

Atmos. Meas. Tech. Discuss., referee comment RC1
<https://doi.org/10.5194/amt-2021-211-RC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on amt-2021-211

Anonymous Referee #1

Referee comment on "Validation and error estimation of AIRS MUSES CO profiles with HIPPO, ATom, and NOAA GML aircraft observations" by Jennifer D. Hegarty et al., Atmos. Meas. Tech. Discuss., <https://doi.org/10.5194/amt-2021-211-RC1>, 2021

The authors present an in-depth error analysis of the AIRS MUSES CO retrieval product. The MUSES algorithm follows the Rodgers optimal estimation approach (OE), which allows them to quantify a smoothing error, measurement noise, systematic uncertainty, cross-state error and retrieval residual. They use these to derive four error terms – theoretical, a-priori, retrieval, and empirical – for MUSES CO retrievals collocated with aircraft observations made during the HIPPO, ATom and NOAA GML campaigns. For each campaign, the authors repeat their analysis and present four figures and a table.

This paper summarizes an extraordinary amount of meticulous work about an important topic in satellite retrieval theory and application today, namely error quantification and validation. The authors acknowledge the importance of understanding all the retrieval error terms if we are to use a satellite record in climate science and the quantification of Earth system change. This is especially true for the AIRS record that now spans two decades. But as this paper makes clear, this is no easy task. I have spent some time with this paper, revisiting sections, and am left with the conclusion that while the authors present clear, detailed results, they fail to communicate what it all means. It would have been very helpful to the reader if the authors explained how these results will be used to improve the MUSES algorithm (or product) going forward, or what this error evaluation means for the AIRS record and its application in Earth system science. After working through all the details, trying to understand the results, I (the reader) am left thinking "so what?". This paper can make a meaningful, even important, contribution if the authors elaborate on the significance of their results in the Discussion/Conclusion.

Review:

- Lines 37-39: "We find mean biases of + 6.6% +/- 4.6%, +0.6% +/- 3.2%, -6.1% +/- 3.0%, and 1.4% +/- 3.6%, for 750 hPa, 510 hPa, 287 hPa, and the column averages, respectively. The mean standard deviation is 15%, 11%, 12%, and 9% at these same

pressure". This sentence is very difficult to read and I suggest rephrasing it to help clarify one of the main results of this paper.

- Line 143: "Atmospheric Infrared Sounder (AIRS)" already defined
- Line 164: "CO is retrieved using the 2181-2200 cm⁻¹ spectral range". Does MUSES use all channels in this spectral range for its CO retrieval?
- What does MUSES use as a-priori for CO? I think it is important to state this clearly in the paper given the results presented later.
- MUSES does not follow the AIRS Science Team approach of deriving an aggregate clear-sky radiance from each 3 x 3 array of AIRS measurements in partly cloudy skies. While cloud clearing reduces the spatial resolution of the radiance measurements ahead of retrieval, it not only allows stable retrievals in complex scenes but also allows the quantification of uncertainty due to clouds (Smith and Barnett, 2019; Susskind et al., 2014; Maddy et al., 2009; Chahine, 1982, 1977). Clouds being one of the primary sources of scene-dependent uncertainty, this is an important source to account for in an error analysis. How does MUSES quantify systematic uncertainty due to clouds? And given the results presented, can the authors draw any conclusions about the validity of retrieving CO from AIRS measurements in the presence of clouds?
- Line 175: "the original non cloud-cleared radiances". It will be more straightforward to simply say "the instrument radiances"
- Line 178: "profiles with thick clouds were also removed from the set." How did the authors distinguish thick (versus thin) clouds?
- Figure 2 needs to be resized.
- It is difficult to make sense of the results in Tables 1 through 3. I think a single figure summarizing the values from all three would have made it easier to inter-compare among latitude zones and aircraft campaigns.
- Lines 285-286: "Beyond examining biases and variability of the retrieved profiles, evaluating the retrieval error estimates is also important, since they provide users with a measure of the reliability of the data". I agree with this statement in principal but "provid[ing] users with a measure of the reliability of the data" is what bias and standard deviation tells the user at first order. What, in the authors experience, do an error analysis contribute over and above a measure of reliability? Perhaps the authors can clarify this point with examples on how such an error analysis influences algorithm design/updates and data application.
- Figures 5, 9, 13:

Is it correct to interpret this figure as meaning that the MUSES retrieval basically added noise to the a-priori between the Earth surface and 600 hPa?

How sensitive are these error values to variation in a-priori error?

What would be an ideal relationships between all these error terms?

Do the authors think that the observation error in the lower troposphere will change if they adjust the a-priori error vertically?

The legend states "mean observation error", "mean a priori error", 'AIRS-AIRCRAFT Std.

Dev" and "A Priori-AIRCRAFT Std. Dev", but in the text associated with these figures, the authors discuss the "theoretical error" and "empirical error". It will help the reader a great deal if the authors maintain consistency in their terminology.

Do these results mean that an end user should use the a-priori instead of the retrieval in the lower troposphere?

- Line 277/Line 355: Figure 3/Figure 7 captions, "The number of profiles and the latitude bands are indicated in the upper left". I only see the sampling size listed, not the latitude band.
- Figure 11 should be resized.
- Line 603: "AIRS MUSES algorithm indicated that the retrieval quality was generally within expected limits." What are these limits?
- Given this in-depth analysis of multiple error sources, what does it mean for AIRS product design? What are the lessons learned for algorithm teams and/or end user applications?

References

Chahine, M. T.: Remote sounding of cloudy parameters.II. Multiple cloud formations, *J. Atmos. Sci.*, 34, 744–757, 1977.

Chahine, M. T.: Remote sensing of cloud parameters, *J. Atmos. Sci.*, 39, 159–170, 1982.

Maddy, E. S., Barnet, C. D., and Gambacorta, A.: A computationally efficient retrieval algorithm for hyperspectral sounders incorporating a-priori information, 6, 802–806, <https://doi.org/10.1109/LGRS.2009.2025780>, 2009.

Smith, N. and Barnet, C. D.: Uncertainty Characterization and Propagation in the Community Long-Term Infrared Microwave Combined Atmospheric Product System (CLIMCAPS), 11, 1227, <https://doi.org/10.3390/rs11101227>, 2019.

Susskind, J., Blaisdell, J. M., and Iredell, L.: Improved methodology for surface and atmospheric soundings, error estimates, and quality control procedures: the atmospheric infrared sounder science team version-6 retrieval algorithm, 8, 084994, <https://doi.org/10.1117/1.JRS.8.084994>, 2014.

