We thank the second reviewer for the constructive and detailed comments. In response, we have carefully edited the text to improve clarity throughout. We respond to each specific comment below. The reviewer's original comments are shown in red. Our replies are shown in black. The corresponding changes in the manuscript are shown in blue.

Reviewer 2:

The revised manuscript by Cao et al. represents a major improvement compared to the previous version. The authors have delivered a serious effort to comply with the criticism of the first report by including comparisons with ground-based observations, and by significantly improving the presentation and consistency. However, there are still weaknesses in the revised manuscript. Page numbers refer to the revised version.

1) Section 3 should present only comparisons with the a priori model, but a posteriori results are also shown in Figure S4 in the same section. This is inconsistent with the section title, please change for consistency.

> Thanks for pointing this out. We understand that the inclusion of the *a posteriori* results in Figures S4 and S5 may cause some confusion in Section 3. However, it was necessary to also show the *a posteriori* model simulation in Figures S4 and S5 to facilitate the visual comparisons between observations, the *a priori*, and the *a posteriori* results in Section 4 (lines 638-648). If we split the a posteriori results into separate figures, there would be a lot of repetition between the figures, and it would be much harder for the readers to make the visual comparison. To reduce the potential confusion to readers, we revised the main text to improve clarity:
>
> Main text: line 475 to line 490: Figure S4 compares the GOME-2A and the model a priori formaldehyde VCDs in 2007 against the multi-year (during the years 2010 to 2016) monthly mean formaldehyde VCD measured by MAX-DOAS at Xianghe (a rural site in the NCP) at GOME-2A overpass time (Vlemmix et al., 2015). The GOME-2A formaldehyde VCDs were consistent with the MAX-DOAS measurements in terms of the seasonal variation (R = 0.95) but showed an annual mean bias of -3.78×10^{15} molecules cm^{-2}. The interannual variability of the local formaldehyde VCDs (as represented by the standard deviation of the MAX-DOAS measurements) was relatively small and thus unlikely to be sole driver for the differences between the GOME-2A observations in 2007 and the MAX-DOAS measurements during 2010 to 2016. The seasonal variation of the model a priori formaldehyde VCDs were less consistent with that of the MAX-DOAS measurements (R = 0.81). Figure S4 also showed that, 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 -0.21×10^{15} molecules cm^{-2}. Figures 3 and 4 show that the differences between the satellite and MAX-DOAS measurements were 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.
>
> Main text: line 638 to line 648: Figure S4 also compared the model a posteriori formaldehyde VCDs in 2007 against the GOME-2 observations, the model a priori formaldehyde VCDs, and the MAX-DOAS measurements (during 2010-2016) at Xianghe at GOME-2 crossing time. Compared to the a priori, our a posteriori formaldehyde VCDs were in better agreement with the seasonal variation of the MAX-DOAS measurements (R values

increased from 0.81 for the a priori to 0.95 for IE-1 and 0.93 for IE-3). During the warm months (May to September), the monthly a posteriori formaldehyde VCDs from IE-1 and IE-3 bracketed the interannual variation of monthly formaldehyde VCDs measured by MAX-DOAS. For the rest of the year, both the GOME-2A observations and the a posteriori formaldehyde VCDs were systematically biased low relative to the MAX-DOAS measurements. As discussed before, these biases could not be fully accounted by the interannual variability of the local formaldehyde VCDs and was thus likely due to sampling or retrieval difference between the MAX-DOAS and the satellite.

Supplementary information: Figure S4

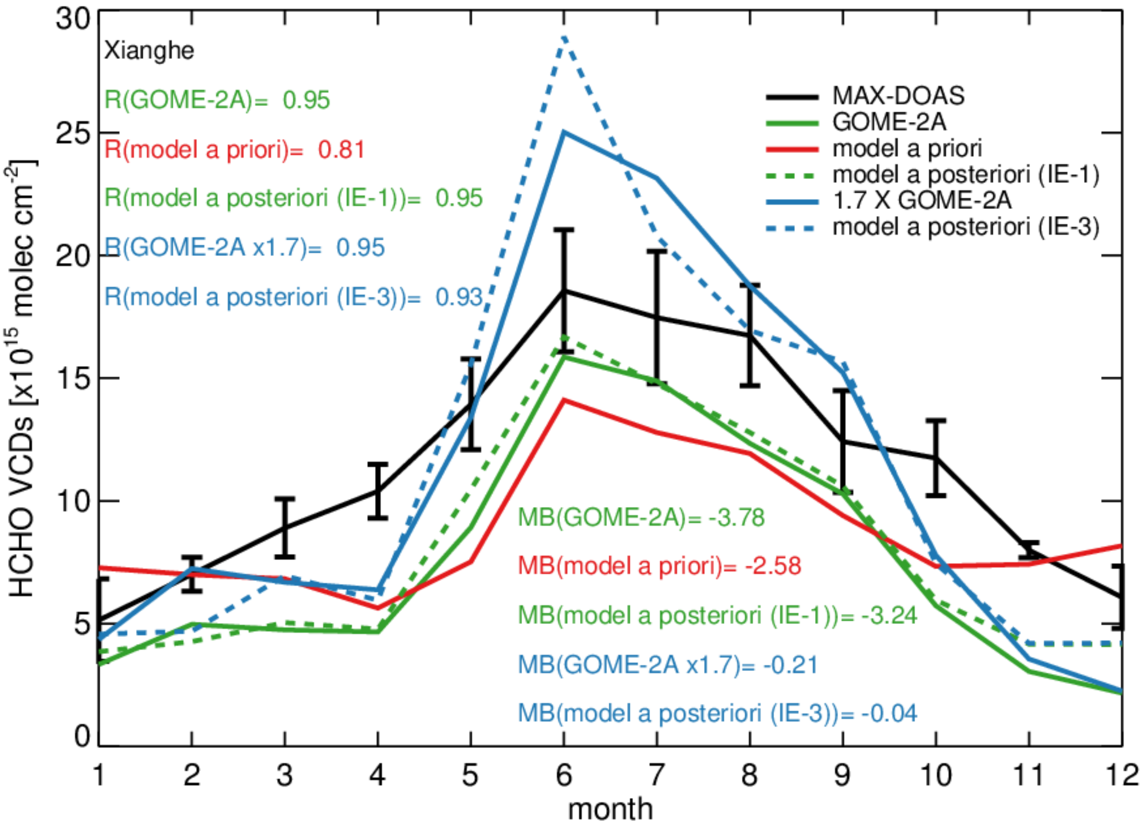

Figure S4. Measured and simulated monthly mean formaldehyde VCDs at Xianghe at GOME-2A overpass time: MAX-DOAS measurements (black line, monthly mean averages for the years 2010 to 2016 from Vlemmix et al., 2015), GOME-2A measurements (green solid line), GOME-2A measurements multiplied by 1.7 (blue solid 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 biases (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

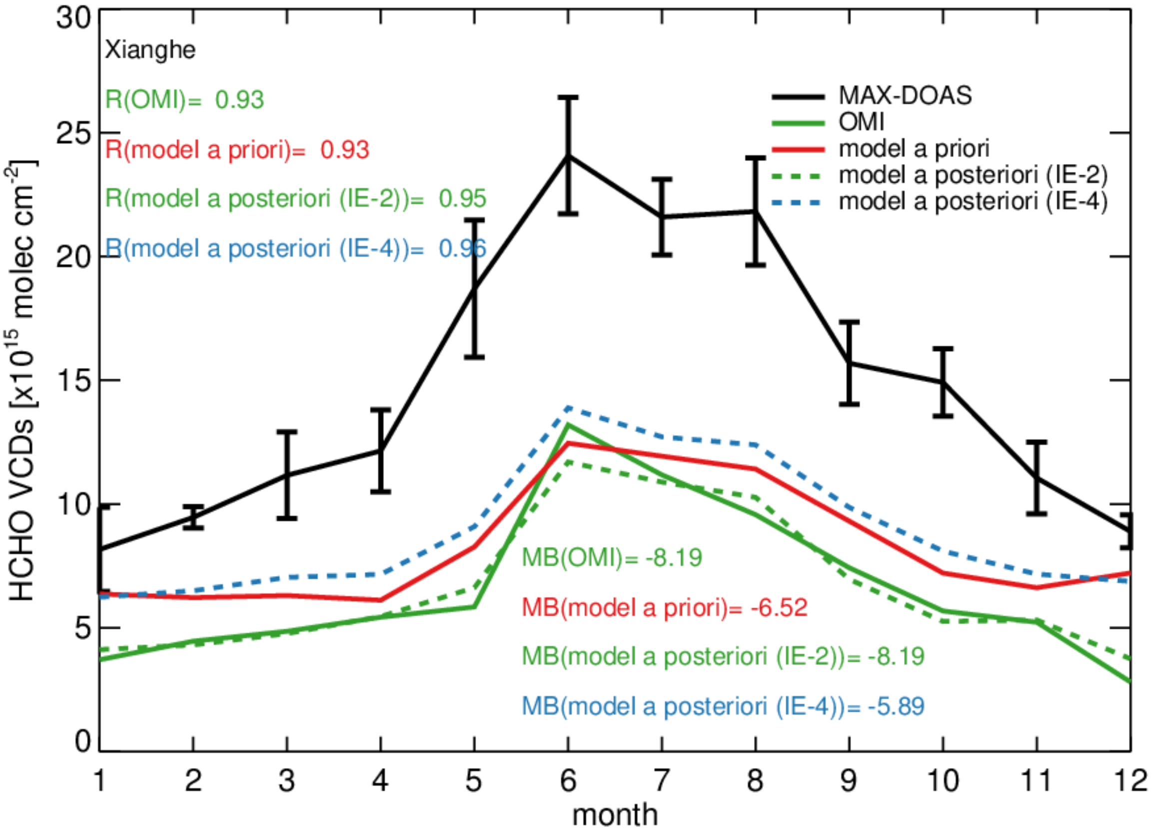

Figure S5 Measured and simulated monthly mean formaldehyde VCDs at Xianghe at OMI overpass time: MAX-DOAS measurements (black line, monthly mean averages for the years 2010 to 2016 from Vlemmix et al., 2015), OMI measurements (green solid line), monthly mean formaldehyde VCDs from the a priori simulation (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 biases (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.

2) The ground-based data are not obtained during the studied year (2007). How can this affect the comparisons? The interannual variability of CH_2O MAX-DOAS data should be discussed. For stations with ground-based data available for more than one year, the model results should be compared to a mean seasonal cycle.

Thank you for the suggestion. As suggested, we compared the satellite-observed and the simulated formaldehyde VCDs against the multi-year (2010 to 2016) monthly average and standard deviations of the MAX-DOAS formaldehyde measurements at Xianghe to represent the multi-year mean seasonal cycle and the interannual variability of the local formaldehyde VCDs. As shown in Table S3, Figure S4, and Figure S5, the differences between the satellite and the MAX-DOAS observations could not be fully accounted by the interannual variability of the local formaldehyde VCDs.

Main text: line 475 to line 490: Figure S4 compares the GOME-2A and the model a priori formaldehyde VCDs in 2007 against the multi-year (during the years 2010 to 2016) monthly mean formaldehyde VCD measured by MAX-DOAS at Xianghe (a rural site in the NCP) at GOME-2A overpass time (Vlemmix et al., 2015). The GOME-2A formaldehyde VCDs were consistent with the MAX-DOAS measurements in terms of the seasonal variation (R = 0.95) but showed an annual mean bias of -3.78×10^{15} molecules cm^{-2}. The interannual variability of the local formaldehyde VCDs (as represented by the standard deviation of the MAX-DOAS measurements) was relatively small and thus unlikely to be sole driver for the differences between the GOME-2A observations in 2007 and the MAX-DOAS measurements during 2010 to 2016. The seasonal variation of the model a priori formaldehyde VCDs were less consistent with that of the MAX-DOAS measurements (R = 0.81). Figure S4 also showed that, 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 -0.21×10^{15} molecules cm^{-2}. Figures 3 and 4 show that the differences between the satellite and MAX-DOAS measurements were 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.

Main text: line 638 to line 648: Figure S4 also compared the model a posteriori formaldehyde VCDs in 2007 against the GOME-2 observations, the model a priori formaldehyde VCDs, and the MAX-DOAS measurements (during 2010-2016) at Xianghe at GOME-2 crossing time. Compared to the a priori, our a posteriori formaldehyde VCDs were in better agreement with the seasonal variation of the MAX-DOAS measurements (R values increased from 0.81 for the a priori to 0.95 for IE-1 and 0.93 for IE-3). During the warm months (May to September), the monthly a posteriori formaldehyde VCDs from IE-1 and IE-3 bracketed the interannual variation of monthly formaldehyde VCDs measured by MAX-DOAS. For the rest of the year, both the GOME-2A observations and the a posteriori formaldehyde VCDs were systematically biased low relative to the MAX-DOAS measurements. As discussed before, these biases could not be fully accounted by the interannual variability of the local formaldehyde VCDs and was thus likely due to sampling or retrieval difference between the MAX-DOAS and the satellite.

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, Hebei (39.75N, 116.96E)	2010-2016	JAN	0.51±0.17	0.82±0.17
			FEB	0.70±0.07	0.95±0.04
			MAR	0.89±0.12	1.12±0.17
			APR	1.04±0.11	1.21±0.16
			MAY	1.39±0.19	1.87±0.28
			JUN	1.86±0.25	2.41±0.24
			JUL	1.75±0.27	2.16±0.15
			AUG	1.67±0.20	2.18±0.22

			SEP	1.24 ±0.21	1.57 ±0.17
			OCT	1.17 ±0.15	1.49 ±0.14
			NOV	0.80 ±0.03	1.11 ±0.15
			DEC	0.61 ±0.13	0.89 ±0.07
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)

3) l.467. 'Nevertheless, these ground-based measurements....', change to 'The measurements...' to avoid repetition with l. 465. Here and elsewhere check the text to avoid repetitions.

Thanks for pointing out this repetition. We changed the expression to:

Main text: line 465 to line 472: 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, the MAX-DOAS 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) in June, the formaldehyde VCDs over Xianghe were the highest among the three MAX-DOAS sites, reflecting the strong emissions from biomass burning in the NCP.

4) l.446. 'reflecting the seasonal biomass burning emissions there', change to 'reflecting the occurrence of seasonal biomass burning'

Changed as suggested. Thank you.

Main text: line 445 to line 446: In spring, GOME-2A formaldehyde VCDs were high over Southwest China and Southeast Asia, reflecting the occurrence of seasonal biomass burning.

5) l.542-548. Needs rewriting. Why 'two possible causes'? Change to 'Possible causes for this apparent contradiction could be...' What kind of errors do you mean in l.545? In the chemical oxidation scheme? Please elaborate on the possible errors.

Thanks for pointing out this lack of clarity. We re-wrote the sentences.

Main text: line 550 to line 556: It thus appeared that the constraints on Chinese NMVOC emissions indicated by the OMI formaldehyde and glyoxal observations were contradictory. Possible causes for this apparent contradiction could be: (1) the chemical production and losses of formaldehyde and glyoxal at different times of the day were not accurately simulated by the model, 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; and (2) it is also possible that there were different inherent biases in the OMI formaldehyde and glyoxal retrievals.

6) l.553. 'disparate, and apparently contradictory'

Changed the the main text as suggested, thankyou.

Main text: line 560 to line 562: The qualitative analyses in Section 3 showed that the GOME-2A and OMI retrievals of formaldehyde and glyoxal VCDs provided disparate, and apparently contradictory information on seasonal Chinese NMVOC emissions.

7) l.559. You mention that the cost function decrease is 8% in one inversion. This reduction is unusually modest. Can you explain in which inversion experiments this occurred and what were the reasons?

Thanks. The one case where the cost function was reduced by only 8% occurred in the optimization in IE-3 for April. In that case, the initial cost function was already small. We added the explaination in the main text.

Main text: line 567 to line 571: Relative to their respective initial cost function values, the optimized cost function values were reduced by 8% to 75% for all four experiments. The unusually modest 8% reduction occurred in the optimization in IE-3 for April. In that case, the initial cost function value was small; i.e., the a priori formaldehyde VCDs were already in good agreement with 1.7 times the GOME-2A formaldehyde VCDs (Figure 3 and Table S5).

8) l.617. 'The impacts...was', correct the verb. This type of error exists in other sentences too, please check carefully the grammar before resubmission.

Thanks for pointing out this type of error. We corrected it thoughout the whole main text.

Main text: line 629 to line 630: The impacts of satellite glyoxal observations on constraining Chinese glyoxal precursors emission estimates were further demonstrated in IE-4.

9) l.742. 'seasonal contrast', replace by 'seasonal amplitude'

Thank you for the suggestion. We revised the wording here to improve clarity and to comply with this comment and the next comment.

Main text: line 764 to line 769: As discussed above, three out of our four inversion experiments showed a stronger summer-versus-winter contrast in the NMVOC emissions, compared to the a priori emissions (Figure 2). We evaluated the impacts of this stronger seasonal amplitude in NMVOC emissions on surface ozone and secondary organic carbon (SOC) aerosol concentrations by driving the GEOS-Chem model with the a priori NMVOC emission estimates and with the average top-down emission estimates from our four inversion experiments, respectively.

10) l.744. 'stronger seasonal contrast' appears also in l.742. Try to avoid repetitions throughout the text.

Thank you for point this out. We revised the wording here to improve clarity and to avoid repetition.

Main text: line 764 to line 769: As discussed above, three out of our four inversion experiments showed a stronger summer-versus-winter contrast in the NMVOC emissions, compared to the a priori emissions (Figure 2). We evaluated the impacts of this stronger seasonal amplitude in NMVOC emissions on surface ozone and secondary organic carbon (SOC) aerosol concentrations by driving the GEOS-Chem model with the a priori NMVOC emission estimates and with the average top-down emission estimates from our four inversion experiments, respectively.

11) l.763-765. Bad phrasing. Start your sentence by e.g. 'The comparisons for ozone corroborate the stronger seasonal amplitude of the top-down NMVOC emissions derived in this study'.

Thanks. We re-wrote the sentence.

Main text: line 788 to line 789: These comparisons for surface ozone corroborated the stronger seasonal amplitude of the top-down NMVOC emissions derived in this study.

12) Section 6. Here comparisons are presented for June and December. Why did you focus only on those months? The year-round ozone variability at sites where data are available should be checked.

We focused on June and December because these were the two months when the differences between our averaged top-down emission estimate and the a priori emission estimate were the greatest. We wanted to evaluate the impacts of top-down VOCs emission on the surface ozone production during these two extreme conditions. Another difficulty was that regionally-representative surface ozone measurements were surprisingly scarce in China for all the other months in the literature. We explain our reasoning in the text:

Main text: line 773 to line 775: We focused here on surface ozone in June and December, when the differences in NMVOC emissions between our averaged top-down estimate and the a priori emission estimate were greatest.

13) Table 1. The last row seems to be equal to the average of the previous rows. Why don't you write this explicitly in the last row instead of 'Our top-down estimates'?

Changed as suggeted, thank you.

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
Average top-down estimates		20.2 (6.5 for aromatics)	19.2 (9.6 for isoprene)	2.48	41.9

[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.