

## Referee report to the revised version of the “Joint spectral retrievals of ozone with Suomi NPP CrIS augmented by S5P/TROPOMI” manuscript by Edward Malina et al.

In the revised manuscript, all errors and incorrect formulations have been eliminated but not all of my previous comments have been addressed in text or answered properly. The manuscript is still very lengthy and lacks a clear focus. It looks rather like the authors made a lot of work within different studies and just put the results together without carrying much about the necessity and logical flow. In my opinion, the manuscript still needs a revision to make it shorter and more concise. The authors should also clarify if the focus of their study is on the profiles or rather at stratospheric and tropospheric columns. The latter is suggested by the presentation strategy of the validation results but is not clearly stated in the text.

### Detailed comments

- The authors ignored my request to show a reasonable time evolution of the ozone data (at least for a year). Instead they only extend the MLS comparison to one month and show one additional day in April in the Supplement. This additional day is presented in an absolutely unreadable way for all comparisons in one plot (Fig. S7). No attempt to analyze vertical profiles for a day in a different month was made. Even a sparse illustration of the time evolution for August presented for MLS in Fig. 10 was not repeated for AIRS-OMI comparisons.
- The comparison of the total ozone column from TROPOMI with the GODFIT OFFL TROPOMI TOC product is motivated by authors by the fact that they use the same Huggins band. This motivation is in my opinion not really convincing. If a comparison of the total ozone is really needed why the comparison data should necessary be derived from the Huggins band? Another question is how this comparison is related to objectives of the paper. The total ozone is not a product analyzed in the manuscript and its validation is not really relevant for the results. If the authors think, the total column comparisons might bring some additional information, they should discuss what exactly is the information and how the obtained results are related to the main topic of the manuscript. I also ask myself why then the total ozone columns from CrIS are not included in comparison, or may be even those from AIRS and OMI if these comparisons are really related to the topic.
- Comparisons with CAMS and MOMO-Chem cannot be treated as validation results. In the presented form the comparisons add nothing to the results already presented for AIRS-OMI. If the authors wish to investigate strengths and weaknesses of the chemical models, this must be done with much more details and stronger focus on the problematic locations after the validation part is completed. For this activity, the usage of all three retrievals is absolutely unnecessary. The comparison should be done

using the retrieval which is rated best based on the validation results. Comparisons with more than one instrument are also useful in this case.

- With respect to Sect. 6.1, I agree that the MUSES retrieval using the joint TROPOMI bands 1 and 2 in combination with CrIS measurements will be most probably substantially different from that discussed by Mettig et al., however, the real retrievals will also be substantially different from the synthetic simulations shown in Sect. 6.1 as the noise characteristics and regularization will need to be adjusted and it is impossible to say for now, how strong this potential adjustment will change the results. The results of Sect. 6.1 would be useful if it was the first publication on this topic, but for now, we already know that adding UV bands increases DOFs and we also know that synthetic retrievals often do not reflect the results which will be achieved in the real retrieval after making it stable. Thus, any qualitative assessment performed with synthetic data and not exactly the same retrieval settings as used for real retrievals should be taken with care. For this reason, I still do not agree Sect. 6.1 is useful.
- Line 206: "... facilitating independent validation of the satellite tropospheric ozone products": a model cannot be seen as a reasonable validation source for measured data. The only exception is a usage of a model to assess an overall plausibility of data if no other validation sources are available. Here, it is, however, not the case.
- Figure 2: showing spectral fits from the retrieval would be useful.
- Line 341: "In general, (TIR) have greater sensitivity in the lower troposphere (surface to 500 hPa) whereas the UV is more sensitive to upper tropospheric ( 500 hPa to tropopause) ozone" - this statement is very difficult to follow as there is no 500 hPa mark in the plots. Generally the statement is a bit confusing as it is unclear what e.g. "(TIR) have greater sensitivity in the lower troposphere (surface to 500 hPa)" refers to. The TIR Jacobians are clearly larger for 8 - 100 hPa level than below 500 hPa. With the values of around -6 the Jacobians for UV are much higher than those for TIR (around -15) even in the lower layers. Please be more precise and mark 500 hPa level in the plot.
- Line 381: "CrIS Jacobians, in contrast, show sensitivity at multiple pressure levels due to the many spectral windows used for the CrIS retrieval." - the statement does not seem to be correct. Jacobians have their maxima just below 10 hPa for all windows. First two windows seem to have a dominating contribution while the contribution from the last four windows is rather minor. No clear differentiation of the sensitivity with pressure for different spectral windows is seen from the plot.
- Line 391: "CrIS-TROPOMI generally shows higher DFS values than CrIS alone in both the troposphere and the stratosphere" - For the troposphere, the differences are hardly visible, the statement is poorly justified.

- Line 394: “Further, there are numerous cases for CrIS and CrIS-TROPOMI where DFS values of 2 are achieved in the whole troposphere (Fig. C1).” - If you mean cases in the 2-2.5 bin, they cannot be really defined as “numerous”, if you mean 1.5 - 2.0 bin, most of them do not reach 2.
- Line 395: “CrIS-TROPOMI and CrIS are highly useful for tropospheric ozone estimation.” - In my understanding, the paper is meant to be focused on highlighting the goals from using TROPOMI, surely CrIS is useful, but does TROPOMI add anything?
- Line 404: “... is largely consistent throughout the atmosphere ...” - what does “consistent” mean in this case?
- Line 405: “... where the magnitude is lower.” - do you mean the magnitude (of uncertainty) is higher? If not, which magnitude is meant here?
- Figure 8: the value of the comparison between the daytime and nighttime results of CrIS is unclear. The nighttime data are clearly irrelevant in the framework of the presented study and the comparison should be removed for the sake of brevity and clarity.
- Figure 8: Information on the top and bottom pressure in the panel (c) needs to be provided. Generally, minor ticks on the pressure axis of all plots would increase the readability of the plots.
- Figure 9: The reason to show the nighttime CrIS measurements is unclear. It is also unclear if these measurements are included into the regression fit.
- Line 495: “ A key statistic is the CrIS-TROPOMI mean difference (0.32 DU) is significantly lower than either CrIS (3.8 DU) or TROPOMI (-2.88 DU) alone, again indicating improved performance from CrIS-TROPOMI.” - here, it should be noted that the improvement of the agreement in the stratospheric columns is most probably caused by a strong negative bias of the CrIS-TROPOMI retrieval above 10 hPa resulting in an error canceling.
- Line 512: “Larger disagreement is apparent above 10 hPa, with differences up to 40% shown outside of the tropics. This is expected as the CrIS-TROPOMI sensitivity decreases above this pressure level, while MLS retains high sensitivity.” - In the same way, it must be expected for CrIS-only retrievals but it does not happen. The explanation seems to be unsuitable.
- Figure 10: It is nice that a temporal evolution within August is now shown, however, it is not sufficient to show it only for one month and only for the global mean differences. The plot needs to be extended to cover at least one year and to show differences for different latitude bands.

- Sect. 5.2 and some other sections: it is not a good idea to present an extensive discussion in the main text of figures which are shown in appendices. Either the plots need to be moved to the main text or their discussions (if considered less relevant) need to be moved to appendices.
- Line 533: “have similar differences between -10 and 10% in” - I do not think it is true, there is a lot of cyan (and partially also blue) color in the mid-latitudes, which marks the differences  $< -10\%$
- Line 536: “... indicate excellent agreement ...”- CrIS - TROPOMI shows a significant negative bias in panel (a) of Fig. B2, the profile comparison in panel (c) looks just terrible for CrIS - TROPOMI. The rating “excellent agreement” is not acceptable here.
- Line 561: with respect to the CAMS and MOMO-Chem models authors state that “these reanalysis datasets are not validation sources, and areas of disagreement should be considered as a point for further discussion and evaluation.”, however, in the following discussion, the models are used in exactly the same way as validation data and no additional discussion and evaluation is provided.
- Line 581: “... with differences  $< 5\%$  ...” - I do not think it is true as there is a lot of cyan, blue and yellow colors in the CrIS-TROPOMI/CrIS-only plots.
- Line 665: “improves comparisons against MLS in the stratosphere” - this is only true for stratospheric columns, comparisons for the profiles look much worse than those for CrIS-only. This must be mentioned in the summary.
- Line 667: “differences with CAMS and MOMO-Chem raise interesting scientific questions for further analysis.” - either the questions need to be formulated or the statement deleted.
- Line 688: “... future MUSES joint CrIS-TROPOMI UV1 & UV2 retrieval will differ from this work.” - most probably it is true but, for now, this “future” retrieval does not yet exist and such a statement is a pure speculation, which should not appear in a scientific manuscript.
- Line 698: “ Cross comparisons of CrIS-TROPOMI/CrIS/TROPOMI, with independent datasets from MLS, MUSES AIRS-OMI, show in general CrIS-TROPOMI has the highest quality performance relative to the other instruments.” - This statement is not correct. With exception of Fig. 10, CrIS-only retrievals in comparison with MLS and AIRS-OMI show at least the same performance and sometimes are even a bit better than CrIS-TROPOMI.

## Technical corrections

- Line 41: “... via chemical means are attached as ...” → “... via chemical means, are attached as ...”
- Line 424: “... having too large a magnitude.” → “... having too large magnitude.”
- Line 528: “... in the tropospheric column. similar magnitudes ...” → “... in the tropospheric column. Similar magnitudes ...”
- Line 576: “CrIS-TROPOM/CrIS-only” → “CrIS-TROPOMI/CrIS-only”