Response to reviewer #2

We really appreciate your constructive comments and suggestions on our manuscript. We have considered every comment carefully, responded on a point to point and marked every change in red in the revised version.

General comments

The paper aims to present an improved TROPOMI HCHO retrieval over China. Compared to the ESA operational product, two major differences are highlighted: (1) the use of BOAS instead of DOAS for the of fit the slant columns, (2) the use of a priori profiles from a regional model in order to recalculate the AMF, with a finer spatial resolution, and optimized emissions over China.

Overall, the paper fails in showing an improvement of the slant columns, the differences between the two products being negligible.

Responses: Thank you very much for this suggestion. We using a different spectral retrieval technique (BOAS method) of HCHO slant columns. Moreover, the background correction has also been improved. Although DOAS and BOAS HCHO DSCDs show a similar spatial pattern. The spatial distribution of BOAS HCHO SCD is expected to be smooth, less noisy. Besides, operational product using different SCD retrieval methods is compared with MAX-DOAS HCHO measurements. Using the BOAS HCHO SCDs reduced the underestimation in summer and overestimation in winter of the operational product. In summer, using different SCD retrieval methods results a difference of 7.00% (± 1.71%, ± Error) from the TROPOMI operational HCHO VCD. The result shows that using the BOAS HCHO SCDs reduced the underestimation in summer and overestimation in winter of the operational product.

The scientific interest of the paper lies in the second improvement. The authors should focus more on this aspect, and go further into a detailed analysis of the spatio-temporal effects of using more precise profiles for satellite HCHO observations. However, it is not demonstrated how the finer spatial resolution of the model improves the validation. Here it could help to show that the improvement is more important over the urban site compared to the sub-urban sites. Or is it more an effect of the different model chemistry/emissions, and not a spatial resolution effect?

Response: Thank you very much for this suggestion. We followed the reviewer's comment. We also compared simulated a priori HCHO profiles and MAX-DOAS HCHO profiles at suburban site (UCAS) in the revised version. The improvement of WRF-Chem simulations at urban site is more significant than suburban site, which is mainly related to the finer spatial resolution and more up to date emission inventory over China used in the simulations. The bias between simulated and measured HCHO profiles at urban site is larger than that at suburban site, which is mainly due to smaller spatial gradient over suburban areas.

Moreover, TROPOMI HCHO VCDs are compared with MAX-DOAS measurements. The results show that the improvements of our VCD retrieval in winter time is more significant than summer and the improvement at urban site is also more significant than suburban sites in winter. The results indicates better anthropogenic emission inventory and the finer spatial resolution over China in WRF-Chem simulations.

Distributions of operational HCHO VCD in four seasons are compared with distributions of the improved HCHO VCD in Section 5.4. Operational product shows similar spatio-temporal distribution with our retrieval over China while the absolute

values are (slight) smaller than our retrieval (Figure 12). In summer, hotspots can be observed over BTH, YRD, PRD, Shandong province, Henan province, Wuhan (Hubei's provincial capital), SCB and cities along Fen nutrient-laden valley in Shaanxi and Shanxi provinces in our retrieval. These hotspot patterns are strongly correlated to the population density and industrial emission pattern indicating a significant anthropogenic contribution. These hotspots are less obvious in the operational product in summer.

In a conclusion, both enhanced emission inventory and resolution of WRF-Chem simulations contribute to improving TROPOMI HCHO retrieval.

Along the paper, the numbers are often used in a quite subjective way. (0.15% difference being called an improvement for exemple). I recommend writing quantitative comparisons with a more rigorous analysis to strengthen the message of the paper.

Responses: Thank you very much for this suggestion. In the revised version, we added the error bar in calculating the NMBs (Table 3). The error bars of NMBs are two times standard error (SE) and is calculated following:

$$\text{Error} = 2 \times \sqrt{\frac{1}{n(n-1)} \frac{\sum_{i=1}^{i=n} \left(V_T(i) - V_M(i) - \overline{V_T(i) - V_M(i)}\right)^2}{\left(\sum_{i=1}^{i=n} V_M(i)/n\right)^2}} \times 100\% \qquad \text{(A1)}$$

The error shows 95% confidence interval (Streiner, 1996) and The NMBs larger than error are statistically significant. The 0.15% difference is statistically insignificant. The sentence "while the SCD retrieval only shows a minor effect of 0.15 %" is deleted.

Changes in manuscript: L302-306, P14-15 in the revised version.

The paper needs major revisions before being published in AMT.

Specific comments

The title does not fairly reflect the contents of the paper. Unless the results are significantly extended, the title of the paper should focus more on the "improved AMF calculation over China".

Responses: Thank you very much for this suggestion. In addition to the improvement of AMF calculation, we are also using a different spectral retrieval technique of HCHO slant columns. Moreover, the background correction has also been improved. We think the title suggested by the reviewer neglected these points, and therefore, would like to keep the title of the manuscript as is.

The section called "Improved HCHO retrieval algorithm" presents the retrieval algorithm developed for this study. As described in this section, it is actually very similar to the ESA operational product. The similarities and the differences need to be clearly explained. For example, the wavelength calibration. The description is the same as for the operational product. Why a specific section dedicated to this aspect? It would be interesting to see a comparison of the calibration results between the 2 products.

Responses: Thank you very much for this suggestion. The theory of wavelength calibration in our retrieval is same with operational product. While the parameters used in the wavelength calibration including the TROPOMI slit function and polynomial orders are different from operational product. The preflight slit function is obtained from the TROPOMI

Calibration Key Data (CKD) (available at http://www.tropomi.eu/data-products/isrf-dataset, last access: 22 May 2019) which is derived from TROPOMI calibration measurements performed in March 2015 at CSL in Liege. Comparing the spectral fit residual of using different versions of preflight slit function in the spectral fitting, we found that using version v3.0.0 results in lowest residual (Fig. A1). The preflight instrument slit function version v3.0.0 is used in our retrieval, while the operational product uses version v1.0.0 preflight slit function. In the operational algorithm, polynomials are not considered in wavelength calibration. In our retrieval, the third order polynomials are selected through sensitivity analysis (Fig. A1). The result shows that using the third order polynomials contributes to reducing residual in wavelength calibration.

Changes in manuscript: L163-171, P7-8 in the revised version.

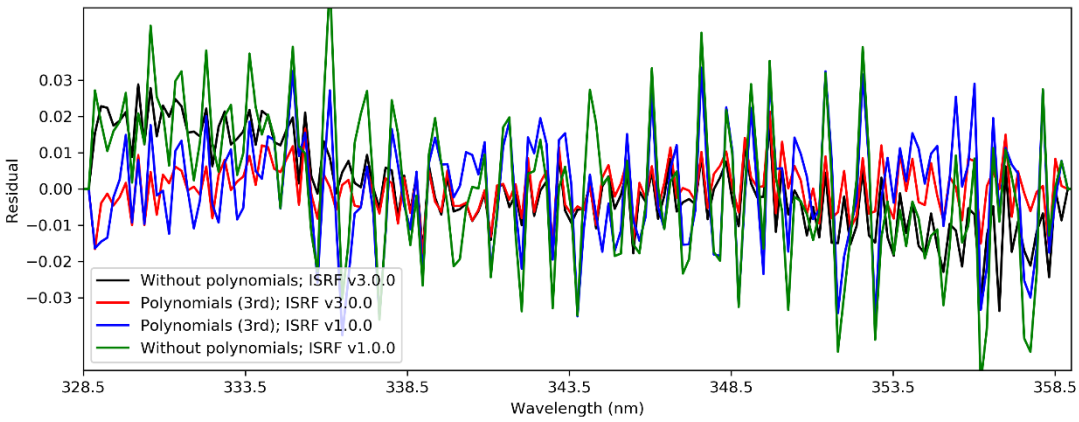

Figure A1. Comparisons of spectral fit residuals using different version of preflight slit function and using different polynomials during wavelength calibration.

The same holds for the AMF calculation part. It is very similar to the Tropomi HCHO ATBD and the differences are not clearly explained, except for the a priori profiles. Same for the reference sector correction.

Responses: Thank you very much for this suggestion. The AMF calculation is improved by using more precise a priori HCHO profiles in my study. Cloud information, surface albedo and surface pressure used in AMF calculations in our retrieval is same with operational product. The similarities and differences are listed in Table 1.

In Reference sector correction, we improved the reference sector correction by considering the variability of M_0/M ratio.

Changes in manuscript: Table 1 and L240, P11 in the revised version.

It is not shown in the paper that the SCDs have been improved. There is a contradiction between the introduction ("BOAS has been reported featured with lower fitting uncertainties to the standard DOAS method") and the result section, where it is stated (page 10) that the RMS are identical. So the "lower fitting uncertainties" of the BOAS technique are not demonstrated. As for Figure 4 with the slant columns, a figure with RMS comparison needs to be added.

Responses: Thank you very much for this suggestion. The sentence at Line 341 in page 15 "On the other hand, RMS of both methods is very similar." is not rigorous. We added RMS comparisons in the Figure 5 in the revised version. RMS comparisons shown in Fig. 5(c) indicate that averaged RMS of the DOAS retrieval ($6.1 \pm 1.56\times10^{-4}$) is slightly higher than that of BOAS

retrieval ($5.95 \pm 1.50\times10^{-4}$). The sentence "BOAS has been reported featured with lower fitting uncertainties to the standard DOAS method" is deleted.

Changes in manuscript: L336-337, P15 in the revised version.

In section 4.1.2, it is explained that the operational product has been updated using a priori profile from WRF model. Does it mean that operational averaging kernel have been used? Or did the authors used their own radiative transfer calculation? It is important to know in order to understand if other sources of differences, such as the albedo, can play a role in the observed vcd differences.

Responses: Thank you very much for this suggestion. The operational averaging kernel (AK) has not been used in AMF calculation. We recalculated AMF using LUT table following Sect 3.2. The parameters used in AMF calculation are same with operational algorithm except for a priori HCHO profiles. Besides higher horizontal resolution, WRF-Chem simulation also has a higher vertical spatial resolution compared to TM5-MP data set. Although the difference in vertical resolution only shows negligible effect on box AMF (below 2%) under clear sky condition (Fig. A2 (b)), lower vertical resolution profiles would cause significant impact for cloudy cases due to the interpolation of coarse grid (Fig. A2 (d)). Therefore, the operational averaging kernel was not used.

Changes in manuscript: L371-374, P17 in the revised version.

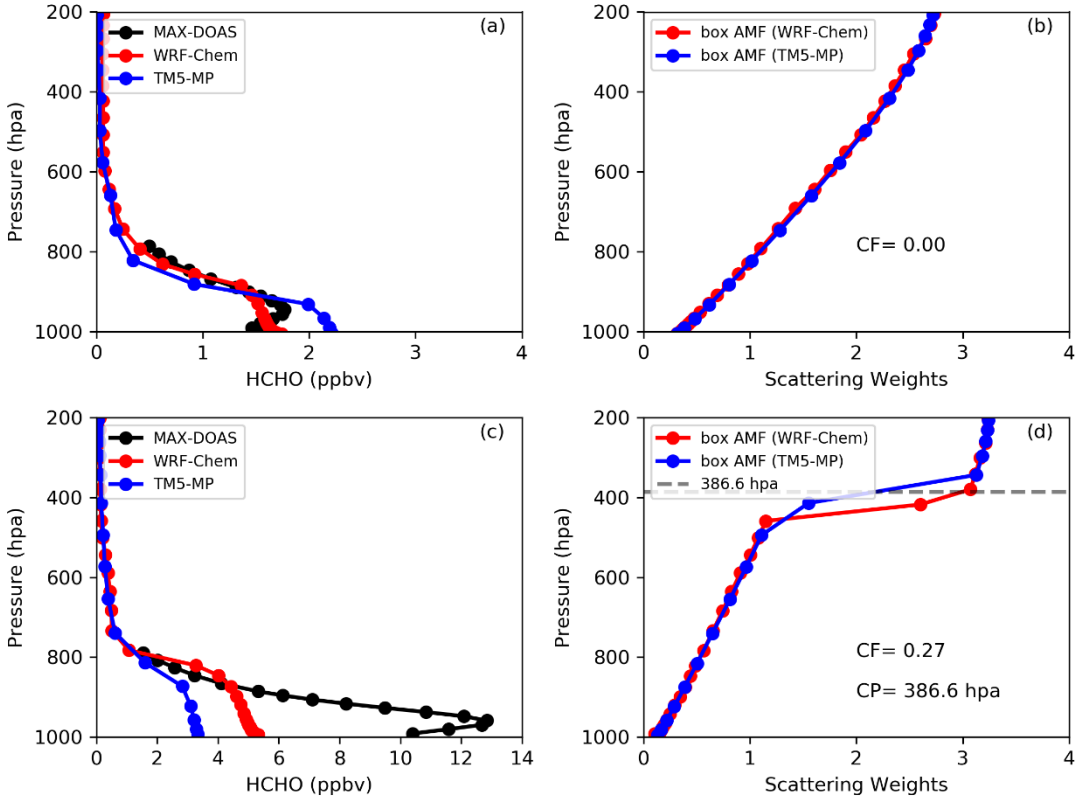

Figure A2. Daily averaged vertical HCHO profiles obtained from MAX-DOAS, WRF-Chem and TM5-MP model in clean case on 03 March 2019 (a) and in polluted case on 26 June 2019 (c). Comparisons of box AMF using WRF-Chem and TM5-MP simulations in clean case with clear sky on 03 March 2019 (b) and in polluted case with cloudy sky on 26 June 2019 (d). The locations of two pixels are within 20 km of the CAMS site.

Improved AMF for China should include some tests about the aerosol effects. They are not even mentioned.

Responses: Thank you very much for this suggestion. Aerosol effect on TROPOMI HCHO retrieval is discussed in Section 5.3. To estimate aerosol effect on TROPOMI HCHO retrieval, we calculate the AMFs using MAX-DOAS measured aerosol extinction profiles using VLIDORT (version 2.6) (Spurr, 2008). The AMFs are applied on the operational product and our retrieval. The TROPOMI HCHO VCDs with and without considering aerosols are compared to MAX-DOAS HCHO VCDs. The comparison results are shown in Figure 10. The results show that considering aerosol in the AMF calculations does not improve the agreement with ground based measurements. Considering aerosol effect in TROPOMI retrieval reduces HCHO VCDs by 11.46% (± 1.48%) for the operational product and 17.61% (± 1.92%) for our retrieval in winter. The reduction over urban site is more significant than suburban sites, mainly due to higher aerosol load. Operational product using both HCHO and aerosol extinction profiles from MAX-DOAS shows underestimation of 8.36 % (± 4.63 %). Our retrieval using MAX-DOAS HCHO and aerosol extinction profiles for AMF calculation underestimates HCHO VCD by 18.53% (± 4.04 %).

Changes in manuscript: L453-466, P20 in the revised version.

When comparing to the MAX-DOAS data in Beijing, MAX DOAS profiles have been used to re-calculate the amf of the improved Chinese HCHO product (Figure 7). For a fair comparison, the same method needs to be applied to the operational product. All the needed information are provided in the operational L2 files.

Responses: Thank you very much for this suggestion. We have recalculated HCHO VCD for the operational product by using MAX-DOAS HCHO profiles as a priori profiles. The Pearson correlation coefficient (R) between the recalculated operational product and MAX-DOAS HCHO VCD decreases by 0.02 to 0.79. The slope of the regression line increases by 0.19 to 0.84 with offset reduces by 0.24×10^{16} molec cm^{-2} to 0.15×10^{16} molec cm^{-2}.

Changes in manuscript: L406-409, P18 and Figure 10 in the revised version.

I have some concerns about the way validation results are presented; The operational product, such as most of existing HCHO satellite products, is rather known to be underestimated over emission regions such as Beijing. See for example Jung et al. 2019 (https://doi.org/10.1029/2019EA000702) or Vigouroux et al. 2020 (https://doi.org/10.5194/amt-2020-30) and references therein. Here the authors claim to find an opposite result. The operational product is overestimated, and the improved Chinese product is lower and in better match with the MAX-DOAS. But actually this result holds for winter time only. The results should be discussed more in terms of low column (winter) or high columns (summer). Finally, a link to previous satellite HCHO validation studies should be made, and the reasons for such different conclusions need to be discussed.

Responses: Thank you very much for this suggestion. The study of Jung et al. 2019 (https://doi.org/10.1029/2019EA000702) shows that excluding the aerosol effect underestimates HCHO VCDs over East China during 2006-2007. While HCHO VCDs were not compared with ground-based instruments.

At CAMS site, HCHO VCDs are underestimated by TROPOMI observations and the underestimation of our retrieval (5.78 ± 3.49 %) is slightly smaller (29.77 ± 22.83 %) than that of the operational product (8.23 ± 3.09 %). At UCAS site, both our retrieval and operational product overestimate HCHO VCDs and the overestimation of our retrieval (5.22 ±3.49 %) is smaller (51.44 ± 33.70 %) than that of the operational product (10.75 ± 3.09 %). The overestimation of operational product at UCAS site is opposite to previous study that TROPOMI operational product underestimated HCHO VCDs in Xianghe located at ~ 50 km southeast of Beijing compared to FTIR measurements (Vigouroux et al., 2020). In order to investigate the reason of the overestimation, we separated the data by seasons for comparison at three MAX-DOAS sites (Table 5). In summer, high HCHO columns are due to oxidation of VOCs related to enhanced biogenic emissions. Both our retrieval and operational product underestimate HCHO VCD in summer which is consistent with previous studies (Vigouroux et al., 2020; Chan et al., 2020). In wither, lower HCHO columns (< 1×10^{16} molec cm^{-2}) are mainly related to anthropogenic emissions including vehicle exhaust and industrial emissions. The vertical HCHO profiles simulated by WRF-Chem model are similar to the one measured by MAX-DOAS in winter, while the TM5-MP profiles show larger difference to the MAX-DOAS measurements. Both our retrieval and operational product overestimate HCHO VCDs in winter. The overestimation of our retrieval (11.23 ± 4.61 %) is 63.18 % (± 22.63 %) smaller than operational product (29.08 ± 5.59 %) and the overestimation in urban area (QKY) is smaller than that in suburban areas (UCAS and NC sites). The improvements of our retrieval in winter time are more significant and the improvement at urban site is also more significant than suburban sites in winter, which are mainly related to better anthropogenic emission inventory and the finer spatial resolution over China in WRF-Chem simulations. In order to eliminate effect from a priori HCHO profile, a priori HCHO profiles from MAX-DOAS measurements are used for AMF calculations for comparison. The overestimations at CAMS and NC sites become less significant. The overestimation at UCAS site of our retrieval reduces to 3.94 % (± 6.87 %), while the overestimation of operational product reduces to 10.60 % (± 8.71 %). Our result shows an overestimation of TROPOMI HCHO VCDs during winter at UCAS site. Our result is different from the previous FTIR comparison study (Vigouroux et al., 2020) which shows TROPOMI underestimated HCHO columns in winter. On the one hand, the pollution conditions at three MAX-DOAS sites are different from the FTIR sites used in Vigouroux et al. (2020). On the other hand, HCHO concentrations in the lower troposphere is lower in winter, resulting a relatively larger portion of HCHO above the MAX-DOAS retrieval height of 3km. Therefore, the MAX-DOAS measurements show an underestimation in winter. Our findings are consistence with the previous study that SCIAMACHY HCHO VCDs are in general lower than FTIR measurements while higher than MAX-DOAS observations (Vigouroux et al. 2009). The remaining overestimation is mainly related to a portion of HCHO above 3 km where the MAX-DOAS is not sensitive. In addition, the TROPOMI retrieval assumes aerosol free atmosphere which might also lead to the overestimations.

Changes in manuscript: L415-424, P18-19, L427-431, P19 and L434-451, P19-20 in the revised version.

The last paragraph of section 4.2 is the most interesting part of the paper and deserves to be extended. It is found that both algorithms remain underestimated in summer time, when the columns are the largest and mainly related to biogenic emissions. Both models simulate profiles not peaked enough near the surface. However, in winter time, when the columns are the lowest (no biogenic emissions), an improvement is observed compared to the MAX-DOAS observations when using WRF-Chem model as a priori profiles. Can you say something about possible reasons for this? Does the WRF-Chem model perform better than TM5 for anthropogenic emissions? Is it related to the spatial resolution or to the chemistry?

Responses: Thank you very much for this suggestion. The anthropogenic emission in WRF-Chem is obtained from The Multi-resolution Emission Inventory for China (MEIC). The MEIC emission inventory has improved the emissions estimation from power plants (Liu et al., 2015), vehicles (Zheng et al., 2014), and residential combustions of non-methane volatile organic compounds (NMVOCs) (Li et al., 2013; Peng et al., 2019).

Comparisons of seasonal averaged a priori HCHO profiles from MAX-DOAS, WRF-Chem and TM5-MP simulations at suburban (UCAS) sites in four seasons are added in the revised version. The improvement of WRF-Chem simulations at urban site is more significant than suburban site, which is mainly related to the finer spatial resolution and more up to date emission inventory over China used in the simulations.

Moreover, TROPOMI HCHO VCDs are compared with MAX-DOAS measurements. The improvements of our VCD retrieval in winter time is more significant than summer and the improvement at urban site is also more significant than suburban sites in winter, which are mainly related to better anthropogenic emission inventory and the finer spatial resolution over China in WRF-Chem simulations.

We think that both enhanced emission inventory and resolution of WRF-Chem simulations contribute to improving TROPOMI HCHO retrieval.

Changes in manuscript: L437-439, P19 in the revised version.

The discussion about seasonal variation of the improved Chinese product, and its spatial distribution over China does not bring anything new about current HCHO satellite observations. I advise to either remove this part, either extend with meaningful observations going much more into details. Comparison maps of SCD and AMF are shown. It would be good to do the same for the final VCD.

Responses: Thank you very much for this suggestion. The maps of operational HCHO VCDs in four seasons are added in the revised version. Spatial distributions of operational VCD and our retrieval are compared in Section 5.3. Operational product shows similar spatio-temporal distribution with our retrieval over China while the values are smaller than our retrieval (Figure 10). In summer, hotspots can be observed over BTH, YRD, PRD, Shandong province, Henan province, Wuhan (Hubei's provincial capital), SCB and cities along Fen nutrient-laden valley in Shaanxi and Shanxi provinces in our retrieval. These hotspot patterns are strongly correlated to the population density and industrial emission pattern indicating a significant anthropogenic contribution. These hotspots is less obvious in map of the operational HCHO VCDs in summer. The distribution of high HCHO VCDs observed in our retrieval over the Xinjiang Uygur Autonomous Region is related to the unique topography and industrial areas. The total industrial VOCs emission over the Xinjiang Uygur Autonomous Region is higher than Shanxi province (Zheng et al., 2017).

Changes in manuscript: L473-474 and L477-480, P21 and Figure 12 in the revised version.

Technical corrections

Abstract

L18: We present ~~the~~ an improved retrieval…

Responses: Thank you very much for this suggestion. Changed.

L19: The new retrieval optimizes the slant column density retrieval: this is not demonstrated in the paper. Please rephrase.

Responses: Thank you very much for this suggestion. We using a different spectral retrieval technique (BOAS method) of HCHO slant columns. Moreover, the background correction has also been improved. Although DOAS and BOAS HCHO DSCDs show a similar spatial pattern. The spatial distribution of BOAS HCHO SCD is expected to be smooth, less noisy. Besides, operational product using different SCD retrieval method is compared with MAX-DOAS HCHO measurements. Using the BOAS HCHO SCDs reduced the underestimation in summer and overestimation in winter of the operational product. In summer, using different SCD retrieval methods results a difference of 7.00% (± 1.71%, ± Error) from the TROPOMI operational HCHO VCD. The result also shows that using the BOAS HCHO SCDs reduced the underestimation in summer and overestimation in winter of the operational product.

L24: MAX-DOAS measurements in ~~China~~ Beijing

Responses: Thank you very much for this suggestion.

L26: while the SCD retrieval only shows a minor effect of 0.15%. This is negligible! We can even talk about a perfect agreement between the SCD retrievals.

Responses: Thank you very much for this suggestion. The sentence is deleted. Using different SCD retrieval methods results a difference of 7.00% (± 1.71%, ± Error) from the TROPOMI operational HCHO VCD in summer. So we cannot talk about a perfect agreement between the SCD retrievals.

L29-30: The last sentence is not demonstrated in the paper.

Responses: Thank you very much for this suggestion.

TROPOMI HCHO VCDs are compared with MAX-DOAS measurements. The improvements of our VCD retrieval in winter time is more significant than summer and the improvement at urban site is also more significant than suburban sites in winter, which are mainly related to better anthropogenic emission inventory and the finer spatial resolution over China in WRF-Chem simulations.

Moreover, the spatio-temporal distribution of the improved and operational HCHO VCDs is compared in Section 5.4 in the revised version. Operational product shows similar spatio-temporal distribution with our retrieval over China while the values are smaller than our retrieval (Figure 10). In summer, hotspots can be observed over BTH, YRD, PRD, Shandong province, Henan province, Wuhan (Hubei's provincial capital), SCB and cities along Fen nutrient-laden valley in Shaanxi and Shanxi provinces in our retrieval. These hotspot patterns are strongly correlated to the population density and industrial emission pattern indicating a significant anthropogenic contribution. These hotspots is less obvious in map of the operational HCHO VCDs in summer. These hotspots are less obvious in the operational product in summer. The distribution of high HCHO VCDs observed in our retrieval over the Xinjiang Uygur Autonomous Region is related to the unique topography and industrial areas. The total industrial VOCs emission over the Xinjiang Uygur Autonomous Region is higher than Shanxi province (Zheng et al., 2017).

These results show that our retrieval is more suitable for the analysis of regional and city scale pollution in China.

Changes in manuscript: L473-474 and L477-480, P21 and Figure 12 in the revised version.

Introduction:

L48: Again, It is not shown in the paper that the SCDs have been improved. The "lower fitting uncertainties" of the BOAS technique are not demonstrated.

Responses: Thank you very much for this suggestion. Thank you very much for this suggestion. We using a different spectral retrieval technique (BOAS method) of HCHO slant columns. Moreover, the background correction has also been improved. Although DOAS and BOAS HCHO DSCDs show a similar spatial pattern. The spatial distribution of BOAS HCHO SCD is expected to be smooth, less noisy. Besides, operational product using different SCD retrieval method is compared with MAX-DOAS HCHO measurements. Using the BOAS HCHO SCDs reduced the underestimation in summer and overestimation in winter of the operational product. In summer, using different SCD retrieval methods results a difference of 7.00% (± 1.71%, ± Error) from the TROPOMI operational HCHO VCD. The result also shows that using the BOAS HCHO SCDs reduced the underestimation in summer and overestimation in winter of the operational product. The sentence "This technique has been reported featured with lower fitting uncertainties compared to the standard differential optical absorption spectroscopy (DOAS) method (Chance and Kurosu, 2003)." is deleted.

L54: the result is expected to be more realistic for the investigation of spatio temporal variation of HCHO over China. ok but this needs to be demonstrated. How the spatio temporal variation of HCHO over China has been improved? In the current version, only a reduction of the bias compared to MAX-DOAS data is shown in winter time.

Responses: Thank you very much for this suggestion.

Operational and improved HCHO VCDs are compared with MAX-DOAS measurements. The improvements of our VCD retrieval in winter time is more significant than summer and the improvement at urban site is also more significant than suburban sites in winter, which are mainly related to better anthropogenic emission inventory and the finer spatial resolution over China in WRF-Chem simulations.

Moreover, the spatial distributions of the improved and operational HCHO VCDs in four seasons are compared in Section 5.4 in the revised version. In summer, hotspots can be observed over BTH, YRD, PRD, Shandong province, Henan province, Wuhan (Hubei's provincial capital), SCB and cities along Fen nutrient-laden valley in Shaanxi and Shanxi provinces in our retrieval. These hotspot patterns are strongly correlated to the population density and industrial emission pattern indicating a significant anthropogenic contribution. These hotspots are less obvious in the operational product in summer.

These results show that our retrieval is expected to be more realistic for the investigation of spatio temporal variation of HCHO over China.

Changes in manuscript: L473-474 and L477-480, P21 and Figure 12 in the revised version.

Figure 1: The scale could be reduced to better show emission spots.

Responses: Thank you very much for this suggestion. The scale is reduced to 1.5.

WRF-model

L79: more up to date emission inventory of China: this is really vague and needs to be explained

Responses: Thank you very much for this suggestion. The anthropogenic and biogenic emissions are obtained from The Multi-resolution Emission Inventory for China (MEIC) and the Model of Emissions of Gases and Aerosols from Nature (MEGAN) (Guenther et al., 2006; Li et al., 2017), respectively. The MEIC emission inventory has improved the emissions estimation from power plants (Liu et al., 2015), vehicles (Zheng et al., 2014), and residential combustions of non-methane volatile organic compounds (NMVOCs) (Li et al., 2013; Peng et al., 2019).
Changes in manuscript: L116-118 in the revised version.

Improved HCHO retrieval algorithm

L133: in Table ~~1~~ 2

Responses: Thank you very much for this suggestion. Changed

Table 2:

● Why the use of DSCD in the caption?

Responses: Thank you very much for this suggestion. Daily detector row averaged radiance over the equatorial Pacific is used as reference spectra. Due to residual HCHO signals in reference, the differential SCD (DSCD) is retrieved in spectra fitting.
Changes in manuscript: L86-87, P4 in the revised version.

● Please indicate the differences compared to the operational product.

Responses: Thank you very much for this suggestion. The similarities and differences are added in Table 1.
Changes in manuscript: Table 1 in the revised version.

● What about the Ring correction?

Responses: Thank you very much for this suggestion. The term $\alpha_r X_r(\lambda)$ in Eq. (3) represents Ring effect. $X_r(\lambda)$ is Raman spectrum calculated in Chance and Spurr (1997).
Changes in manuscript: Table 1 in the revised version.

● Do you include corrections for non-linear Ozone absorption effects?

Responses: Thank you very much for this suggestion. In our fitting, wavelength dependency of O_3 SCDs are not considered. We added the uncertainty analysis from wavelength dependency of O_3 SCDs. The uncertainty it causes on SCD is about 3.49%.
Changes in manuscript: L267-269, P12 and Table 3 in the revised version.

● How is the radiance reference sector calculated? Per instrument row? Per day?

Responses: Thank you very much for this suggestion. Radiances measured 1 day before the processing day over the Pacific with latitudes ranging from 30°S to 30°N and longitude ranging from 180°W to 140°W are averaged and used as reference in the spectral fit.
Changes in manuscript: L179-180, P8 in the revised version.

● Why this particular choice of O_3 and BrO cross-sections? Can these choices explain the differences with the operational product?

Responses: Thank you very much for this suggestion. The O_3 and BrO cross-sections are chose following the study of González Abad et al. (2015). The biases between DSCDs and RMS in operational product and our retrieval are mainly related to the difference of retrieval method, retrieval settings and selection of reference.

To eliminate the impact from retrieval settings, BOAS HCHO DSCDs using same retrieval settings with operational DSCD retrieval are compared with DOAS HCHO DSCDs (Figure. A3). Using same retrieval settings, difference between DOAS HCHO DSCDs and BOAS HCHO DSCDs (27.33%) is significantly reduced and the remaining difference is due to retrieval algorithm. Besides, smaller difference (4.41%) in HCHO SCDs indicates that reference sector correction reduces the effect from retrieval method (Figure. A3 (b)).

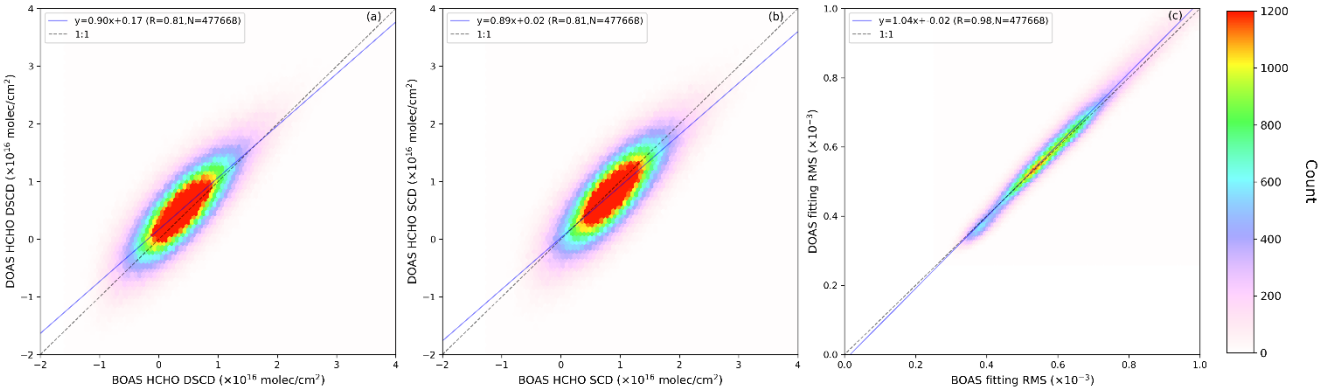

Figure A3. (a) Pixel to pixel comparisons of DOAS and BOAS HCHO DSCDs, (b) DOAS and BOAS HCHO SCDs and (c) DOAS and BOAS fitting RMS using same retrieval settings on 06 August 2018 in the region between 73° E and 130° **E**, and 18° N and 54° N.

Changes in manuscript: L345-349, P15-16 in the revised version.

L168: the surface albedo is obtained from the S5P operational cloud product. This is a bit surprising. Please specify the wavelength.

Responses: Thank you very much for this suggestion. The surface albedo is extracted from the S5P operational HCHO product in which surface albedo is from OMI-based monthly minimum LER at 342 nm for HCHO fitting window. We deleted the sentence and added the information used in AMF calculation in Table 1.

Changes in manuscript: Table 1, P12 in the revised version.

L187: specify the meaning of k and m

Responses: Thank you very much for this suggestion. In the old version, k and m are the position along track and across track of satellite pixels. The expression in the old version maybe is difficult to understand. We have changed the expression of the Eq. (10) in the revised version.

Changes in manuscript: L239, P11 in the revised version.

Results and discussions

L196: VT and Vm is the average tropospheric CHO VCD measured by TROPOMI and MAX-DOAS. How are the data averaged in space / time?

Responses: Thank you very much for this suggestion. MAX-DOAS measurements are temporally averaged within ±1 h around the TROPOMI overpass time, while TROPOMI pixels within 20 km of the MAX-DOAS site are spatially averaged for comparison. TROPOMI pixels in our retrieval and operational product are both filtered for intensity-weighted cloud fraction smaller than 0.3, root mean square of spectral fit residual smaller than 10^{-3}, AMF larger than 0.1 and SZA smaller than 70°, quality assurance value (QA value) larger than 0.55 and successful SCD retrieval.

Changes in manuscript: L307-311, P14 in the revised version.

Figure 3: Do the maps show SCD, DSCD (as mentioned in Table 2) or corrected SCDs? It would be good to show corrected SCDs (with a color scale including negative values), since an offset is found in the SCDs. It would help to better see differences in the two spatial distributions.

Responses: Thank you very much for this suggestion. Figure 3 in the old version shows the maps of DSCDs which are not corrected. Figure 3 in the old version is Figure 4 in the revised version. We added the maps of SCDs which are calculated by applying reference sector correction on DSCDs (Fig. 4 (b) and (d)). The color scale in Fig. 4 is changed to include negative values.

Changes in manuscript: L352-355, P16 and Figure 4 in the revised version.

L206: Please compare numbers for the corrected slant columns over Tibet.

Responses: Thank you very much for this suggestion. The numbers of valid satellite measurements over Tibet for BOAS and DOAS retrieval are 22244 and 21987, respectively.

Changes in manuscript: L325-326, P15 in the revised version.

L218: Please compare the RMS.

Responses: Thank you very much for this suggestion. Figure 5 in the revised version adds pixel to pixel comparison of RMS on 06 August 2018 in the region between 73° E and 130° E, and 18° N and 54° N. The result indicate that averaged RMS of the DOAS retrieval ($6.1 \pm 1.56\times10^{-4}$) is slightly higher than that of BOAS retrieval ($5.95 \pm 1.50\times10^{-4}$).

Changes in manuscript: L336-337, P15 and Figure 5 in the revised version.

L218-219: This sentence is vague. Please be more specific

Responses: Thank you very much for this suggestion. The biases between DSCDs and RMS in operational product and our retrieval is mainly related to the difference of retrieval method, retrieval settings and selection of earthshine radiance reference. The impact of selection of earthshine radiance reference and retrieval method are also investigated in the revised version.

Changes in manuscript: L339-349, P15-16 in the revised version.

L223: Please give the numbers in brackets for the Chinese product as well.

Responses: Thank you very much for this suggestion. Averaged SCD taken from the operational product on 06 August 2018 over China ($0.85 \pm 0.69\times10^{16}$ molec cm^{-2}) is on average 4.49 % lower than our retrieval ($0.89 \pm 0.61\times10^{16}$ molec cm^{-2}).

Changes in manuscript: L353-355, P16 in the revised version.

L228. It is not clear how using the BOAS HCHO SCDs reduces the overestimation if changing SCD retrieval method only shows a tiny effect of 0.15%? There is a contradiction here.

Responses: Thank you very much for this suggestion. We added the error bar in Table 3 and find the conclusion that changing SCD retrieval method only shows a tiny effect of 0.15% is less rigorous. The conclusion is changed into "In summer, using different SCD retrieval methods results a difference of 7.00% ($\pm$ 1.71%, $\pm$ Error) from the TROPOMI operational HCHO VCD. The result shows that using the BOAS HCHO SCDs reduced the underestimation in summer and overestimation in winter of the operational product (Table 5)."

Changes in manuscript: L359-362, P16 in the revised version.

L228: The mean random errors relative to BOAS are mentioned. Can you give a definition? And where are those errors presented in the paper?

Responses: Thank you very much for this suggestion. Uncertainty analysis is added in Sect. 4 in the revised version. The Random uncertainties can be approximated by the root mean square (RMS) of spectral fitting residual, the degrees of freedom, and the diagonal term of the covariance matrix for HCHO ($C_{j,j}$):

$$\sigma^2_{N_{S,rand}} = RMS^2 \frac{m}{m-n} C_{j,j} C_{j,j} \qquad (A2)$$

Where m is the number of spectral pixels and n is the number of fitted parameters.

Changes in manuscript: L255-259, P12 in the revised version.

AMF calculation

Table 3: This table is difficult to understand. The presentation of the numbers can be improved. The legend says that NMBs between satellite and MAX-DOAS are provided, but it seems to be more than that (NMB s1, s2). Error bars should be added. It would be more relevant to separated numbers for winter and summer periods.

Responses: Thank you very much for this suggestion. The legend of Table 3 is changed into "NMBs between TROPOMI HCHO VCDs with different retrieval settings (NMB $_{S1,S2}$) and NMBs between TROPOMI and MAX-DOAS observations (NMB$_{S1,M}$). TROPOMI HCHO VCDs are calculated with four different settings, (1) operational retrieval setting (2) replacing DOAS SCDs using BOAS SCDs in the operational product, (3) changing the a priori profiles from TM5 to regional WRF-Chem simulations in the operational product and (4) both (2) and (3) changes in the operational product. The error bars are also presented. All values are in %.".

The standard error (SE) of the NMBs is calculated by dividing the standard deviation (SD) by the square root of day numbers. The two times the standard deviation (95% confidence interval (Streiner, 1996)) is regarded as the error of the NMBs in this study. The NMBs larger than the error are considered statistically significant. The error of NMB is calculated following Eq.

(A1). The NMBs in summer and winter are also presented in Table 5.

Changes in manuscript: L302-306, P14 and Table 5 in the revised version.

Figure 5:

● Profiles are shown at the more urban CAMS station. It would be interesting to also show a suburban station, in order to detect the gain in spatial resolution.

Responses: Thank you very much for this suggestion. Profiles comparison at UCAS site is added in Figure 8 in the revised version. The improvement of WRF-Chem simulations at urban site is more significant than suburban site, which is mainly related to the finer spatial resolution and more up to date emission inventory over China used in the simulations. The bias between simulated and measured HCHO profiles at urban site is larger than that at suburban site, which is mainly due to smaller spatial gradient over suburban areas.

Changes in manuscript: L378-381, P17 and Figure 8 in the revised version.

● How many profiles are averaged? What is the spatial resolution?

Responses: Thank you very much for this suggestion. In AMF calculations, both WRF-Chem and TM5-MP simulations are interpolated to TROPOMI spatial resolution. Interpolated WRF-Chem and TM5-MP simulations within 20 km of the MAX-DOAS site are spatially averaged to compare with MAX-DOAS profiles. MAX-DOAS profiles are temporally averaged in the period 13:30-14:30 (Local Time) within ±1 h around the TROPOMI overpass time.

Changes in manuscript: L366-370, P16 in the revised version.

L249: The operational data are filtered using the QA value. Is the same selection applied to the improved Chinese product? If not, which selection is applied?

Responses: Thank you very much for this suggestion. TROPOMI pixels in our retrieval and operational product are both filtered for intensity-weighted cloud fraction smaller than 0.3, root mean square of spectral fit residual (RMS) smaller than 10^{-3}, AMF larger than 0.1 and SZA smaller than 70°, quality assurance value (QA value) larger than 0.55 and successful SCD retrieval.

Changes in manuscript: L308-311, P14 in the revised version.

L255: Validation results are discussed at the 3 sites using correlation, slope and offset. Looking at those 3 parameters, mainly the offset is improved compared to the operational product. Correlations are almost identical. This needs to be discussed more in detail, related to the observed offset in the AMFs.

Responses: Thank you very much for this suggestion. The difference in validation results of our retrieval and operational product is caused by not only the AMFs but also the SCD retrieval. The SCD retrieval can effect HCHO VCDs largely by 7.00% (± 1.71%, ± Error) in summer (Table 3 in the revised version). Using the BOAS HCHO SCDs help reduce the underestimation in summer and overestimation in winter of the operational product (Table 5).

L263: Please explain how the MAX-DOAS are used to recompute the AMFs. Do you use the averaging kernels? The same

needs to be done with the operational product.

Responses: Thank you very much for this suggestion. We didn't use the averaging kernels in calculation of AMFs using MAX-DOAS measurements as a priori file. We have recalculated HCHO VCD for the operational product by using MAX-DOAS HCHO profiles as a priori profiles. The Pearson correlation coefficient (R) between the recalculated operational product and MAX-DOAS HCHO VCD decreases by 0.02 to 0.79. The slope of the regression line increases by 0.19 to 0.84 with offset reduces by 0.24×10^{16} molec cm^{-2} to 0.15×10^{16} molec cm^{-2}.

Changes in manuscript: L406-409, P18 and Figure 10 in the revised version.

L272: The vertical profiles simulated by WRF-Chem are similar to the one measured by MAX-DOAS in ~~summer~~ winter !

Responses: Thank you very much for this suggestion. Changed

L273: The underestimation of both retrievals in summer time are similar. 9.96% versus 10.88% is not significant. Please add error bars. Only the differences in winter time are significant.

Responses: Thank you very much for this suggestion. We recalculated the NMBs in summer and added the error bars calculated following Eq. A(1). The sentence is changed to "Therefore, our retrieval shows slightly better agreement with the ground based measurements (underestimation of 15.81 ± 2.71 %) compared to the operational product (underestimation of 18.33 ± 3.10 %) during summer."

Changes in manuscript: L427-428, P19 in the revised version.

Section 4.3

Not much useful information is given in this short paragraph. I suggest extending with a comparison with maps of VCD from the operational product, for the 4 seasons.

Responses: Thank you very much for this suggestion. The spatial distribution of the improved and operational HCHO VCDs is compared in Section 5.4 in the revised version. Operational product shows similar spatio-temporal distribution with our retrieval over China while the absolute values are (slight) smaller than our retrieval (Figure 12). In summer, hotspots can be observed over BTH, YRD, PRD, Shandong province, Henan province, Wuhan (Hubei's provincial capital), SCB and cities along Fen nutrient-laden valley in Shaanxi and Shanxi provinces in our retrieval. These hotspot patterns are strongly correlated to the population density and industrial emission pattern indicating a significant anthropogenic contribution. These hotspots are less obvious in the operational product in summer. The distribution of high HCHO VCDs observed in our retrieval over the Xinjiang Uygur Autonomous Region is related to the unique topography and industrial areas. The total industrial VOCs emission over the Xinjiang Uygur Autonomous Region is higher than Shanxi province (Zheng et al., 2017).

Changes in manuscript: L473-474 and L477-480, P21 and Figure 12 in the revised version.

Conclusion

As for the abstract and the title, the conclusions need to be redirected towards the real content of the paper, which is the use of a regional model to compute the AMFs, and the validation at 3 sites in Beijing.

Responses: Thank you very much for this suggestion. The abstract and conclusions are rewrote.

Chance, K. V., and Spurr, R. J.: Ring effect studies: Rayleigh scattering, including molecular parameters for rotational Raman scattering, and the Fraunhofer spectrum, Appl Optics, 36, 5224-5230, https://doi.org/10.1364/AO.36.005224, 1997.

Li, M., Zhang, Q., Kurokawa, J., Woo, J., He, K., Lu, Z., Ohara, T., Song, Y., Streets, D. G., and Carmichael, G. R.: MIX: a mosaic Asian anthropogenic emission inventory under the international collaboration framework of the MICS-Asia and HTAP, Atmos. Chem. Phys., 17, 935-963, 2017.

Liu, F., Zhang, Q., Tong, D., Zheng, B., Li, M., Huo, H., and He, K. B.: High-resolution inventory of technologies, activities, and emissions of coal-fired power plants in China from 1990 to 2010, Atmos. Chem. Phys., 15, 13299-13317, 2015.

Peng, L., Zhang, Q., Yao, Z., Mauzerall, D. L., Kang, S., Du, Z., Zheng, Y., Xue, T., and He, K.: Underreported coal in statistics: A survey-based solid fuel consumption and emission inventory for the rural residential sector in China, Applied Energy, 235, 1169-1182, 2019.

Streiner, D. L.: Maintaining Standards: Differences between the Standard Deviation and Standard Error, and When to Use Each, The Canadian Journal of Psychiatry, 41, 498-502, https://doi.org/10.1177/070674379604100805, 1996.

Zheng, C., Shen, J., Zhang, Y., Huang, W., Zhu, X., Wu, X., Chen, L., Gao, X., and Cen, K.: Quantitative assessment of industrial VOC emissions in China: Historical trend, spatial distribution, uncertainties, and projection, Atmos Environ, 150, 116-125, https://doi.org/10.1016/j.atmosenv.2016.11.023, 2017.

Zheng, B., Huo, H., Zhang, Q., Yao, Z., Wang, X., Yang, X., Liu, H., and He, K. B.: High-resolution mapping of vehicle emissions in China in 2008, Atmos. Chem. Phys., 14, 9787-9805, 2014.