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 $\sim$ 2002 through $\sim$ 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.