Response to Reviewer 3

Dear Reviewer,

Thank you for the review, we respond to your comments below. We have kept your original text in black, our responses are in blue and specific changes are in underlined blue.

This study employs machine learning (ML) to predict retrieval failures based on measured radiances from sensors (CrIS and AIRS+OMI). By using ML as a pre-processing tool, computational resources can be utilized more effectively, as retrieval algorithms are often computationally intensive due to the high precision and accuracy they require. The manuscript is well-written and provides a good analysis of the ML algorithm's ability to filter spectra based on measured radiances from various instruments.

Thank you for the general positive comments.

I believe the study is suitable for publication after addressing the following minor points:

1. Line 200: Regarding PCA, how much of the variance is explained by the 30 components?

We added the following sentence to provide the requested detail:

"We perform PCA to reduce the number of dimensions to 30, or the dimensionality of the dataset, whichever is lower. 30 principal components account for 98.9%-99.6% of explained variance for the full spectrum, and 99.99% for the fitting spectral regions."

2. Section 4.2: When evaluating feature importance, the conclusion suggests that features outside a given window are as important, if not more so (depending on the retrieved species), as those within the window. If the master quality flag includes information from all windows, this would logically increase the importance of information outside the window. Additionally, the spectrum contains information about O_3, CO, and TATM outside their respective windows. Could the ML algorithm be sensitive to these regions as well? If so, could future work explore developing an ML algorithm to extract O_3, CO, and TATM directly from the spectra? This can be addressed as potential future work in section 6.

This point is very true, it may be possible to develop a ML algorithm to derive trace gas concentrations directly from measured spectra, as the principle is the same as that shown in this paper. We think it is important to note that, the accuracy of this method remains to be improved, and direct applications of this method to trace gas retrieval

could lead to significant uncertainties that are not traceable due to the black-box nature of ML. We have added a discussion on this point in the paper, as bellow:

"One of the implications of this paper is that ML models can differentiate different atmospheric conditions from measured spectra. This implies that an appropriately trained ML model may be able to infer trace gas concentrations directly from measured spectra, as opposed to using the OEM or other retrieval methods. While this will form future interesting work, the risks of all ML methods, such as appropriate training sets and unintended biases would apply, which would add uncertainties to any retrievals derived from this method."

3. In Section 5, O_3 and CO are shown for different ML threshold values. This section could be strengthened by comparing the different filtering thresholds to a truth proxy. This would allow for a clear presentation of how the filtering threshold impacts bias and precision. The current quality filter would be a base line for comparison.

This is a good and valuable point the reviewer makes, and certainly would help strengthen this work. However, we feel that this would be better placed in a follow up paper, that further explores this concept in more depth, after applying additional training methods as described in the discussion section of this paper. Further, this paper already contains extensive technical detail, and we think that adding an additional section as described would make this paper very long.