In response to some concerns of the first reviewer, this revision of our paper has a significant change to the retrieval algorithm in that we removed the initial, "tentative" retrieval of cloud optical depth (τ_{cld}) prior to the simultaneous retrieval of τ_{cld}, T_{cldtop}, r_{eff}, T_{sfc}, T_{atm}, H_2O, O_3 and CO_2. Changes to results were comparatively minor, but we thought significant enough to modify our algorithm. Conclusions are for the most part unchanged. Figures have been updated, as well as quantitative comparisons to radiosondes, AIRS-V6 and CloudSat/CALIPSO. Please see our response to the first reviewer for more details.

Overall assessment. This manuscript tackles one of the most fundamental issues in the field of hyperspectral IR remote sensing, that is the problem of clouds in an inversion scheme. The text is very well written and provides a way forward to the advancement of the inversion problem. All assumptions in the treatment of the cloud a priori are well explained. This lays out the foundation for a constructive conversation on the future directions that IR retrieval developers may want to consider when it comes to address the presence of clouds in an inversion methodology. I suggest this manuscript to be published, pending minor corrections as outlined below.

Minor comments: 1. The authors should add that, besides not needing absorption and scattering in the forward model, another advantage of cloud clearing is that it also does not require a formal cloud geophysical a priori and its error co-variance; it is computationally fast; it enables full column retrievals. A limitation, besides assuming constant water vapor in the cluster of FOVs, comes from also assuming uniform surface properties over the cluster of FOVs, which can be challenging over coastal regions.

We agree that there are significant differences in the benefits/drawbacks of single-footprint vs. cloud-cleared spectra, which for brevity we did not go into. Instead, we discuss what we thought the most important difference, namely the horizontal resolution, which shows up most significantly in the water vapor and relative humidity.

2. Could the authors explain in few sentences how their work differs from the cited existing methods employed for direct use of cloudy infrared spectra in atmospheric retrievals (Liu, 2009, Blackwell 2005 and Kulawik, 2006a).

We hesitate to do this as differences between AIRS-OE and the Liu, Blackwell and Kulawik methods would be fairly lengthy even if described cogently, and our paper is already quite long. We thought readers would be better served by going to those papers directly, and more important (and perhaps more interesting from the reader's perspective) to contrast AIRS-OE with the AIRS operational retrieval, AIRS-V6.

3. Can the authors say anything about using a monthly averaged surface emissivity, is this a robust approximation for a given day of the month? Also, have they evaluated the inter-annual variability in the data set to derive a rough assessment of how good of an approximation is to apply the 2003 climatology to the year before? Have they *attempted a comparison with the existing surface emissivity regression solution of AIRS v6?*

In Sec. 3.2.3, we note that the emissivity is not retrieved in this version of AIRS-OE, but we noted an emissivity retrieval may happen in future versions. We have added a note about this in comparing AIRS-OE to AIRS-V6 in the introduction. In Sec. 3.7.2, we also note that using a fixed emissivity is an error source that we have not yet quantified. The use of a fixed emissivity can be an admitted deficiency of the algorithm in its current state. At this stage of AIRS-OE development, it may be premature to make a robust evaluation of the emissivity as its calculation may change in future versions.

If not, can they say few words about their motivation to replace the AIRS v6 method? Also, for completeness, they should include this as part of the list of differences with respect to AIRS v6 on page 2 -3

This effort was a proof-of-concept not meant to "replace" AIRS-V6, but rather to provide a means where studies, in particular those of water vapor, can be done on the native resolution of the AIRS infrared observations. In the introductory comparison with AIRS-V6, we've noted that AIRS-OE does not retrieve emissivity, although that might change.

4. Do the MODIS cloud parameters come with a formally computed error covariance? Otherwise, what is the source for this?

The MODIS Level 2 files used have quality flags, but do not have an error covariance provided; we do not use the MODIS data directly in the retrieval, but rather a weighted average of them. We have used ad hoc error covariances in the retrieval, as described in Table 1.

5. Was there a specific geophysical regime where the use of the 9.6 micron band was problematic, for example the tropical region, or desert areas? The broader scientific community would benefit from a more detailed explanation of what the sentence "We therefore retrieve O3 only as an "interferent" gas within the 14 μm CO2 region, and avoid the 9.6 μm band" means. Did the authors only use channels in the CO2 band for the retrieval of ozone? If not, can they explain more?

In Sec. 3.5, we have expanded our explanation of why we didn't use the 9.6 micron band, note that we are not currently retrieving ozone as a "primary" product, and caution against using our O3 retrievals for scientific investigation. In the conclusions section, we briefly discuss extended the retrieval to retrieve ozone using the 9.6 micron band.

6. Section 3.6 Is the approximation of leaving all other variables fixed taken into account in the retrieval equation, for example, similarly to what is done in AIRS v6? Also, in the list of differences from the AIRS v6 approach, the authors stated that this is a simultaneous retrieval approach. Can they more fairly say that this is a two-step sequential approach?

We note in Sec. 3.7.2 the error from parameters not retrieved, and how future versions will incorporate an estimate from these. As noted earlier, and for reasons we have written in the response to the first reviewer, we have revised our procedures to go directly into the simultaneous retrieval.

7. Also about section 3.6, if in within the page limits, can the authors say few words on the convergence criteria, besides mentioning the reference to Bowmann et al, 2006? Are those the same as outlined later in the text? If yes, for clarity, the authors should refer the reader to that section.

We've put some numbers on convergence and exit criteria in the solver, but in the interests of brevity, we refer the reader to the Bowman and Moré papers.

8. What enters in the measurement noise covariance term, only instrument noise? Can the authors specify this?

Currently only the diagonal of the instrument noise is in the measurement error covariance term. We discuss expanding this term in future version to include correlated noise and random error from the forward model.

9. As an alternative method to the median of the medians, the authors should try the methodology described in Nalli et al., "Validation of satellite sounder environmental data records: Application to the Cross-track Infrared Microwave Sounder Suite", JGR, 2013, where each layer statistic is weighted by the mean water vapor quantity in the layer. This method was also used in Tobin et al., 2005, JGR and seems to be a well established methodology in the satellite retrieval field.

At this stage of AIRS-OE development (and validation), the difficulty was not so much from statistics skewing for H_2O in dry regions, as discussed by Nalli et al. in their Sec 3.1.2. Given that our retrievals were quite different than those on cloud-cleared radiances, the problem was how to combine retrieval statistics when there are widely differing results because of cloud cover, and still make a fair a comparison to AIRS-V6. An important consideration was how to avoid skewing the statistics so that (near-)clear scenes are would not be overrepresented. The "median of the medians" approach was developed to succintly indicate how well the AIRS-OE retrieval was doing without getting too granular, and see how it compares to the current AIRS-V6 retrieval. As the retrieval matures, and more data are analyzed, we'll reconsider our validation using something more akin to Nalli et al.

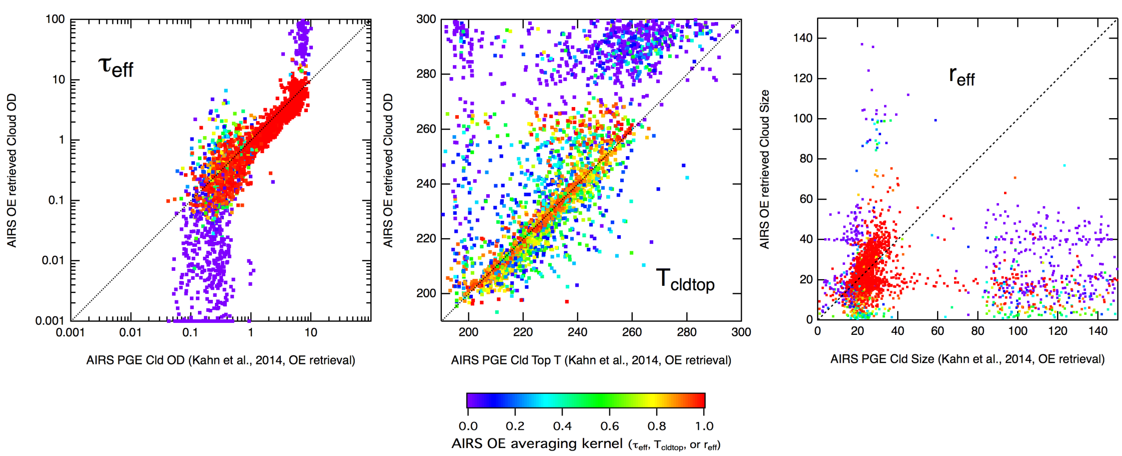
10. Summary and discussion section. How are the authors planning to extend this retrieval method to multi-layer cloud retrievals. Does the MODIS retrieval output contain multi-layer clouds that can be used as a priori?

The MODIS data can tell us if there are clouds at different cloud-top temperatures in the scene, but their vertical extents, particularly at night, are non-trivial issues we would need to work on. We briefly mention using weather prediction for this in the conclusion, as did DeSouza-Machado et al. (2017).

11. Could the authors also consider comparing their cloud retrieval products to the AIRS v6 cloud retrieved products?

We have looked at comparisons between the Kahn et al. (2014) optical estimation retrieval of cirrus (Kahn2014), which uses the AIRS-V6 cloud-top temperature as a priori, against the AIRS-OE retrieval. (Note that the Kahn2014 retrieval does not include liquid water clouds.) We include a comparison figure below. We did not include it for brevity's sake as we felt that validation of the temperature and water vapor (and relative humidity) was more needed, as their increased horizontal resolution was what was new about this work (at least compared to AIRS-V6). We also feel that a detailed cloud comparison needs to go well beyond a simple comparison such as we show below, thus we did not include in the manuscript. We did include the comparison with CloudSat/CALIPSO as an initial check against a completely different instrument suite, however.

Comparisons for Granule 44, Sept 6, 2002 are shown below, colored by the AIRS-OE averaging kernels (AKs) for each parameter. Results for τ_{eff} are well correlated for AKs close to unity, but the AIRS-OE results are biased low (~25%) compared to Kahn2014. Cloud-top temperature shows mostly good agreement for higher AKs. There is more disagreement for effective radius, r_{eff}, but we note that the areas where they differ greatly (Kahn2014 r_{eff} > ~60 microns in the lower right), almost all the matchups have Kahn2014 quality flags of 2 ("Do Not Use"). The r_{eff} retrieval is sensitive to the "tilt" of the spectrum, so some of the r_{eff} difference may be related to the spectral range of channels used; the upper range of the Kahn2014 channels is ~1200 cm^{-1}, while AIRS-OE goes to ~1600 cm^{-1}. This will require more investigation, and we will continue to compare results as the AIRS-OE retrieval evolves, and we process more data.



Technical comments: 1. The broader science community might not be familiar with what "L1b" data are. The author should either keep using the "Level 1" definition for radiance data or explain what the "b" stands for.

We’ve added some language clarifying this in the first paragraph of the introduction.

Additionally, we found a programming error that led to an undercount of successful AIRS-OE retrievals matched to sondes (Figures 14 and 15 in the revised draft). This error has been corrected.

We thank the reviewer for perceptive comments and good questions.