Dear reviewer,

Thank you for reviewing our paper, we appreciate that with its length, it was a significant time commitment.

In order to respond to your comments we have kept your original comments in black, our responses are in blue, and changes to the paper are identified with underlined blue.

Reviewer 1:

Referee Report to "Joint spectral retrievals of ozone with Suomi NPP CrIS augmented by S5P/TROPOMI" by Edward Malina et al. The manuscript presents an algorithm to retrieve vertical profiles of ozone by combining IR measurements from the CrIS instrument with UV measurements from TROPOMI. The idea has already been explored and results has been published by several authors. The authors highlight that their algorithm differs from the previous ones by a usage of a reduced spectral information, which makes the algorithm less demanding with respect to the calibration of the data reducing, however, substantially the information content in the stratosphere. As a result, the obtained improvement over the pure CrIS retrieval is often vanishing small. The authors must be honest and clearly state this in the abstract. Although the data from other combined CrIS-TROPOMI retrievals are available, the authors make no attempts to make comparisons. Instead many of quite useless comparisons are presented, e.g. with the total ozone product from GODFIT OFFL TROPOMI or ozone field from chemical reanalysis models. The paper is very lengthy and quite boring because of many similar plots (Figs. 10–14), which do not provide much information. In general, however, the obtained results might be useful for certain applications and the manuscript can be published after a major revision.

Major comments

• Please be objective in the abstract and conclusions. You completely ignore the fact that for some conditions the results from CrIS-TROPOMI are worse compared to those from CrIS retrieval. Considering only minor differences between the results from CrIS-TROPOMI and CrIS, as presented in the manuscript, the statement "These results demonstrate that CrIS/TROPOMI retrievals have the potential to substantially improve our understanding of ozone." is a clear over-rating.

We have removed over-rated statements, replacing them with more moderated statements, for example.

"These results demonstrate that CrIS and CrIS/TROPOMI retrievals have the potential to improve global satellite ozone retrievals, especially with further future developments"

• The message of the comparisons with a lot of different data products for one single day performed in Sect. 5 (Figs. 10–14) is unclear. As the declared objective of the paper is to present advantages of the CrIS-TROPOMI retrieval, 2-3 comparisons of this kind would be sufficient. Much more interesting would be an analysis of the time evolution, which is completely missing in this part of the paper. Please provide some time evolution plots or at least the plots for different seasons as it is done for ozonesonde comparisons.

Additional comparisons were provided in the supplementary material, however further analysis is now provided. We provide a month long comparison of MLS vs CrIS-TROPOMI/CrIS/TROPOMI, which shows consistent results.

We have also provided a summary of the results from the cross-comparisons and validations to help highlight the take home messages. Which reads as follows:

"The results shown in this paper and the supplementary material shows that CrIS-only agrees well with all datasets compared against it, both in the troposphere and the stratosphere. The addition of the short TROPOMI window to form CrIS-TROPOMI, improves comparisons against MLS in the stratosphere, with all results shown in the main paper and the supplementary material having the lowest mean difference bias of CrIS-TROPOMI/CrIS/TROPOMI. Differences between CrIS-TROPOMI and CrIS-only are less clear in other stratospheric comparison cases, however, differences with CAMS and MOMO-Chem raise interesting scientific questions for further analysis. Comparisons with satellite data and reanalysis in the troposphere do not show clear differences between CrIS and CrIS-TROPOMI, however, CrIS-TROPOMI shows better performance than CrIS-only against ozonesondes, which is indicative of improved performance in the troposphere through joining CrIS and TROPOMI. "

• The goal of the comparisons with chemical reanalysis models is totally unclear. The authors state these comparisons cannot be treated as validation, so what exactly is the purpose of this comparisons? The purpose of the comparisons with the total ozone data from GODFIT OFFL TROPOMI is also unclear. Why do you compare the total ozone but not compare the profiles from other CrIS-TROPOMI retrievals? I think the manuscript would largely benefit if you make it shorter by removing Sects. 5.3 and 5.4.

The purpose of the comparison is to show how CrIS-TROPOMI compares with a chemical reanalysis that in turns has been informed by data. Given the comparisons of CrIS/TROPOMI with independent data, e.g., sondes, this comparison shows qualitatively the information that CrIS/TROPOMI could provide.

Thank you for the points, however we are confused by the reviewer's statement that the comparisons with chemical reanalysis models is unclear.

With respect to GODFIT retrieval, our TROPOMI ozone retrieval is similar to that retrieval, in that the same spectral window is employed. It is our aim therefore to prove that our TROPOMI-only retrieval is comparable to what is currently operationally in use and add confidence to its use in the joint retrievals. This is why do not compare GODFIT to our other retrievals. When we presented our work at conferences, we were specifically recommended to include this analysis, and we therefore think it remains important in this work.

We have added the following text to section 2.3.3 to further clarify.

The GODFIT and the MUSES TROPOMI retrieval algorithms aim to quantify similar quantities (identified in detail in Sect.3.3) through the use of similar spectral windows. It is therefore important to benchmark the MUSES TROPOMI algorithm against the GODFIT algorithm.

• The comparison with ozonesondes is shown only up to 100 hPa although the ozonesondes deliver reliable ozone profiles up to about 20 hPa. Please present the comparison for the entire altitude range covered by ozonesondes.

We have now added this comparison.

• The information presented in Sect. 6.1 has already been published by Mettig et al. (2022) and does not need to be repeated. The fact that the information content increases when using UV bands is generally known. Furthermore, the reliability of an investigation with any assumed settings without having a working retrieval is questionable as adding new spectral ranges requires often an optimization of the settings to keep the retrieval stable, i.e. it is unknown if the retrieval of the real data using the assumed settings is possible. A pre-condition of having a perfect calibration accuracy, as assumed by authors, is never satisfied in the reality. For this reasons, I recommend to skip Sect. 6.

Thank you for this assessment, we accept your point about the perfect calibration not yielding the results as indicated in the paper, however we disagree with the section being unnecessary. We indicate several times that the performance of our CrIS retrieval is significantly different to that shown by Mettig et al. (2022), most notably in the stratosphere. Therefore our theoretical joint TROPOMI bands 1 & 2 – CrIS retrievals are potentially substantially different from those shown by Mettig et al, and are important to show.

• Line 659: ”Focusing on comparisons with MLS the stratospheric ’gold standard’ on August 12th 2020, a linear slope of 1.029, intercept of -7.9 DU (∼3%) and correlation coefficient of 0.952 are found, highlighting the quality of the retrievals.” - A comparison for one day is definitely not enough to make any robust conclusion. Please extend the comparison with MLS to a longer period similar to the comparison with ozonesondes.

In the supplementary material to this paper additional MUSES CrIS-TROPOMI/CrIS/TROPOMI vs MLS comparisons for a number of days are shown, along with further comparisons against the other data sources identified in this paper. We accept that these comparisons were not obvious from the paper, we have now added some additional text to make it obvious these comparisons exist, in the introduction to the ‘Validation and cross-comparison’ section.

“Cross comparisons for additional days, focusing on August 2020, but with additional months are shown in the supplementary material.”

Further, we have now included a comparison of mean differences between MLS and CrIS-TROPOMI/CrIS/TROPOMI over the month of August 2020, please see the new Figure 10. The results showing an improvement in the CrIS-TROPOMI retrievals are consistent over the month, and add weight to our statements.

Minor issues

• Line 233: Please define the scaling matrix D

We have added the following definition:

“the calculation of these values are described in Sects. 5.5 and 6.3 of More (1978)”.

• Line 234: “ with large λ values prioritising the speed of the convergence, but making the steps more non-linear, while small values reduce the speed of the convergence, and is more linear in the iteration, similar to the conjugate gradient method.” - In my opinion it is other way around. Small values of λ result in the Gauss-Newton method, which converges faster but is more non-linear, while large values of λ result in a gradient descent update.

Yes, thank you for correcting us, this has been swapped.

• Figure 1: Please comment why single CrIS and TROPOMI retrievals are necessary before the joint retrieval.

These steps are not necessary for the CrIS-TROPOMI retrieval, but can be used if an updated initial guess is desired by the user. The following text has been inserted.

“Note that Fig 1 shows CrIS and TROPOMI ozone retrievals feeding into the CrIS-TROPOMI retrieval, these steps are not necessary for the CrIS-TROPOMI retrieval, but may be used to provide an updated ozone initial guess depending on user needs.”

• Line 255: “...as only one FoV per observation from the CrIS cross track position is used in the processing, the impact of clouds will be less pronounced.”- please clarify why the impact of clouds is less in this case

We have deleted this passage, which now just reads as:

“The a priori cloud properties come from an 'initial guess' refinement step using brightness temperature differences”.

More details about the cloud clearing methods of the MUSES CrIS-TROPOMI algorithm are provided in section 3.4 with the appropriate context as to why the impact of clouds is less severe than other CrIS retrieval methods.

• Table 6: Please explain how the cloud fraction is used in the retrieval. Please give some details how albedo is retrieved and what the orders mean, is it wavelength dependence?

We added the following sentence with respect to cloud fraction:

"If the cloud fraction is determined to be >0.3, then the retrieval is flagged as poor quality and not considered in further analysis."

The section referring to the albedo values has been modified as follows:

"The a priori for the zero-order albedo term is taken from OMI climatology (Kleipool et al., 2008), with the first and second order albedo terms added to allow for non-linear variation of albedo across the spectral band. Thus the effective albedo forms the quadratic equation,

$$A(\lambda) = A_0 + A_1(1 - \lambda/\lambda_0) + A_2(1 - \lambda/\lambda_0)^2,$$

where A is the effective surface albedo at wavelength λ, A_0, A_1 and A_2 are the zero, first and second order parameters fit by MUSES, and λ_0 is the first wavelength."

• Line 277: "...pixels within a 20 minute time frame (where Suomi-NPP and S5P pass the same scene within 10 mins)" - Please clarify why passing the same scene within 10 minutes results in 20 minutes time frame.

We have added the following clarification.

", allowing for some variation in scene overpass times".

• Line 279: "From the current sounding subset, select all pairs that are within < 50 km distance, and 4) select the pair that has the minimum distance." - what is the reason first to select all pairs within 50 km distance and then select the pair with a minimum distance? I expect that selecting a pair with minimum distance skipping the intermediate selection of all pairs within 50 km distance should have the same result. For the pair with minimum distance it can then be checked if the distance is within 50 km.

The sentence has been changed to:

"check all pairs that are within < 50 km..."

• Line 283: "... additional steps with respect to other target gases do occur in the pipeline, but are not highlighted here." - This text does not provide any information. Please skip it if the over gases are not related to the ozone retrieval or provide more details otherwise.

This sentence has been deleted.

• Line 286: "Relevant cloud properties..." - Please clarify how the clouds are handled within the ozone retrieval. Is scattering within clouds considered?

This sentence has been rewritten as follows:

"Relevant cloud properties (e.g., cloud top height and extinction) for the FoVs are retrieved and passed into the ozone retrieval for quality control, with too optically thick clouds being flagged as poor quality".

• Figure 2: Please explain how NESR is defined.

We have added the following text:

"as defined in Zavyalov et al. (2013)".

• Line 316: "...suggesting that CrIS are subject to larger fit errors,..." - This is not a correct conclusion as mean RMSE for CrIS is the smallest.

This sentence has been replaced with the following:

"The RMSE values indicate larger standard deviations in the CrIS-TROPOMI and CrIS cases, while CrIS has the lowest mean RMSE. This suggests more variability in the CrIS fits which is understandable given the wider retrieval windows, while TROPOMI has a more constant bias."

• Figure 3 and all figures below having the pressure as the vertical axis: 1) upper limit of the pressure axis must be indicated 2) providing a second y-axis in km would help the interpretation of the results and facilitate the comparison with the results of previous publications;

Thank you for this point, we have made these changes.

• Line 321: "In general, longer wavelengths have greater sensitivity in the lower troposphere whereas shorter wavelengths are more sensitive to upper tropospheric ozone." - Please precise which wavelengths and pressure levels you are talking about.

This sentence has been replaced with the following:

"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."

• Equation 3: In the formulation chosen by authors the term "δcs" must not appear as it is implicitly contained in the A[xtrue − xa] term. The citation to (Fu et al., 2018) is inappropriate as they just use the formula form (Worden et al., 2007) without any explanation and misinterpret the notations of (Worden et al., 2007), where the main term was written as Axx[xtrue − xa], i.e. included only a sub-matrix of A related to the main parameter. In accordance with Sect. 3.1 of the manuscript under review both A and x contain all retrieved parameters. Thus no additional cross-term must appear.

Thank you for pointing this out, we have removed all references to cross state error in the equation.

• Equation 6: As above, the last term is a natural part of the first term and must not appear here.

Removed, as per comment above.

• Line 363: “For example, focusing on the Atlantic Ocean, there are regions with clearly improved DFS values from CrIS-TROPOMI, as opposed to CrIS.” - please explain how this sentence follows from Fig. C1.

Fig. C1 refers to the following sentence, which we have rewritten to remove potential misunderstanding.

“Further, there are numerous cases for CrIS and CrIS-TROPOMI where DFS values of ~2 are achieved in the whole troposphere (Fig. C1).”

• Line 364: “DFS values of 2 are achieved” - “values between 1.5 and 2” would be more correct.

Thank you for this point, however, this was not the point we are trying to make. We are identifying that these instruments/combinations can achieve 2 DFS in the troposphere, which is highly significant.

• Line 365: “This suggests that CrIS-TROPOMI and CrIS are highly useful instruments for tropospheric ozone estimation.” - the notation “instrument” is incorrect if applied to CrIS-TROPOMI.

We removed the term “instrument”.

• Figure 6 caption: “the measurement or precision error” - do you mean “measurement noise error”?

Yes, sometimes these are used interchangeably, however we have changed the caption to “measurement noise error”.

• Line 382: “with the most reduction at the tropopause.” - at which pressure level is the tropopause?

We clarified this statement as follows:

“with the most reduction above the tropopause (~100 hPa)”.

• Line 383: “Given that the majority of the DFS are contained within the stratosphere for CrIS (Fig. 5), this is the expected result.” - why the reduction of the uncertainty is expected at the tropopause and not in the stratosphere, where the majority of the DFS are contained?

The previous sentence was supposed to read as:

"Comparisons of the total uncertainty with the a priori uncertainty shows a general reduction in the uncertainty, with the most reduction above the tropopause region"

With 'above' replacing 'at'.

• Line 383: "except in the lower and upper stratosphere" - please specify pressure levels.

We have changed this to:

"…above 10 hPa…"

• Line 383: "... the variability of the total uncertainty is smaller than that of CrIS, suggesting that the inclusion of the TROPOMI radiances reduces the uncertainty of the CrIS retrievals." - reducing the variability of the uncertainty does not necessary mean reducing the uncertainty itself.

We have re-written this sentence as follows:

"The key difference is that the variability of the total uncertainty is smaller than that of CrIS, and the total/smoothing error is slightly smaller. Suggesting that the inclusion of the TROPOMI radiances reduces the uncertainty of the CrIS retrievals slightly."

• Line 424: "30◦ and 50◦" - I guess you mean northern latitudes, please precise.

Yes, thank you, corrected.

• Line 443: "For example, in Mongolia" I am not sure everybody can easily find Mongolia on your maps, please provide lat/lon.

Rough coordinates were added.

• Sect. 5: when discussing stratospheric and tropospheric columns vertical ranges must be specified.

At the beginning of section 3.5. we have added a paragraph explaining our pressure level sub-column definitions.

"The following analysis compares ozone retrievals in sub-columns and individual pressure levels. The sub-columns are defined as the troposphere (surface to the tropopause), lower troposphere (surface to 500 hPa), upper troposphere (500 hPa to the tropopause) and the stratosphere (tropopause to 1 hPa)."

• Line 510: " These results highlight the utility of the CrIS-TROPOMI retrievals in the stratosphere" - It is seen from Fig. B2a that the bias of CrIS-TROPOMI results is larger than for CrIS, i.e. the black solid line is shifted more upwards from the blue dashed line for CrIS-TROPOMI case. Is the intercept calculated correctly? If yes, is it representative for a bias? Also from Fig B2b the bias for CrIS-TROPOMI seems larger (green-blue color) as compared to that for CrIS (green-yellow colors). Please re-calculate the bias using the conventional definition as a mean difference and check your conclusions.

Thank you for pointing this out, we have re-written this section to take this point into account.

"CrIS-TROPOMI indicates a high degree of linearity (slope = 0.997), while CrIS-only and TROPOMI-only show similar linearity to a lesser magnitude. Comparable R^2 values are shown for all three cases, while the main difference is the mean difference (bias) with AIRS-OMI, where CrIS-only shows the lowest magnitude difference (-0.03), significantly lower than the other cases."

The mean difference is indicated on all plots, and is now referenced in all discussions relating to these plots.

• Line 586: "...e.g., the Atlantic ocean." - please indicate the latitude region you are talking about.

Added the latitude range 0-30°N, 0-30°W.

• Figure 15: Please add a comparison for a priori.

A priori comparison added to Figure 15, and the percentage difference added to the statistics.

• Figure 15: The range of x-axis is unnecessary wide. Please reduce to ±50%.

Corrected.

• Line 622: "as well as the other presented pressure levels across all seasons." - do you mean that RMS for CrIS-TROPOMI/CrIS-only is lower at all pressure levels in all seasons? Please reword this part of the sentence to make it more clear.

The sentence has been rephrased to:

"The percentage RMS difference for CrIS-TROPOMI/CrIS-only is generally lower than for AIRS-OMI/AIRS/OMI."

• Line 629: " ... as well as the stratosphere." - stratospheric results are not considered in Sect. 5.5. Thus, the statement about the stratosphere is not appropriate here.

Phase has been removed.

• Table 9: Please provide statistics for the tropospheric column.

We have added this in Table 8, but not in Table 9, we do not see the value in this since Table 9 is provided to show the impact of not applying the instrument operator to the ozonesondes, the science in of itself is not that useful, and we therefore do not see the value in adding additional information.

• Line 659: "In the stratosphere we find improved performance..." - the performance is improved only slightly, this must be said. An open question remains if the scatter plot really representative for the bias. This needs to be checked by calculating the mean difference over all data. If needed, conclusions have to be adjusted.

We have changed this sentence to following:

"In the stratosphere we find modest improved performance"

• Line 660: "Cross comparisons of CrIS-TROPOMI/CrIS/TROPOMI, with independent datasets from MLS, MUSES AIRS-OMI, JPL MOMO-Chem and CAMS, show in general CrIS-TROPOMI has the highest quality performance relative to the other instruments" - comparisons with JPL MOMO-Chem and CAMS cannot say anything about performance of one retrieval with respect to the other as it is unclear how to rate the agreement or disagreement with the model data.

We have changed the sentence as follows:

"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"

• Line 664: "By contrast, despite being a TIR instrument CrIS shows close linear correlation with MLS, indicating the utility of CrIS by itself" - The vertical region should be mentioned, which this statement is applicable to.

We now indicate in the previous sentence that the stratospheric column is compared, with stratospheric column being defined earlier in the paper.

• Line 686: "MUSES will immediately be able to take advantage of any improvements." - The statement is questionable as including additional spectral ranges often requires an adjustment of the retrieval parameters and subsequent tests and validation. The statement has to be removed.

We replaced "immediately" with "quickly", we feel justified in this statement due to significant other retrieval mechanics already in place, so we will not be started from 0.

• DFS for CrIS retrieval is significantly larger in Fig. 16 in comparison to Fig. 4 (4.12 vs. 3.62), although the scenario is expected to be the same. Please explain why it is the case?

Thank you for pointing this out, this difference is because the AKs in Figure 16 are calculated from a forward model run only, which accounts for the difference in DFS values. We have added the following sentence.

"We also only run the forward model to generate these AKs, which causes a minor disparity between the CrIS DFS values in this analysis, and those shown in Fig 4.".

Technical corrections

• Lines 126 and 128: Corrupted citations

Corrected.

• Line 279: remove ">", remove "4)"

Done.

• Lines 352 and 356: Corrupted citations • Line 379: "is a maximum" - do you mean "is at maximum" or "is maximum"?

Thank you, corrected to "is maximum".

• Line 381: "Comparisons of the total uncertainty with the a priori uncertainty shows.." - the noun is in plural while the verb is in singular

Corrected.

• Line 402: "failures are by and large due to" −→ "failures are by large due to"

Corrected.

• Line 403: "having too large a magnitude" −→ "having too large magnitude"

Corrected.

• Line 443: "For example, in Mongolia" −→ "For example, over Mongolia"

Corrected.

• Line 443: "retrievals had ∼ ×2 greater magnitude" - please replace "had" by "have" and write ∼ ×2 in words.

Corrected to:

"have roughly two times".

• Line 541: “Yet, considering Eq. 7, when using the TROPOMI AK, (which based on Fig. 4 is effectively 0 in the troposphere).” - incomplete sentence

Replaced full stop with comma.

• Line 543: “troposphere retrievals” –-> “tropospheric retrievals”

Corrected.

• Line 543: “This was confirmed when we compared the TROPOMI-only tropospheric column results against CAMS, unmodified by the observational operator. Where the a priori and TROPOMI-only tropospheric column show almost identical comparisons.” - Either it should be one sentence or the second sentence should be reworded to avoid a suboptimal beginning with “where”.

This sentence has been modified as follows:

“This was confirmed when we compared the TROPOMI-only tropospheric column results against CAMS unmodified by the observational operator, given the a priori and TROPOMI-only tropospheric column show almost identical comparisons.”

• Line 582: “... spatial differences ...” - “spatial distribution of differences” would be more appropriate

Agreed, changed.

• Line 594: “both CAMS have MOMO-Chem” - should it be “both CAMS and MOMOChem”?

Yes, thank you, changed.

• Line 622: “lower for CrIS-TROPOMI/CrIS-only than AIRS-OMI/AIRS/OMI” –-> “lower for CrIS-TROPOMI/CrIS-only than for AIRS-OMI/AIRS/OMI”

Changed.

• Line 637: Corrupted citation

Corrected.

• Line 647: Corrupted citation

Corrected.

• Line 685: Corrupted citations

Corrected.

References

Fu, D., Kulawik, S. S., Miyazaki, K., Bowman, K. W., Worden, J. R., Eldering, A., Livesey, N. J., Teixeira, J., Irion, F. W., Herman, R. L., Osterman, G. B., Liu, X., Levelt, P. F., Thompson, A. M., and Luo, M.: Retrievals of tropospheric ozone 6 profiles from the synergism of AIRS and OMI: Methodology and validation, Atmospheric Measurement Techniques, 11, 5587-5605, https://doi.org/10.5194/amt-11-5587-2018, 2018.

Mettig, N., Weber, M., Rozanov, A., Burrows, J. P., Veefkind, P., Thompson, A. M., Stauffer, R. M., Leblanc, T., Ancellet, G., Newchurch, M. J., Kuang, S., Kivi, R., Tully, M. B., Van Malderen, R., Piters, A., Kois, B., St¨ubi, R., and Skrivankova, P.: Combined UV and IR ozone profile retrieval from TROPOMI and CrIS measurements, Atmos. Meas. Tech., 15, 2955-2978, https://doi.org/10.5194/amt-15-2955-2022, 2022.

Worden, H. M., Logan, J., Worden, J. R., Beer, R., Bowman, K., Clough, S. A., Eldering, A., Fisher, B., Gunson, M. R., Herman, R. L., Kulawik, S. S., Lampel, M. C., Luo, M., Megretskaia, I. A., Osterman, G. B., and Shephard, M. W.: Comparisons of Tropospheric Emission Spectrometer (TES) ozone profiles to ozonesodes: methods and initial results, J. Geophys. Res., 112, https://doi.org/10.1029/2006JD007258, 2007.