Review of Su et al., 2020: “An improved TROPOMI tropospheric HCHO retrieval over China”.

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.

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?

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.

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”.

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. 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.

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.

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.

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

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.

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.
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?

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.

Technical corrections

Abstract

L18: We present ~~the~~ an improved retrieval…
L19: The new retrieval optimizes the slant column density retrieval: this is not demonstrated in the paper. Please rephrase.
L24: MAX-DOAS measurements in ~~China~~ Beijing
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.
L29-30: The last sentence is not demonstrated in the paper.

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.
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.

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

WRF-model

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

Improved HCHO retrieval algorithm

L133: in Table ~~1~~ 2
Table 2:

- Why the use of DSCD in the caption?
- Please indicate the differences compared to the operational product.
- What about the Ring correction?

- Do you include corrections for non-linear Ozone absorption effects?
- How is the radiance reference sector calculated? Per instrument row? Per day?
- Why this particular choice of O3 and BrO cross-sections? Can these choices explain the differences with the operational product?

L168: the surface albedo is obtained from the S5P operational cloud product. This is a bit surprising. Please specify the wavelength.
L187: specify the meaning of k and m

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?
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.
L206: Please compare numbers for the corrected slant columns over Tibet.
L218: Please compare the RMS.
L218-219: This sentence is vague. Please be more specific
L223: Please give the numbers in brackets for the Chinese product as well.
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.
L228: The mean random errors relative to BOAS are mentioned. Can you give a definition? And where are those errors presented in the paper?

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.
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.
- How many profiles are averaged? What is the spatial resolution?

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?
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.

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

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.

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.