Response to Reviewer #2

We are grateful to the reviewer for the valuable comments that facilitate the important improvements of the original manuscript. We list the point-by-point responses below. The reviewer's comments are marked black and our responses are marked dark blue. Line numbers refer to the discussion paper acp-2021-820. We attach the updated figures and supplementary information in the end.

This paper presents model evaluations of HCHO against a combination of satellite, ground, and aircraft observations in a very sensitive area but rarely studied in terms of atmospheric composition. It is in general well written and fits well within the scope of ACP. It will add important insights regarding HCHO source and variability at high latitudes. I'd recommend it for publication. A few concerns and comments are listed below for considerations for clarification or potential improvement:

1. Isoprene (and monoterpenes) may not be good tracers to directly evaluate their emissions given its short lifetime. Do MVK+MACR observations available during Atom? They would provide a more regionally representative signal for isoprene emissions.

We now include MVK+MACR measurements from ATom-1 aircraft campaign for our model evaluation. Figure 1 is updated as follows.
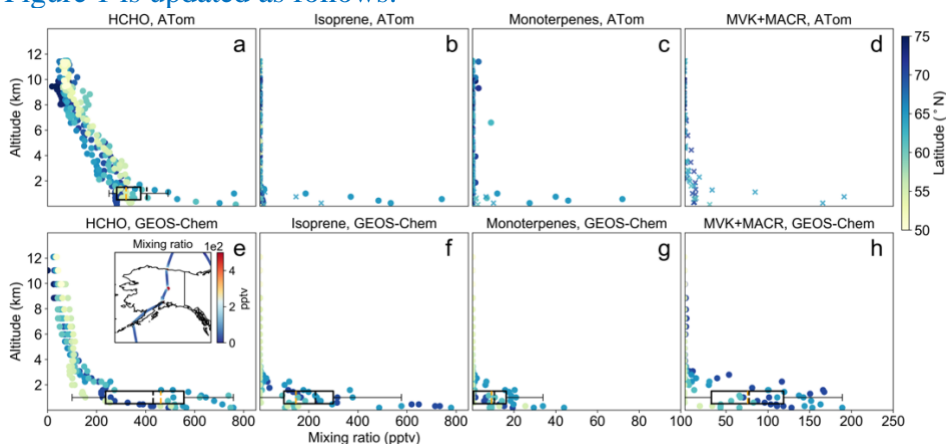
We add a sentence in model-ATom comparison section 3: "In addition, modeled MVK+MACR shows average mixing ratio of 78 pptv while observations show 38 pptv, providing additional constraints on isoprene oxidation."

2. For model evaluation with Atom-1: why was 1 hour averaged model output used? The model was running using 10 min /20 min time steps? Would a higher time resolution comparison better help resolve the vertical profiles?

We now change the model output to 1-min time resolution using Planeflight diagnostics. We revised the text as:
"For the comparisons between observations and model shown below, we sample the model output along the flight track at the flight time with 1-min time resolution."

3. Figure 1: The model shows a somewhat large underestimate of HCHO in the free troposphere and the boundary layer? It seems so too for isoprene and monoterpenes? It would be worth emphasizing as they reflect some knowledge gaps that might be the first time shown in the literature.

Based on ATom and the model PlaneFlight diagnosis (mentioned in comment 2), we discussed the model underestimation above boundary layer (>2km). The HCHO VCD derived from model and ATom are comparable. We reorganized two paragraphs in ATom-model comparison section 3:

"Our model shows reasonable agreement with measurements in the boundary layer (<2 km). Modeled HCHO has a mean mixing ratio of 431 pptv, slightly higher than the observed value (405 pptv). Modeled isoprene has a mean mixing ratio of 227 pptv in the boundary layer, in agreement with observed values given the large variability of observations. Both observations and model show significantly less monoterpenes compared to isoprene, on the order of tens of pptv. In addition, modeled MVK+MACR shows average mixing ratio of 78 pptv while observations show 38 pptv, providing additional constraints on isoprene oxidation.

Our model tends to underestimate HCHO above boundary layer (>2 km). We show in Figure 1 that mean modeled HCHO is 98 pptv at 3-6 km, and ~46 pptv at 6-10 km, compared to observed values of 212 pptv and 104 pptv respectively. The reason is unknown, but could be related to the large underestimate of CH_3OH in the same region (Bates et al., 2021). As a result, the model-derived HCHO VCD is likely lower than that calculated from ATom measurements, by 2.5×10^{15} molecules cm^{-2}. Such bias may lead to a systematic bias on our estimate of background HCHO VCD_0 in this region."

4. Figure 2: why were the model results for the regional average used? Any justification that it should be this way? Or is it better than using the model output for the grid cell containing the station? Given the high-resolution model, I don't quite understand why such a regional average is needed.

The ground-based MAX-DOAS part is now removed in the revised text.

5. a) Figure 2 shows some interesting features of enhancements captured by the model. MAX-DOAS however seems quite a noise although it is hourly data. Can the comparison be done more quantitatively while still being able to factor in MAX-DOAS instrument uncertainty? How does the model perform in non-fire conditions vs fire influence conditions? Can any quantitative results be interpreted here? Would the MAX-DOAS be useful to compare to the TROPOMI HCHO products directly?

The ground-based MAX-DOAS part is now removed in the revised text.

b) Line 385: if the detection limit of MAX-DOAS is 1e15 molecules cm-2, then there'd be only a few data points above the detection limit in Figure 2? Am I interpreting it correctly?

The ground-based MAX-DOAS part is now removed in the revised text.

      c) Figure 2: some panels lack y scale.

The ground-based MAX-DOAS part is now removed in the revised text.

6. Fire influence in the model: Is that the fire influence in the nested domain, or is that global? Fire smoke in other regions may transport to AK and affect 2019 summer? Depending on how this sensitivity was set up (does it reflect the fire influence within the AK domain, or globally), it may be the reason why the fire VOC emission within AK is only a factor 1 to 2 higher than biogenic emissions, but $dVCD_{GC, \, Fire}$ is 10 X higher than $dVCD_{GC, \, Bio}$? i.e., Lines 510-513

The nested simulations are based on boundary conditions from a global run with wildfire and biogenic emission turned on. We made another nested simulation using a new boundary condition from a global run with wildfire and biogenic emission turned **off**.

We add a sentence at section 4.3: "Due to model sensitivity tests, intercontinental transport of wildfire emissions contributes a minor part of $dVCD_{GC,Fire}$ (~1% in interior Alaska, ~10% in southwest Alaska for 2019 July)."

7. From the comparison of TROPOMI HCHO VCD with the GEOS-Chem HCHO VCD, it seems the model is predicting HCHO well and there is no significant knowledge gap regarding HCHO from biogenic VOCs or fire smoke in Alaska? But from the comparison with Atom observations, the model seems underpredicting HCHO, while the MAX-DAOS comparison may not be too quantitative? How would these be reconciled, particularly regarding Atom and TROPOMI evaluations? Overall, I was hoping to see those evaluations could be done more quantitatively. How exactly does the model HCHO compare to observations? Does the model underpredict HCHO at the surface or throughout the troposphere, which seems to be the case when compared to Atom?

First, currently we cannot take advantage of the vertical profile optimal estimation of MAX-DOAS, so we have to remove this part.

Second, GEOS-Chem does have discrepancies on VOC vertical profiles comparing to ATom measurements, depends on VOC species. But for HCHO, only 35% of the difference is in <2km layer where dVCD mainly loads. Model sensitivity test shows that >90% of the HCHO total column discrepancy is on background.

We quantitatively discussed the underestimation of model HCHO in above 2 km layers, and the reasonable agreement with ATom in <2km layer. Please see the two paragraphs in our response for the comment 3.

At the last paragraph of satellite-model comparison section 4.3, we discussed the uncertainty that difference sources can introduce to TROPOMI VCD, including the reprocessing based on our model: "Satellite retrievals of HCHO in wildfire region remains as a major challenge. One source of uncertainty stems from *a priori* profiles used in AMF calculation (Kwon et al., 2017). We find that for regions with heavy smoke, our calculated GEOS-Chem AMF_{GC} is 50% lower than the AMF_{SAT} in the operational product, due to the difference in HCHO vertical profiles (Figure S3). As a result, our reprocessed HCHO VCD product, $VCD_{SAT,GC}$, is higher than the operational product by $3\text{-}5\times10^{15}$ molecules cm^{-2} in heavy smoke regions in July 2019 (Figure S10). Another uncertainty lies in the aerosol optical properties. Wildfire smoke is a major source of brown carbon (June et al., 2020). As the current retrieval algorithm for HCHO does not account for absorbing aerosols, smoke can reduce the sensitivity of satellite measurements to atmospheric layers below and above the aerosol layer, leading to a smaller AMF by 20-30% (Jung et al., 2019; Martin et al., 2003)."

We discuss the challenges of comparing model with satellite HCHO in conclusion section 5: "Quantifying HCHO at northern high latitudes can be further improved in several aspects. First, we show that background signal, often taken from model output, can be dominant in final product of HCHO VCD. However, model results differ significantly on HCHO even over the Pacific Ocean (Figure S8), leading to a large uncertainty in the final satellite product in this region. Second, reference sector correction represents another major uncertainty (Zhu et al., 2020). This is particularly a problem for Alaska, as it lies in the reference sector defined by most retrieval algorithms (González Abad et al., 2015; De Smedt et al., 2018). Any systematic bias in Alaska can propagate to retrievals in other regions. Third, pristine regions can also be influenced by wildfire plumes, which can largely impact HCHO retrieval. Future work is warranted to improve HCHO retrieval and therefore our understanding of HCHO at northern high latitudes."

8. Lines 460-465: Here and a few other places claim the VCD is mostly driven by wildfire direct emission, rather than secondary production during fire conditions, but it is according to model sensitivity tests. The more quantitative comparison between model and observation may show the model is underpredicting HCHO vertical distribution (Item 7), and the satellite data comparison approach may be biased since it uses the model information for reprocessing (Item 9). I wonder if the observations and the model evaluations have any evidence to support that the direct fire emission of HCHO drives its VCD, rather than secondary productions.

Our discussion on the difference between model and ATom can refer to our response for comment 3.

Our discussion on the uncertainty in model-satellite comparison can refer to our response for comment 7.

We add recent in-situ measurements of HCHO emission factors in fire smokes, which is consistent to model fire HCHO emission factor, to introduction section 1: "The GFED4s inventory reports the HCHO emission factor to be 1.86g/kg dry matter for boreal forest fires and 2.09 g/kg dry matter for temperate forest fires, consistent with recent field measurements (Liu et al., 2017; Permar et al., 2021)"

We add a sentence in the conclusion section 5: "Model sensitivity suggests the direct emission of HCHO from wildfires accounts for the majority of HCHO VCD. While the emission factor of HCHO from wildfires (1.86 g/kg dry matter for boreal forest) applied in our model largely agree with field measurements, the role of secondary production of HCHO is likely underestimated due to unaccounted VOCs and underrepresented plume chemistry."

9. a) I am a bit confused about the reprocessed TROPOMI HCHO VCD. My understanding is that it also uses information from GEOS-Chem (for a priori, background column, and AMF), and later the paper compares this reprocessed product with GEOS-Chem. Wouldn't that model information used to reprocess TROPOMI VCD cause some internal biases to the new data, so that the reprocessed product would be essentially similar to and dependent on the model? Can authors explain how it would or would not be the case, and would it affect the interpretation of HCHO VCD evaluation? In other words, is it a fair and independent comparison? The authors seem to agree with that by stating the TROPOMI products are a 'semi-quantitative tool' to constrain fire emission, which should be further clarified

The satellite retrieval indeed relies on model information to provide the final product. We here replace the model information in TROPOMI L2 product with a high-resolution model with year-specific wildfire emissions, for a more realistic representation of HCHO in the atmosphere.

The reprocess error from model includes two parts: background and AMF.

For background error, we discussed the model underestimation comparing to ATom, which is mainly in background. Our modification in the paper can refer to our response for comment 3.

For error from AMF and other sources, we discussed it at the last paragraph of satellite-model comparison section 4.3, which can refer to our response for comment 7.

b) Some common practices of evaluating satellite retrievals include smoothing the model with satellite averaging kernels so that they have the same vertical sensitivity or reprocessing the satellite data with a certain priori profile so that they reflect the measurements, rather than a priori information. It seems the model and satellite data in the work both use the same a priori and the AMF. Am I understanding it correctly? If so, how often the a priori is updated in the reprocessed product? Overall, it would be great if the method for reprocessed data can be further clarified, i.e., the exact difference between the reference sector correction of this study and the default.

Yes. We add a sentence at TROPOMI introduction section 2.1: "GEOS-Chem vertical profiles are updated hourly with collocated TROPOMI HCHO pixels."

The GEOS-Chem and S5P AMF are compared in Figure S7. The GEOS-Chem background versus S5P background is shown in Figure S8. Reprocessed TROPOMI HCHO VCD and S5P HCHO VCD are compared in Figure S10.

> c) The model seems to underpredict the HCHO vertical distribution relative to Atom field data, while the model is used to reprocess TROPOMI HCHO VCD. How does the HCHO underprediction relative to ATom affect the reprocessed VCD?

Our discussion on the model underestimation relative to ATom can refer to our response for comment 3.

The possible error sources of reprocessed VCD, including the model, can refer to our comment 7.

> 10. Line 100. The paper cites Liu et al. 2017 for HCHO EF. There are some new studies from recent aircraft campaign and they seem to support around 2 g/kg for HCHO EF (i.e., WE-CAN VOC emission paper Permar et al 2021 and recent FIREX-AQ AGU conference talks?). Would EF HCHO used in the model/GFED be consistent with those recent studies? It may be able to support that the fire emission in the model is simulated well.

We add this sentence to the introduction section 1: "Several studies have reported a similar level of HCHO emitted from wildfire plumes. Liu et al (2017) found formaldehyde as the second most abundant NMVOC from wildfires in western US, with an emission factor of 2.3 (±0.3) g/kg dry matter for temperate forests and a similar emission factor for boreal forest fires. WE-CAN aircraft measurement reports the HCHO emission factor in near-fire smoke plume to be 1.9 (±0.43) g/kg (Permar et al., 2021). "

The GFED HCHO emission factor is added in GEOS-Chem introduction section 2.3: "The GFED4s inventory reports the HCHO emission factor to be 1.86g/kg dry matter for boreal forest fires and 2.09 g/kg dry matter for temperate forest fires, consistent with recent field measurements (Liu et al., 2017; Permar et al., 2021)."

> 11. Lines 550 -555: I am not entirely sure how the uncertainties of reprocessed VCD were calculated by reading this part. Can the authors clarify it?

All the uncertainty values are directly from TROPOMI L2 HCHO ATBD.

We reorganized the uncertainty part by:
1. Add a paragraph at the end of section 2.1: "We estimate the total uncertainty of reprocessed TROPOMI HCHO vertical column to be ≥ 90% for fire free region (TROPOMI L2 HCHO Algorithm Theoretical Basis Document, https://sentinels.copernicus.eu/documents/247904/2476257/Sentinel-5P-ATBD-HCHO-TROPOMI.pdf/db71e36a-8507-46b5-a7cc-9d67e7c53f70?t=1646910030856, and references therein). This includes 75% of uncertainties from the AMF_{SAT}, 25% from $dSCD_{SAT}$ and 40% from $VCD_{0,SAT}$. The uncertainties in regions with strong fire are

estimated to be ≥ 35%, including 30% of uncertainties from AMF$_{SAT}$, 15% from dSCD$_{SAT}$ and 10% from VCD$_{0,SAT}$. The relative lower uncertainties reflects much stronger VCDs in these wildfire regions. "

2. Remove the uncertainty section 4.4 and add a paragraph at the end of section 4.3: "Satellite retrievals of HCHO in wildfire region remains as a major challenge. One source of uncertainty stems from *a priori* profiles used in AMF calculation (Kwon et al., 2017). We find that for regions with heavy smoke, our calculated GEOS-Chem AMF$_{GC}$ is 50% lower than the AMF$_{SAT}$ in the operational product, due to the difference in HCHO vertical profiles (Figure S3). As a result, our reprocessed HCHO VCD product, VCD$_{SAT,GC}$, is higher than the operational product by 3-5×10^{15} molecules cm^{-2} in heavy smoke regions in July 2019 (Figure S10). Another uncertainty lies in the aerosol optical properties. Wildfire smoke is a major source of brown carbon (June et al., 2020). As the current retrieval algorithm for HCHO does not account for absorbing aerosols, smoke can reduce the sensitivity of satellite measurements to atmospheric layers below and above the aerosol layer, leading to a smaller AMF by 20-30% (Jung et al., 2019; Martin et al., 2003)."

12. Section 2.2: how many vertical profiles were used from ATom. How could AWAS do 3-5 minutes average for isoprene or monoterpenes. That's be lots of samples for AWAS? A bit of clarification would be good.

We reorganized the ATom data introduction section 2.2: "During ATom-1, two flights sampled eight vertical profiles over Alaska during 1-3 August 2016. We make use of 1-minute averaged measurements of HCHO, isoprene, monoterpenes (α-pinene and β-pinene) and the sum of methyl vinyl ketone and methacrolein (MVK+MACR). HCHO measurements sampled in 1-Hz frequency were made by laser induced fluorescence by the NASA In Situ Airborne Formaldehyde (ISAF) instrument (Cazorla et al., 2015). Isoprene and monoterpenes were measured by two instruments: the University of Irvine Whole Air Sampler WAS) followed by laboratory Gas Chromatography (GC) analysis, sampled every 3-5 minutes (Simpson et al., 2020), and the National Center for Atmospheric Research (NCAR) Trace Organic Gas Analyzer (TOGA), sampled every 2 minutes with a 35-second integrated sampling time (Apel et al., 2021). MVK and MACR were also measured by TOGA. These measurements are interpolated to 1-minute time resolution for model comparison. Within our study domain, there are 341 1-minute averaged mixing ratio values for HCHO, 101 and 231 for isoprene and α-pinene/β-pinene from WAS, 337 for isoprene, α-pinene/β-pinene and MVK/MACR from TOGA. The reported measurement uncertainties are $\pm$10% for HCHO, $\pm$10% for WAS isoprene and monoterpenes, $\pm$15% for TOGA isoprene and $\pm$30% for TOGA monoterpenes, $\pm$30% for MVK and $\pm$20% for MACR."

13. Line 523: ATBD not defined until next page.

Fixed.