Review of

"**Monitoring and assimilation tests with TROPOMI data in the CAMS system. Part 1: Near-real time total column ozone**"

By Inness et al., ACPD 2018

This manuscript describes the first testing of TROPOMI TCO3 NRT retrievals version v1.0.0 as provided by DLR in the CAMS system. They describe in detail the actual testing procedure that has been performed so far with intention to arrive at operational application of the TROPOMI O3 data within the CAMS system.

Even though significant biases have been detected in this early retrieval product, the authors show that once applied in the CAMS data-assimilation system the automatic bias correction scheme is able to remove a large portion of these biases, which then brings the TROPOMI data within magnitudes such that no deterioration, and even a small improvement is obtained when compared to independent observations. Also this configuration shows the strength of having the monitoring capabilities in the CAMS system to be able to quickly assess potential issues in the retrieval product (already during commissioning phase). This is valuable information to the retrieval product developers.

I agree such careful assessment of any retrieval product is a very valuable application of the CAMS system. However, the scientific interest of current testing may be limited, considering that the authors have clearly used a preliminary retrieval product (details, as will described in Loyola et al., 2019, are still "in preparation"), which is not expected to be used within the community. Hence, their assessment, even though very relevant for the developers of the retrieval algorithm and for its potential impacts when applied operationally within CAMS, may be of limited general interest to deserve documentation in a scientific publication of this kind. As result, the number of figures could possibly be reduced in the main text, and shifted to supplementary material, with exception of a selection of key figures, for which the figure quality could be improved. Furthermore, the evaluation period (effectively only about three months of data) is really short and also, for instance, not including the austral spring, which makes it hard to make an overall judgement.

Consequently. the description of the procedure to introduce new satellite instruments in the CAMS system is still valuable information, but may fit better in a GMD-type of publication.

Furthermore, the authors conclude that the TROPOMI data do not degrade the ozone in the current configuration and therefore are intending to include full assimilation of TROPOMI NRT TCO3 data in the operational CAMS analysis soon. Their argument is that this adds redundancy and resilience in case that some of the older instruments stop working. Here I have severe problems. It is true that the significant biases, as visible from the current, preliminary, version of the retrieval algorithm are more or less efficiently removed in the full data-assimilation system, but this is exactly owing to the presence of the existing observation system, including, for example, MLS and OMI onboard of AURA, which has already exceeded its foreseen lifetime. When these observations stop to be delivered, the bias correction algorithm may not work as efficiently, giving more weight to the (currently erroneous) TROPOMI retrievals. The magnitude of this effect has not been evaluated. Hence, to validate their statement that

adding TROPOMI with its current retrieval algorithm indeed adds the necessary resilience I suggest the authors to include two additional sensitivity experiments:

- One where observations from (for instance) the AURA instrument (MLS and OMI) are no longer included, and also no TROPOMI data is included
- One without the AURA observations, but now including TROPOMI observations to take over.

Such a study would indeed make a clearer case as to resilience of the system in the case of a sudden failure of these important instruments in the CAMS system, which would be a valuable assessment of the CAMS system that is also of quite more scientific interest.

More detailed comments

p1, l. 19 “agree well”: I do not agree here: the TROPOMI data show biases wrt the CAMS system of up to 50 DU locally. Please reformulate.

p1, l.28 “less” should be “more”?

Abstract: this is too lengthy to my taste, and authors should try to condense. It doesn’t need to be complete.

P1, l.39 (and also p11, l6): “After more tests”: which kind of testing is meant here?

P2, l.34: the authors describe here the importance of the ozone hole in the earth system. This subject is however not covered in the assessment of the ozone assimilation experiment. The issue is that the reader is directed to wrong expectations.

P3, l.7: the authors discuss the important issue of resilience of the system. However, in my opinion they don’t fully test this in the current manuscript, see my comments above. Please consider revision of statements, or inclusion of additional sensitivity experiments.

P3, l.22: “Because the departures …”: I don’t understand this sentence, please consider re-formulation.

Introduction: While authors make the case for the importance of ozone assimilation in CAMS, I miss references to other examples where new satellite data has been tested in early phase, e.g. in terms of composition, or possibly meteorology.

P3, l.31: “are not too large”: this sounds like a very subjective statement. Can this be quantified in any way?

P4, l.3: for more clarity, suggest to rewrite “… as originally implemented in…” or “… as originating from the Chemistry…”

P4, l.30: here the authors state why the application of averaging kernels is not needed, either because they are not needed (having profile data as in MLS) or not available. This I find questionable, and I wonder if averaging kernel aspects could be important to reconciliate differences found between, for

instance, the GOME-2 and OMI retrievals as seen from Figure 5. At least, it would be good if the authors can back-up their statements with literature on the subject.

P6, l.12 “that all four instruments agree well”: I find this statement with the current dataset a bit problematic, particularly when authors note differences in total columns of “up to 60 DU” (p6, l17). Please consider re-formulation

P6, l.30: “Tropics”,“problem with the OMI retrievals”: The signal appears actually mostly isolated over the tropical continents. Therefore, couldn’t it be a sign of something physical, associated to tropospheric chemistry, which is indeed picked up in the OMI retrieval and not in GOME-2, e.g. because of different sensitivities towards the lower altitudes? Evaluation with independent data shows indeed a negative model bias over West Africa (Fig. 16), although the model also shows positive bias averaged over the tropics when evaluated against sondes (Fig. 15). As bottom line, are you confident that the bias correction functioning properly when bias is removed from OMI and not from GOME?

P7, l.26: “they are less than 1% in the NH and SH”, but this includes compensating errors, or? Please comment.

P7, l.37: about “60S and 50N” should be “about 60S and 40N”.

P8, l.2: “destriping correction”: although indeed a sensitivity to the scan position is shown, I find it confusing to refer to the ‘destriping correction’ here. I so far understood that this de-striping has been introduced to handle, for instance, row-anomaly issues with particular affected rows, as is the case with OMI, but not so much with changing signal towards the edges of the scan, which may require a more physical explanation in the retrieval algorithm. Can you comment?

P8, l.10, “blacklisting poleward of 60 degree”. Seeing the emerging biases at 40 degree north, it would make sense to blacklist data from this latitude onwards, or?

P9, l.1, “are less than 2%”. This refers to the average bias for the 26 Nov – 3 May time period. But as is clear from Figure 10, there is only little data in the December-January time period, which means that the difference for this period with/without TROPOMI data is negligible. Therefore, I believe it is more meaningful to present & report these biases for the 28 January – 3 May period, as done also in Figure 13.

P9,l.5: ”26 November”: Figure caption of Fig. 13 writes “ 28 January”. Please check.

P9, l.10: “the impact would be larger”: Yes, I agree, and this is worrying, particularly as the positive impact, seen so far, may no longer be the case. Could you please comment?

P9, l.35 “observation**s**”

P9, l.39: “Loyola et al., 2019”: the current data-stream appears preliminary, considering that the algorithm description has not been published yet. This is important to stress.

P10, l.20 “solar elevation”: this seems only partly an appropriate predictor. Would the CAMS system be flexible enough to add different predictors in the system for its bias correction scheme, such as albedo?

P10, l.34 “to add redundancy”, This is actually not tested, please see my comments above. Please comment.

Figure 10: Figure quality appears insufficient for publication, and legend seems incomplete. Please consider improved legibility.