Dear Editor and Reviewer,

Thank you for your continuing efforts in reviewing our paper, in order to respond to your comments we have kept your original comments here in black, our responses are in blue, and specific changes are in underlined blue.

We thank the associate editor and reviewer for their patience in dealing with this paper, and we hope that this response will answer all remaining criticisms.

Editor Comments

Dear Ed,

Reviewer #1 (R#1) provided very detailed comments on your paper and suggested major revisions to your manuscript. I agree with him/her that the paper should be shortened and the focus of your paper sharpened. I suggest the following major modifications mainly based upon R#1 to shorten the paper.

Thank you for the review, we respond in line below.

* removal of the total ozone comparison

This comparison has been removed.

* removal of the comparisons with CAMS and MOMO-Chem as these comparisons, given the unknown uncertainties of the models, do not provide clues on the quality of the retrievals

We have removed these comparisons.

* removal of Section 6.1 (synthetic retrievals) as the settings for a retrieval from real data generally differ significantly

Upon further discussion with you, we have decided upon the following action which we hope the reviewer will agree with. We accept the argument from the reviewer that the synthetic retrieval will be different to an eventual operational retrieval, we do not agree that it will be as drastically different to the synthetic retrievals as stated by the reviewer. We feel these retrievals are significantly different from the Mettig el al paper, and are therefore of interest. In addition, the use of the UVIS and CrIS bands isn't shown elsewhere, and is therefore unique. Based on this argument, we have kept this study in the paper.

R#1 made additional valuable comments, which shall be addressed

Please find our responses to the review below.

Please provide a revised manuscript with track changes and an item-by-item reply to the reviewer's and my comments.

Kind regards,
Mark Weber

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.

In accordance with your and the associate editors recommendations, we have shorted the length of paper by removing the sections relating to the total column comparison, and the comparisons with the Chemistry Transport Models. The appendix has been shorted, and the supplementary materials removed. We highlight at the beginning of the validation/intercomparison section that from this point the study is focused on column intercomparisons.

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.

We understand the reviewers point, and we do accept that an ideal scenario would have 1-year of comparisons for the satellite retrieval cross-comparisons. Unfortunately, this is not practical at this time, currently we do not have 1-year of CrIS-TROPOMI/CrIS/TROPOMI or AIRS-OMI retrievals readily available. The processing required to create this amount of data, at the resolution generated in the last update requires a significant amount of time, especially for AIRS-OMI retrievals. Due to current HPC scheduling and staffing conflicts at the first author's previous institution where this work was undertaken, creating a 1-year dataset is unfortunately not practical at this time. What we have done to address the reviewer's concerns is extend the CrIS-TROPOMI/CrIS/TROPOMI vs MLS comparisons to two months, and included a couple of days of comparisons in different months to include some information on seasonal bias. We have also added CrIS-TROPOMI/CrIS/TROPOMI vs AIRS-

OMI comparisons for one month and several days in different seasons. We have removed the results in the supplementary material, and included this data into the time evolution plots of Figures 10 and 13.

- 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.

In accordance with your recommendation and the directive of the associate editor, we have removed the comparisons with the Total Column product.

- 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 intrument are also useful in this case.

In accordance with your recommendation and the directive of the associate editor, we have removed the comparisons with CAMS and MOMO-Chem.

- 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.

We accept the argument from the reviewer that the synthetic retrieval will be different to an eventual operational retrieval, we do not agree that it will be as drastically different to the

synthetic retrievals as stated by the reviewer. We feel these retrievals are significantly different from the Mettig el al paper, and are therefore of interest. In addition, the use of the UVIS and CrIS bands isn't shown elsewhere, and is therefore unique. Based on this argument, and upon further discussion with the associate editor, we have decided to keep this part of the paper.

- 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.

The comparisons with the CTMs CAMS and MOMO-Chem have been removed, therefore this sentence has been removed.

- Figure 2: showing spectral fits from the retrieval would be useful.

The red lines on Figure 2 are actually the spectral fit of the fit windows we use, this was not obvious from the figure caption, this has been changed.

- 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.

We have modified this section to read as follows:

For the CrIS Jacobians shown in Fig. 3 b) peak ozone sensitivity is found in the 9.6 and 10.43 micron bands primarily between 100 and 10 hPa, but does extend through to the lower troposphere (1000 hPa) and the upper stratosphere (1 hPa), indicating sensitivity through the whole atmosphere, as opposed to TROPOMI-only.

- 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.

We have modified to read as follows:

CrIS Jacobians, in contrast, show sensitivity at multiple pressure levels through the troposphere and stratosphere.

- 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.

We have replaced this sentence with the following:

Note that according to the histograms in Fig. B1 b) and c) CrIS-TROPOMI can exhibit higher DFS values than CrIS alone in the stratosphere, although as indicated by Fig.4 the differences can be subtle. The differences are very minor for the troposphere, so this is not obvious from the histograms.

- 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.

We have changed this sentence to the following, to reduce the strength of the statement:

Further, there are a small number of cases for CrIS and CrIS-TROPOMI where DFS values of more than 2 are achieved in the whole troposphere

- 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?

Thank you for this point, a significant aspect of this paper is to highlight the capability of CrIS-only, as well as CrIS-TROPOMI. We note that according to the DFS figures, indeed TROPOMI does not add that much in terms of DFS, but according to the validation with the ozonesondes, CrIS-TROPOMI does show an improved performance. We would therefore argue that TROPOMI does add value.

- Line 404: “... is largely consistent throughout the atmosphere ...” - what does “consistent” mean in this case?

Upon reviewing this sentence, we decided that as written it does not add much value, and we have therefore removed it.

- Line 405: “... where the magnitude is lower.” - do you mean the magnitude (of uncertainty) is higher? If not, which magnitude is meant here?

As with the above comment, we have removed this sentence as it does not add much value.

- 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.

Part of the aim of this paper is to inform potential users of what data is available from this algorithm. We therefore believe it is important to highlight the existence of night-time only data, even if it does not play a significant role in the remainder of the paper.

- 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.

We have added additional pressure ticks into this plot.

- 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.

We address this question in a previous comment, the measurements are not included in the regression fit, but are included in the earlier aspects of the paper such as the information content and uncertainty analyses.

- 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.

The following text has been added alongside the longer term comparison figure.

In addition, Fig. 10 further highlights the point shown in Fig. 9 that there is a strong likelihood that the improved performance of the CrIS-TROPOMI stratospheric column values being the result of bias cancellation in different parts of the stratosphere.

- 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.

We agree with this statement, we have modified this sentence to read as follows:

Larger disagreement is apparent above 10 hPa, with differences up to 40% shown outside of the tropics, that remain to be investigated.

- 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.

As discussed above in the reviewer's previous comment, we regret that we are limited in the amount of processing that can be done for this study. We have added another month (July) for extended comparisons, and additional days outside of the summer. Due to the relatively sparse nature of these retrievals, and of the other instruments, co-locations are limited and we do not feel there will be additional benefit to comparisons within 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.

We agree with this point, the AIRS-OMI figure that was previously in the Appendix has been moved into the main text and discussed in context there. All other instances (Figures A1 and B1) are only discussed briefly in the main text, and we have kept as in the previous version.

- 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%

We have modified this sentence as follows:

Both CrIS-TROPOMI and CrIS have similar differences generally between -10 and 10% in the mid-latitudes through the troposphere, although there are cases with larger differences >40%.

- 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.

We have reduced the strength of the language here to good agreement, as we feel this is what is shown in panel (a), the bias is quite low (1-3%). We agree in panel (c) that the differences are more pronounced, but these are discussed further down in the text, and we added the text:

contrasting the overall results shown in (a)

- 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.

All references to CAMS and MOMO-Chem in this paper have now been removed.

- 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.

All references to CAMS and MOMO-Chem in this paper have now been removed, so this correction is no longer necessary.

- 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.

We have modified a sentence in the summary to read as follows:

The addition of the short TROPOMI window to form CrIS-TROPOMI, improves comparisons against MLS stratospheric columns, although challenges remain for stratospheric profile comparisons, requiring further investigations.

- 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.

All references to CAMS and MOMO-Chem in this paper have now been removed.

- 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.

Statement modified to:

Room for further investigation.

- 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.

We have modified this paragraph to read as follows:

Cross comparisons of CrIS-TROPOMI/CrIS/TROPOMI, with independent datasets from MLS, MUSES AIRS-OMI, show some positive results for CrIS-TROPOMI, especially focusing on stratospheric column comparisons with MLS the stratospheric 'gold standard' on August 12th 2020, a linear slope of 1.029, bias of -0.32 DU and correlation coefficient of 0.952 are

found, highlighting the quality of the retrievals. A month long comparison in August 2020 shows a constantly lower bias between MLS and CrIS-TROPOMI and either CrIS or TROPOMI alone. Despite being a TIR instrument CrIS shows high linear correlation with MLS, indicating the utility of CrIS by itself. CrIS-only shows a linear slope of 0.921, bias of 3.8 DU and a correlation coefficient of 0.926. TROPOMI shows a significant bias, with a slope of 0.898. However, room for improvement is also identified, with large biases in the upper stratosphere identified with CrIS-TROPOMI.

Technical corrections

- Line 41: "... via chemical means are attached as ..." −→ "... via chemical means, are attached as ..."

Corrected.

- Line 424: "... having too large a magnitude." −→ "... having too large magnitude."

Corrected.

- Line 528: "... in the tropospheric column. similar magnitudes ..." −→ "... in the tropospheric column. Similar magnitudes ..."

Corrected.

- Line 576: "CrIS-TROPOM/CrIS-only" −→ "CrIS-TROPOMI/CrIS-only"

We couldn't find this particular correction, it may have already been corrected.