Review of "An improved TROPOMI tropospheric HCHO retrieval over Chine" submitted by Wenjing Su et al., to AMT

General comments

This paper presents TROPOMI formaldehyde retrievals over China. The main difference with respect the operational HCHO retrievals distributed by ESA is the use of specific a-priori information during the air mass factor (AMF) calculation to much better the current knowledge of HCHO distributions over China.

Overall, the results presented here are interesting but the paper lacks details essential to understand the value and implications of the new retrieval. The specific comments section suggests may aspects where the manuscript could improve by including further details. Definitely providing information about TROPOMI level 1 radiances and irradiances is absolutely necessary since they are the base for the retrievals described in section 3. Likewise, for the S5P cloud operational product whose uncertainties will become of paramount importance while assessing the uncertainties of the HCHO retrieval presented here.

The description of the retrieval process also lacks many details. To start with, Figure 1 is included but never referenced in the text. Figure 2 shows two HCHO signals superimposed with the spectral fit residual. It will be more informative to include a similar figure showing the contributions from each one of the parameters included in Table 2. It is very difficult to estimate the quality of this new retrieval without seen the results for one full orbit at least (where issues such as stripes and noise are easier to appreciate). While current figure 3 does this in part, the color scale employed masks may of the possible issues (leaving out negative values) and providing little sensitivity to the region between 0.5 and 1.5 x 10^{16} molecules cm^{-2} most relevant for HCHO retrievals. It will also be very helpful to have plots showing the WRF-Chem simulations used to extract a priories as well as the GEOS-Chem simulation used for background corrections. Finally and of outmost importance, a discussion of the retrieval uncertainties is completely missing. While section 4 provides some bias estimates, the retrieval section should include a description of the random and systematic uncertainties linked to spectral fit and AMF calculation.

Because there is no description of the operational HCHO TROPOMI retrieval in detail, it is impossible to assess the weight and validity of the conclusions derived from the comparison exercise between both products (the one presented here and the operational one). The authors focus on the impact due to the ΔSCD retrieval and the AMF calculation but do not devout any time evaluating the impact of using different earthshine radiances. It is also necessary to include a discussion of the MAX-DOAS errors. Figure 5 shows some vertical profiles. What are they? Seasonal averages, daily averages,… Are they representative of similar periods of time and time of day? This results should be place into context with literature recently published (https://www.atmos-meas-tech-discuss.net/amt-2020-30/). Also keep in mind the discussion about QA values in that paper. Given the statistical uncertainties and the methodology used, an affirmation such as "These results suggest that our retrieval is better than the operational product both in urban and suburban in China" is probably over confident. First the MAX-DOAS measurements are representative of a small region in China dominated by large urban development. Second, small differences in the SCDs and the use of a higher spatial resolution model in the new retrieval indicates that both perform similarly. If the authors could provide some details about

the appearance and differences of box-AMFs between both retrievals it will be possible to get a better idea of the impact of the a priori profiles.

The organization of the paper is clear (despite missing important aspects). However, while the use of English in the paper is good enough to be understandable it could benefit from further proofreading and grammar correction. Some language specific suggestions provided in the technical corrections section represent few of the possible improvements. While this reviewer will love to provide detailed language corrections it is time consuming and out of the scope of the reviewer duties.

For the reasons mentioned above the paper needs several major revisions and needs to expand to include essential details. The results shown are promising but in its current form, it will be difficult for anyone outside the authors to use the presented retrievals for a scientific study.

Specific comments

Abstract

A reference to the paper describing the operational TROPOMI HCHO retrieval by De Smedt et al., should be added here.

Recently Vigouroux et al., have published an excellent operational TROPOMI HCHO retrieval validation paper (https://www.atmos-meas-tech-discuss.net/amt-2020-30/). To cite in the introduction will provide a nice platform to place the results from this study into context later on.

Data sets

Overall, the description of the datasets lacks details that will help the reader to understand better the results of this paper. The authors should expend sometime increasing the description of the datasets employed to ensure that their results can be reproduced by other groups. The retrieval of slant column densities requires the use of calibrated TROPOMI radiances. This is maybe, the most important dataset that necessary to obtain the results presented in this work. Its description and how the authors obtained it should be a section in data sets.

These are some of the questions that would be nice to have described in the paper:

1. The original spatial resolution of TROPOPMI observations at NADIR was 3.5 km x 7 km but it improved to 3.5 km x 5.5 km in August 2019.
2. Fitting parameters considered in the spectral fitting of the operational product?
3. How is the reference sector correction applied in the operational product?
4. Procedures used to generate earthshine radiance over the remote Pacific in the operational HCHO retrieval.
5. Besides a-priori profiles and observation geometry, information about clouds, aerosols and surface reflectance play important roles in AMFs calculations. How is the operational product accounting for them?
6. Does the model simulation include information about pyrogenic sources?
7. What was the spin off period of the simulations?
8. It will be interesting to have a reference for the MEIC inventory.

9. What are the uncertainties associated with TROPOMI HCHO retrievals and MAX-DOAS observations. Table 1 could be expanded with information about the estimated uncertainties for each site as well as the dates and amount of available data.

HCHO SCD retrieval: wavelength calibration

One has to assume that S represents the preflight instrument slit function. It will be very beneficial to discuss the behavior of the preflight slit function. Are they available to the public? How stable the instrument slit function has been after launch?

HCHO SCD retrieval: Radiance fitting

As mentioned above, there is lack of detail in the description of the methodology employed. Please explain the following questions: (1) Methodology employed to calculate the daily average earthshine radiance over the Pacific. (2) Any I_0 corrected cross sections in the spectral fit. (3) What is the impact of not including SO_2 when if SO_2 optical thickness becomes relevant. (4) What is the impact of not including water vapor? (5) Which method is employed to estimate Raman spectra? (5) How was the fitting window selected? HCHO retrievals show big dependencies with fitting windows. (6) New O_3 cross sections have become available in recent years (for example Serdyuchenko et al., 2014 (https://www.atmos-meas-tech.net/7/609/2014/)). (7) Have the authors taken into account the effect of ozone in the fitted slant columns as described by Pukite et al., 2008 (https://www.atmos-meas-tech.net/3/631/2010/amt-3-631-2010.pdf).

AMF calculation

Details missing in the AMF calculation include: (1) Terrain height and surface pressure corrections. (2) Origin of cloud information. (3) Descriptions of the nodes of the box AMF look up table. (4) VLIDORT set up. (4) Impact of aerosols. (5) Error analysis.

Reference sector correction of SCDs

Add description of the GEOS-Chem configuration employed in the reference sector correction. Which longitudes define remote Pacific? Is the correction applied only to -30 degrees to 30 degrees? That region is used to calculate the earthshine radiance. Any contributions outside -30 to 30 degrees will not be correcting for the residual HCHO column but for biases of different origin.

Comparison of operational and improved HCHO product

What are the coincidence criteria to match TROPOMI and MAX-DOAS measurements? To use daily averages in the case of MAX-DOAS seems inappropriate considering the diurnal variations of HCHO columns.

Comparisons of SCD retrievals

This discussion applies to SCDs or ΔSCDs? The result of the spectral fit is in both cases the differential slant column but the title of the section indicates the comparison of slant columns. In that case, differences are not only due to the spectral fit but also to the reference sector correction. What is the impact of using different earthshine radiances?

What are the reasons to filter out retrievals with RMS higher than 10^{-3}. What is the percentage of retrievals with RMS above the threshold? What is the definition of outliers? Showing a histogram of the distributions of both datasets will help to understand the outlier definition.

The authors make a distinction between non-corrected and corrected SCDs. It is not clear in the text what corrected implies. One has to assume we are talking about reference sector corrected and non-corrected SCDs. If that is the case, one could argue that part of the differences observed between SCDs in the non-corrected case are due to the different selection (or calculation) of earthshine radiance reference. To make a real comparison of the performance of booth fitting algorithms they should be using consistent earthshine radiances. The results seem to indicate that part of the biases are due to using different earthshine radiances since the correlation for the non-corrected and corrected SCDs columns is similar for both products but the bias between them is significantly reduced after applying the correction.

Comparisons of retrievals after AMF calculations

One question that raises this section and always permeates the use of high-resolution models is how much information is folded back from the model in the retrieval. To better understand this question it will be very useful to know more about the WRF-Chem simulations and how they were used. The operational product employs daily forecast interpolated in time and space. This procedure should be added to the methodology. How do the box-AMFs of the different retrievals compare?

Comparison between HCHO VCDs observed by MAX-DOAS and TROPOMI

Blue lines in figure 7 right panels, more than speaking of the retrieval, speak about the difference between the a priori profiles for both models and the retrieved MAX-DOAS profiles. Slopes and correlation coefficients are similar (within error estimates) for both retrievals when using model a priori. It is easier to imagine how the comparison of the operational retrieval with MAX-DOAS observations will also suffer a dramatic improvement if MAX-DOAS a priori were to be used in the AMF calculations.

Technical corrections

Line 18 (grammatical suggestion (GS)): Since what is improved is the HCHO retrieval a more correct grammatical structure will be “We present an improved retrieval of formaldehyde (HCHO) over China from the TROPO…”

Line 23 (GS): remove “the” from “agreement with the ground based…” since it is possible there are more than MAXDOAS measurement than those used in this study.

Line 25 (GS): add “s” to “profile” “…higher resolution a priori profiles”.

Line 26: The percentage of what is reported by 61.11% and 0.15%? With the current text it is impossible to know if it refers to the change in the mean VCD, or the percentage of the bias correction attributed to each step. Please specify.

Line 29 (GS): Change “indicating” to “indicate”.

Line 45 (GS): Add “to” after compared “Compared to its predecessor”

Line 53 (GS): Add “s” to “profile” and “a” after “from”. Which regional model is employed?

Line 64 (GS): Change “is” to “are” “.., which are devided in …”.

Line 91 (GS): Add “the” in “… is located in the Chinese…”.

Line 105: The first step of the methodology explained in this paper is the calculation of ΔSCD, not SCD.

Line 111: There can be other causes for wavelength miss-alignment can be Doppler shift, non-uniform illumination of the slit due to presence of clouds or other high reflectance surface in the pixel.

Line 115: What is the benefit of having two closure polynomials during wavelength calibration considering that it is done using TOA irradiances and there for are not affected by the presence of clouds or aerosols that may introduce low-frequency structures?

Line 142 (GS): Add “presence of” to “… atmosphere (presence of clouds, vertical HCHO distribution…”

Line 143: What is a comprehensive radiative transfer model?

Line 150 (GS): Add “on” as “The box AMF depends on wavelength, …”

Line 173 (GS): Add “be” to “… is known to be caused by …”

Line 173: Could the authors provide a reference for the known cause of stripping?

Line 175: AMFs are not retrieved; they are calculated and are independent of a remote Pacific earthshine radiance.

Line 204 & 209 (GS): Change “outliners” to “outliers”.

Line 216 & 217: What is the error associated to the average SCDs? Standard deviations?

Line 222: How is the 32.32% lower calculated? A difference of 0.02 x 10^{16} molecules cm^{-2} looks rather small.

Line 223-226 (GS): Please rewrite sentence.

Line 228: How does BOES compute mean random errors? Does this calculation use the operation product random error definition? Without describing both is impossible to interpret BOAS random errors lower than DOAS by 22%.

Figure 1: What is figure 1 trying to illustrate. It is not mention in the text.

Table 1: What is the definition of viewing azimuth angle? Probably clock wise with respect to North.

Table 2: What is the methodology employed to calculate the Raman spectra? What is the definition of Pacific (remote)? Will it be possible to provide a reference for the pre-flight instrument slit function?

Figure 2: What the figure shows is the ΔSCD not SCD. Please correct. The size of the residuals in plot (b) seem to be large considering an RMS of 2.78 x 10^{-4}.

Figure 3: A different color map, extending to negative values, will provide a better picture of the BOAS retrieval performance since there is a significant number of pixels with negative columns. Which orbits contribute to these plots?

Figure 4: Does plot (a) show ΔSCDs and plot (b) SCDs? Please clarify. Does this regression consider an independent variable? Probably is better to consider the errors of both variables in the linear fit.

Table 3: It is very difficult to interpret. How can be the NMBs between satellite and MAX-DOAS be 0% for the operational product? The caption needs to be re-written to provide a proper description of what the table is showing.

Figure 5: What is the methodology used to calculate the vertical profiles? Time averaging, filtering of MAX-DOAS observations… Co-registration of models and models with in-situ measurements.

Figure 6: Using a divergent color map scale in plot © will help to appreciate the positive and negative differences. What is the effect of clouds and aerosols? Some of the big differences resemble cloud structures. What is the correlation between both AMF calculations?

Figure 8: Do these plots show improve HCHO data? It seems to show gridded data. What is the spatial sampling of the grid? Which methodology was used to calculate the gridded fields.