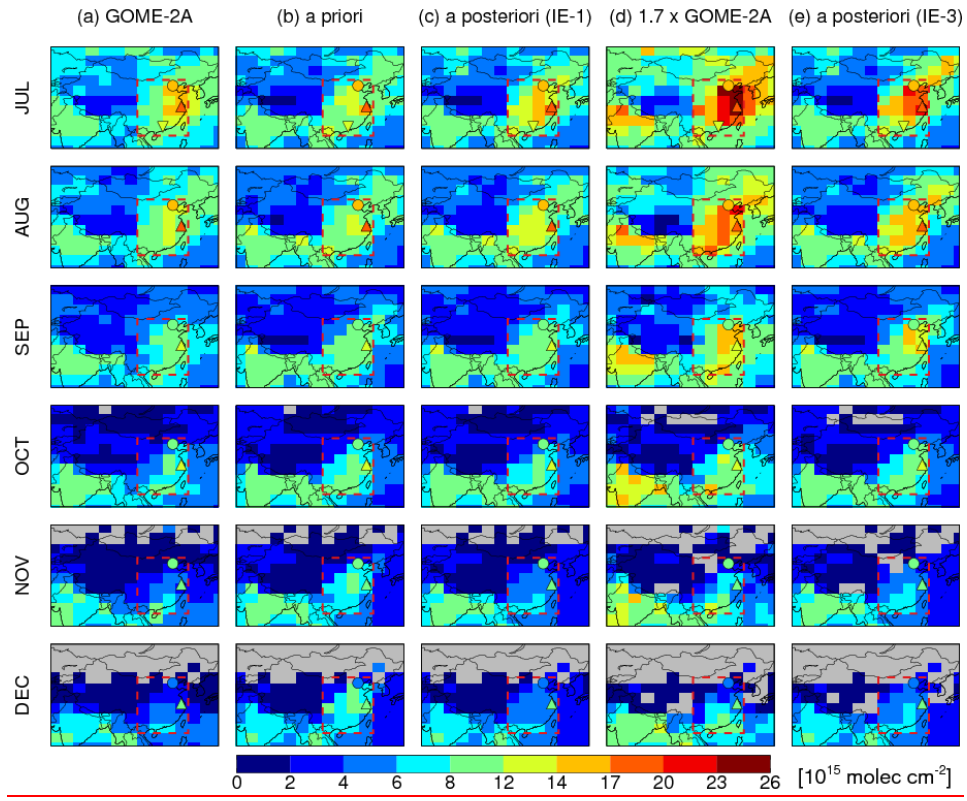


Figure 4. Same as Figure 3 except for July to December. Also shown are ground-based MAX-DOAS measurements at 9:30 local time at Back Garden (July mean, inverted triangles).

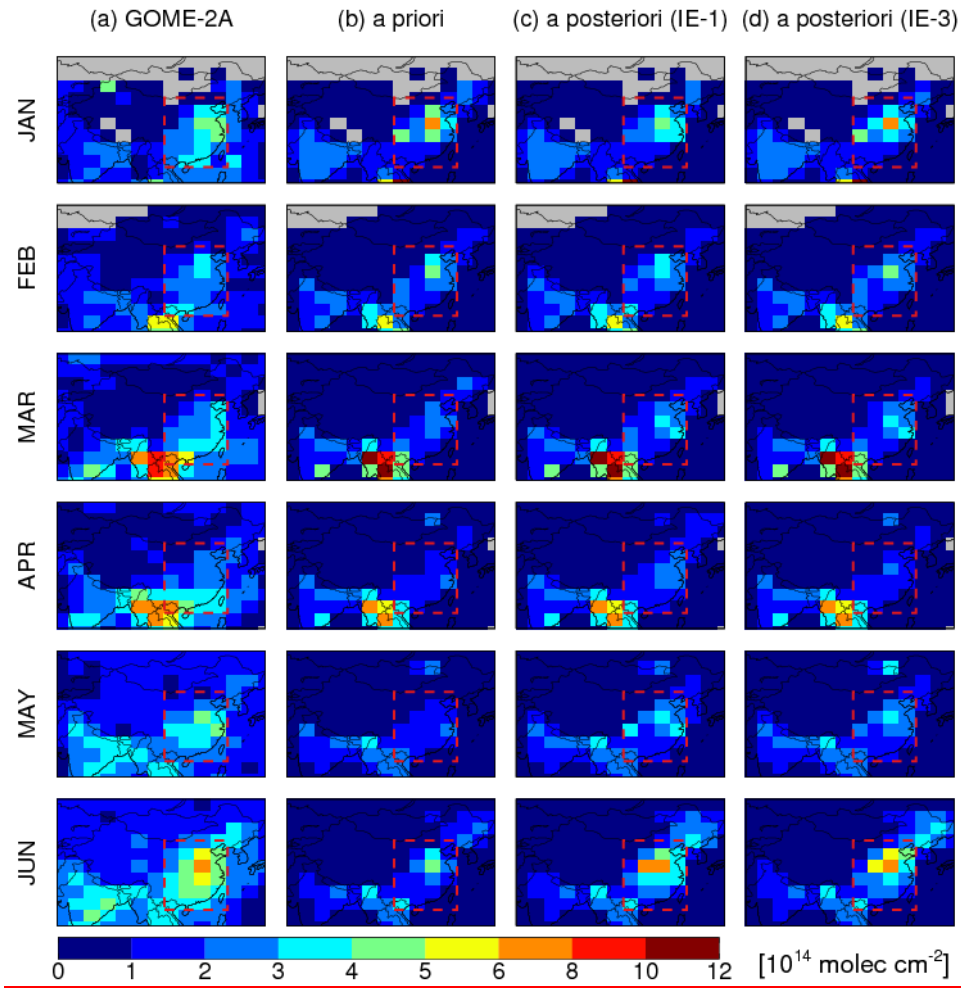


Figure 5. Monthly mean glyoxal VCDs over China from January to June. For each month, the panels from left to right show: (a) glyoxal VCDs observed by GOME-2A, (b) glyoxal VCDs simulated by the model using *a priori* emission estimates, (c) the *a posteriori* glyoxal VCDs from IE-1, and (d) the *a posteriori* glyoxal VCDs from IE-3. All model results were sampled at GOME-2A overpass time.

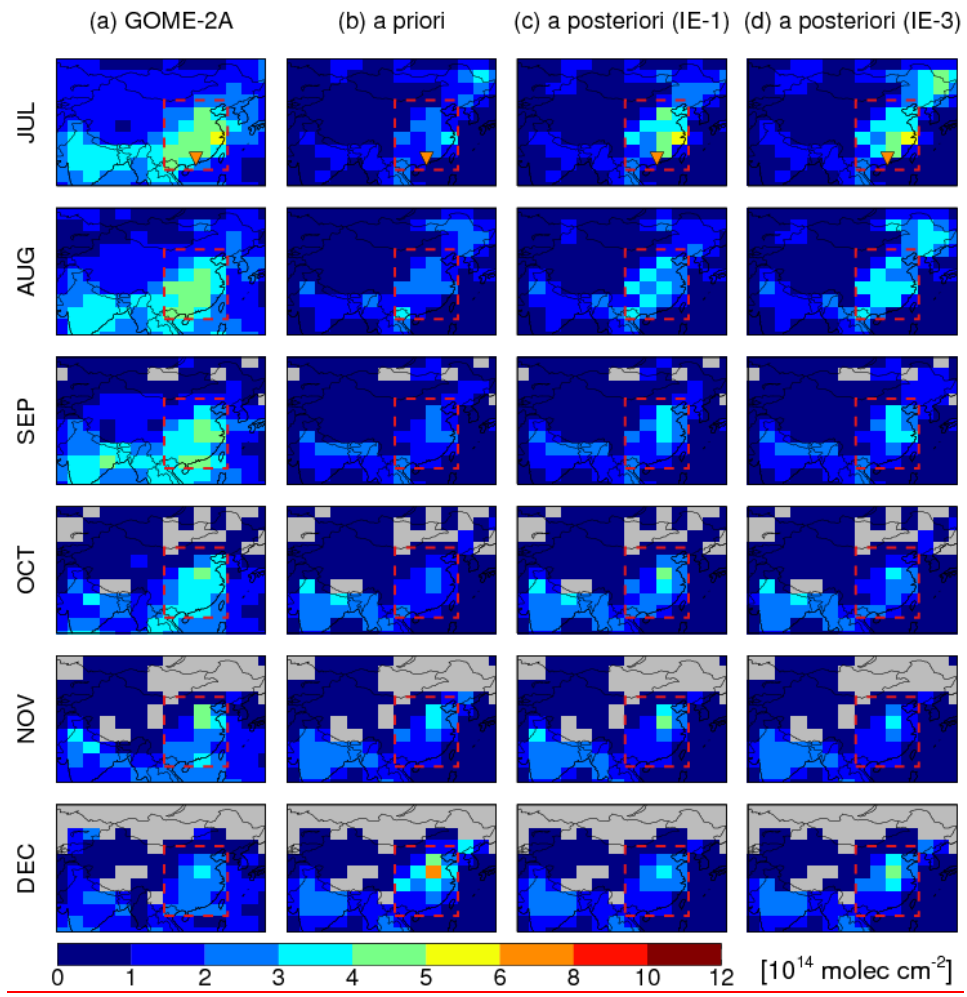


Figure 6. Same as Figure 5 except for July to December. Also shown are ground-based MAX-DOAS measurements at 9:30 local time at Back Garden (July mean, inverted triangles).

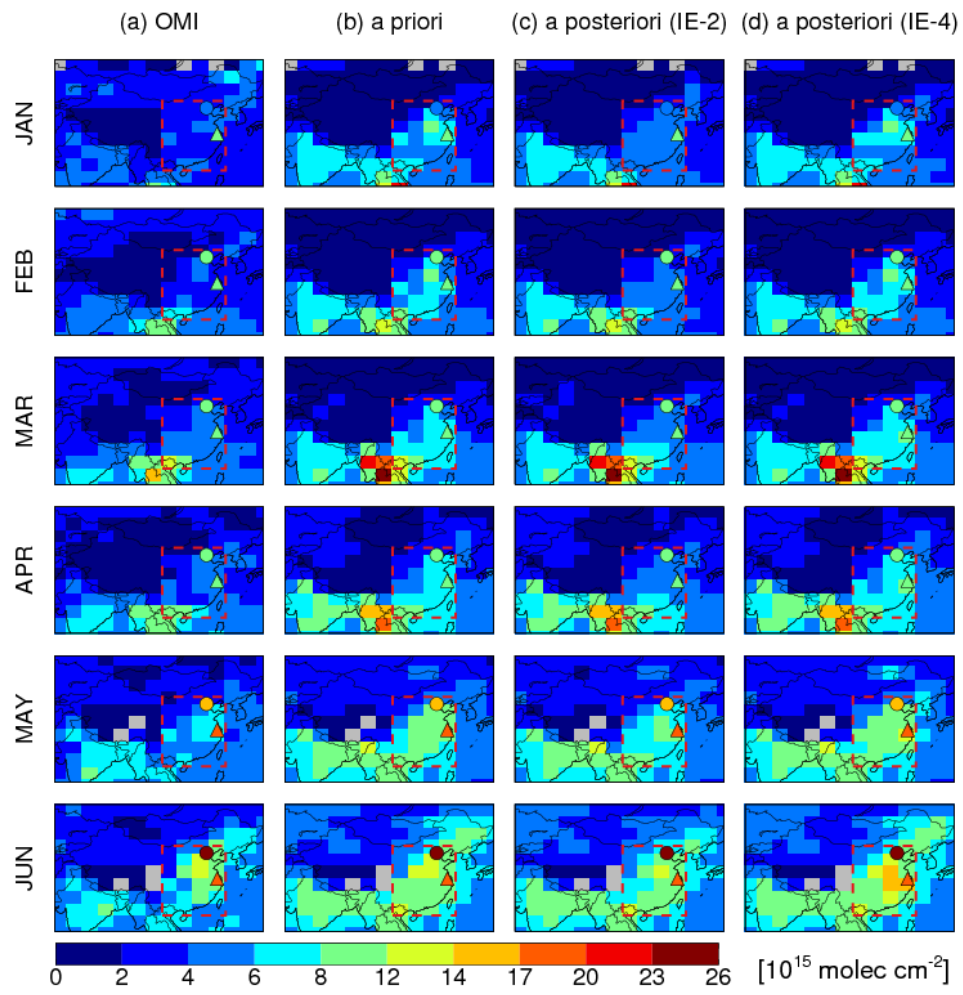


Figure 7. Monthly mean formaldehyde VCDs over China from January to June. For each month, the panels from left to right are: (a) formaldehyde VCDs observed by OMI, (b) formaldehyde VCDs simulated by the model using *a priori* emission estimates, (c) the *a posteriori* formaldehyde VCDs from IE-2, and (d) the *a posteriori* formaldehyde VCDs from IE-4. All model results were sampled at OMI overpass time. Also shown are ground-based MAX-DOAS measurements at 13:30 local time at Xianghe (monthly mean, circles) and Wuxi (bimonthly mean, upward triangles).

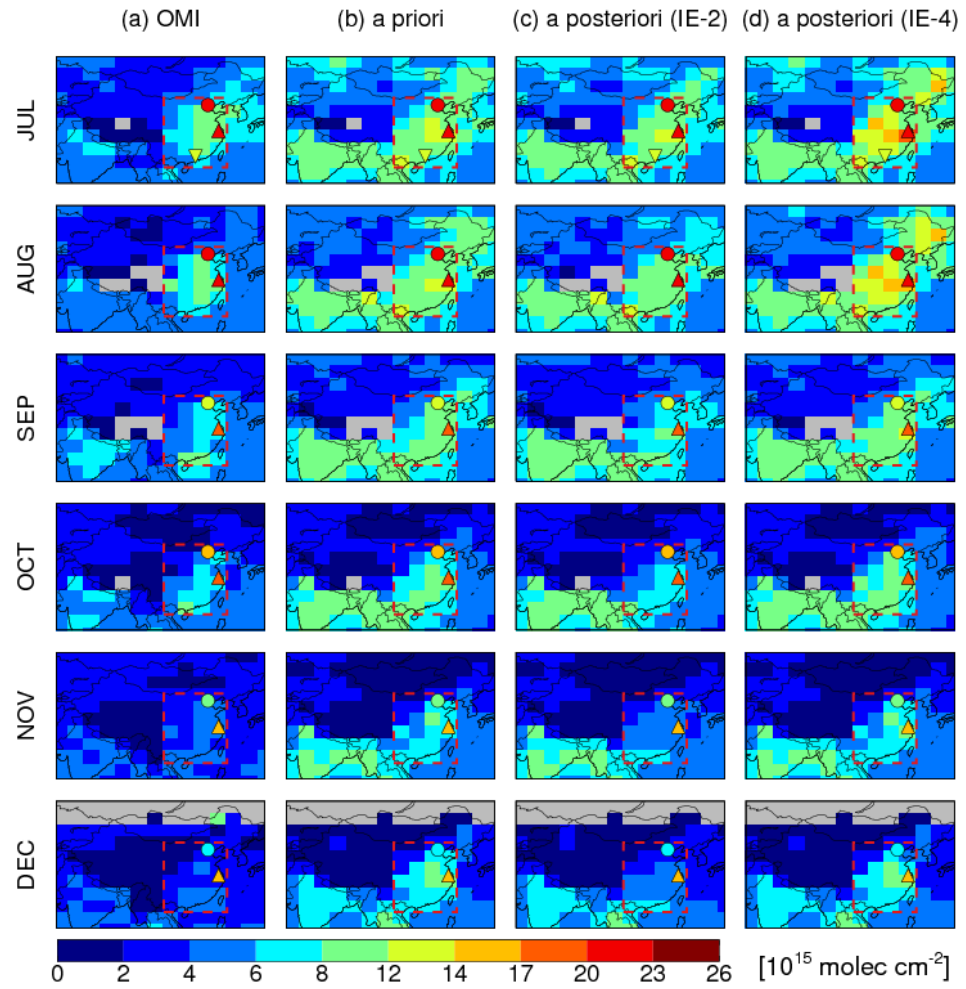


Figure 8. Same as Figure 7 except for July to December. Also shown are ground-based MAX-DOAS measurements at 13:30 local time at Back Garden (July mean, inverted triangles).

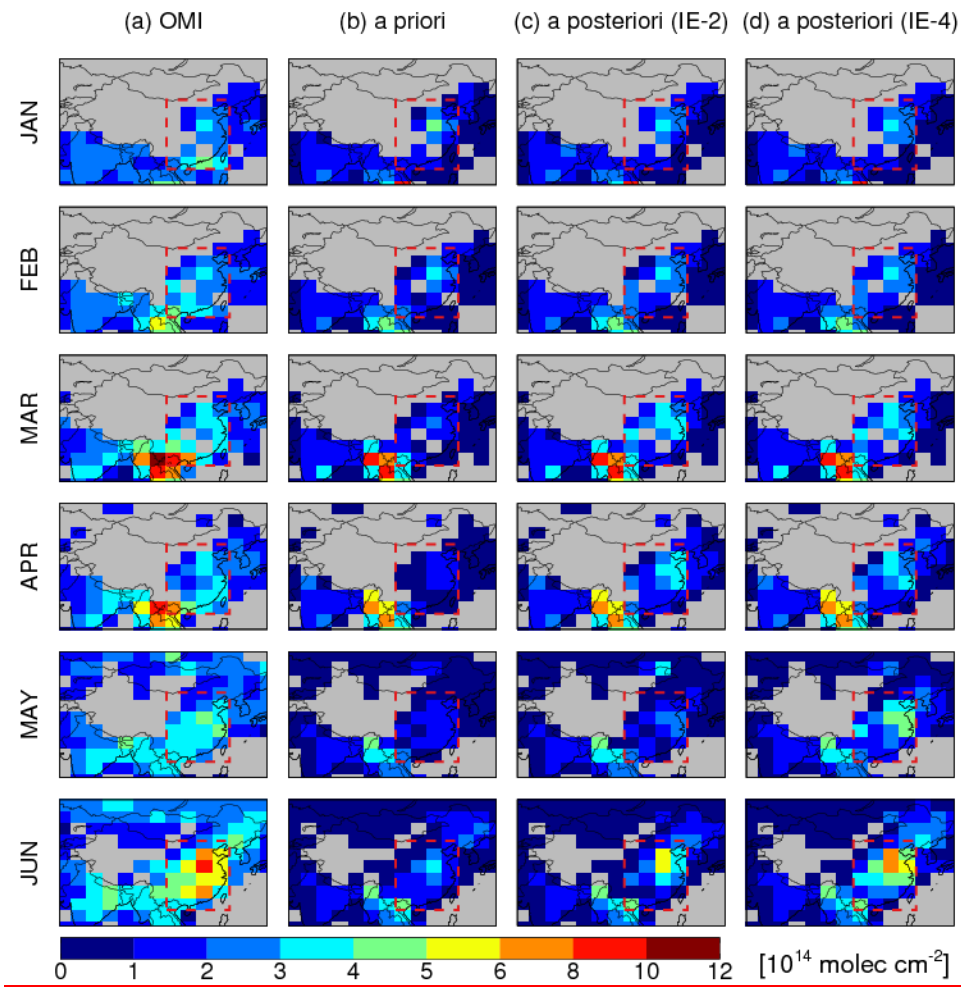


Figure 9. Monthly mean glyoxal VCDs over China from January to June. For each month, the panels from left to right are: (a) glyoxal VCDs observed by OMI, (b) glyoxal VCDs simulated by the model using *a priori* emission estimates, (c) the *a posteriori* glyoxal VCDs from IE-2, and (d) the *a posteriori* glyoxal VCDs from IE-4. All model results were sampled at OMI overpass time.

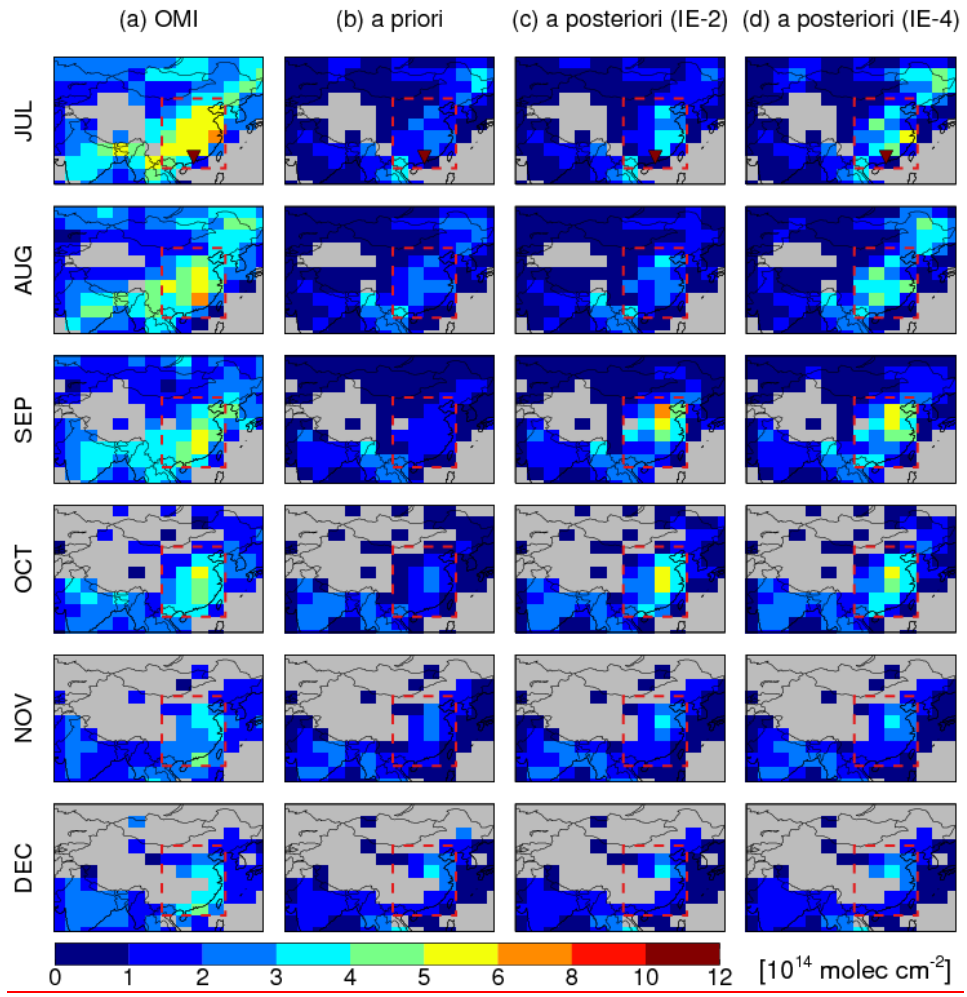


Figure 10. Same as Figure 9 except for July to December. Also shown are ground-based MAX-DOAS measurements at 13:30 local time at Back Garden (July mean, inverted triangles).

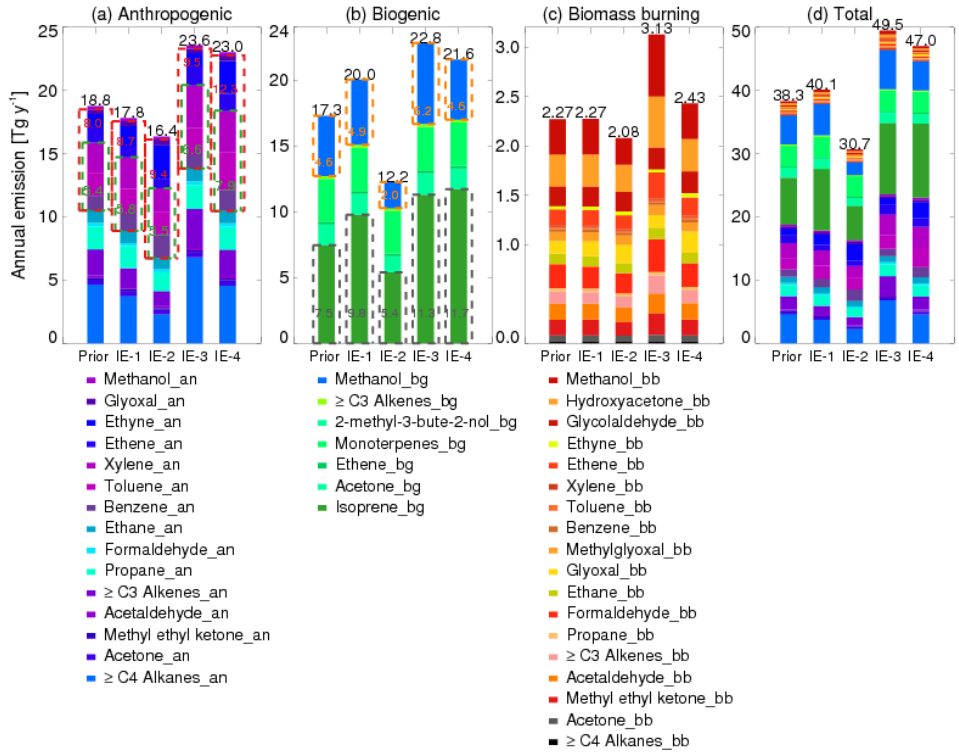


Figure 11. Comparison of estimates of annual Chinese NMVOC emissions from (a) anthropogenic, (b) biogenic, (c) biomass burning, and (d) total sources. For each panel, the bars from left to right are the *a priori* estimates and the *a posteriori* estimates from IE-1, IE-2, IE-3, and IE-4. Annual total NMVOC emissions are shown in black numbers on top of each bar. The red dashed boxes and red numbers in (a) indicate annual emissions of anthropogenic glyoxal precursors. The green dashed boxes and green numbers in (a) indicate annual emissions of anthropogenic aromatics. The grey dashed boxes and grey numbers in (b) indicate annual biogenic isoprene emissions. The orange dashed boxes and orange numbers in (b) indicate annual biogenic methanol emissions. Color keys to NMVOC species are shown at the bottom, with suffixes of 'an', 'bg', 'bb' indicating anthropogenic source, biogenic source, and biomass burning source, respectively.

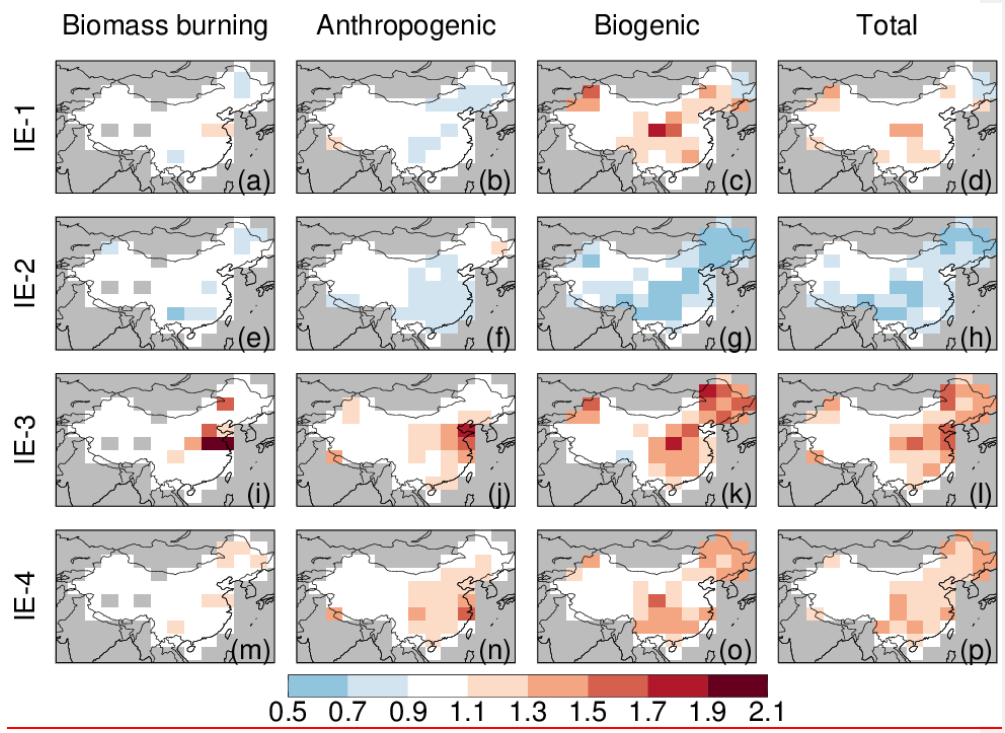


Figure 12. Spatial distributions of the optimized scale factors for Chinese annual NMVOC emissions, relative to the *a priori* emission estimates, for the four inversion experiments.

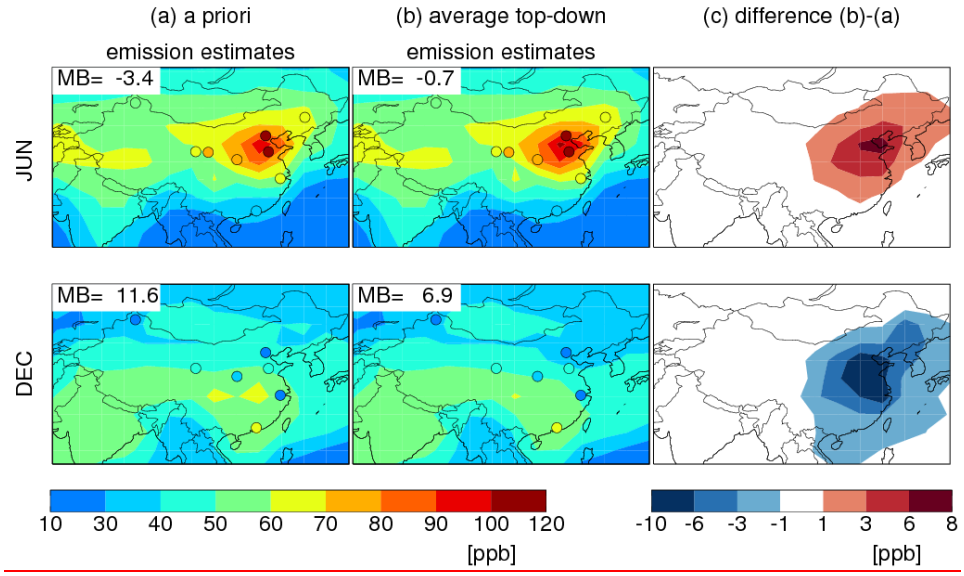


Figure 13. Simulated monthly mean afternoon (13:00-17:00 LT) surface ozone concentrations in June and December 2007 driven by (a) the *a priori* emissions and (b) our average top-down emissions, respectively, as well as (c) the differences. Filled circles show the afternoon surface ozone observations at several sites in China (Table S9). Mean biases (MB) of the simulated concentrations relative to surface measurements are shown inset.

Formatted: Normal

Formatted: Font: 9 pt, Bold

Table 1. Comparison of recent estimates for Chinese annual NMVOC emissions

Literature [a]	Target year	NMVOC [Tg C y^{-1}]				
		Anthropogenic		Biogenic		Biomass burning
		Total	Aromatics	Total	Isoprene	
Bottom-up						
Bo et al. (2008) [b]	2005	10.7				2.2 [e]
Zhang et al. (2009) [b]	2006	19.7 (± 68%)	2.1			
Cao et al. (2011) [b]	2007	29.8				
Kurokawa et al. (2013) [b]	2008	22.8 (± 46%)				
Li et al. (2014) [b]	2010	19.8	4.9			
Wu et al. (2016) [b]	2008	15.6				2.2 [e]
	2009	18.3				1.9 [e]
	2010	20.0				2.1 [e]
	2011	20.8				2.1 [e]
	2012	21.5				2.4 [e]
Huang et al. (2012) [b]	2006					1.3 (0.62-2.0)
van der Werf et al.(2010)	2007					0.24
Guenther et al.(2006)	2007			10.8 [f]	6.6 [f]	
Top-down						
Liu et al. (2012) [c]	2007	28.7	11.3			
Stavrakou et al. (2015) [d]	2010	(17.3-20.7)			(5.0-5.5)	(1.1-1.5)
This work	2007	17.0 [g] (13.5-19.7)	6.1 [g] (5.0~7.3)	12.6 [g] (8.9-14.8)	8.4 [g] (4.9-10.5)	1.22 [g] (1.06 – 1.47)

[a] Emission estimates from literature were originally in units of Tg y^{-1}. We converted the units to Tg C y^{-1} using to carbon-to-organic-compound mass ratios (0.84 for anthropogenic VOCs, 0.57 for biomass burning VOCs, and 0.85 for biogenic VOCs based on the *a priori* emission estimates).

[b] These emission estimates included NMVOC species that were not included in this work. See color-keys in Figure 2 for NMVOC species whose emissions were included in this work.

[c] Used SCIAMACHY-observed glyoxal VCDs as constraints.

[d] Used GOME-2A-observed and OMI-observed formaldehyde VCDs as constraints.

[e] Consisted of emissions from open burning of crop residues and from biofuel burning.

[f] Calculated by the GEOS-Chem model using GEOS-5 meteorological data.

[g] Average of top-down estimates from four inversion experiments.

Formatted: Font: Bold, Check spelling and grammar

Table 2. Inversion experiments to constrain Chinese NMVOC emissions

Inversion experiments	Observational constraints from satellites [± uncertainties]	Annual Chinese NMVOC emission estimates [Tg C y^{-1}]			
		Anthropogenic	Biogenic	Biomass burning	Total
		A priori emission estimates [± uncertainties]			
		15.5 (4.9 for aromatics) [a] [±200%]	10.8 (6.6 for isoprene) [b] [±55%]	1.10 [±300%] [c]	27.4
		A posteriori emission estimates [range]			
IE-1	GOME-2A formaldehyde [±90%] and glyoxal [±150%]	15.7 (5.9 for aromatics)	12.5 (8.2 for isoprene)	1.13	29.3
IE-2	OMI formaldehyde [±90%] and glyoxal [±150%]	13.5 (5.0 for aromatics)	8.9 (4.9 for isoprene)	1.06	23.4
IE-3	GOME-2A formaldehyde × 170% [±90%]	19.2 (6.0 for aromatics)	14.8 (10.5 for isoprene)	1.47	35.4
IE-4	OMI glyoxal [±150%]	19.7 (7.3 for aromatics)	14.1 (9.9 for isoprene)	1.24	35.1
Our top-down estimates		17.0 [d] [13.5–19.7] (6.1 [d] [5.0–7.3] for aromatics)	12.6 [d] [8.9–14.8] (8.4 [d] [4.9–10.5] for isoprene)	1.2 [d] [1.1–1.5]	30.8 [d] [23.4–35.4]

Formatted: Normal, Don't keep with next

Formatted: Font: +Headings (Cambria), 10 pt

Formatted Table

[a] From Li et al. (2014).

[b] From Guenther et al. (2006).

[c] Compiled from the emission estimated by van der Werf et al. (2010) plus a scaling of the emission estimated by Huang et al. (2012). See text (section 2.2) for details.

[d] Average of top-down estimates from the four inversion experiments.

Table 3. Technical details for GOME-2A and OMI formaldehyde and glyoxal retrievals used in this study

Technical details	GOME-2A Formaldehyde [a]	OMI				
Onboard satellite		Glyoxal [b]		Formaldehyde [c]		Glyoxal [d]
	European Metop-A	NASA Aura				
Operation time						
October 2006-present						
July 2004-present						
Overpass time	20059:30 LT	13:30 LT				
Global coverage	1.5 days [e]	1 day				
Spatial resolution	20078 0 km × 40 km	13 km × 24 km				
Spectral window	240-790 nm	270-500 nm				
Spectral resolution		0.26-0.5 nm	0.42 nm and 0.63			

Formatted: Keep with next
Formatted: Font: Bold
Formatted: Keep with next
Formatted: Justified, Keep with next
Formatted: Justified, Keep with next
Formatted: Justified, Keep with next
Formatted Table
Merged Cells
Formatted: Right: 0 cm, Keep with next
Merged Cells
Inserted Cells
Formatted: Justified, Keep with next
Inserted Cells
Inserted Cells
Formatted: Justified, Keep with next, Position: Horizontal: Left, Relative to: Column, Vertical: 0 cm, Relative to: Paragraph, Horizontal: 0.32 cm, Wrap Around
Formatted: Justified, Keep with next
Formatted: Keep with next
Formatted: Keep with next, Position: Horizontal: Left, Relative to: Column, Vertical: 0 cm, Relative to: Paragraph, Horizontal: 0.32 cm, Wrap Around
Formatted: Font: 9 pt
Inserted Cells
Formatted: Justified, Keep with next
Inserted Cells
Inserted Cells
Inserted Cells
Formatted: Keep with next, Position: Horizontal: Left, Relative to: Column, Vertical: 0 cm, Relative to: Paragraph, Horizontal: 0.32 cm, Wrap Around
Formatted: Justified, Keep with next
Formatted ...
Formatted: Font: Not Bold
Formatted: Justified, Keep with next
Formatted ...
Formatted: Justified, Keep with next
Formatted ...
Inserted Cells
Formatted: Justified, Keep with next
Formatted ...
Formatted: Justified, Keep with next

			nm			
Selected absorption bandkawa	328.5 – 346 nm	435 – 460 nm	328.5 – 356.5 nm	435 – 461 nm		
Retrieval algorithm	DOAS fitting	Direct fitting				
Cloud parameters Surface albedo Air mass factor calculation Extinction by aerosols	FRESCO+ (Wang et al., 2008)	OMCLDO2 (Acarreta et al., 2004)				
	Kleipool et al. (2008)	Kleipool et al. (2008)				
	Radiative transfer model 20	LIDORT (Spurr, 2008)	VLIDORT (Spurr, 2006)			
	Tracer gas profiles	IMAGE v2 (Stavrakou et al., 2009b)	GEOS Chem (González Abad et al., 2015)			
	Considered implicitly via cloud correction	Considered implicitly in the cloud retrieval				
Discarded pixels	2006Cloud fraction > 40% or zenith angles >60[e]	Cloud fraction > 40%	Impacted by random telegraph signals (RTS)[f]			

[e] From De Smedt et al. (2012)

Formatted: Keep with next, Position: Horizontal: Left, Relative to: Column, Vertical: 0 cm, Relative to: Paragraph, Horizontal: 0.32 cm, Wrap Around
Formatted: Keep with next
Formatted: Justified, Keep with next
Formatted: Justified, Keep with next, Tab stops: Not at 1.88 cm
Inserted Cells
Inserted Cells
Formatted: Keep with next, Position: Horizontal: Left, Relative to: Column, Vertical: 0 cm, Relative to: Paragraph, Horizontal: 0.32 cm, Wrap Around
Formatted: Keep with next, Tab stops: Not at 3.16 cm, Position: Horizontal: Left, Relative to: Column, Vertical: 0 cm, Relative to: Paragraph, Horizontal: 0.32 cm, Wrap Around
Formatted: Keep with next, Position: Horizontal: Left, Relative to: Column, Vertical: 0 cm, Relative to: Paragraph, Horizontal: 0.32 cm, Wrap Around
Formatted: Justified, Keep with next
Formatted: Justified, Keep with next, Tab stops: Not at 1.88 cm
Merged Cells
Formatted: Keep with next
Formatted: Keep with next, Position: Horizontal: Left, Relative to: Column, Vertical: 0 cm, Relative to: Paragraph, Horizontal: 0.32 cm, Wrap Around
Formatted: Justified, Keep with next
Formatted: Justified, Keep with next, Tab stops: Not at 1.88 cm
Formatted: Justified, Keep with next
Formatted: Justified, Keep with next
Formatted: Justified, Keep with next, Tab stops: Not at 1.88 cm
Formatted: Justified, Keep with next
Formatted: Justified, Keep with next, Tab stops: Not at 1.88 cm
Formatted: Justified, Keep with next
Formatted: Justified, Keep with next, Tab stops: Not at 1.88 cm
Inserted Cells
Formatted: Justified, Keep with next
Formatted: Justified, Keep with next, Tab stops: Not at 1.88 cm
Formatted ...
Formatted: Justified, Keep with next
Formatted: Keep with next

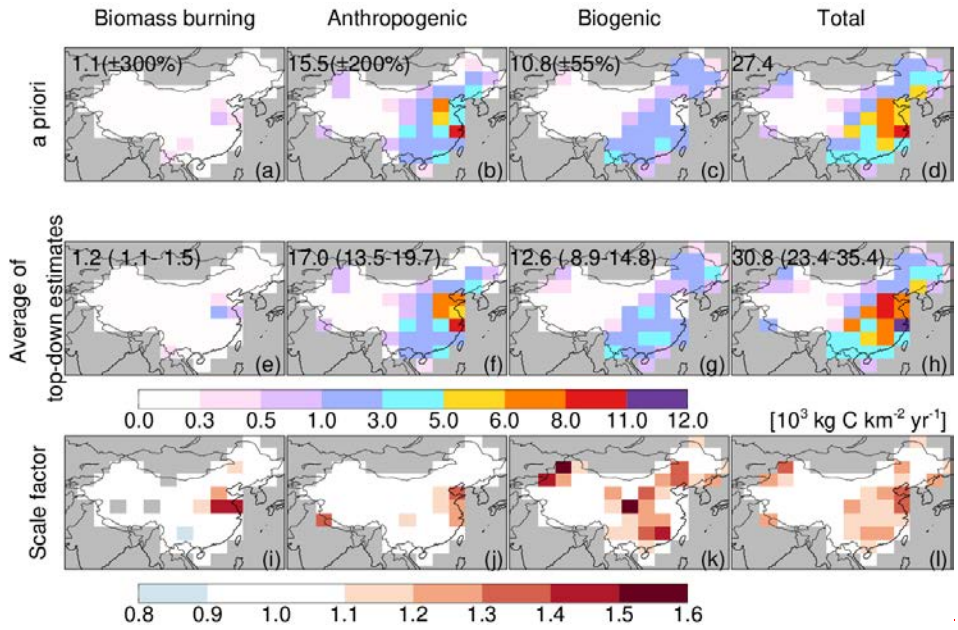
[b] From Lerot et al. (2010)

[c] From González Abad et al. (2015)

[d] From Chan Miller et al. (2014)

[e] Before the swath was narrowed in June 2013. After that, the global coverage is achieved every 3 days.

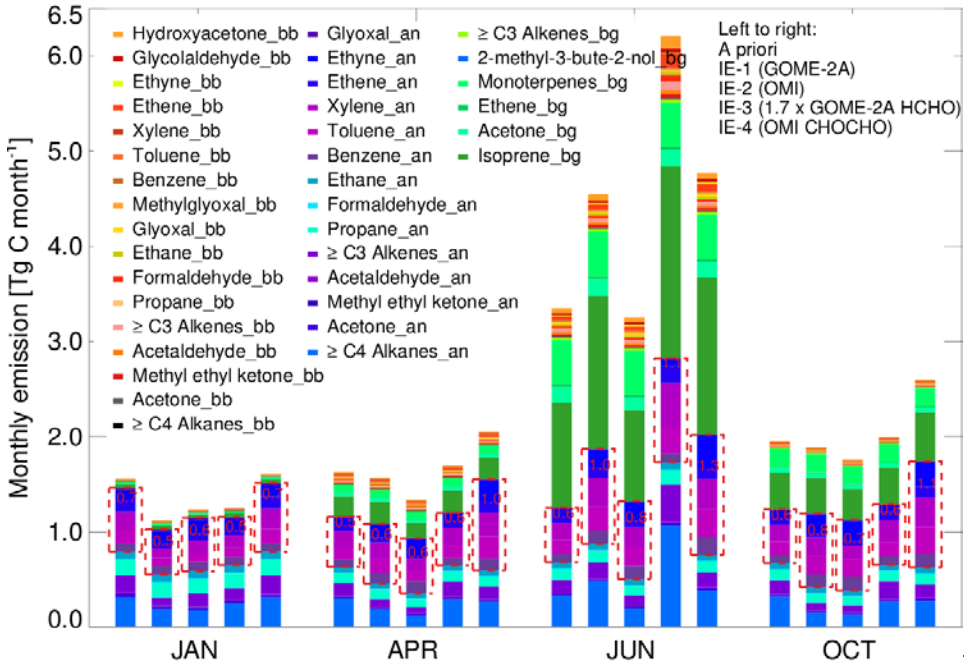
[f] Pixels that have been flagged as RTS in the level 1B product (Kleipool, 2005).



See color keys in Figure 2 for NMVOC species whose emissions were included in this work.

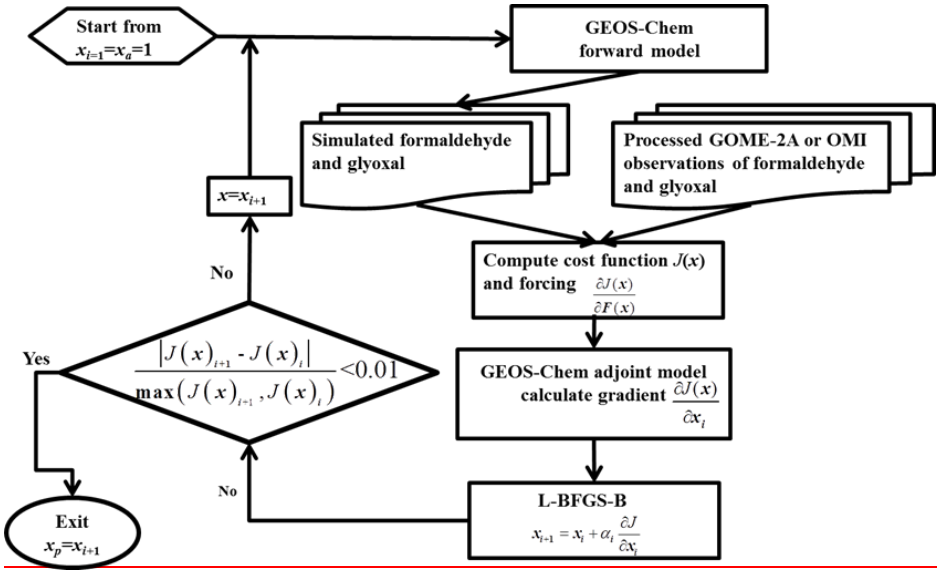
Figure 1. Spatial distributions of annual NMVOC emissions from China. (a)-(d): the *a priori* annual NMVOC emission estimates from (a) biomass burning, (b) anthropogenic, (c) biogenic, and (d) total sources. (e)-(h): averaged top-down estimates of annual NMVOC emissions. Annual Chinese total emission estimates are shown inset in units of [Tg C y^{-1}]. The uncertainties of the *a priori* emission estimates and the range of top-down emission estimates are shown in parentheses. (i)-(l): scale factors for our averaged top-down estimates relative to the *a priori* estimates.

Formatted: Keep with next



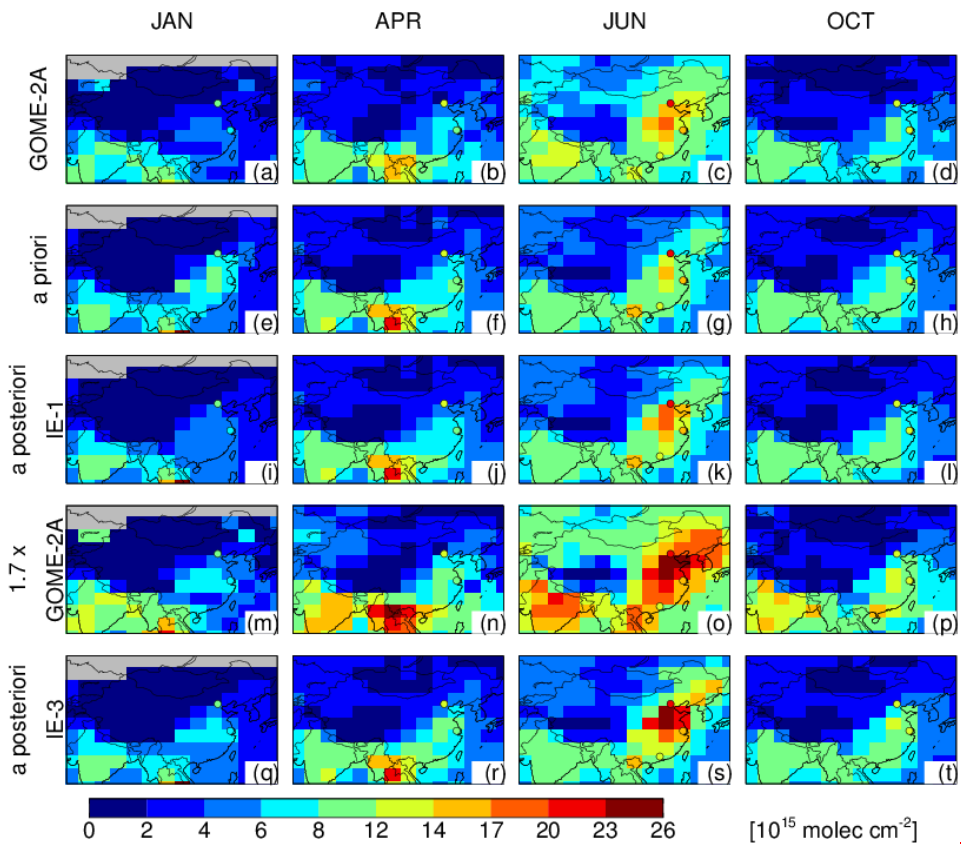
~~**Figure 2. Estimates of monthly Chinese NMVOC emissions for January, April, June, and October 2007. For each month, the bars from left to right represent: the *a priori* emission estimates and the *a posteriori* emission estimates from IE-1, IE-2, IE-3, and IE-4. The red dashed boxes and red numbers indicate monthly emissions of anthropogenic glyoxal precursors. Color keys for NMVOC species are shown inset, with the suffixes of 'bb', 'an' and 'bg' indicating emissions from biomass burning, anthropogenic, and biogenic activities, respectively.**~~

Formatted: Keep with next



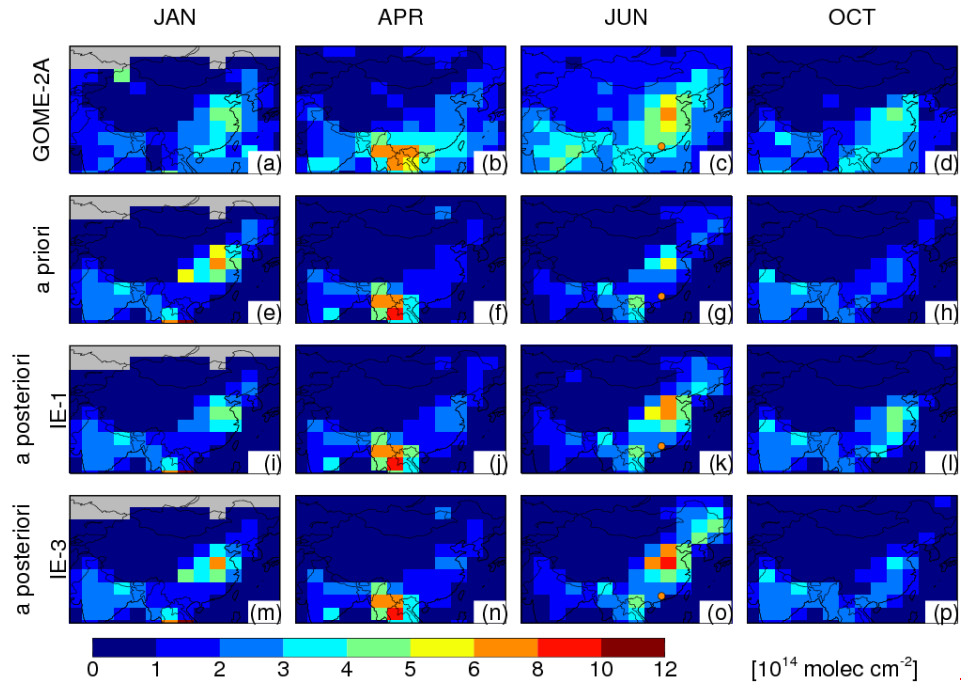
~~**Figure 3. Protocol for the adjoint inversion experiments.**~~

Formatted: Keep with next



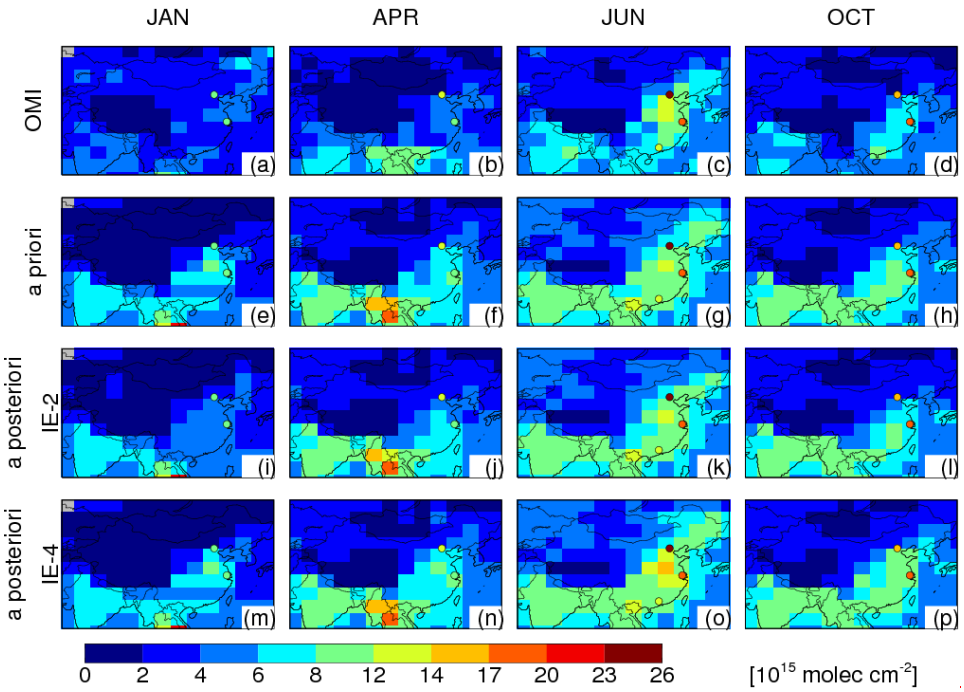
~~**Figure 4. Monthly mean formaldehyde VCDs over China. (a-d): GOME-2A observed formaldehyde VCDs and (m-p) GOME-2A formaldehyde VCDs scaled by a factor of 1.7. (e-h): Formaldehyde VCDs simulated by the model using *a priori* emission estimates; (i-l) the *a posteriori* formaldehyde VCDs from inversion IE-1;(q-t) the *a posteriori* formaldehyde VCDs from the inversion IE-3. Also shown are ground-based MAX-DOAS measurements at 9:30 LT (circles) at Beijing (De Smedt et al., 2015), Wuxi (Wang et al, 2017), and Back Garden (Li et al, 2013).**~~

Formatted: Keep with next

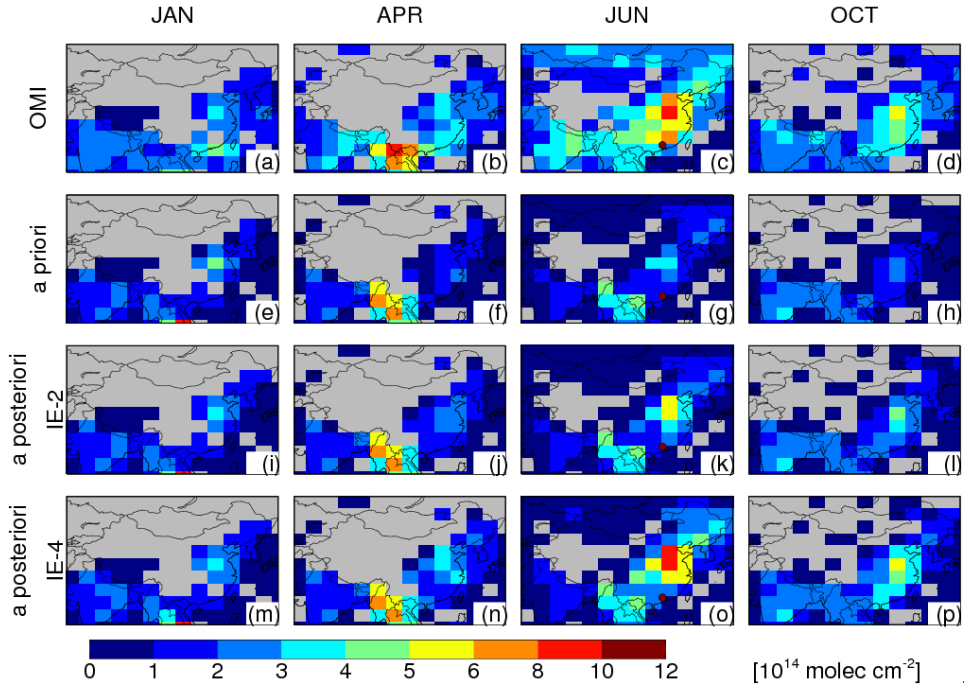


~~Figure 5. Monthly mean glyoxal VCDs over China (a-d) observed by the GOME-2A instrument, (e-h) simulated by the model using the *a priori* emission estimates, (i-l) obtain from inversion IE-1, and (m-p) obtain from inversion IE-3. Also shown are ground based MAX-DOAS measurements at Back Garden in July 2006 (Li et al, 2013).~~

Formatted: Keep with next



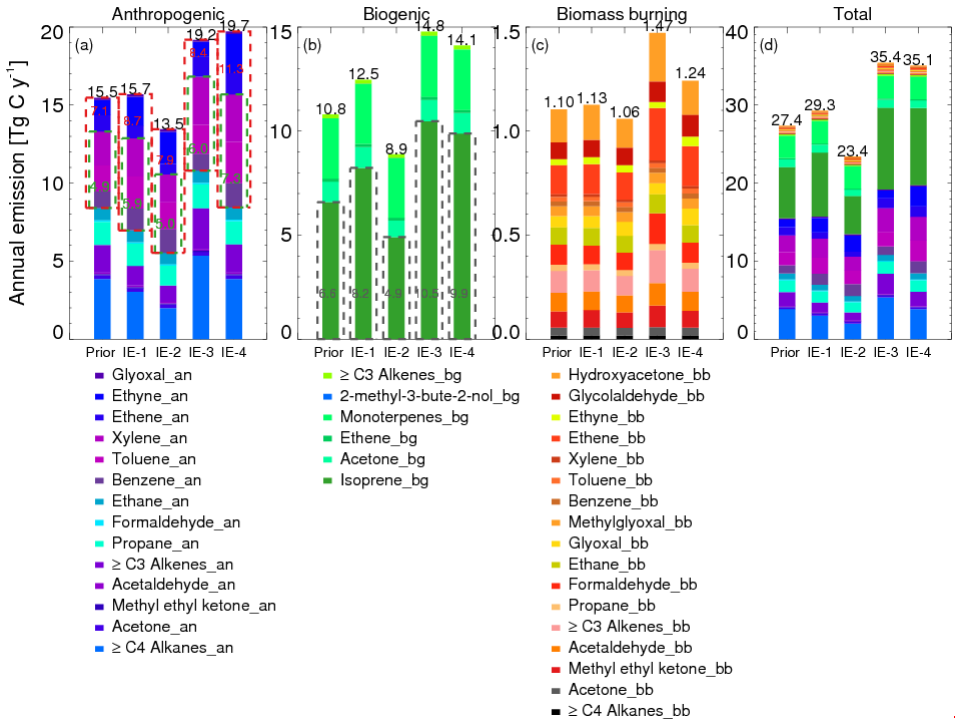
~~Figure 6. Monthly mean formaldehyde VCDs over China. (a-d): formaldehyde VCDs observed by the OMI instrument. (e-h): formaldehyde VCDs simulated by the model using the *a priori* emission estimates. (i-l): the *a posteriori* formaldehyde VCDs from inversion IE-2. (m-p): the *a posteriori* formaldehyde VCDs from inversion IE-4. Also shown are ground-based MAX-DOAS measurements at 13:30 LT (circles) at Beijing (De Smedt et al., 2015), Wuxi (Wang et al, 2017), and Back Garden (Li et al, 2013).~~



~~Figure 9. Monthly mean glyoxal VCDs over China (a-d) observed by the OMI instrument, (e-h) simulated by the model using the *a priori* emission estimates, (i-l) obtained from inversion IE-2, and (m-p) obtained from the inversion IE-4. Also shown are ground-based MAX-DOAS measurements at 13:30 LT (circles) at Back Garden in July 2006 (Li et al, 2013).~~

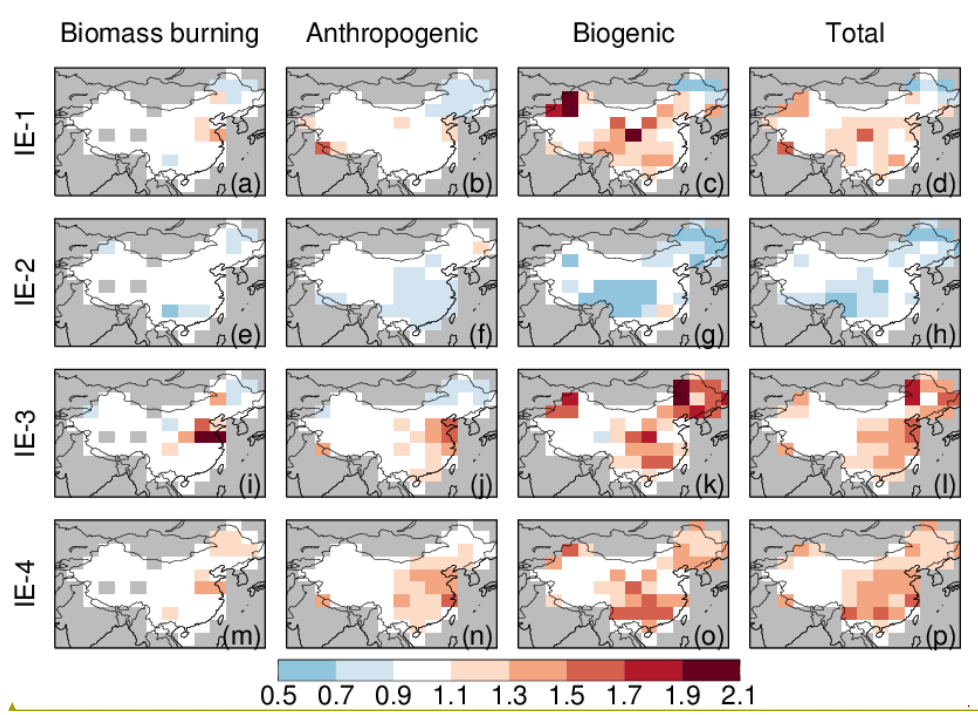
Formatted: Keep with next

Formatted: Keep with next



~~**Figure 11. Comparison of estimates of annual Chinese NMVOC emissions from (a) anthropogenic, (b) biogenic, (c) biomass burning, and (d) total sources. For each subfigure, shown from left to right are the *a priori* estimates and our *a posteriori* estimates from IE-1, IE-2, IE-3, and IE-4. Annual total NMVOC emission estimates are shown in black numbers on top of each bar. The red dashed boxes and red numbers in (a) indicate annual emissions of anthropogenic glyoxal precursors. The green dashed boxes and green numbers in (a) indicate annual emissions of anthropogenic aromatics. The grey dashed boxes and grey numbers in (b) indicate annual biogenic isoprene emissions. Color keys to NMVOC species are shown at the bottom, with suffixes of 'an', 'bg', 'bb' indicating anthropogenic source, biogenic source, and biomass burning source, respectively.**~~

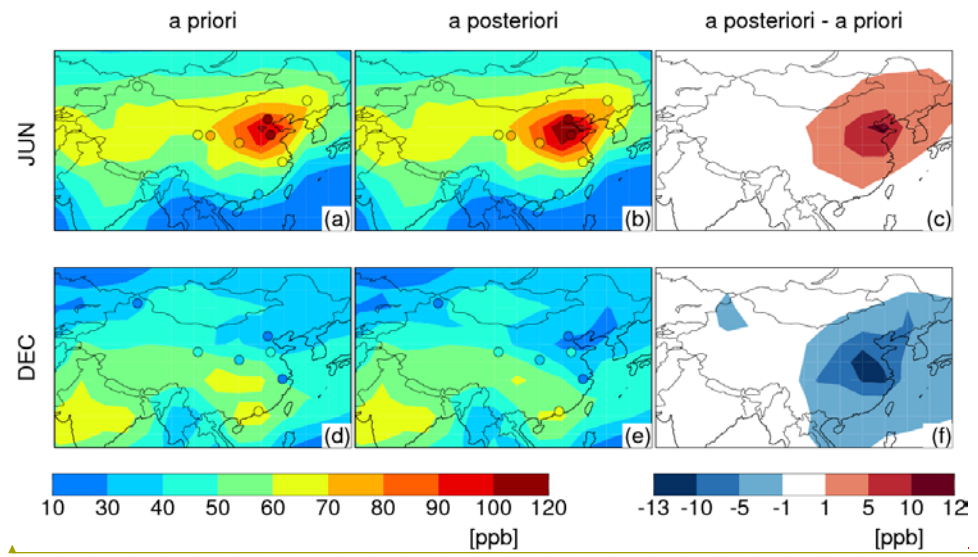
Formatted: Normal, Keep with next



Formatted: Font: (Default) Times New Roman, 9 pt

~~**Figure 12. Spatial distributions of the optimized scale factors for Chinese annual NMVOC emissions, relative to the *a priori* emission estimates, for the four inversion experiments.**~~

Formatted: Normal, Keep with next



Formatted: Font: (Default) Times New Roman, 9 pt

~~**Figure 13. Simulated monthly mean afternoon (13:00-17:00 LT) surface ozone concentrations driven by the *a priori* emissions and average of our top-down emissions, respectively, as well as corresponding difference (*a posteriori* - *a priori*) in June and December 2007. Filled circles overlaid on the contour maps represent surface ozone observations at several sites of China (Table S2).**~~

Formatted: Normal, Keep with next

Formatted: Font: +Headings (Cambria), 10 pt

Formatted: Font: +Headings (Cambria), 10 pt, Not Bold

Formatted: Font: Not Bold

Supplementary Information

Adjoint inversion of Chinese non-methane volatile organic compound emissions using space-based observations of formaldehyde and glyoxal

Hansen Cao[1], Tzung-May Fu[1,*], Lin Zhang[1], Daven K. Henze[2], Christopher Chan Miller[3], Christophe Lerot[4], Gonzalo González Abad[3], Isabelle De Smedt[4], Qiang Zhang[5], Michel van Roozendael[4], Francois Hendrick[4], Kelly Chance[3], Jie Li[6], Junyu Zheng[7], Yuanhong Zhao[1]

[1]Department of Atmospheric and Oceanic Sciences and Laboratory for Climate and Ocean-Atmosphere Studies, School of Physics, Peking University, Beijing, China,100871
[2]Department of Mechanical Engineering, University of Colorado, Boulder, USA
[3]Atomic and Molecular Physics Division, Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts, USA
[4]Belgian Institute for Space Aeronomy ~~(BIRA-IASB)~~, Brussels, Belgium
[5]Center for Earth System Science, Tsinghua University, Beijing, China
[6]Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
[7]College of Environmental Science and Engineering, South China University of Technology, Guangzhou, China

Correspondence to: Tzung-May Fu (tmfu@pku.edu.cn)

Table S1 Ultimate yields of formaldehyde and glyoxal from the oxidation of NMVOC precursors by OH in our model under high-NO_x and low-NO_x conditions

NMVOCs	Formaldehyde (molecules per C)		Glyoxal (molecules per C)	
	High-NO_x [a]	Low-NO_x [b]	High-NO_x [a]	Low-NO_x [b]
Ethene	0.995	0.366	0.0665	0.067
Glycolaldehyde	0.366	0.366	0.067	0.067
Isoprene	0.436	0.38	0.0255	0.073
2-methyl-3-bute-nol (MBO)	0.092	0.092	0.0168	0.0168
Benzene	0.001	0.001	0.0555	0.0555
Toluene	0.198	0.18	0.037	0.037
Xylenes	0.269	0.155	0.026	0.026
Monoterpenes (lumped)	0.006	0.006	0.005 [c]	0.005 [c]
Ethyne	-	-	0.318	0.318
Methanol	1.0	1.0	-	-
Ethane	0.5	0.5	-	-
Acetaldehyde (lumped)	0.5	0.5	-	-
Propane	0.49	0.317	-	-
≥C_3 alkenes (lumped)	0.657	0.333	-	-
Acetone	0.64	0.383	-	-
Hydroxyacetone	0.333	0.333	-	-
Methyglyoxal	0.333	0.333	-	-
≥C_4 alkanes (lumped)	0.578	0.187	-	-
Methy ethyl ketone (lumped)	0.465	0.25	-	-

[a] Yields under high-NO_x conditions were calculated assuming that all RO_2 radicals from the oxidation of the NMVOC precursor reacted with NO.

[b] Yields under low-NO_x conditions were calculated assuming RO_2:HO_2 concentration ratio of 1:1.

[c] Glyoxal produced from the oxidation of monoterpenes by ozone

Table S2 Technical details for the GOME-2A and OMI formaldehyde and glyoxal observations used in this study

Technical details		GOME-2A		OMI	
		Formaldehyde	Glyoxal	Formaldehyde	Glyoxal
Product reference		De Smedt et al. (2012)	Lerot et al. (2010)	González Abad et al. (2015)	Chan Miller et al. (2014)
Platform		European MetOp-A satellite		NASA Aura satellite	
Operation time		October 2006 – present		July 2004 – present	
Overpass time		9:30 local time		13:30 local time	
Global coverage		Every 1.5 days before June 2013; every 3 days after June 2013		Every 1 day	
Spatial resolution		80 km × 40 km		13 km × 24 km	
Spectral window		240-790 nm		270-500 nm	
Spectral resolution		0.26-0.5 nm		0.42 nm and 0.63 nm	
Selected absorption band		328.5 - 346 nm	435 - 460 nm	328.5 - 356.5 nm	435 - 461 nm
Retrieval algorithm		Differential Optical Absorption Spectroscopy (DOAS) fitting		Direct fitting	
Cloud parameter data		FRESCO+ (Wang et al., 2008)		OMCLDO2 (Acarreta et al., 2004)	
Surface albedo data		Kleipool et al. (2008)		Kleipool et al. (2008)	
Air mass factor calculation	Radiative transfer model	LIDORT (Spurr, 2008)		VLIDORT (Spurr, 2006)	
	Tracer gas profiles	IMAGES model outputs (Stavrakou et al., 2009b)		GEOS-Chem model outputs (González Abad et al., 2015)	
Extinction by aerosols		Considered implicitly via cloud correction (Boersma et al., 2004)		Considered implicitly in the cloud retrieval (Acarreta et al., 2004)	
Discarded pixels		Pixels with cloud fraction >40% or zenith angles >60° were discarded		Pixels with cloud fraction > 40% were discarded	Pixels flagged as impacted by random telegraph signals were discarded [a]

[a] Some pixels were flagged as impacted by random telegraph signals in the level 1-B product (Kleipool, 2005).

Table S3 Ground-based MAX-DOAS measurements of formaldehyde and glyoxal vertical column densities in China at GOME-2A and OMI overpass times

Reference	Location	Time of measurement		Vertical column densities	
				9-10 local time	13-14 local time
Formaldehyde [10^{16} molecules cm^{-2}]					
Vlemmix et al. (2015)	Xianghe, Heibei (39.75N, 116.96E)	2011	JAN	0.24	0.54
			FEB	0.78	0.99
			MAR	0.77	0.95
			APR	0.99	0.98
			MAY	1.08	1.53
			JUN	2.06	2.67
			JUL	1.49	2.10
			AUG	1.47	2.03
			SEP	1.05	1.36
			OCT	1.11	1.64
			NOV	0.85	1.18
		2010	DEC	0.49	0.79
Lee et al. (2015)	Beijing (39.59°N, 116.18°E)	August 16 to September 11, 2006		-	1.79
Wang et al. (2017)	Wuxi, Jiangsu (31.57°N,120.31°E)	2011 – 2014	JF	0.7 [a]	0.8 [a]
			MA	0.9±0.15 [a]	1.1±0.26 [a]
			MJ	1.5±0.12 [a]	1.9±0.15 [a]
			JA	1.7±0.10 [a]	2.2±0.26 [a]
			SO	1.2±0.12 [a]	1.7±0.12 [a]
			ND	0.8±0.30 [a]	1.4±0.32 [a]
Li et al. (2013)	Back Garden, Guangdong (23.50°N, 113.03°E)	July 2006		1.3±1.0 [b]	1.3±0.7 [b]
Glyoxal [10^{14} molecules cm^{-2}]					
Li et al. (2013)	Back Garden, Guangdong (23.50°N, 113.03°E)	July 2006		6.8±5.2 [c]	11.4±6.8 [c]

[a] From Figure 12 of Wang et al. (2017)

[b] From Figure 4 of Li et al. (2013)

[c] From Figure 5 of Li et al. (2013)

Table S4 Statistical comparisons of the *a priori* and *a posteriori* (from IE-1) simulated formaldehyde VCDs against the formaldehyde VCDs observed by GOME-2A over eastern China [a]

Month	Formaldehyde VCD comparisons (model against GOME-2A observations)					
	NMB [b]		R [b]		RMSE [b] ($\times10^{15}$ molecules cm^{-2})	
	a priori	*a posteriori* IE-1	*a priori*	*a posteriori* IE-1	*a priori*	*a posteriori* IE-1
JAN	0.43	0.066	0.80	0.92	2.12	0.645
FEB	0.13	-0.036	0.80	0.92	1.59	0.947
MAR	-0.024	-0.064	0.94	0.97	1.29	1.10
APR	0.11	0.047	0.93	0.97	1.25	0.83
MAY	-0.099	0.044	0.87	0.97	1.50	1.21
JUN	-0.11	0.005	0.80	0.87	2.27	1.93
JUL	-0.064	0.042	0.81	0.87	1.87	1.71
AUG	0.014	0.074	0.85	0.87	1.31	1.46
SEP	0.017	0.051	0.83	0.87	1.25	1.15
OCT	0.13	0.04	0.90	0.95	1.31	0.809
NOV	0.45	0.13	0.74	0.94	2.28	1.00
DEC	0.67	0.17	0.51	0.85	3.27	1.19

[a] The eastern China domain is defined as the area within the red dashed box (20^oN-42^oN,103^oE-123^oE) in Figure 3 of the main text.

[b] NMB: normalized mean bias; R: Pearson correlation coefficient; RMSE: root mean square error

Table S5 Statistical comparisons of the *a priori* and *a posteriori* (from IE-3) simulated formaldehyde VCDs against the 1.7 times the formaldehyde VCDs observed by GOME-2A over eastern China [a]

Month	Formaldehyde VCDs (model against GOME-2A observations ×1.7)					
	NMB [b]		R [b]		RMSE [b] ($\times10^{15}$ molec cm^{-2})	
	a priori	*a posteriori* IE-3	*a priori*	*a posteriori* IE-3	*a priori*	*a posteriori* IE-3
JAN	0.092	-0.054	0.71	0.89	1.56	0.880
FEB	-0.20	-0.21	0.76	0.82	2.75	2.63
MAR	-0.32	-0.22	0.93	0.88	4.30	3.79
APR	-0.19	-0.18	0.92	0.94	2.96	2.78
MAY	-0.39	-0.15	0.85	0.92	5.12	2.51
JUN	-0.41	-0.15	0.78	0.93	7.64	3.84
JUL	-0.37	-0.19	0.80	0.92	6.79	3.62
AUG	-0.31	-0.17	0.85	0.90	4.69	2.77
SEP	-0.28	-0.11	0.82	0.93	3.64	1.73
OCT	-0.16	-0.12	0.89	0.92	2.46	1.98
NOV	0.12	-0.10	0.62	0.90	2.55	1.99
DEC	0.30	-0.048	0.38	0.82	3.14	1.80

[a] The eastern China domain is defined as the area within the red dashed box (20°N-42°N,103°E-123°E) in Figure 3 of the main text.

[b] NMB: normalized mean bias; R: Pearson correlation coefficient; RMSE: root mean square error

Table S6 Statistical comparisons of the *a priori* and *a posteriori* (from IE-1 and IE-3) simulated glyoxal VCDs against the glyoxal VCDs observed by GOME-2A over eastern China [a]

Month	Glyoxal VCDs (model against GOME-2A observations)								
	NMB [b]			R [b]			RMSE [b] ($\times10^{14}$ molec cm^{-2})		
	a priori	*a posteriori* IE-1	*a posteriori* IE-3	*a priori*	*a posteriori* IE-1	*a posteriori* IE-3	*a priori*	*a posteriori* IE-1	*a posteriori* IE-3
JAN	-0.075	-0.21	-0.14	0.65	0.80	0.61	1.25	0.900	1.23
FEB	-0.15	-0.19	-0.16	0.67	0.76	0.67	0.900	0.723	0.892
MAR	-0.44	-0.33	-0.36	0.67	0.64	0.65	1.59	1.43	1.48
APR	-0.55	-0.44	-0.55	0.85	0.78	0.85	1.67	1.47	1.66
MAY	-0.59	-0.32	-0.39	0.80	0.83	0.79	1.64	1.02	1.18
JUN	-0.55	-0.32	-0.33	0.80	0.85	0.86	2.14	1.52	1.51
JUL	-0.53	-0.31	-0.29	0.89	0.87	0.88	1.90	1.23	1.21
AUG	-0.52	-0.34	-0.33	0.74	0.77	0.75	1.82	1.31	1.30
SEP	-0.56	-0.41	-0.41	0.77	0.73	0.62	1.85	1.47	1.54
OCT	-0.48	-0.33	-0.44	0.85	0.83	0.83	1.44	1.09	1.33
NOV	-0.25	-0.26	-0.40	0.72	0.82	0.77	0.94	0.853	1.10
DEC	0.079	-0.21	-0.15	0.60	0.79	0.69	1.18	0.745	0.876

[a] The eastern China domain is defined as the area within the red dashed box (20°N-42°N,103°E-123°E) in Figure 3 of the main text.

[b] NMB: normalized mean bias; R: Pearson correlation coefficient; RMSE: root mean square error

Table S7 Statistical comparisons of the *a priori* and *a posteriori* (from IE-2 and IE-4) simulated formaldehyde VCDs against the formaldehyde VCDs observed by OMI over eastern China [a]

Month	Formaldehyde VCDs (model against OMI observations)								
	NMB [b]			R [b]			RMSE [b] ($\times10^{15}$ molec cm^{-2})		
	a priori	*a posteriori* IE-2	*a posteriori* IE-4	*a priori*	*a posteriori* IE-2	*a posteriori* IE-4	*a priori*	*a posteriori* IE-2	*a posteriori* IE-4
JAN	0.38	0.12	0.40	0.85	0.86	0.84	1.94	0.78	2.01
FEB	0.46	0.18	0.49	0.85	0.94	0.83	1.90	0.87	2.04
MAR	0.22	0.081	0.29	0.93	0.97	0.90	1.37	0.72	1.74
APR	0.38	0.22	0.45	0.88	0.93	0.86	1.96	1.24	2.26
MAY	0.47	0.29	0.56	0.94	0.89	0.95	2.56	1.71	3.04
JUN	0.25	0.16	0.39	0.81	0.84	0.81	2.45	1.82	3.39
JUL	0.27	0.19	0.41	0.81	0.84	0.77	2.49	1.96	3.69
AUG	0.38	0.24	0.58	0.85	0.85	0.80	2.82	1.96	4.36
SEP	0.36	0.15	0.48	0.84	0.83	0.82	2.29	1.30	3.10
OCT	0.29	0.10	0.38	0.94	0.94	0.92	1.59	0.85	2.08
NOV	0.36	0.16	0.42	0.84	0.86	0.84	1.78	0.97	2.02
DEC	0.70	0.23	0.72	0.83	0.92	0.83	2.66	1.00	2.75

[a] The eastern China domain is defined as the area within the red dashed box (20°N-42°N,103°E-123°E) in Figure 3 of the main text.

[b] NMB: normalized mean bias; R: Pearson correlation coefficient; RMSE: root mean square error

Table S8 Statistical comparisons of the *a priori* and *a posteriori* (from IE-2 and IE-4) simulated glyoxal VCDs against the glyoxal VCDs observed by OMI over eastern China [a]

Month	Glyoxal VCDs (model against OMI observations)								
	NMB			R			RMSE ($\times10^{14}$ molec cm^{-2})		
	a priori	*a posteriori* IE-2	*a posteriori* IE-4	*a priori*	*a posteriori* IE-2	*a posteriori* IE-4	*a priori*	*a posteriori* IE-2	*a posteriori* IE-4
JAN	-0.32	-0.33	-0.29	0.12	0.077	0.10	1.40	1.34	1.28
FEB	-0.46	-0.36	-0.32	0.45	0.38	0.33	1.49	1.32	1.26
MAR	-0.60	-0.38	-0.33	0.69	0.39	0.38	2.41	2.00	1.92
APR	-0.63	-0.33	-0.30	0.83	0.46	0.48	2.14	1.63	1.53
MAY	-0.63	-0.58	-0.26	0.82	0.80	0.80	2.04	1.88	1.09
JUN	-0.66	-0.51	-0.29	0.64	0.80	0.88	3.24	2.50	1.60
JUL	-0.65	-0.58	-0.45	0.78	0.79	0.83	2.92	2.63	2.10
AUG	-0.60	-0.50	-0.32	0.68	0.78	0.80	2.35	1.96	1.45
SEP	-0.65	-0.17	-0.16	0.68	0.61	0.75	2.35	1.35	1.05
OCT	-0.61	-0.14	-0.15	0.87	0.90	0.89	1.96	0.733	0.73
NOV	-0.46	-0.34	-0.28	0.53	0.48	0.52	1.39	1.19	1.09
DEC	-0.35	-0.38	-0.30	0.28	0.38	0.37	1.44	1.38	1.28

[a] The eastern China domain is defined as the area within the red dashed box (20°N-42°N,103°E-123°E) in Figure 3 of the main text.

[b] NMB: normalized mean bias; R: Pearson correlation coefficient; RMSE: root mean square error

74 **Table S9 Comparison of measured and simulated surface ozone concentrations over China**

Reference	Location	Platform	Time	Mixing ratio (ppb)			Bias (model - observation)	
				observations	*a priori* emissions [a]	average top-down emissions [a]	*a priori* emissions [a]	average top-down emissions [a]
Wang et al. (2012)	Beijing (39.8°N, 116.47°E)	Ozonesonde	14:00 LT, June 2002-2010	100 to 120	97	103	-23 to -3	-17 to 3
			14:00 LT, December 2002-2010	0 to 30	46	38	16 to 46	8 to 38
Sun et al. (2016)	Mt. Tai (36.25°N, 117.10°E, 1533 m a.s.l.)	Ground-based	Maximum daily 8h-average, June 2006-2015	108	97	103	-11	-5
Li et al. (2007)	Mt. Tai (36.25°N, 117.10°E, 1533 m a.s.l.)	Ground-based	13-17 LT, December 2004	46	46	38	0	-8
Li et al. (2007)	Mt. Hua (110.09°E, 34.49°N, 2064 m a.s.l.)	Ground-based	13-17 LT, June 2004	76	74	78	-2	2
			13-17 LT, December 2004	38	56	51	18	13
Xu et al. (2008)	Lin'an (30°3N, 119°7E, 139 m a.s.l.)	Ground-based	13-17 LT, June 2006	62	57	59	-5	-3
			13-17 LT, December 2005	27	56	48	29	21
Xu et al. (2016)	Waliguan (36.28°N, 100.9°E, 3816 m a.s.l.)	Ground-based	11-16 LT, June 1994-2013	61	60	61	-1	0
			11-16 LT, December 1994-2013	41	47	47	6	6
Zheng	Huizhou	Groun	13-17 LT,	34	36	36	2	2

et al. (2010)	(114.4°E,23.09°N)	d-based	June 2007					
			13-17 LT, December 2007	66	61	59	-5	-7
J.M. Zhang et al. (2009)	Lanzhou (36.13°N, 103.69°E, 1631m a.s.l.)	Ground-based	13-17 LT, June 2006	74	67	68	-7	-6
Li et al. (2015)	Changchun (43.9°N, 125.2°E, 237 m a.s.l.)	Ozonesonde	14 LT, June 13, 2013	62	66	69	4	7
Wang et al. (2015)	Akedala (47.1°N, 87.5°E, 502 m a.s.l.)	Ground-based	13-17 LT, July 2013	53	56	56	3	3
			13-17 LT, November 2013	21	36	36	15	15

[a] Simulated surface ozone concentrations were sampled from 13:00 to 17:00 local time.

Table S10 Surface measurements of SOC concentrations in June during 2006 and 2007 (Zhang et al., 2012)[a] and comparison to simulated SOC concentrations

Site	Site type	SOC concentration (μg C m^{-3})			Bias (model - measurement)	
		measurement	*a priori* simulation	average top-down emission estimates simulation	*a priori* simulation	average top-down emission estimates simulation
Chengdu (30.65°N, 104.03°E)	urban	3.79	1.31	1.61	-2.49	-2.18
Dalian (38.9°N, 121.63°E)	urban	2.64	1.32	2.09	-1.32	-0.55
Dunhuang (40.15°N, 94.68°E)	regional	2.51	0.38	0.41	-2.13	-2.11
Gaolanshan (36.0°N, 105.85°E)	regional	1.29	0.73	0.97	-0.56	-0.32
Jinsha (29.63°N, 114.2°E)	regional	1.81	1.40	1.85	-0.42	0.03
Lhasa (29.67°N, 91.13°E)	regional	2.34	0.47	0.48	-1.88	-1.86
LinAn (30.3°N, 119.73°E)	regional	2.51	0.95	1.29	-1.55	-1.22
Longfengshan (44.73°N, 127.6°E)	regional	1.89	0.85	1.09	-1.04	-0.79
Nanning (22.82°N, 108.35°E)	urban	1.70	0.72	0.74	-0.98	-0.96
Taiyangshan (29.17°N, 111.71°E)	regional	1.11	1.38	1.72	0.27	0.61
XiAn (34.43°N, 108.97°E)	urban	5.41	1.70	2.39	-3.71	-3.02
Zhengzhou (34.78°N, 113.68°E)	urban	2.78	1.59	2.17	-1.19	-0.62

Average		2.48	1.07	1.40	-1.42	-1.08

[a] SOC concentrations were computed using organic carbon measurements (μgC m^{-3}) and the EC-tracer approach (Zhang et al., 2012).

Table S1. Ground-based MAX-DOAS measurements of formaldehyde and glyoxal vertical column densities over China

Reference	Location		Time	Vertical column densities
			– 9-10 LT	13-14 LT
Formaldehyde [10^{16} molecules cm^{-2}]				
2010Wang et al. (2017)	Wuxi (31.57°N,120.31°E)	2011 – 2014	JF	0.7 [a]	0.8 [a]
			MA	0.9±0.15 [a]	1.1±0.26 [a]
			MJ	1.5±0.12 [a]	1.9±0.15 [a]
			JA	1.7±0.10 [a]	2.2±0.26 [a]
			SO	1.2±0.12 [a]	1.7±0.12 [a]
			ND	0.8±0.30 [a]	1.4±0.32 [a]
Lee et al. (2015)	Beijing (39.59°N, 116.18°E)	August 16 to September 11, 2006			1.79
De Smedt et al., (2015)	Beijing (39.98°N,116.38° E)	2008 – 2013	DJF	0.9±0.2 [b]	0.8±0.2 [b]
			MAM	1.3±0.3 [b]	1.2±0.2 [b]
			JJA	2.0±0.6 [b]	2.5±0.5 [b]
			SON	1.3±0.3 [b]	1.6±0.3 [b]
Li et al. (2013)	Back Garden, Guangdong (23.50°N, 113.03°E)	July 2006		1.3±1.0 [c]	1.3±0.7 [c]
Glyoxal [10^{14} molecules cm^{-2}]					
Li et al. (2013)	Back Garden, Guangdong (23.50°N, 113.03°E)	July 2006		6.8±5.2 [d]	11.4±6.8 [d]

[a] Bimonthly mean computed from Figure 12 of Wang et al. (2017)

[b] From hourly data in Figure 10 of De Smedt et al. (2015)

[c] From Figure 4 of Li et al. (2013)

[d] From Figure 5 of Li et al. (2013)

Split Cells
Split Cells
Formatted Table
Deleted Cells
Formatted: Font: Not Bold
Formatted: Font: Not Bold
Split Cells
Formatted: Font: Not Bold
Split Cells
Formatted: None, Space Before: 0 pt, After: 0 pt, Line spacing: single, Don't keep with next, Don't keep lines together
Formatted: Font: Not Bold
Formatted: Font: Not Bold
Formatted Table
Formatted Table
Formatted: Justified, No widow/orphan control

Table S2. Ground-based and ozonesonde measurements of surface ozone concentrations over China.

Reference	Location	Platform	Time	Mixing ratio (ppb)
Wang et al. (2012)	Beijing (39.8°N, 116.47°E)	Ozonesonde	14:00 LT, June 2002–2010	100–120
			14:00 LT, December 2002–2010	<30
Sun et al. (2016)	Mt. Tai (36.25°N, 117.10°E, 1533m a.s.l.)	Ground-based	Maximum daily 8h-average, June 2006–2015	108
Li et al. (2007)	Mt. Tai (36.25°N, 117.10°E, 1533m a.s.l.)	Ground-based	13–17 LT, December 2004	46
Li et al. (2007)	Mt. Hua (110.09°E, 34.49°N, 2064m a.s.l.)	Ground-based	13–17 LT, June 2004	76
			13–17 LT, December 2004	38
Xu et al. (2008)	Lin'an (30°3N, 119°7E)	Ground-based	13–17 LT, June 2005–2006	62
			13–17 LT, December 2005–2006	27
Xu et al. (2016)	Waliguan (36.28°N, 100.9°E, 3816m a.s.l.)	Ground-based	11–16 LT, June 1994–2013	61
			11–16 LT, December 1994–2013	41
Zheng et al. (2010)	Huizhou (114.4°E,23.09°N)	Ground-based	13–17 LT, June 2007	34
			13–17 LT, December 2007	66

Formatted ... Formatted Table ... Formatted ... Deleted Cells ... Merged Cells ... Formatted ... Deleted Cells ... Deleted Cells ... Split Cells ... Split Cells ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted ... Split Cells ... Formatted ... Formatted ... Split Cells ... Formatted ... Formatted ... Inserted Cells ... Inserted Cells ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted ... Split Cells ... Split Cells ... Split Cells ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted ... Inserted Cells ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted ... Formatted

J.M. Zhang et al. (2009)	Lanzhou (36.13°N, 103.69°E, 1631m a.s.l.)	Ground-based	13-17 LT, June 2006	74		
Li et al. (2015)	Changchun (43.9°N, 125.2°E)	Ozonesonde	14 LT, June 13, 2013	62		
Wang et al. (2015)	Akedala (47.1°N, 87.5°E, 502m a.s.l.)	Ground based	13-17 LT, July 2013	53		
			13-17 LT, November 2013	21		

Formatted: Font: Not Bold
Formatted: Left
Formatted: Font: Not Bold
Formatted: Font: Not Bold
Formatted: Font: Not Bold
Formatted: Font: Not Bold
Formatted: Font: Not Bold
Formatted: Left
Formatted: Font: Not Bold
Formatted: Font: Not Bold

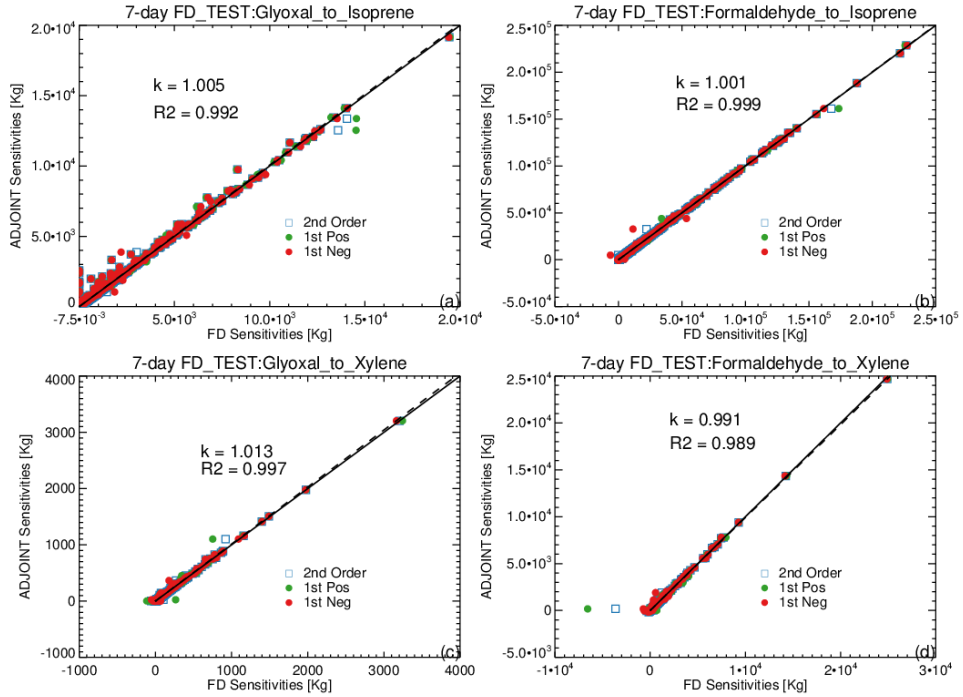


Figure S1. Finite difference test ~~(July 1th to 7th, 2007)~~ for the GEOS-Chem adjoint model~~. (a): sensitivities for the period between July 1st and July 7th, 2007. (a): Sensitivities~~ of global glyoxal burden to biogenic isoprene emission scale factor; (b): sensitivities of global formaldehyde burden to biogenic isoprene emission scale factor; (c) sensitivities of global glyoxal burden to anthropogenic xylene emission scale factor; (d): sensitivities of global formaldehyde burden to anthropogenic xylene emission scale factor. ADJOINT sensitivities and FD sensitivities were calculated by the adjoint model and the forward model, respectively. ~~'k' and 'R2' represent regression slope and square of correlation coefficient, respectively.~~ '2nd Order', '1st Pos' and '1st Nes' represent sensitivities calculated by central, forward, backward finite difference methods, respectively. The slopes of the regression lines (k) and the correlations (R^2) are shown in set.

Formatted: Left, Widow/Orphan control

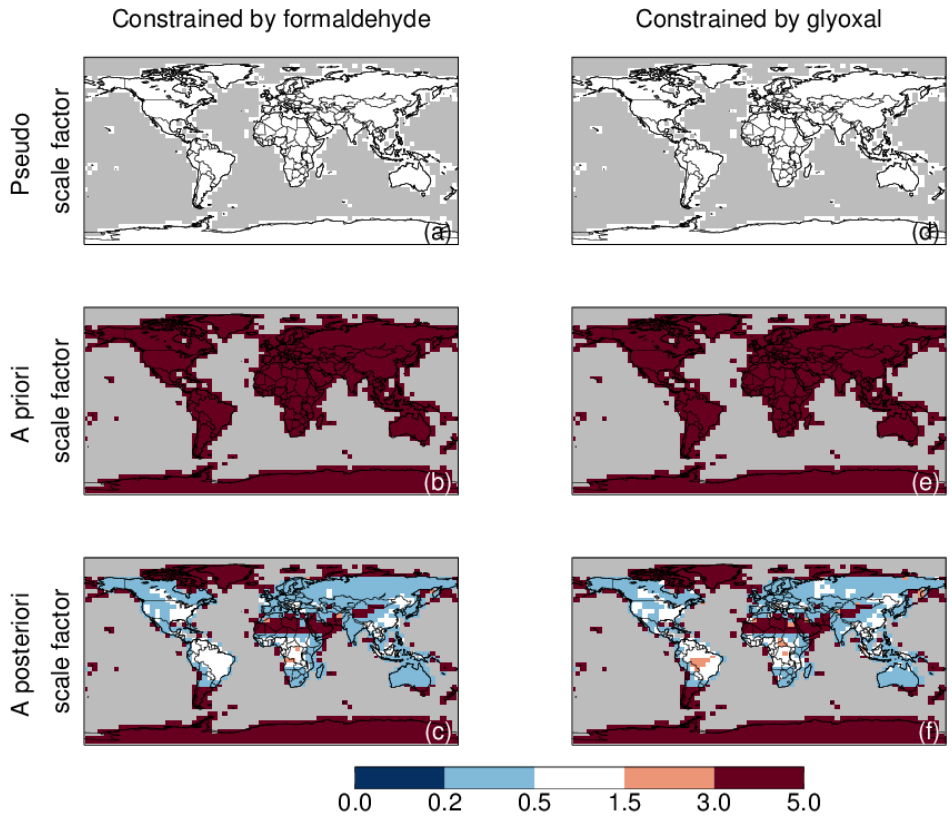
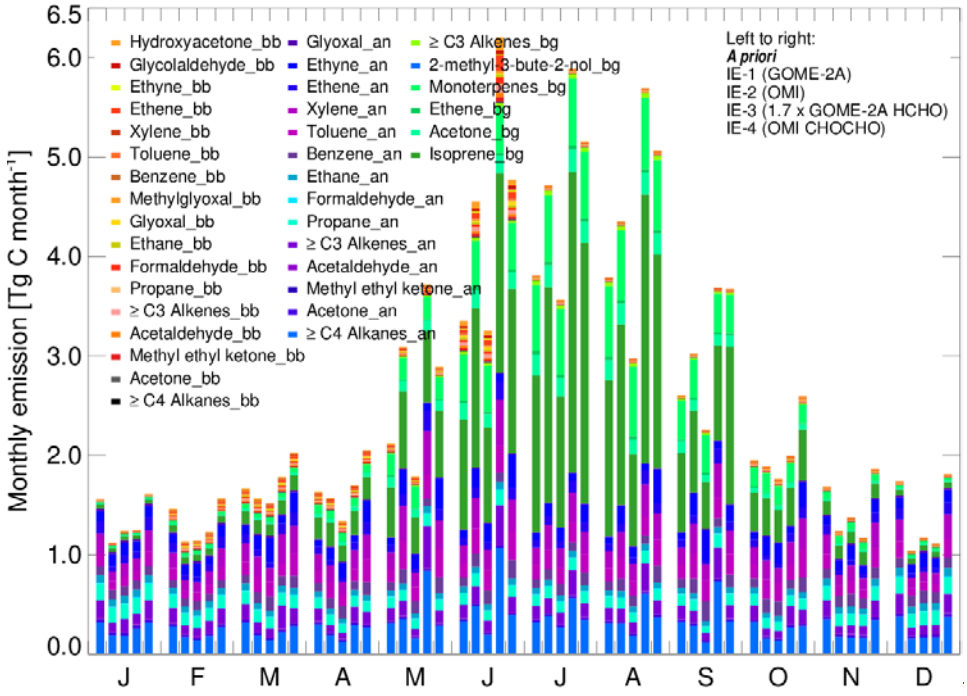


Figure S2. Pseudo isoprene emission scale factor ((a) and (d), uniformly set to 1.0 to generate pseudo observations), the *a priori* isoprene emission scale factor ((b) and (e), uniformly set to 5.0), and the *a posteriori* isoprene emission scale factor ((c) and (f)) in inversion tests (July 1th to 7th, 2007) constrained by pseudo observations of formaldehyde and glyoxal, respectively.

Formatted: Don't keep with next



~~Figure S3. Comparison of the *a priori* and *a posteriori* monthly Chinese NMVOC~~

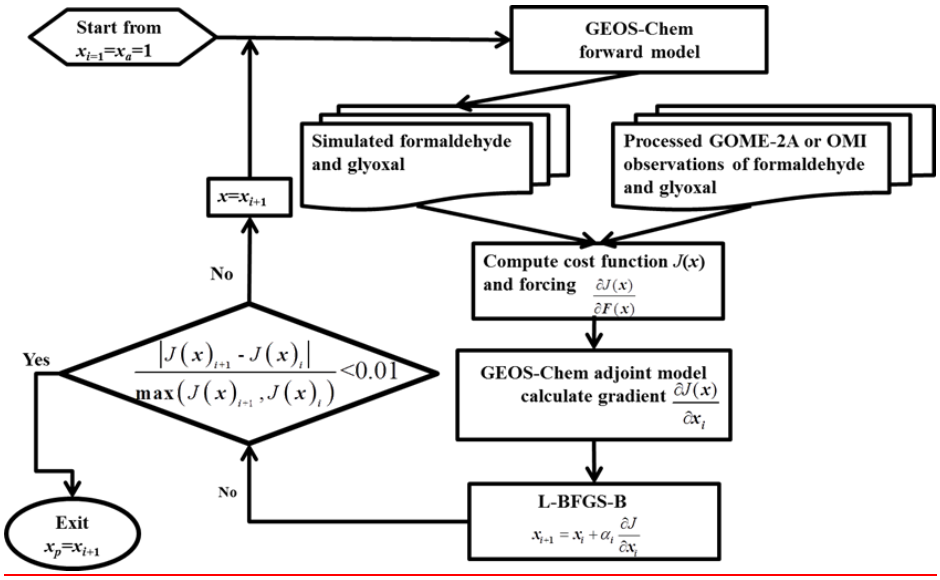


Figure S3. Protocol for the inversion experiments. For each month, we began by driving the GEOS-Chem forward model with the *a priori* emissions ($x_{i=1} = x_a = 1$) to simulate the monthly mean formaldehyde and glyoxal VCDs at satellite-crossing time. The simulated and satellite-observed VCDs were used to calculate the cost function, $J(x)$, and the forcing arrays ($\frac{\partial J(x)}{\partial F(x)}$). The adjoint of GEOS-Chem was then used to compute the cost function gradient ($\frac{\partial J(x)}{\partial x}$), and the next guess of the emission scale factor (x_{i+1}) was calculated using the Quasi-Newton L-BFGS-B algorithm (Byrd et al., 1995; Zhu et al., 1997), subject to the bounds $0.32 \leq x \leq 10$. These bounds were selected based on the largest uncertainties quoted in the literature on Chinese NMVOC emission estimates (Q. Zhang et al., 2009; Liu et al., 2012). The process was then iterated until the incremental relative reduction of the cost function ($\frac{|J(x)_{i+1} - J(x)_i|}{\max(J(x)_{i+1}, J(x)_i)}$) was less than 1% after at least five iterations. We took x_{i+1} from the last iteration as the optimized emission scale factor (x_p) and applied it to calculate the top-down emission estimate.

Field Code Changed

Field Code Changed

Field Code Changed

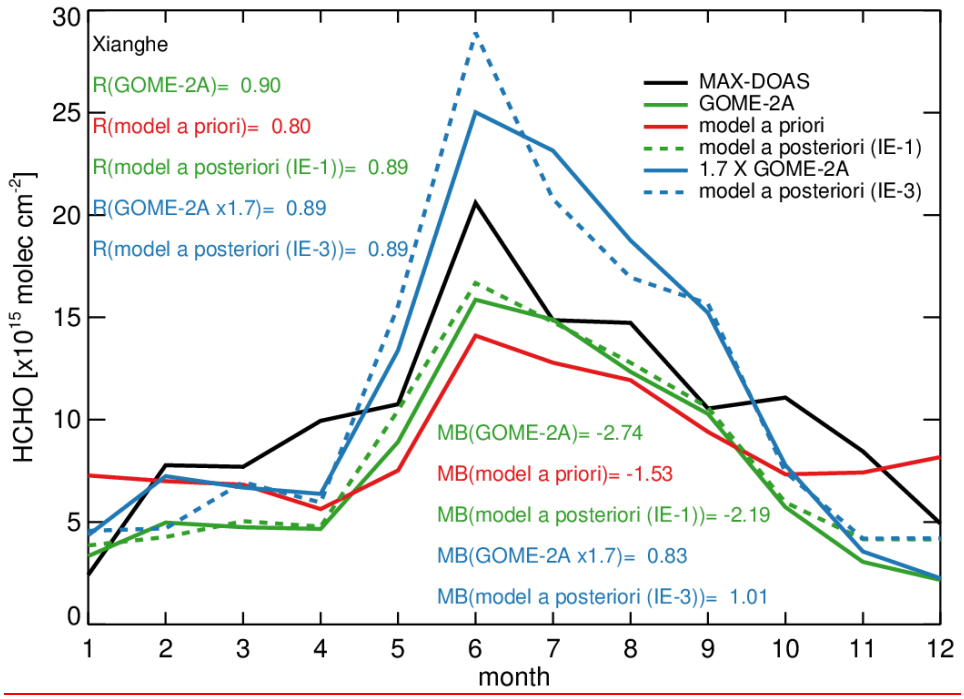


Figure S4. Measured and simulated monthly mean formaldehyde VCDs at Xianghe at GOME-2A overpass time: MAX-DOAS measurements (black line, Vlemmix et al., 2015), GOME-2A measurements (green solid line), GOME-2A measurements multiplied by 1.7 (blue solid line), monthly mean formaldehyde VCDs from the *a priori* simulation (red line), the IE-1 *a posteriori* simulation (green dashed line), and the IE-3 *a posteriori* simulation (blue dashed line). Pearson correlation coefficients (R) of the satellite-observed and simulated formaldehyde VCDs against the MAX-DOAS measurements are shown in the top left. Annual mean bias (MB, in units of 10^{15} molecules cm^{-2}) of the satellite-observed and simulated formaldehyde VCDs against the MAX-DOAS measurements are shown in the bottom right.

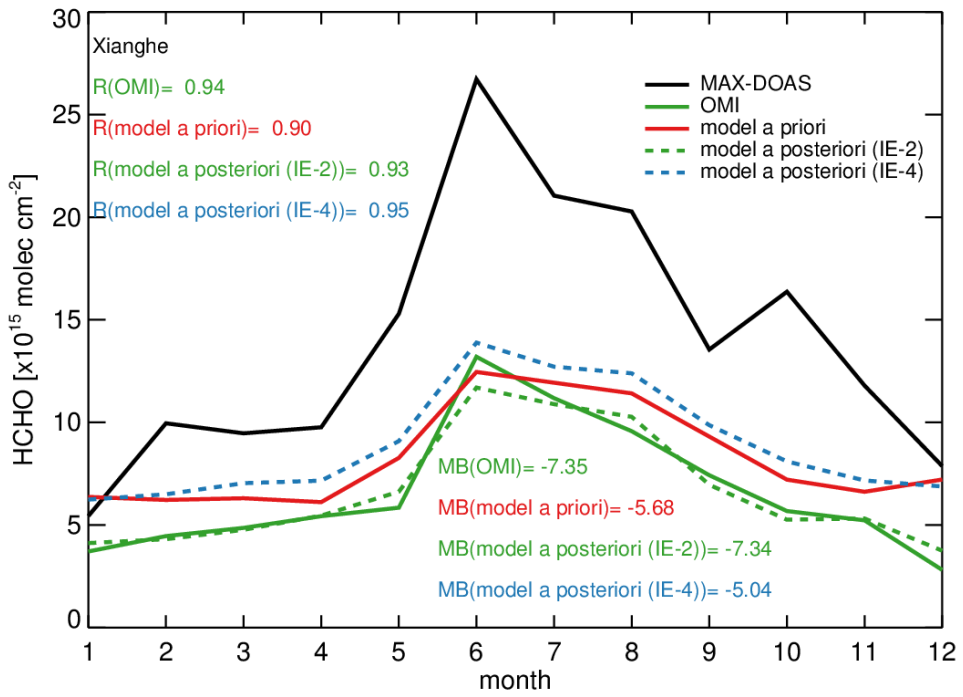


Figure S5 Measured and simulated monthly mean formaldehyde VCDs at Xianghe at OMI overpass time: MAX-DOAS measurements (black line, Vlemmix et al., 2015), OMI measurements (green solid line), monthly mean formaldehyde VCDs from the *a priori* simulation (red line), the IE-2 *a posteriori* simulation (green dashed line), and the IE-4 *a posteriori* simulation (blue dashed line). Pearson correlation coefficients (R) of the satellite-observed and simulated formaldehyde VCDs against the MAX-DOAS measurements are shown in the top left. Annual mean bias (MB, in units of 10^{15} molecules cm^{-2}) of the satellite-observed and simulated formaldehyde VCDs against the MAX-DOAS measurements are shown in the bottom right.

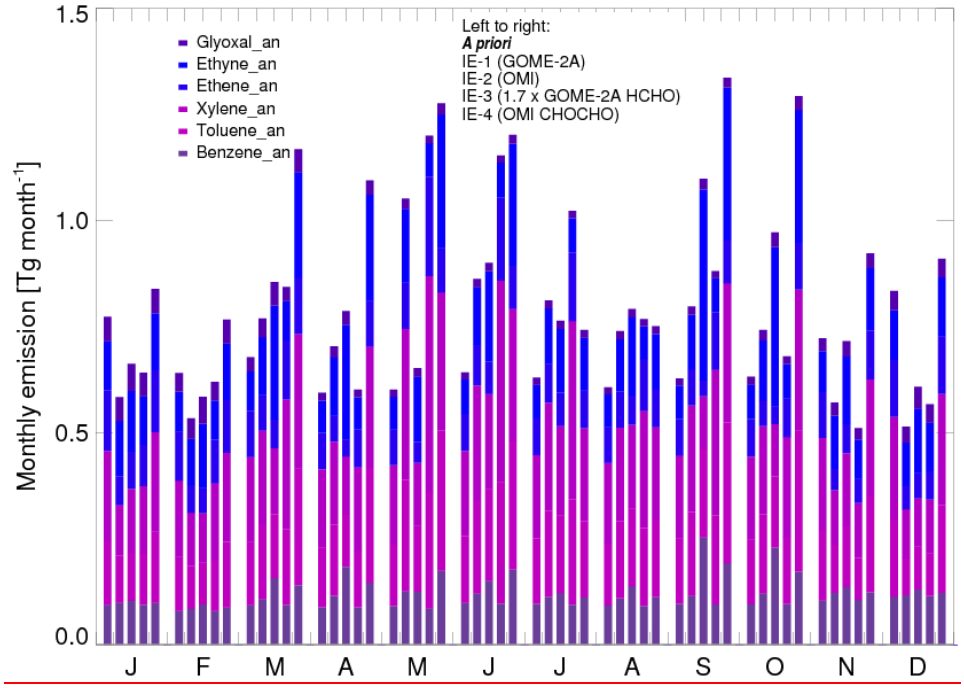
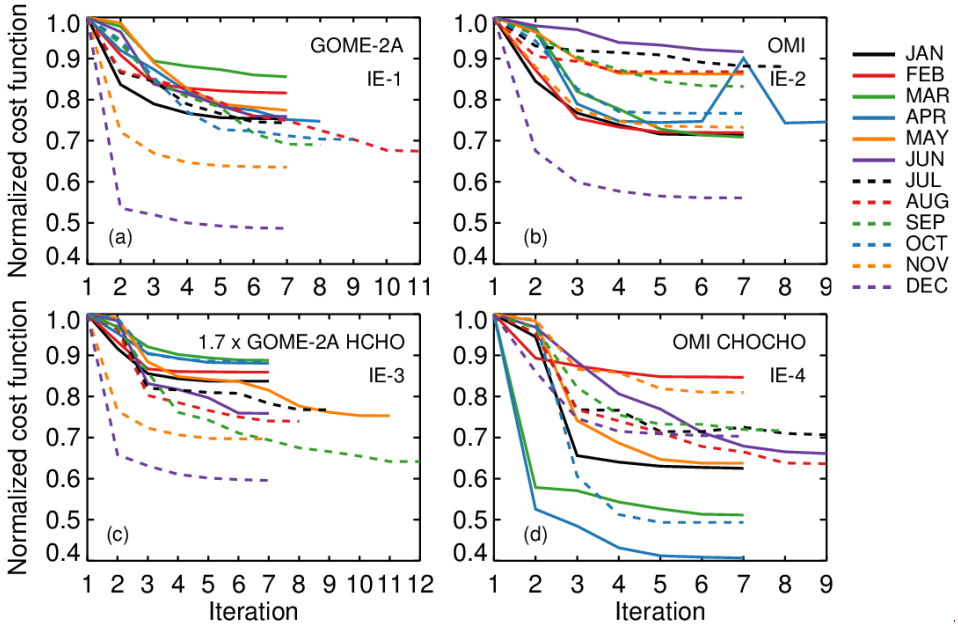


Figure S6. Comparison of the *a priori* and *a posteriori* monthly Chinese anthropogenic glyoxal precursors emission estimates for the year 2007. The bars from left to right for each month represent the *a priori* emission estimates, and the *a posteriori* emission estimates from IE-1, IE-2, IE-3, and IE-4, respectively. Color keys for the NMVOC species are shown inset; the ~~suffixes~~suffix 'an'~~, 'bb', and 'bg' indicate~~ indicates the anthropogenic source~~, biomass burning source, and biogenic source, respectively~~.


Normalized cost function
GOME-2A
IE-1
(a)
OMI
IE-2
(b)
1.7 x GOME-2A HCHO
IE-3
(c)
OMI CHOCHO
IE-4
(d)
Iteration
Iteration
JAN
FEB
MAR
APR
MAY
JUN
JUL
AUG
SEP
OCT
NOV
DEC

158

159

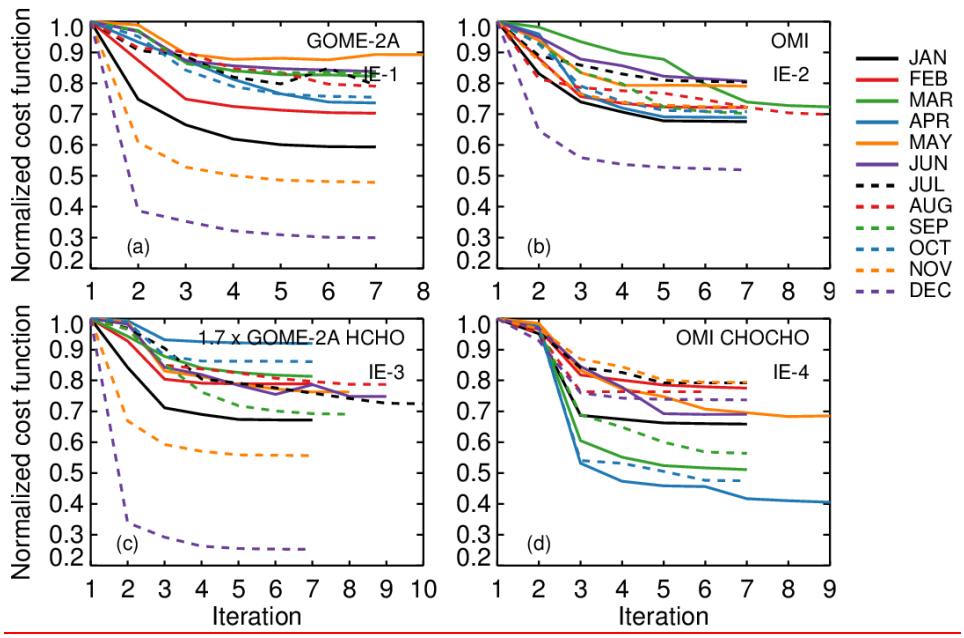


Figure ~~S4.~~S7. Change in the normalized cost function ($J(x)_i$ / $J(x)_{i=1}$) over China in the four inversion experiments: (a) IE-1, (b) IE-2, (c) IE-3, and (d) IE-4.

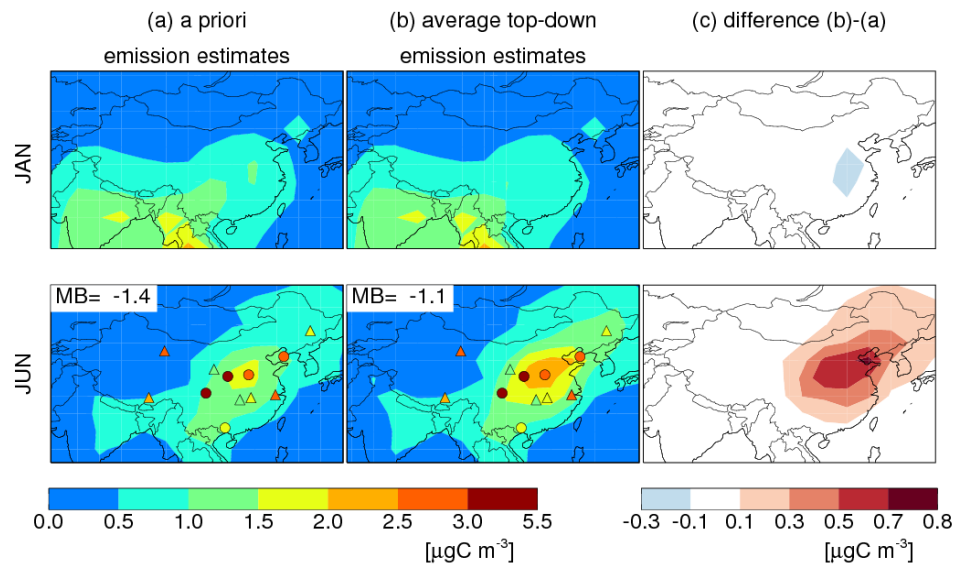


Figure S8. Simulated monthly mean surface secondary organic carbon (SOC) concentrations in June and December 2007 driven by (a) the *a priori* emissions and (b) our average top-down emissions, respectively, as well as (c) the differences. Overlaid symbols show the SOC measurements at 12 urban (circles) and regional (triangles) sites in China in June (Table S10). Mean biases (MB) of the simulated concentrations relative to surface measurements in June are shown inset.

Formatted: Font: +Headings (Cambria), 10 pt

Reference

Acarreta, J. R., De Haan, J. F., and Stammes, P.: Cloud pressure retrieval using the O2-O2absorption band at 477 nm, J. Geophys. Res., 109, doi: 10.1029/2003jd003915, 2004.

Boersma, K. F., Eskes, H. J., and Brinksma, E. J.: Error analysis for tropospheric NO2retrieval from space, Journal of Geophysical Research: Atmospheres, 109, D04311, doi:10.1029/2003jd003962, 2004.

Byrd, R. H., Lu, P. H., Nocedal, J., and Zhu, C. Y.: A Limited Memory Algorithm for Bound Constrained Optimization, Siam J. Sci. Comput., 16, 1190-1208, doi: 10.1137/0916069, 1995.

Chan Miller, C., Gonzalez Abad, G., Wang, H., Liu, X., Kurosu, T., Jacob, D. J., and Chance, K.: Glyoxal retrieval from the Ozone Monitoring Instrument, Atmos. Meas. Tech., 7, 3891-3907, doi:10.5194/amt-7-3891-2014, 2014.

De Smedt, I., Van Roozendael, M., Stavrakou, T., ~~Hendrick, F., Danckaert, T., Vlemmix, T., Pinardi, G., Theys, N., Lerot, C., Gielen, C., Vigouroux, C., Hermans, C., Fayt, C., Veefkind, P.,~~ Müller, J. F., Lerot, C., Theys, N., Valks, P., Hao, N., and van der A, R.: Improved retrieval~~F., and Van Roozendael, M.: Diurnal, seasonal and long-term variations~~ of global tropospheric formaldehyde columns ~~inferred from combined~~from GOME-2/MetOp-A addressing noise reduction and instrumental degradation issues, Atmos. Meas. Tech., 5, 2933-2949, doi:10.5194/amt-5-2933-2012, 2012.

González Abad, G., Liu, X., Chance, K., Wang, H., Kurosu, T. P., and Suleiman, R.: Updated Smithsonian Astrophysical Observatory Ozone Monitoring Instrument (SAO OMI~~ and GOME-2 observations~~) formaldehyde retrieval, Atmos. ~~Chem. Phys., 15, 12519-12545~~Meas. Tech., 8, 19-32, doi: 10.5194/~~acp-15-12519~~amt-8-19-2015, 2015.

Kleipool, Q. L., Dobber, M. R., de Haan, J. F., and Levelt, P. F.: Earth surface reflectance climatology from 3 years of OMI data, Journal of Geophysical Research, 113, doi: 10.1029/2008jd010290, 2008.

Kleipool, Q. L.: Transient signal flagging algorithm definition for radiance data, Tech. Rep. TN-OMIE-KNMI-717 TN-OMIEKNMI-717 TN-OMIE-KNMI-717 TN-OMIE-KNMI-717 TNOMIE-KNMI-717, Royal Netherlands Meteorological Institute, De Bilt, the Netherlands, 2005.

Lee, H., Ryu, J., Irie, H., Jang, S.-H., Park, J., Choi, W., and Hong, H.: Investigations of the Diurnal Variation of Vertical HCHO Profiles Based on MAX-DOAS Measurements in Beijing: Comparisons with OMI Vertical Column Data, Atmosphere, 6, 1816-1832, doi: 10.3390/atmos6111816, 2015.

Lerot, C., Stavrakou, T., De Smedt, I., Muller, J. F., and Van Roozendael, M.: Glyoxal vertical columns from GOME-2 backscattered light measurements and comparisons with a global model, Atmos. Chem. Phys., 10, 12059-12072, doi: 10.5194/acp-10-12059-2010, 2010.

Li, D., and Bian, J. C.: Observation of a Summer Tropopause Fold by Ozonesonde at Changchun, China: Comparison with Reanalysis and Model Simulation, Adv. Atmos. Sci., 32, 1354-1364, doi:

10.1007/s00376-015-5022-x, 2015.

Li, J., Wang, Z. F., Akimoto, H., Gao, C., Pochanart, P., and Wang, X. Q.: Modeling study of ozone seasonal cycle in lower troposphere over east Asia, J. Geophys. Res. Atmos., 112, doi: 10.1029/2006JD008209, 2007.

Li, X., Brauers, T., Hofzumahaus, A., Lu, K., Li, Y. P., Shao, M., Wagner, T., and Wahner, A.: MAX-DOAS measurements of NO2, HCHO and CHOCHO at a rural site in Southern China, Atmos. Chem. Phys., 13, 2133-2151, doi: 10.5194/acp-13-2133-2013, 2013.

Liu, Z., Wang, Y., Vrekoussis, M., Richter, A., Wittrock, F., Burrows, J. P., Shao, M., Chang, C.-C., Liu, S.-C., Wang, H., and Chen, C.: Exploring the missing source of glyoxal (CHOCHO) over China, Geophys. Res. Lett., 39, doi:10.1029/2012gl051645, 2012.

Spurr, R.: LIDORT and VLIDORT: Linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems, in: Light Scattering Reviews, edited by: Kokhanovsky, A., Springer, 3, 229–275, 2008.

Spurr, R. J. D.: VLIDORT: A linearized pseudo-spherical vector discrete ordinate radiative transfer code for forward model and retrieval studies in multilayer multiple scattering media, J. Quant. Spectrosc. Radiat. Transf., 102, 316-342, doi: 10.1016/j.jqsrt.2006.05.005, 2006.

Stavrakou, T., Muller, J. F., De Smedt, I., Van Roozendael, M., van der Werf, G. R., Giglio, L., and Guenther, A.: Global emissions of non-methane hydrocarbons deduced from SCIAMACHY formaldehyde columns through 2003-2006, Atmos. Chem. Phys., 9, 3663-3679, doi:10.5194/acp-9-3663-2009, 2009b.

Sun, L., Xue, L. K., Wang, T., Gao, J., Ding, A. J., Cooper, O. R., Lin, M. Y., Xu, P. J., Wang, Z., Wang, X. F., Wen, L., Zhu, Y. H., Chen, T. S., Yang, L. X., Wang, Y., Chen, J. M., and Wang, W. X.: Significant increase of summertime ozone at Mount Tai in Central Eastern China, Atmos. Chem. Phys., 16, 10637-10650, doi: 10.5194/acp-16-10637-2016, 2016.

Vlemmix, T., Hendrick, F., Pinardi, G., Smedt, I., De Fayt, C., Hermans, C., Piters, A., Wang, P., and Levelt, P.: MAX-DOAS observations of aerosols, formaldehyde and nitrogen dioxide in the Beijing area: comparison of two profile retrieval, Atmos. Meas. Tech., 2, 941–963, doi:10.5194/amt-8-941-2015, 2015.

Wang, H. Q., Ma, J. M., Shen, Y. J., and Wang, Y. A.: Assessment of Ozone Variations and Meteorological Influences at a Rural Site in Northern Xinjiang, Bull. Environ. Contam. Tox., 94, 240-246, doi: 10.1007/s00128-014-1451-y, 2015.

Wang, Y., Beirle, S., Lampel, J., Koukouli, M., De Smedt, I., Theys, N., Li, A., Wu, D. X., Xie, P. H., Liu, C., Van Roozendael, M., Stavrakou, T., Muller, J. F., and Wagner, T.: Validation of OMI, GOME-2A and GOME-2B tropospheric NO2, SO2 and HCHO products using MAX-DOAS observations from 2011 to 2014 in Wuxi, China: investigation of the effects of priori profiles and

aerosols on the satellite products, Atmos. Chem. Phys., 17, 5007-5033, doi: 10.5194/acp-17-5007-2017, 2017.

Wang, P., Stammes, P., R., v. d. A., Pinardi, G., and Roozendael, M. V.: FRESCO+: an improved O2 A-band cloud retrieval algorithm for tropospheric trace gas retrievals, Atmos. Chem. Phys., 8, 6565-6576, doi: 10.5194/acp-8-6565-2008, 2008.

Wang, Y., Konopka, P., Liu, Y., Chen, H., Muller, R., Ploger, F., Riese, M., Cai, Z., and Lu, D.: Tropospheric ozone trend over Beijing from 2002-2010: ozonesonde measurements and modeling analysis, Atmos. Chem. Phys., 12, 8389-8399, doi: 10.5194/acp-12-8389-2012, 2012.

Xu, W. Y., Lin, W. L., Xu, X. B., Tang, J., Huang, J. Q., Wu, H., and Zhang, X. C.: Long-term trends of surface ozone and its influencing factors at the Mt Waliguan GAW station, China - Part 1: Overall trends and characteristics, Atmos. Chem. and Phys., 16, 6191-6205, doi: 10.5194/acp-16-6191-2016, 2016.

Xu, X., Lin, W., Wang, T., Yan, P., Tang, J., Meng, Z., and Wang, Y.: Long-term trend of surface ozone at a regional background station in eastern China 1991-2006: enhanced variability, Atmos. Chem. Phys., 8, 2595-2607, doi: 10.5194/acp-8-2595-2008, 2008.

Zhang, J. M., Wang, T., Ding, A. J., Zhou, X. H., Xue, L. K., Poon, C. N., Wu, W. S., Gao, J., Zuo, H. C., Chen, J. M., Zhang, X. C., and Fan, S. J.: Continuous measurement of peroxyacetyl nitrate (PAN) in suburban and remote areas of western China, Atmos. Environ., 43, 228-237, doi: 10.1016/j.atmosenv.2008.09.070, 2009.

Zhang, Q., Streets, D. G., Carmichael, G. R., He, K. B., Huo, H., Kannari, A., Klimont, Z., Park, I. S., Reddy, S., Fu, J. S., Chen, D., Duan, L., Lei, Y., Wang, L. T., and Yao, Z. L.: Asian emissions in 2006 for the NASA INTEX-B mission, Atmos. Chem. Phys., 9, 5131-5153, 10.5194/acp-9-5131-2009, 2009.

Zhang, X. Y., Wang, Y. Q., Niu, T., Zhang, X. C., Gong, S. L., Zhang, Y. M., and Sun, J. Y.: Atmospheric aerosol compositions in China: spatial/temporal variability, chemical signature, regional haze distribution and comparisons with global aerosols, Atmospheric Chemistry and Physics, 12, 779-799, doi:10.5194/acp-12-779-2012, 2012.

Zheng, J. Y., Zhong, L. J., Wang, T., Louie, P. K. K., and Li, Z. C.: Ground-level ozone in the Pearl River Delta region: Analysis of data from a recently established regional air quality monitoring network, Atmos. Environ., 44, 814-823, doi: 10.1016/j.atmosenv.2009.11.032, 2010.

Zhu, C. Y., Byrd, R. H., Lu, P. H., and Nocedal, J.: Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization, ACM T. Math. Software, 23, 550-560, doi: 10.1145/279232.279236, 1997.

Formatted: Font: 10 pt, Not Bold

Formatted: Font: 10 pt