Responses to Referee's Comments

We appreciate the reviewers' careful reading and insightful comments. We have carefully checked and addressed all comments and questions.

We have written the referee's comments in black and our responses to the comments in blue and italics. The revised sentences and paragraphs related to these comments are shown in red.

Referee #1:

1. Figure 8 suggests that retrieved slant columns using irradiance as source term are smaller than differential slant columns retrieved using radiance reference. This result is contradictory with physics retrievals and with figure 8 of the original submission. This results need to be further investigated or clarified.

Thank you for the comments. We conduct a de-striping process for the irradiance reference retrieval, which substracts latitudinally median columns of background regions (120-150°E) from each retrieved slant column, resulting in lower SCDs than the retrieved dSCDs based on radiance reference. However, the GEMS irradiance retrieval is not 100% mature and needs further improvement, especially for a de-striping process.

2. Throughout the text and in the relevant figures, please be accurate and refer to slant column when using irradiance source term and differential slant column when using radiance reference.

As you pointed out, we used SCDs from the irradiance retrieval and dSCDs from the radiance retrieval in the revised manuscript.

Referee #2:

I have read the revision and the responses to the reviewers, and think the authors have properly addressed the reviewers' comments. I have a few additional comments.

1. The authors have considered the a priori profile difference between GEMS and ground-based retrievals, but why the a priori profile difference between GEMS and TROPOMI retrievals are not considered in the comparison of VCDs from the two satellite sensors. The a priori profile difference can lead to AMF difference and finally contribute to VCD differences.

Thanks for the constructive comment, which we agree with. The comparison of two satellite products with identical a priori profiles is our ongoing work for future publication, and it takes some time for a long-term period of validation. In this study, we focused on comparing the operational products between GEMS and TROPOMI as they are.

2. The caption for Fig. 5c should be "absolute difference (c) (a-b)" rather than "relative difference (c) ((a-b)/b)". The unit of the colorbar of Fig. 5c is missing.

Thank you for pointing out the typo. We have replaced the caption of "relative difference (c) ((a-b)/b)" with "absolute difference (c) (a-b)". Also, we have added the unit of the colorbar.

Referee #3:

The paper describes the use and improvements of the GEMS formaldehyde retrievals. Also, it shows the results of comparison between GEMS and TROPOMI, as well as ground observations. The paper is well written and provides useful information for the scientific community. Therefore, I recommend its publication, after the minor comments below are addressed.

Specific comments

1. Line 27: Is the agreement substantially higher in terms of correlation coefficient only ? or the bias is much lower in this case? Any reason why the agreement is better over Northeast Asia?

Both the correlation coefficient and NMBs are better in Northeast Asia. In Northeast Asia, there are many anthropogenic sources for HCHO, so the spectral fitting can be more accurate than clean regions. In addition, considering that GEMS is located at 127°E, viewing geometric angles in Northeast Asia are smaller than those in western areas of the GEMS domain. That could be the reason why the agreement is better over Northeast Asia.

2. Line 41: Could you provide the ranges for the emissions and the uncertainties?

We added the range of global biogenic VOCs emission estimates for 2000 to 2019 (Sindelarova et al., 2022).

Line 40: NMVOCs also play a critical role in the formation of secondary organic aerosols (DiGangi et al., 2012). They are emitted from both anthropogenic and biogenic sources (Vrekoussis et al., 2010). The latter is more significant globally but has significant uncertainty of a priori emission estimates (424–591 Tg C yr^{-1}) (Abbot et al., 2003; Palmer et al., 2001; Sindelarova et al., 2022).

3. Line 100: How do you define clean?

As shown in line 126, we defined clean pixels where cloud radiance fraction is lower than 0.4 over the background region including Pacific Ocean. We have added the term "background" into the sentence at line 100.

We have tested the cloud masking criterion of 0.3 and 0.2, but there are no significant improvements in the retrieved dSCDs quality. To conserve stable number of observed clen pixels sampled in radiance references, we decided to use 0.4 as an optimum criterion of the clean pixel.

Line 100: In spectral fitting, the measured radiances over clean background regions, referred to as radiance references, can be used instead of the solar irradiance.

4. Line 110: I was wondering how sensitive the background correction is to the selection of the transport chemical model?

Background HCHO concentrations are mainly derived by the oxidation of CH_4 over the Pacific Ocean. The model does not show dramatic differences in background columns since the simulations have no anthropogenic or biogenic emission sources over the ocean. Figure S1 from Kwon et al. (2019) shows background HCHO columns derived from the initial version of the GEMS a priori profile (Table 2). By comparing Fig. S1 (Kwon et al., 2019) with Fig. S2 (this study), background columns do not vary sensitively by the selection of the model.

5. Line 127: Why did you select 0.4 as a threshold for the cloud radiance fraction for the clear-sky pixels? What is the % of the pixels taken into account after masking in Fig 1?

Please refer to our response about the cloud fraction threshold above. The radiance reference pixels with cloud masking (cloud fraction < 0.4) are about ~80 % of the total radiance references.

6. Line 153: There is an updated version for CEDS available. Are the emissions much different compared to the version used in the study?

McDuffie, E. E., Smith, S. J., O'Rourke, P., Tibrewal, K., Venkataraman, C., Marais, E. A., Zheng, B., Crippa, M., Brauer, M., and Martin, R. V.: A global anthropogenic emission inventory of atmospheric pollutants from sector- and fuel-specific sources (1970–2017): an application of the Community Emissions Data System (CEDS), Earth Syst. Sci. Data, 12, 3413–3442, https://doi.org/10.5194/essd-12-3413-2020, 2020.

Thank you for pointing out the updates in the latest reference paper of CEDS. Our a priori profile simulations are based on the CEDS version you mentioned in the comment. We have revised the reference literature and CEDS version referred to in the paper.

Table 2. Summary of the input options of a priori profiles for the GEMS HCHO algorithm.

Version	Initial	Operational
Model	GEOS-Chem (v9-01-02) (Bey et al., 2001)	GEOS-Chem (v13) (Bey et al., 2001)
Period	2014	August 2020–July 2021
Horizontal resolution	2° × 2.5°	0.25° × 0.3125°
Vertical layers	47	47
Meteorology	Modern-Era Retrospective Analysis for Research and Applications (Rienecker et al., 2011)	GEOS-FP (Goddard Earth Observing System -Forward Processing) assimilated meteorology
Emission inventory	**Biogenic** - Model of Emissions of Gases and Aerosols from Nature (MEGAN) version 2.1 (Guenther et al., 2006) **Anthropogenic** - Database for Global Atmospheric Research version 2.0 inventory (Olivier et al., 1996) - Mosaic fashion with the Intercontinental Chemical Transport Experiment Phase B (Zhang et al., 2009) **Monthly biomass burning** Global Fire Emissions Database (GFED) version 3 inventory (van der Werf et al., 2010)	**Biogenic** - MEGAN version 2.1 (Guenther et al., 2006) **Anthropogenic** - Community Emissions Data System v2020-08 (McDuffie et al., 2020) - KORUS version 5 over Asia (Woo et al., 2020) **Monthly biomass burning** GFED version 4 inventory (Giglio et al., 2013)

7. Line 165: the meaning of the sentence is not very clear to me.

We have revised the sentence for better readability.

Line 165: We considered the polarization sensitivity vectors of the instrument as an additional absorption cross-section, termed a pseudo absorber, in the spectral fitting.

8. Figure 5&7&8: The orientation of the labels in these plots is different from the others. The labels in the colorbal are written from top to bottom, whereas in these plots they are written from bottom to top.

We modified the orientation of labels in Fig. 5, 7, and 8 as written from bottom to top format.

9. Figure 5: panel c shows the relative or the absolute difference (a-b as indicated in the caption)?

Thank you for pointing out the typo. As mentioned in the Referee#2's comment, we revised it.

10. Line 177-182: Could you explain better and elaborate about the determination of the fitting window?

We added the sentence to elaborate the determination of our fitting window selection.

Line 180: Theoretically, retrieved dSCDs over the reference sector should be zero. We selected the fitting window of 329.3–358.6 nm for the GEMS HCHO operational retrieval based on the dSCDs closest to zero, lowest fitting residuals, and lowest fitting uncertainty (Fig. 6).

11. Line 256 and line 257 : Any reference for the reason for the highest HCHO? The peak for Hanoi by GEMS is not in spring but around September.

Following the statement of Baek et al. (2014), the high HCHO VCDs over Hanoi are mainly due to the emissions from anthropogenic sources. These emissions consistently elevate the HCHO concentrations throughout the year, resulting in small fluctuations from September to March. Furthermore, cloud radiance fractions continuously exceed 0.3 throughout March. Elevated low-level cloud fractions could lead to increased AMF and lower VCDs.

12. Line 283: Why was the comparison done only for this site? Why in the comparison with TROPOMI you used the averaged values over pixels within a 20 km x 20 km grid box centered on the center of each city, while now you set a grid of 0.4 0.4 degrees?

1. The available ground-based HCHO observation in East Asia is limited during the first year of the GEMS operation. We decided to use Xianghe MAX-DOAS and FTIR products because they cover most of these periods and provide averaging kernel and a priori profile variables, which enable precise intercomparison in satellite validation.

2. We acknowledge the inconsistency of the collocation criteria and have adjusted the spatial criterion for ground-based validation from a grid size of 0.4° × 0.4° to 20 km × 20 km. The general evaluation results in the updated ground validation (Figures 11, 12, S8, S10, S11) are similar to the previous comparison.

13. Line 295: is the agreement better just because GEMS and ground obs use the same info (same a priori profile)?

The improved agreement between GEMS and ground-based observations could be from sharing the same a priori profile and the smoothing effect of averaging kernel. We have added brief descriptions for the potential *discrepancies from the vertical sensitivity.*

Line 298: In addition, different vertical sensitivities of satellite and ground-based observations could cause inevitable discrepancies in their vertical columns.

14. Line 342 and 345: How much is the high positive bias?

We have added the exact number of the high positive biases in the sentence.

Line 347: The initial algorithm caused high positive biases (10–40 %) in the slant columns from the spectral fitting, primarily due to radiance references constructed under cloudy conditions with high reflectance.

15. Line 358: I would give numbers for the correlation coefficient and the biases in the conclusions to summarise the main findings.

We have added the numbers for the correlation coefficient and the biases in the conclusions.

Line 363: We found high correlations between GEMS and TROPOMI HCHO VCDs and a good representation of seasonality with the regional characteristics of GEMS HCHO among the major cities (r=0.58–0.82), showing active emissions from biogenic and anthropogenic sources over East Asia.

Line 370: GEMS produced approximately 30 % lower VCDs than MAX-DOAS but showed high correlations (r=0.77) and consistent seasonality with MAX-DOAS during the year.

Line 372: The MAX-DOAS- and FTIR-recalculated VCDs with the a priori profiles of GEMS decreased, showing reduced NMBs (−47.4 % to −31.5 % and −38.6 % to −26.7 % for MAX-DOAS and FTIR, respectively) against GEMS.

16. Figure S4: The figure can be read easier if the sites are ordered by increasing longitude in the legend.

We have modified the order of legend by longitude.

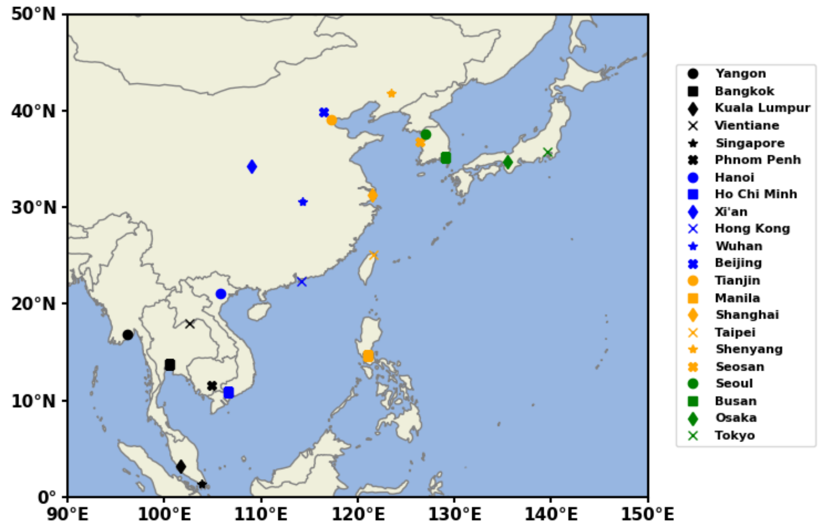

Fig. S4. Regions selected for the comparison between GEMS and TROPOMI.

17. Figures: The use of italics in the figures labels and units is not consistent in the main paper and the supplement.

We used labels with the italic style in Fig. 4, Fig. 11, Fig. S5, and Fig. S9

References

Baek, K. H., Kim, J. H., Park, R. J., Chance, K., & Kurosu, T. P. (2014). Validation of OMI HCHO data and its analysis over Asia. Science of the Total Environment, 490, 93–105.

McDuffie, E. E., Smith, S. J., O'Rourke, P., Tibrewal, K., Venkataraman, C., Marais, E. A., Zheng, B., Crippa, M., Brauer, M., and Martin, R. V: A global anthropogenic emission inventory of atmospheric pollutants from sector-and fuel-specific sources (1970–2017): an application of the Community Emissions Data System (CEDS), Earth Syst Sci Data, 12, 3413–3442, 2020.

Sindelarova, K., Markova, J., Simpson, D., Huszar, P., Karlicky, J., Darras, S., and Granier, C.: High-resolution biogenic global emission inventory for the time period 2000–2019 for air quality modelling, Earth Syst. Sci. Data, 14, 251–270, https://doi.org/10.5194/essd-14-251-2022, 2022.