We would like to thank the reviewers for their well-considered and constructive comments and questions, which have definitely helped us to improve this paper. We have made extensive updates to the paper as a consequence of the comments. Our responses to reviewer comments are in italicized text below.

Reviewer 2

GENERAL COMMENTS

==========

The paper describes a new level 2 data set for the CrIS instrument. The retrieval process is briefly described and an in depth comparison to in situ data gathered by the ATom campaign(s) is given. Due to the number of CrIS instruments in orbit and in planning, this is an important data product. The paper identifies a strong bias of the provided data, which is larger than the assumed uncertainty. A H2O-VMR-based bias correction is suggested in the User Data documentation, but not discussed in the paper itself. The paper should address the bias more explicitly and discuss causes and corrections. Ideally, the root causes for the bias should be identified and the data product improved.

I recommend publication after revising the paper to discuss these points in detail and answering the other comments below.

We thank the reviewer for the detailed and insightful review and for the positive statement about the importance of this data product.

We appreciate the reviewer taking the time to check the user guide against the information presented in this manuscript. We agree that the paper ought to reflect the information in the user guide.

Updates to the user guide are underway. In the time between when the first version of the user guide was made available and when this paper was submitted, we did some more thinking about problems associated with water vapor interference in the CrIS PAN retrievals. As we have discussed in the paper, water vapor is a strong interferent in the spectral region used for the CrIS PAN retrievals, and is retrieved separately in a step before the PAN retrieval step. However, the master quality flag in the PAN products that are in the forward stream CrIS PAN dataset (cite doi, access date) does not include a check on the quality of the water vapor retrieval from that previous step. We find that a large number of cases in the Tropics with "bad" quality for the water vapor step were associated with strongly negative PAN retrievals. Those strongly negative PAN retrievals in the Tropics that were having a large impact on the H2O-VMR-based bias correction described in that initial version of the user guide. If we screen out cases where the water vapor retrieval step fails quality control, then the H2O-VMR-based bias correction (that had been based on "bad" cases) is less severe. In the version of the dataset that is now currently available, we recommend that the user use the H2O product information for screening the PAN. In future algorithm updates, it will make sense to directly account for the quality flag from the H2O step in the quality control for the PAN step. In the revised version of the

manuscript, we now include information on an updated formula for bias correction and we are working to get a revised version of the user guide posted with the TROPESS PAN products.

MAJOR COMMENTS

==========

line 132

The paper identifies a bias of -100 pptv in the derived data, which is larger than the supplied uncertainty in the data (80 pptv) derived from the standard deviation computed from differences to in situ measurements. This suggests that the bias is real and significant, particularly for non-polluted airmasses outside of plumes. The employed spectral region is full of emission signatures of a wide range of trace gasses. It seems as, e.g., CCl4 could still have an effect, but also other CFCs, or ClONO2 emit in this region. While the strong H2O emission line at 785 has been avoided, weaker lines are certainly present in the left window. The User Guide for the data even provides a bias correction formula depending on water vapour.
I question the usefulness of the data set in the current state.

1) Why was the obvious and *astonishingly* stable bias not corrected in the data set?

We have now extensively updated the manuscript to include discussion of this water vapor dependent bias as well as discussion of additional possible sources of systematic errors in the PAN products. Please see the response to reviewer 1 for discussion of the water vapor dependent bias. We have now included discussion of the bias correction formula in the paper.

The water vapor-dependent bias discussed above is by far the dominant source of purely systematic bias. The impact of uncertainties in specification of CCl4 is small (~0.01 ppbv), and the optical depth contributions of the CFCs and CLONO2 are smaller than CCl4. Note that we have now also included discussion of the propagation of estimated retrieval errors in temperature, water vapor and ozone (which are retrieved in previous steps) into the PAN retrieval. These can be regarded as "pseudo-random" contributions to the observation error, and we find that including these terms can at least partially account for the discrepancy between the observation error from instrument noise and the standard deviation of the differences between satellite and aircraft values (theoretical vs empirical).

2) Why was the retrieval not improved to the point, where no bias correction is necessary?

We strongly agree with the reviewer that the ideal scenario would be to address the root causes of any biases in remotely-sensed products and eliminate them completely. However, this idealized goal may or may not be attainable. While further improvements to future versions of the product are desirable, the current situation is that version of this product is available to the public and the CrIS PAN data are already being utilized for science studies. This paper describes the product and documents the observed bias.

3) Why was the bias correction formula of the User guide not mentioned or applied for the comparison?

We agree with the reviewer that the bias correction referred to in the User Guide should be discussed and applied here for the comparison. This has now been addressed in Section 4, in what are now Figures 9 and 8 and in the abstract.

Figure 2

High PAN VMRs occur often at higher tropospheric altitudes (particularly due to the longer lifetime at colder temperatures) close to the tropopause. The used aircraft data rarely go above 12km. Biomass burning plumes reach higher than 12km, particularly in the tropics. The given altitude range of 800hPa to 300hPa is key here, as 300hPa corresponds roughly to 10km.

How does this limited altitude range affect the accuracy of estimating total PAN in the UTLS?

If we understand correctly, the reviewer is asking about the component of the error budget associated with what we have assumed for the profile above the top of the aircraft measurements. This is a good point, and something that was not discussed in the initial manuscript version. As stated in the initial version, GEOS-Chem model output for runs specific to the time period was appended above the uppermost and below the lowermost altitudes spanned by the ATom PAN profiles. Due to the vertical sensitivity of the CrIS PAN retrieval, the assumption about what to append above the top of the aircraft profile is far more important than what is appended at the bottom, (provided what we append at the bottom is reasonable!) If we did not have these targeted model runs available, an alternative crude approach could have been to simply append the retrieval prior to the top of the aircraft profiles. The difference between these two approaches provides some estimate of the uncertainty associated with the assumed profile above the top of the aircraft measurement. We find a 20 % reduction in the aircraft/satellite slope between the case where we append the prior and the case where we append the dedicated GEOS-Chem runs. We can think of this as a pessimistic estimate of the error associated with the assumption of the profile above the top of the aircraft profiles. We have now included this information in Section 4.

Why is the instrument not sensititve (at all? enough?) to high PAN VMRs closer to the tropopause? Is this related to the low temperature at this altitudes?

CrIS is sensitive to PAN throughout the atmospheric column. We have now added a figure that shows a representative averaging kernel and additional discussion of vertical sensitivity in Section 3.2. Please also see responses to Reviewer 1's question about the DOFS.

SPECIFIC COMMENTS

==========

line 135

Particularly in the face of the discovered systematic error, a discussion on the sensitivity of the retrieved PAN VMRs on the previously derived quantities (i.e. the 'b' vector) might be interesting. It is mentioned that the retrieval processor under-estimates the "observation error", without detailing what exactly this entails. Often this only contains - for practical reasons - an estimate of the noise induced error, not the systematic errors. How does the identified systematic bias relate to the error diagnostics for systematic (b-related)
errors?

This is a good suggestion. In response to comments from both reviewers, we have now made extensive updates to the manuscript to include discussion of systematic errors.

line 165

Please show a (representative set of) averaging kernels to show the region of sensitivity.

We have now added a figure that shows an example averaging kernel and added some discussion of this figure to the text.

Figure 4

The residual shows structure beyond the noise level (blue lines). The caption indicates that this spectrum was computed with a zero PAN profile. Please show both a spectrum with the derived PAN profile and with a zero profile to show the improvement and PAN signal as well as quality of fit of the used spectra (similar to Glatthor et al., 2007)

Note that the figure in question shows sample residuals over a large range (760-860 cm-1), but that the PAN windows are pretty small. PAN is only fitted within the small red windows. For retrieval development purposes, we had set up runs where the "pre-PAN" state information (including retrieved temperature, water vapor and ozone profiles) is run through the forward model for the 760-860 cm-1 range to generate these wide filter residuals. We could, in principle, update our system to run a "post-PAN" step to generate the wide filter residuals after the PAN retrieval. However, this would take some effort. We think that the figure shown is sufficient to show the expected PAN signal and so we would rather leave this figure as it is.

We also note that one benefit of the figure as it stands is that it does show an example of the radiance offset issue that exists before the PAN step that leads to the water vapor dependent bias.

Figure 5

The paper identifies a low bias of 100 pptv causing many VMRs to be negative as shown in Fig. 7. Figure 5 shows only positive VMRs. Please explain the discrepancy.

This was an oversight that has now been corrected. The data used to generate Figure 5 did include negative values, but we had set the color scale to bottom out at zero. The black region at the bottom of the original bar was misleading. We have updated the color scale to show the regions where PAN remains systematically negative, even after the water vapor dependent bias correction has been applied, and we have updated the caption to make it clear that the color corresponding to the lowest box on the color bar includes points that are more negative than the lowest marked values. We also include an additional figure that shows the actual range of values for this day.

MINOR REMARKS

==========
line 108

A big X with a hat was not in (1). Maybe big-hat-x -> hat-x and hat-x-a -> hat_x ?

Yes. Now fixed. Thank you.

line 113

\Delta f should be 'bold'.

Fixed.

line 128

---
an approximate solution?

We prefer to leave this the way it is.

line 134

CCl_4 (small l)
line 139
Fixed.

It is not clear from the context what the "forward stream" is. The given reference distinguishes a "reanalysis stream" without being clear on the difference. I suppose it has something to do with using (forward) extrapolation of calibration data in contrast to interpolation using (later) data. This is probably a very common term in certain scientific communities. Maybe explain it in a brief sentence.

The distinction between "forward stream" and "reanalysis" in the TROPESS datasets is that the forward stream represents low latency processing with the most recent algorithm version available at the time of the measurements, meaning that the data version can change in time, whereas the "reanalysis" represents a processing of a long-term dataset with a consistent version of the retrieval algorithm. We have now added some explanation of the forward stream and reanalysis to this paragraph.