An algorithm to retrieve peroxyacetyl nitrate from AIRS

Response to reviewer 1

Joshua L. Laughner, Susan S. Kulawik, and Vivienne H. Payne

October 7, 2025

We thank the reviewer for their comments. In particular, their question about whether a stricter threshold for the PC-based filter would filter out the spurious signal seen in the original Australian Fires case led us to identify a way to correctly filter out those soundings. (It was not as simple as increasing the threshold, but was straightforward and appears to be effective.)

Below we respond to the individual comments. The reviewer's comments will be shown in red, our response in blue, and changes made to the paper are shown in black block quotes. Unless otherwise indicated, page and line numbers correspond to the original paper. Figures, tables, or equations referenced as "Rn" are numbered within this response; if these are used in the changes to the paper, they will be replaced with the proper number in the final paper. Figures, tables, and equations numbered normally in our responses refer to the numbers in the revised paper.

Major issues

However, while the technical implementation appears robust, I find the scientific contribution limited in its current form. The novelty primarily lies in extending the CrIS-based PAN retrieval strategy to AIRS, but the manuscript stops short of fully exploiting this opportunity. In particular, the analysis is largely confined to comparisons with CrIS over a few case studies. The broader potential of the AIRS PAN dataset to provide scientific insight remains underexplored. The scientific impact would be significantly enhanced if the authors included additional analyses, such a preliminary global climatology or seasonal cycle of AIRS PAN, assessments of long-term or interannual variability, or regional investigations beyond biomass burning plumes.

It is computationally expensive to generate a long timeseries of data with this retrieval. Therefore, it is only practical to carry out these sorts of analyses after the retrieval has been incorporated into an operational processing environment with sufficient computational resources to process larger batches of data. This analysis was restricted to that which was practical to process in our scientific computing facility. We have added a pair of sentences to the end of the introduction to explain this:

> "Due to the computational cost of this retrieval, our analysis focuses on a few days with significant variation in PAN from major fires in the US and Australia. This product will be incorporated in the operational TRopospheric Ozone and its Precurors from Earth System Sounding (TROPESS) data processing in the future (`https://disc.gsfc.nasa.gov/information/mission-project?keywords=tropess&title=TROPESS`, last accessed 11 Sep 2025), which will enable analysis on a longer timeseries of data."

The authors note that low, warm clouds over oceans can be misinterpreted as PAN. Yet, similar clouds exist over land (e.g., tropical forests). Could the authors clarify why this misinterpretation would be less problematic over land?

We have added a new paragraph with our hypotheses to the end of Sect. 3.2:

> "Our hypothesis is that the reason the AIRS retrieval is affected by the low, warm clouds and CrIS is not is due either the difference in spectral windows used between the retrievals (Fig. 1), the difference in radiance noise between the instruments, or a combination of the two. Further, our hypothesis for why land soundings are much less impacted than ocean soundings is that it is more difficult to distinguish a low, warm cloud from an underlying ocean surface than a land surface."

Why are these low, warm clouds an issue for AIRS but apparently not for CrIS? Are CrIS retrievals performed "above clouds"? If so, wouldn't that introduce a bias in retrieved PAN due to lack of surface contribution? More clarification on this aspect is needed.

CrIS retrievals are not performed specifically "above clouds," though, of course the effect of a cloud must be captured in the radiative transfer. The only difference between the CrIS and AIRS retrievals are the selection of spectral windows for the retrieval and minor details of the strategy table. Our hypothesis is included in the new paragraph quoted in response to the previous comment: this is likely due to some combination of the specific spectral windows used in the retrieval and the different noise characteristics of the instruments.

In Figure 3, some PAN features seen by AIRS (e.g., near 45°N, 145°W and 50°N, 130°W) are absent in CrIS. Are these retrievals cloud-contaminated? A short discussion of these discrepancies would improve interpretation.

We have inserted a few additional sentences discussing those discrepancies:

> "Similarly, in the plume around 50°N, AIRS sees enhanced X_{PAN} further west than CrIS (around 150° W) and more to the northwest of the state of Washington (near 50°N, 125°W). From the cloud properties shown in Fig. 4, these are also potential cases of erroneous impact from clouds."

The paper would benefit from a clearer and more detailed description of how the decision tree was designed, trained, and applied. Specifically, is the nearest CrIS PAN value used only during the training phase, or is it required systematically for each AIRS retrieval at the application stage?

We clarified that the CrIS X_{PAN} is only used in training in the discussion around Table 5:

> "Note that the CrIS X_{PAN} is not an input to the decision tree; it is used only in training. This permits the decision tree to be applied to AIRS soundings without a coincidence CrIS sounding."

If the quality filtering process requires CrIS data on a systematic basis (i.e., for each AIRS retrieval), then the utility of the AIRS PAN retrievals becomes restricted to the CrIS era (i.e., post-2012). This undermines one of the main potential advantages of using AIRS — the opportunity to generate a long-term PAN time series starting from 2002.

As described in the response to the previous comment, the filtering process does not rely on CrIS in this way and thus can be applied to AIRS data before CrIS was launched.

By tailoring the AIRS quality filtering strictly based on CrIS, there is a risk of overly aligning the two datasets. This may introduce biases or lead to the rejection of potentially valid AIRS PAN retrievals in cases where CrIS retrievals are biased, noisy, or simply absent. For instance, even over land, the quality filtering seems to restrict useful retrievals close to strong emission sources or fires.

In principle, we agree that it would be preferable to validate the AIRS product against an independent dataset, as Payne et al. (2022) did with the ATom flight campaign for the CrIS PAN product. However, the majority of the ATom profiles were flown over ocean. Given our need to filter out soundings impacted by the low clouds discussed above and the greater sounding-to-sounding noise seen in the AIRS product, this made both ATom and a similar campaign (HIPPO) difficult to use for our study. Further, other in situ profiles of PAN over land are sparse enough that, given the large sounding-to-sounding noise in our product, individual comparisons with the limited number of over-land profiles would not be meaningful. These issues are already described at the start of Sect. 3.3. However, since CrIS has been validated against in situ data, we believe that tying to CrIS is the best approach to mitigate biases.

The current implementation applies the AIRS AVKs to the CrIS retrievals to enable direct comparison. However, in my understanding of Rodgers, applying the AVKs from one instrument to retrievals from another is generally appropriate only when the second instrument has significantly higher vertical resolution and information content. In that case, it can reasonably serve as a "truth" profile.

While AKs are indeed commonly used to compare remotely sensed data against higher resolution in situ profiles, they can also be used to address different vertical sensitivities between two remote sensing instruments. This exact case is covered in Sect. 4.3 of (Rodgers and Connor, 2003). In our case, since both the AIRS and CrIS retrievals use the same a priori profile, using this profile as the "comparison ensemble," $\mathbf{x}_c$, from Rodgers and Connor (2003) is appropriate, as both retrievals will be optimal with respect to that profile.

However, both AIRS and CrIS PAN retrievals have limited vertical sensitivity, with DOFS that would typically be well below 1, indicating no vertical information.

The averaging kernels shown incorporate the pressure weighting function. Since each level is weighted by its contribution to the total column, that reduces the overall AKs. We have added additional panels to Fig. 15 that show the sum of the rows of the profile AK matrix, which are not weighted by the column operator, and so give a sense of the total sensitivity of each level of the retrieval to all levels of the true profile. We have also added a new figure (Fig. 16) that compares the overall DOFs between AIRS and CrIS. While the AIRS DOFs are lower than those of CrIS, they indicate there is sensitivity to column PAN, especially in scenes with large enhancements.

Fig. 12 shows that both AIRS and CrIS exhibit heterogeneous and situation-dependent vertical sensitivities (their AVKs diverge markedly when surface temperature decreases). Given this, the assumption that AIRS AVKs alone can transform CrIS data into something comparable is questionable. Ideally, a symmetric or "two-way" treatment accounting for both sets of AVKs would be required for this inter-comparison (yet this is practically challenging and still not guaranteed to yield equivalence in a formal sense).

Although the CrIS AKs do not show up explicitly in our Eq. (2), Eq. (26) in Sect. 4.3 of Rodgers and Connor (2003) shows that both instruments' AKs are present when we take the difference of the AIRS columns with respect to the CrIS columns adjusted with our Eq. (2), as the CrIS AKs are implicitly in $\hat{\mathbf{x}}_{\mathrm{CrIS}}$ (i.e., $\hat{\mathbf{x}}_2$ in Rodgers and Connor (2003)).

I find it unfortunate that the discussion and analysis of the AIRS PAN product is currently limited to land. Such limitation significantly reduces its utility in key applications, such as tracing fire plumes, where a large fraction of the signal occurs over oceans. In the case of the Australian bushfires, for example, nearly the entire plume over the ocean is lost.

Fortunately, when we were double checking on the effect of the filter for one of your other comments, we found a way to improve the PC-based filter and allow ocean data through, and then update the decision tree filter to also handle ocean soundings. We have updated our conclusions to reflect this, including that we now believe users can use the ocean data for large PAN plumes, with the potential to apply custom filtering when needed.

Lines 132-142: This section is difficult to follow without prior knowledge of the MUSES algorithm. I recommend expanding the explanation with more technical details to make it more self-contained and accessible to readers unfamiliar with previous TROPESS-related publications.

We have expanded the paragraph discussing the difference between the initial state and a priori constraint:

> "An important distinction within MUSES is the difference between the a priori (or constraint) state vector and the initial state vector. The former is $\mathbf{x}_a$ in Eq. (1) and is a mathematical constraint on the optimal state vector, the latter is the starting point of $\mathbf{x}$ before the first iteration of the Levenberg-Marquardt solver. **This distinction is important within MUSES because it is a multi-step retrieval. The strategy table, mentioned above, defines which elements of the state vector will be retrieved in each step and whether or not the retrieved state for step i becomes the initial state for step $i+1$. For example, the retrieval may begin with an H_2O profile taken from**

a meteorological reanalysis as both the initial guess and the a priori constraint. An early step in the retrieval can then retrieve a new H_2O profile which is more consistent with the observed radiances. This new H_2O profile can then be used as an initial state for later steps (whether or not those steps retrieve H_2O). This can be important for weak absorbers, such as PAN, which need the profiles of strong thermal IR absorbers to be accurate for the scene in question so that the relatively small absorption feature of the weak absorber can be identified. We note that, for a given step, the initial state and a priori constraint can be the same but do not need to be. For later steps of the retrieval, the initial state will have been set by earlier retrieval steps **(as in the example given with H_2O) but the a priori constraint will remain the same for all steps. Or, the a priori constraint may be chosen to be a relatively simple profile to avoid imposing undue assumptions, while the initial state may be chosen to reflect a better estimate of the atmospheric state in that location to attempt to minimize the number of steps needed by the solver.**"

For OSS, we have added a sentence summarizing the benefits and drawbacks of OSS:

"This allows OSS to efficiently simulate the radiances a specific instrument would observe by reducing the number of monochromatic wavelengths that must be modeled for a given instrument channel, but means that OSS must be trained for each instrument used in a retrieval separately."

However, we do recommend that readers interested in the details of OSS read the Moncet papers rather than relying on our description.

Section 3.4: Although I understand that deriving uncertainty estimates for retrieved quantities from satellite measurements is challenging, I remain unconvinced by the authors' approach. The reported uncertainty value (0.5 ppb) is based solely on the difference in NESR between AIRS and CrIS. However, the uncertainty should realistically vary significantly with factors such as thermal contrast, cloud coverage, PAN abundance, and others.

Our conclusion that 0.5 ppb was a reasonable estimate of the uncertainty was not solely due to the NESR analysis. As we said in the first paragraph of Sect. 3.4, it is consistent with the analysis in the previous section that showed the spread in AIRS-CrIS correlation was approximately 0.5 ppb. To reinforce this point, we have moved the original Fig. 15 forward to be Fig. 14, showing that a similar 0.5 ppb spread is seen when individual AIRS and CrIS soundings are correlated. We have also added an acknowledgement that the individual sounding errors will vary, but that we believe the 0.5 ppb estimate to be a reasonable average for the AIRS PAN product:

"While we expect the error of individual soundings to vary depending on the specific atmospheric and surface conditions for each sounding, we believe 0.5 ppb to be a reasonable estimate of the typical uncertainty in the AIRS X_{PAN} data."

Section 3.4: I find the discussion on vertical sensitivity rather brief. For example, what is the typical DOFS of the AIRS PAN retrievals in fire plume regions versus remote areas? How do these values compare to those from CrIS?

We have added a new figure (Fig. 16) comparing the DOFS of the two retrievals and a new paragraph comparing the DOFS values:

> "Figure 16 shows the overall degrees of freedom (DOF) of signal for both the AIRS and CrIS products in the 2020-09-11 US West Coast Fire scene. From Fig. 16 panels a and b, we can see that the DOFs for the CrIS PAN product are grouped around 1, indicating that there is essentially always enough information to retrieval a single piece of vertical information in the form of a column average. In contrast, Fig. 16 panels c and d show that the AIRS DOFs are lower (centered around ~0.5) with a wider distribution. Greater AIRS X_{PAN} values do tend to be associated with greater DOFs. This implies that the AIRS product will retain influence from the prior, particularly in background conditions, but can detect sufficiently large PAN enhancements."

Minor comments

Lines 27-31: Do the authors have an estimate of what fraction of the total APNs signal in the retrievals corresponds specifically to PAN? Given its longer lifetime relative to other APNs, could one expect its share to increase in aged plumes or background air.

We have added a number of references pointing to PAN as typically comprising 75% to 90% of APNs, with a caveat about wildfire plumes:

> "However, PAN typically comprises the majority (75% to 90%) of APNs in both remote areas (Roberts et al., 1998, 2002; Wolfe et al., 2007; Fischer et al., 2014) and urban plumes (LaFranchi et al., 2009). The fraction may be lower in wildfire plumes; Peng et al. (2021) hypothesize that an unknown APN could explain discrepancies in NO_x/CO ratios between their observations and model."

Section 2.1 would benefit from more technical information about the AIRS instrument, especially in relation to its suitability for PAN retrieval (spectral resolution, radiometric noise characteristics (especially compared to CrIS), spatial sampling and footprint size).

We have added this information as Table 1.

Lines 150-153: The manuscript mentions a "global survey" sampling approach with TROPESS products. It would be important to clarify what proportion of soundings are included in the final products. For instance, is it 1 out of 2 soundings, 1 out of 3, etc.? This has implications for data representativity.

We added a sentence describing the thinning strategy:

> "The default survey strategy processes one sounding in each $x^\circ \times x^\circ$ box over land and one out of every four such boxes over ocean. For the current products, x is either 0.7° or 0.8°."

CCl4 is not mentioned in the strategy table (Table 2), yet it has notable absorption features in the thermal infrared that could affect PAN retrievals, especially the spectral baseline. Is CCl4 explicitly fitted in the retrieval process? If not, how is its temporal variability accounted for?

We have added an explaination of how CCl_4 is accounted for:

> "CCl_4 is not retrieved (Table 3) but is simulated in the radiative transfer as an interferent, using climatological profiles scaled by yearly scale factors derived from ground based observations. The base climatological profiles vary with latitude and longitude in 30° and 60° bins, respectively, and were developed from MOZART model output (Brasseur et al., 1998)."

Lines 183-186: Could the use of different a priori profiles across regions introduce discontinuities in the retrieved PAN abundances at regional boundaries?

In principle, yes. But this is not unique to this product; many remote sensing products use priors that vary in space and can introduce such discontinuities. Correct application of the averaging kernels and a priori constraints when comparing to other datasets will account for the influence of the prior.

Lines 211-213: Are the threshold criteria used for AIRS the same as for CrIS? If so, is this appropriate given the different instrument characteristics (spectral resolution, sensitivity, etc.)?

The criteria are not the same as for CrIS. In any case, these criteria are not what determine the filtering of the actual product; that is the machine learning model. We have clarified this:

> "The AIRS data shown in Fig. 3 are those soundings which pass **prototype** quality flags **chosen based on quality flags for other thermal retrievals**, including sufficiently small radiance residual, surface temperature > 265 K, cloud top pressure (as retrieved in our algorithm) below the tropopause, and the quality of the H_2O retrieval in step 4 of Table 3. **(Note that these quality flags were for prototyping purposes only, and are not those used in the final product.)**"

Line 232 ("the filtering approach failed..."): Could stricter filtering criteria resolve this issue?

Yes. In fact, when we checked the PC values to respond to this question, we realized that using PCs derived from a ~ 100 cm^{-1} window around the PAN feature were more reliable than using just radiances in the PAN retrieval windows, and that increasing the threshold from -10 to 0 for PC2 would address the clouds in the Australian case. We have adjusted our conclusions to reflect this, though we do still advise caution as the variation in the effective threshold between these two test cases suggests there may be more variation in the appropriate threshold across time and space.

Lines 302-304: These statements could benefit from clarification in the case of the Australian Bush Fires. Much of the plume appears to be missing over the ocean, and the soundings over land seem relatively noisy. The observation that AIRS shows no PAN enhancement, similarly to CrIS, should be interpreted with caution. The agreement between the two instruments in this case does not necessarily validate the accuracy of the AIRS retrievals, especially in light of the limited data coverage.

With the improved PC-based filter, we are now able to show in the new Fig. 10 that AIRS does see a PAN plume heading towards New Zealand.

In Fig. 8 (and similar figures), it is difficult to assess the differences in spatial sampling and resolution between the CrIS and AIRS soundings. Including a zoomed-in view might help better illustrate these differences.

The spatial sampling of the two instruments is not the point of these figures. We have added Table 1 that shows how the two instruments compare in that regard.

Lines 321–326: Could you clarify whether the intention is to recommend that the AIRS PAN product be used primarily as $10^{\circ} \times 10^{\circ}$ spatial averages? If so, this seems quite restrictive.

We have added a cross reference to Sect. 4 where we discuss this in more detail. In brief, for background concentrations of PAN, we recommend averaging approximately 140 soundings to reduce the random uncertainty sufficiently, but for strong plumes, individual soundings can be used.

Lines 327–330: Would it be feasible to implement a similar H_2O bias correction for the AIRS PAN retrievals as is done for CrIS?

The significant random uncertainty in the individual AIRS soundings makes doing so difficult. We have added text acknowledging this:

> "Payne et al. (2022) were able to derive the CrIS bias correction through comparison between CrIS and in situ background X_{PAN} values. The need to average a significant number of AIRS soundings to reduce the random sounding-to-sounding noise makes it difficult to identify any relationship between AIRS X_{PAN} values and H_2O column amounts."

Technical comments

All technical comments have been addressed, thank you for catching the typos.

References

Brasseur, G. P., Hauglustaine, D. A., Walters, S., Rasch, P. J., Müller, J., Granier, C., and Tie, X. X.: MOZART, a global chemical transport model for ozone and related chemical tracers: 1. Model description, Journal of Geophysical Research: Atmospheres, 103, 28 265–28 289, https://doi.org/10.1029/98jd02397, 1998.

Fischer, E. V., Jacob, D. J., Yantosca, R. M., Sulprizio, M. P., Millet, D. B., Mao, J., Paulot, F., Singh, H. B., Roiger, A., Ries, L., Talbot, R., Dzepina, K., and Pandey Deolal, S.: Atmospheric peroxyacetyl nitrate (PAN): a global budget and source attribution, Atmospheric Chemistry and Physics, 14, 2679–2698, https://doi.org/10.5194/acp-14-2679-2014, 2014.

LaFranchi, B. W., Wolfe, G. M., Thornton, J. A., Harrold, S. A., Browne, E. C., Min, K. E., Wooldridge, P. J., Gilman, J. B., Kuster, W. C., Goldan, P. D., de Gouw, J. A., McKay, M., Goldstein, A. H., Ren, X., Mao, J., and Cohen, R. C.: Closing the peroxy acetyl nitrate budget: observations of acyl peroxy nitrates (PAN, PPN, and MPAN) during BEARPEX 2007, Atmospheric Chemistry and Physics, 9, 7623–7641, https://doi.org/10.5194/acp-9-7623-2009, 2009.

Payne, V. H., Kulawik, S. S., Fischer, E. V., Brewer, J. F., Huey, L. G., Miyazaki, K., Worden, J. R., Bowman, K. W., Hintsa, E. J., Moore, F., Elkins, J. W., and Juncosa Calahorrano, J.: Satellite measurements of peroxyacetyl nitrate from the Cross-Track Infrared Sounder: comparison with ATom aircraft measurements, Atmos. Meas. Tech., 15, 3497–3511, https://doi.org/10.5194/amt-15-3497-2022, 2022.

Peng, Q., Palm, B. B., Fredrickson, C. D., Lee, B. H., Hall, S. R., Ullmann, K., Campos, T., Weinheimer, A. J., Apel, E. C., Flocke, F., Permar, W., Hu, L., Garofalo, L. A., Pothier, M. A., Farmer, D. K., Ku, I.-T., Sullivan, A. P., Collett, J. L., Fischer, E., and Thornton, J. A.: Observations and Modeling of NOx Photochemistry and Fate in Fresh Wildfire Plumes, ACS Earth and Space Chemistry, 5, 2652–2667, https://doi.org/10.1021/acsearthspacechem.1c00086, 2021.

Roberts, J. M., Williams, J., Baumann, K., Buhr, M. P., Goldan, P. D., Holloway, J., Hübler, G., Kuster, W. C., McKeen, S. A., Ryerson, T. B., Trainer, M., Williams, E. J., Fehsenfeld, F. C., Bertman, S. B., Nouaime, G., Seaver, C., Grodzinsky, G., Rodgers, M., and Young, V. L.: Measurements of PAN, PPN, and MPAN made during the 1994 and 1995 Nashville Intensives of the Southern Oxidant Study: Implications for regional ozone production from biogenic hydrocarbons, Journal of Geophysical Research: Atmospheres, 103, 22 473–22 490, https://doi.org/10.1029/98jd01637, 1998.

Roberts, J. M., Flocke, F., Stroud, C. A., Hereid, D., Williams, E., Fehsenfeld, F., Brune, W., Martinez, M., and Harder, H.: Ground-based measurements of peroxycarboxylic nitric anhydrides (PANs) during the 1999 Southern Oxidants Study Nashville Intensive, Journal of Geophysical Research: Atmospheres, 107, https://doi.org/10.1029/2001jd000947, 2002.

Rodgers, C. D. and Connor, B. J.: Intercomparison of remote sounding instruments, J. Geophys. Res. Atmos., 108, https://doi.org/10.1029/2002jd002299, 2003.

Wolfe, G. M., Thornton, J. A., McNeill, V. F., Jaffe, D. A., Reidmiller, D., Chand, D., Smith, J., Swartzendruber, P., Flocke, F., and Zheng, W.: Influence of trans-Pacific pollution transport on acyl peroxy nitrate abundances and speciation at Mount Bachelor Observatory during INTEX-B, Atmospheric Chemistry and Physics, 7, 5309–5325, https://doi.org/10.5194/acp-7-5309-2007, 2007.

An algorithm to retrieve peroxyacetyl nitrate from AIRS

Response to reviewer 1

Joshua L. Laughner, Susan S. Kulawik, and Vivienne H. Payne

October 7, 2025

We thank the reviewer for identifying important points to clarify in our manuscript. We will also note up front that, during our efforts to address the first reviewer's comments, we found a way to improve the filter for clouds over oceans. The revised manuscript now shows that data over oceans correctly identifies plumes in e.g., the Australian Bush Fires.

Below we respond to the individual comments. The reviewer's comments will be shown in red, our response in blue, and changes made to the paper are shown in black block quotes. Unless otherwise indicated, page and line numbers correspond to the original paper. Figures, tables, or equations referenced as "Rn" are numbered within this response; if these are used in the changes to the paper, they will be replaced with the proper number in the final paper. Figures, tables, and equations numbered normally in our responses refer to the numbers in the revised paper.

Major issues

The title is misleading since the primary focus of the paper is not in the discussion of a novel algorithm for the retrieval of PAN from AIRS, but rather in how an existing algorithm can be adopted for a new set of instrument measurements. I strongly recommend adjusting the title to more accurately reflect the goal and content of the paper.

We respectfully disagree. While the machine learning filter is an important component of this product, it was not the sole component needed to produce this retrieval. Since the manuscript includes information on the other components (e.g., the sequence of retrieval steps and spectral windows chosen), we prefer to retain the current title.

There needs to be a sentence contrasting AIRS with CrIS, especially as far as instrument noise and spectral range goes, to help the reader understand the goal of this paper and why retrieving PAN from AIRS is more challenging than PAN from CrIS.

We have added a new Table 1 that compares the key characteristics of AIRS and CrIS.

Line 5: Here the authors state that they retrieve PAN from AIRS but omit all retrievals "from low, warm clouds over ocean", but this is misleading because in Section 3.2, Line 245, the authors conclude that the AIRS PAN product needs to exclude all retrievals over ocean since they struggle to isolate only those cases with interference from low, warm clouds. The abstract needs to correctly reflect their conclusions. Moreover, it will help the reader (and promote the validity of this work) if the authors state in the abstract that the CrIS PAN product does not need the same type of land/ocean filtering as the AIRS PAN product.

During our work to respond to the other reviewer's comments, we identified a way to improve the filtering over ocean to correctly remove the cloud-impacted soundings in both test cases. The abstract has been updated to reflect the improved results.

Line 5: "...we...develop a decision tree quality filter trained to predict whether a PAN value retrieved from AIRS..." The title should reflect this primary goal and outcome. Suggested new title: A quality filter for PAN retrievals from AIRS.

We prefer to retain the current title for the reasons given above.

Line 7: "We show that AIRS is capable of retrieving PAN plumes..." I've studied the figures and reread the paper, but remain unconvinced that the authors succeeded in demonstrating this. At best, the results show just how challenging it can be to design an algorithm for retrieving trace gases from two instruments as disparate as AIRS and CrIS.

This is better shown in the revised paper with more reliable filtering over ocean. In the new Fig. 10 for instance, you can clearly see a PAN plume approaching the northern island of New Zealand and PAN enhancements throughout the SW US which are in the same location as in the CrIS retrievals. Likewise, the new Fig. 7 shows that AIRS captures the PAN plume between Australia and New Zealand, similarly to CrIS.

The authors list many other PAN studies and products, but omit mentioning other successful AIRS+CrIS long-term products. This effort to retrieve a trace gas species from AIRS and CrIS is not the first of its kind. Others have successfully addressed instrument differences between AIRS and CrIS (especially with respect to interference from clouds) to generate consistent long-term records for a host of other trace gas species. Perhaps the authors can contrast their approach to other AIRS+CrIS records to help the reader better understand the authors' algorithm choices and subsequent challenges.

We apologize, we were focused on other PAN retrievals in the interest of brevity. We added paragraphs to the introduction referencing the CLIMCAPS algorithm and a few examples of TROPESS products apply a consistent retrieval to AIRS and CrIS data:

> "Consistent records of atmospheric trace gas concentrations are essential to monitor how air quality is changing over time. A major challenge in this respect is addressing instrument differences among satellites to produce records spanning multiple decades. The Community Long-term Infrared Microwave Combined Atmospheric Product System (CLIMCAPS) product (Smith and Barnet, 2020) invested significant effort in applying a consistent retrieval to radiances from both the Atmospheric Infrared Sounder (AIRS) and the various CrIS instruments as well as minimizing cross-correlations between retrieved variables (Smith and Barnet, 2019). CLIMCAPS produces records spanning the more than two decades since AIRS launched in 2002 that include profiles of atmospheric temperature, H_2O, CO, O_3, CO_2, HNO_3, and CH_4, but does not include PAN.
>
> The TRopospheric Ozone and its Precurors from Earth System Sounding (TROPESS)

project also focuses on applying a consistent retrieval algorithm for various trace gases to radiances from a variety of instruments. This includes thermal radiances observed by AIRS and CrIS, as well as radiances in other parts of the electromagnetic spectrum from the Ozone Monitoring Instrument (OMI) and, in the future, the TROPOspheric Monitoring Instrument (TROPOMI). Cady-Pereira et al. (2024) demonstrated the capability with TROPESS to retrieve NH_3 from both AIRS and CrIS. They validated NH_3 from both instruments against aircraft data and found that, although the retrievals from the two instruments are broadly similar, there are differences in the agreement with aircraft profiles. However, after accounting for the smoothing errors, the biases fall below 1 ppb. Pennington et al. (2025) evaluated O_3 trends in three TROPESS products using thermal radiances from AIRS and CrIS and combined thermal and ultraviolet radiances from AIRS and OMI. They compared these products to ozonesonde data, and found that trends in the bias of the retrieved O_3 was significantly less than the reported O_3 trends."

We also added text to Sect. 3.2 acknowledging that cloud clearing, as done by CLIMCAPS, would be one potential approach to mitigate the impact of ocean clouds on the retrieval, but that the MUSES algorithm is geared towards retrieving one sounding at a time:

"...we tested whether an EOF decomposition could identify the low, warm clouds causing the spurious PAN signal in our AIRS PAN retrieval. **We do note that a cloud-clearing approach, like that used in CLIMCAPS (Smith and Barnet, 2020), could be one approach to address this issue. Such an approach combines radiances from multiple soundings to yield radiances unimpacted by clouds. However, the MUSES algorithm is designed to operate on individual soundings. Therefore, we focused our efforts on the EOF decomposition as a way to screen out these cloud-affected soundings.**"

Line 100: "..the OE algorithm calculates uncertainty from noise only." As the authors well know, OE is a generalized retrieval framework, not a universal retrieval algorithm. The way that noise and uncertainty are quantified in practice vary significantly across the many OE products in operation today. I strongly encourage the authors to rephrase this statement (and similar ones throughout the manuscript) to clarify such characteristics as their own algorithm choices instead of attributing them to the OE framework in general. Again, it may be helpful for the authors to consult and mention other OE retrieval implementations that quantified noise, error and uncertainty in different ways that could help inform their results.

We have reworded this section to clarify that it is the MUSES algorithm's calculation we mean:

"This was larger than the uncertainty calculated by the **MUSES** optimal estimation (OE) algorithm, but Payne et al. (2022) attribute the discrepancy to pseudo-random error contributions from the retrieval of interfering species or the temperature profile. Such interferent-driven error was not included in the uncertainty calculated by the **MUSES** algorithm, as **for PAN retrievals, the algorithm** calculates uncertainty from noise only."

First paragraph of Section 2.4: The summary of the TROPESS product presented here is confusing. Many of the phrases reads more like jargonn than scientific explanations, e.g., what is a "global survey sampling approach"? And, can the authors clarify what they mean with a "forward" and "reanalysis" stream? Why not process the full record (2002 to present) with "the latest version of the MUSES algorithm"? If two different MUSES algorithms are used to process the full record (2002 to 2021 versus 2002 to present), could the resulting PAN product really be considered a consistent record?

We have expanded these paragraphs to better explain the terms used, and clarify that these two streams are not intended to be used together as a consistent record. (The retrospective stream is meant to be that consistent record; the forward stream is more geared towards analyses of episodic events.)

> "The TROPESS project focuses on applying the MUSES algorithm to retrieve a range of atmospheric trace gases from a variety of space-based instruments, including AIRS, OMI, CrIS, and TROPOMI to date. **Operational processing for TROPESS is set up to accommodate two distinct goals. The first is to provide a global record of ozone and related trace gases for the first ∼20 years of the 21st century. The second is to support rapid iteration on and improvement of the underlying level 2 algorithms for application to more recent data. Due to the computational cost of these retrievals, meeting both goals requires two separate data streams.**"
>
> "The first is a "retrospective" or "reanalysis" stream that retrieves trace gas amounts from ∼ 2002 through ∼ 2021. **This stream is processed with a version of the MUSES algorithm frozen at the time the retrospective processing began.** The second is a "forward" stream that processes new radiances as they become available with the latest version of the MUSES algorithm, **including updates to the algorithm made after the retrospective processing began.** The forward stream serves the dual purpose of monitoring significant events affecting air quality and serving as a test bed for improvements to the MUSES algorithm. Due to the difference in the algorithm versions, users must take care not to misinterpret changes in trends between the two streams."
>
> "Both streams use a "global survey" sampling approach to process a subset of all available soundings yet provide global coverage, which allows a balance between computational cost and spatial coverage. **The default survey strategy processes one sounding in each $x^\circ \times x^\circ$ box over land and one out of every four such boxes over ocean. For the current products, x is either 0.7° or 0.8°.** In addition, TROPESS produces special collections with full data density for high interest events (e.g., the 2019–2020 Australian Bush Fires and 2020 US West Coast Fires) and a set of megacities around the world."

Does the TROPESS MUSES PAN product from CrIS cover the full global range of CrIS measurements on a twice daily basis? This is not clear in the text.

Yes, it does. We have clarified this as follows:

> "...is now routinely produced as part of both the reanalysis (Bowman, 2023) and forward (Bowman, 2022) TROPESS streams, **as well as special products. The reanalysis and forward streams provide twice daily (day and night) global coverage, using the global survey strategy described in the previous paragraph.**"

Line 212: How did the authors decide on a surface temperature threshold of 265 K?

This is a carryover from the TES retrieval, which originally intended to avoid frozen surfaces, and was set somewhat arbitrarily below the typical freezing point of water to avoid issues with depressed freezing points. Since this is not used in the final product, we prefer not to confuse the issue by describing this heritage in the paper. Instead, we clarify that this threshold was only use for preliminary investigations:

> "...and the quality of the H_2O retrieval in step 4 of Table 3. **(Note that these quality flags were for prototyping purposes only, and are not those used in the final product.)**"

Lines 274–275: "different vertical sensitivity between CrIS and AIRS." What exactly is the difference? There are many published texts contrasting and quantifying the main instrument differences between AIRS and CrIS. I strongly recommend that the authors add the appropriate citations as well as summarize a few of them in this manuscript, specifically with respect to instrument noise, spectral coverage and resolution.

We have added a cross reference to Fig. 15 to point the reader to an example of how the vertical sensitivity differs between the two instruments in our retrieval. To the latter point, as stated previously we added a new Table 1 that summarizes key instrument characteristics with appropriate citations.

On page 15, the authors conclude that it is best to exclude AIRS PAN retrievals over ocean and deserts from the final product, but I wonder if this is sufficient given the results they present. How do the authors know that their PAN retrievals over land-based low, warm clouds are more accurate than over ocean-based low, warm clouds?

We have added a new Fig. 8 that shows, for our Amazon case, there are similarly low, warm clouds over the Amazon on that day, and we do not see the same clear correlation between the presence of such clouds and erroneously enhanced X_{PAN}.

Lines 313–315: The authors communicate that elevated PAN values are present in both the AIRS and CrIS products presented in Figure 8, but I fail to see this. The AIRS PAN product has a significant speckle effect (random distribution of high and low values) that is mostly absent in the CrIS PAN product. The CrIS PAN product indicates an elevated plume over the region centered on 10S, in contrast to much lower values throughout the rest of the mapped region. The AIRS PAN product, on the other hand, has a speckled distribution of PAN throughout the southern African region without any obvious featured plumes. As this work is currently presented, the conclusion is not supported by the results. I suggest the authors either rethink (and rephrase) their conclusion, or present results in support of their current statements. I have the same concerns for results communicated in Figure 9.

We have qualified this specific comparison:

> "The Amazon hotspot in western Brazil cannot be seen in AIRS due to the swath gap. The PAN hotspot seen by CrIS in the African test over Angola, Zambia, and the Democratic Republic of the Congo **is not as apparent in the AIRS PAN; however, AIRS does appear to capture some enhancement in that area, particularly compared to further north, near the equator.**"

With the improvements made to the ocean filtering since the discussion paper, we also draw attention to the agreement in spatial distribution of enhanced PAN in the other two test cases:

> "In both cases, we can see that the AIRS PAN product matches the location of enhanced PAN plumes seen in the CrIS data very well. **In the US West Coast Fires case, the large X_{PAN} values in Arizona, central/southern California, and northwestern Mexico are all in the same region where CrIS sees high X_{PAN} values. Likewise, in the Australian fires case, AIRS captures the PAN plume approaching New Zealand's northern island, though compared to CrIS, more of the plume is removed by our filtering criteria.**"

Lines 315–320: While I appreciate the authors' attempt to communicate the practical interpretation of their product in downstream applications, I feel this section is a bit muddled. Does the co-located CO product need to be from the same TROPESS MUSES suite, or can an independent CO product serve to confirm elevated PAN retrievals?

The CO product does not need to be the TROPESS product. We have added a sentence to Sect. 4 (as that section covers general recommendations for use), and added a cross references to the lines identified by the reviewer.

> "These two criteria should help users filter out false positive high X_{PAN} values. **When using other species of interest, users need not restrict themselves to TROPESS products—any good-quality dataset will be useful in this regard.** We also encourage..."

Line 344: Why would AIRS maximum sensitivity decrease more quickly as surface temperature decreases?

Our hypothesis is that this is due to the greater noise in AIRS, which we now state:

> "We suspect this is due to the greater noise present in the AIRS radiances, with AIRS sensitivity decreasing more with reduced thermal contrast due to the greater noise. However, we have not confirmed this hypothesis."

Section 4: "AIRS and CrIS is ¡ 0.1 ppb when averaged to a 10 x 10 box", which suggests only spatial aggregation. Yet later in the paragraph the authors suggest that users choose to average 250 PAN retrievals over an unspecified "spatiotemporal window". This is confusing (even misleading) as the authors do not present or discuss whether the AIRS PAN product quantifies small change over time. It appears the authors simply assume that averaging over time (days? Weeks?) will yield the same results as averaging over space.

Yes, we are assuming that averaging over time will produce the same result as averaging over space. We do not envision a scenario, other than a major wildfire or other extreme event, where this would not be true. We have clarified that this is an assumption, and one that can be tested after the algorithm is applied to more data. (Note that, with the filtering improvements since the discussion, we have lowered the minimum number to 140 soundings.)

> "In principle, it should not matter whether the 140 soundings are accumulated by averaging in time or space, as **we assume** the AIRS-CrIS X_{PAN} differences **are similarly uncorrelated in time as in space. We expect this assumption to hold true as long as episodic events that significantly perturb PAN concentrations (such as wildfires) are not included in the time period averaged. We will test this assumption in the future as more data becomes available.**"

Line 377: "We recommend averaging 250 AIRS soundings which will result in a 0.1 ppb error." Why is this type of averaging not recommended for CrIS PAN retrievals? I.e., why does the CrIS PAN product not display the same speckled pattern? Also, on Line 352 the authors state that the 0.1 ppb value should not be interpreted as an overall error, yet here they state it as an overall error. Please clarify.

The CrIS PAN product benefits from the lower noise in the CrIS radiances, which makes retrieval of a weakly absorbing species such as PAN significantly easier. To the second point, we clarified that the 0.1 ppb error is relative to the CrIS product, not an overall error:

> "...∼0.1 ppb error **relative to the existing CrIS PAN product.**"

Minor issues

Line 95: ". . . GEOS-Chem profiles appended to the top." This is not sufficiently descriptive. What do the authors mean by "append" and by "top"?

Line 96: "aircraft free tropospheric PAN column averages" What does this mean?

For both of these comments, we have added a sentence directing the reader to Payne et al. (2022) for details. As we cannot use the same method for AIRS, we prefer not to go into detail in this manuscript.

Figure 8: What does the box over the southwestern region represent?

This is the area with a silicate surface feature that biases our PAN retrievals. We have clarified that in the caption.

Lines 282–289: This discussion is confusing. E.g., “However, we found that either caused the filter to screen out soundings with enhanced Xpan.”, “...to account for these someone uncommon cases”, etc.

We have attempted to clarify this section and included a reference to methods of pruning decision trees for further reading:

> “Typically, it is important to “prune” decision trees **(Esposito et al., 1997)** by limiting the number of decision nodes it can include in order to prevent overfitting to the training data. We tested pruning by limiting both the maximum depth (i.e., the number of nodes along any one path) and maximum number of leaf nodes (i.e., the number of end points for the model). However, we found that **either method of pruning the decision tree** caused the filter to screen out soundings with enhanced X_{PAN}. Our hypothesis is that, because these soundings are still in the minority of all soundings in the training data, limiting the decision tree's size gave it too little flexibility to account for these somewhat uncommon cases. **That is, because soundings with enhanced X_{PAN} are in the minority, a model limited in size lacked the flexibility to develop useful rules for these soundings, and instead was able to achieve better accuracy by simply classifying all such soundings as bad quality.** Therefore, we proceed without limiting the model size.”

Line 332: “The CrIS radiance noise is lower than the AIRS radiance noise.” Can the authors quantify this difference and provide references to text that demonstrate it?

Figure 13 (previously Fig. 11) referenced in the next sentence quantifies the difference. As mentioned previously, we also added a new Table 1 that summarizes characteristics such as these with references.

References

Bowman, K.: TROPESS CrIS-JPSS1 L2 Peroxyacetyl for Forward Processing, Summary Product V1, https://doi.org/10.5067/6HTQB4F81S08, 2022.

Bowman, K.: TROPESS CrIS-SNPP L2 Peroxyacetyl Nitrate for Reanalysis Stream, Summary Product V1, https://doi.org/10.5067/VU8MI4QOA4NI, 2023.

Cady-Pereira, K. E., Guo, X., Wang, R., Leytem, A. B., Calkins, C., Berry, E., Sun, K., Müller, M., Wisthaler, A., Payne, V. H., Shephard, M. W., Zondlo, M. A., and Kantchev, V.: Validation of MUSES NH3 observations from AIRS and CrIS against aircraft measurements from DISCOVER-AQ and a surface network in the Magic Valley, Atmospheric Measurement Techniques, 17, 15–36, https://doi.org/10.5194/amt-17-15-2024, 2024.

Esposito, F., Malerba, D., Semeraro, G., and Kay, J.: A comparative analysis of methods for pruning decision trees, IEEE Transactions on Pattern Analysis and Machine Intelligence, 19, 476–493, https://doi.org/10.1109/34.589207, 1997.

Payne, V. H., Kulawik, S. S., Fischer, E. V., Brewer, J. F., Huey, L. G., Miyazaki, K., Worden, J. R., Bowman, K. W., Hintsa, E. J., Moore, F., Elkins, J. W., and Juncosa Calahorrano, J.: Satellite measurements of peroxyacetyl nitrate from the Cross-Track Infrared Sounder: comparison with ATom aircraft measurements, Atmos. Meas. Tech., 15, 3497–3511, https://doi.org/10.5194/amt-15-3497-2022, 2022.

Pennington, E. A., Osterman, G. B., Payne, V. H., Miyazaki, K., Bowman, K. W., and Neu, J. L.: Quantifying biases in TROPESS AIRS, CrIS, and joint AIRS+OMI tropospheric ozone products using ozonesondes, Atmospheric Chemistry and Physics, 25, 8533–8552, https://doi.org/10.5194/acp-25-8533-2025, 2025.

Smith, N. and Barnet, C. D.: Uncertainty Characterization and Propagation in the Community Long-Term Infrared Microwave Combined Atmospheric Product System (CLIMCAPS), Remote Sensing, 11, https://doi.org/10.3390/rs11101227, 2019.

Smith, N. and Barnet, C. D.: CLIMCAPS observing capability for temperature, moisture, and trace gases from AIRS/AMSU and CrIS/ATMS, Atmospheric Measurement Techniques, 13, 4437–4459, https://doi.org/10.5194/amt-13-4437-2020, 2020.