

In the revision of our paper, we have added more detail per the reviewer's concerns. Most importantly, we made a significant change to the retrieval algorithm in that we no longer have 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 . Figures and quantitative comparisons to radiosondes, AIRS-V6 and CloudSat/CALIPSO are updated. Details of this and other specific areas are addressed in the responses below.

General comments:

The authors describe a 1-D variational retrieval of temperature, atmospheric constituent gases, and cloud properties from AIRS observations (AIRS-OE) and co-located operational MODIS cloud property retrievals. Multiple retrieval strategies are employed based on the availability of MODIS cloud properties and their sub-pixel characteristics within an AIRS FOV. The paper fits well within the scope of AMT and provides a sufficient contribution to scientific progress in the field of remote sounding from hyper- spectral IR measurements.

The layout of the paper is logical and the algorithm flow is well described; however, there are significant technical and material deficiencies that need addressed before acceptance to AMT.

For instance, some important details of algorithm settings (e.g., measurement noise) are missing from the text and the authors' description of the retrieval information content/error estimation is sometimes confusing. In addition, the sections on the discussion of results, validation of the AIRS-OE algorithm against matched radiosonde, and comparisons to operational AIRS v6 (and conclusions drawn from the results) are lacking enough detail to fully assess the comparative pros, cons, and skill of the AIRS-OE retrieval relative to existing products.

Specific comments:

"QA" is not defined (pages 13, 14, 15) in the text.

We've changed "QA" to "QC", and spell out this abbreviation in Sec 3.8.

Page 1, Line 20: "higher vertical resolution of retrieved temperature and water vapor" There are other advantages of thermal infrared data – e.g., trace gas sensitivities, sensitivity to aerosol, . . .

We agree, but trace gases were retrieved here only insofar as they improved the temperature, water vapor and cloud retrievals. We didn't get into the advantages of the infrared with respect to trace gases or aerosol because we (mostly) did not retrieve them in this initial effort. We have changed the phrase "The advantage ..." to "An advantage ..."

Page 2, Line 10: . . . the Stand-alone AIRS Radiative Transfer Algorithm, Delta-Four Stream (SARTA-D4S; Ou et al., 2013). There is another extension to the SARTA package

that enables the simulation of outgoing radiance in the presence of cloud and aerosol and that is directly applicable to AIRS studies: DeSouza-Machado, S., Strow, L. L., Imbiriba, B., McCann, K., Hoff, R., Hannon, S., Martins, J., Tanré, D., Deuzé, J., Ducos, F., and Torres, O.: Infrared retrievals of dust using AIRS: comparisons of optical depths and heights derived for a North African dust storm to other collocated EOS A-Train and surface observations, J. Geophys. Res., 115, D15201, doi:10.1029/2009JD012842, 2010.

We didn't reference the DeSouza-Machado et al. (2010) paper since they it was mostly about dust retrieval (something we didn't try in AIRS-OE). However, the referee is correct in that it was a significant extension to the forward modelling in AIRS, and this is noted in the conclusions. We have also made note of the recent DeSouza-Machado et al. paper currently in discussion in AMT.

Page 4, Line 2: How is the radiance noise covariance prescribed? AIRS detector modules have significant correlated noise among channels within each module (e.g., <http://onlinelibrary.wiley.com/doi/10.1256/qj.03.93/pdf>, page 1480 and Tobin, et al., "Hyperspectral data noise characterization using principal component analysis," J. Appl. Remote Sensing, v1, 2006). Does the algorithm account for the correlation in the observation covariance, or does it only use a diagonal matrix? How is the noise estimated or prescribed? Are forward model error components estimated or used in addition to instrument noise? If so, how? Do you use a bias correction between obser- vations minus calculations; if so, how is it estimated?

Only the diagonal error is reported in the AIRS Level 1b data, and we have not estimated off-diagonal elements. This should have been noted, and we thank the reviewer for pointing this out. We have added some comments about this, and a reference to the Tobin et al. paper. We have not put in estimates of the systematic error components for the forward model (this was noted in 3.7.2), but we expanded on this a little and w.r.t. bias corrections, and we note tuning done in SARTA and refer to the validation paper by Strow et al. (2006). We now note that random error from the "clear" part of the forward model is expected to be less than random errors from the observed radiance, but it should be included in future versions.

Page 5, line 35: "use extrapolated cloud absorption and scattering parameters, and may not be reliable." This statement is not clear. I assume that the cloud absorption and scattering parameters are restricted in the Forward Model lookup tables between 5-85 microns, but it is not clear from the text.

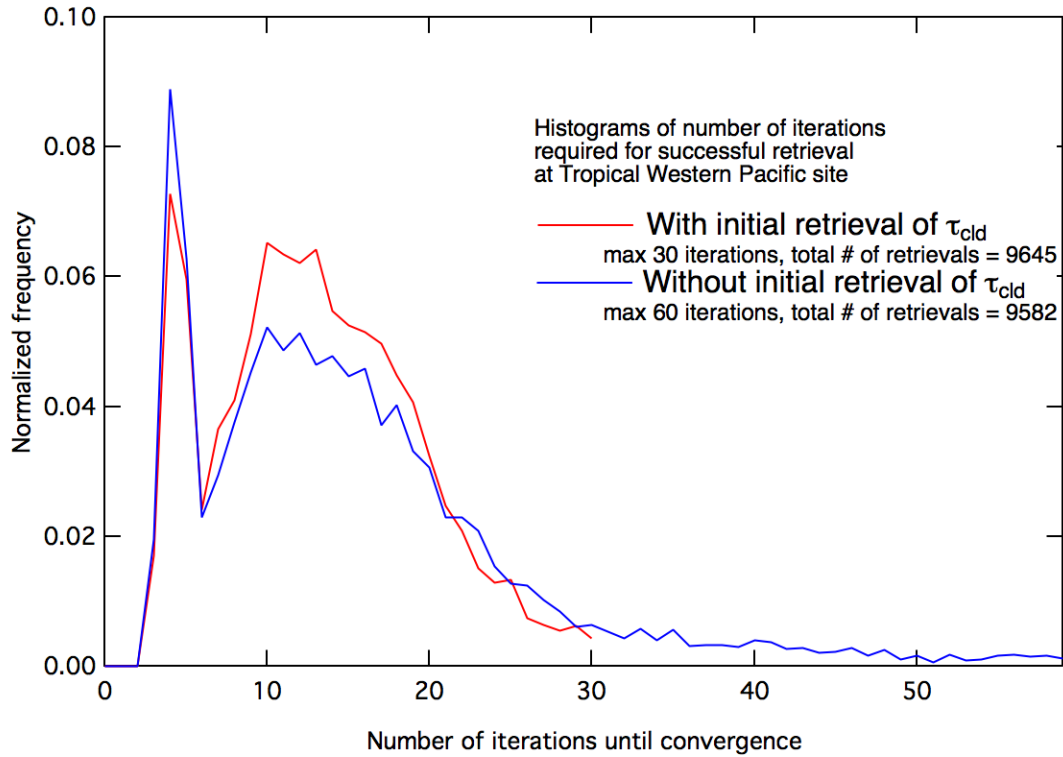
We have changed to line to read, "Lookup tables for cloud absorption and scattering parameters had a particle radii range from 5 to 85 μm ; reported cloud absorption and scattering parameters outside this range rely on extrapolated parameters, and results may not be reliable."

Page 8, line 2: "...of uncertainties in the scattering/absorption ratio (Nakajima and King, 1990; Nakajima et al., 1991)" ... perhaps also due to the 4 stream approximation of the forward model employed in this study.

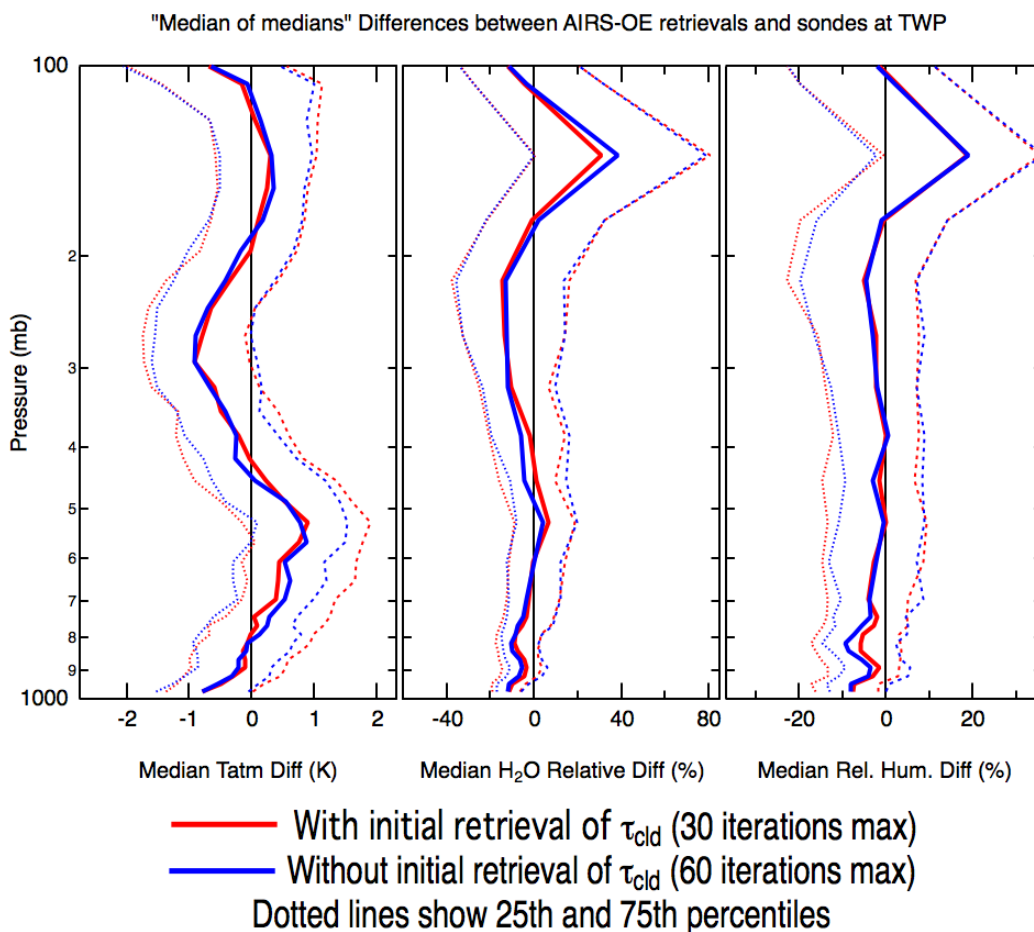
This is possible, but we also encountered difficulties in clear scenes when using the shortwave channels, and so we think it likely that the problems of reflected sunlight contribute substantially to this. Whether this is an issue with just SARTA, or with D4S as well needs investigation. We also did not adjust the noise in the shortwave channels as a function of channel radiance, as illustrated in the Tobin et al. paper. As we used the L1b errors directly (which tend to be less than those of cloud-cleared radiances in the shortwave), we tended to rely more strongly on the accuracy of the forward model in the shortwave (and the fidelity of the reflected sunlight calculation) than would be done in AIRS-V6. Profile biases and failed retrievals when using the shortwave were non-trivial problems that we could not quickly solve, so we decided at this stage of development to avoid the issue altogether and just use longwave channels.

Page 8, Line 13: "...First, leaving other variables fixed, only τ_{cld} is retrieved...The resultant spectral fit from this initial retrieval can be poor." If the spectral fits are poor after this first attempt how is convergence/iteration stopping criteria determined? This particular detail of the algorithm flow seems important and critical to the success or failure of the algorithm to converge to the optimal solution in subsequent steps.

We note that a retrieval can "converge" such that subsequent iterations would not lower the cost function any further; the fit can be poor (as measured by the chiSquare), but better than one calculated directly from the a priori, and that we did not change x_0 in the cost function or the a priori covariance of τ_{cld} from the first step to the simultaneous step – perhaps avoiding the warning in Sec. 5.6.2 of Rodgers [2000]. Nonetheless, we reviewed this aspect of the algorithm, and reprocessed validation data from the TWP, SGP and NSA sites, skipping the initial retrieval of cloud optical depth and going directly to the simultaneous retrieval. We found that to get close to the earlier yields, we had to double the number of allowable iterations. The graph below is a histogram of the number of iterations needed with and without the initial retrieval of cloud optical depth over TWP.



Comparisons against radiosondes (using the “median-of-the-medians” approach) shows effectively no differences at NSA, some differences at SGP, and the largest differences at TWP. The TWP comparisons (below) show small and isolated biases with each other, but we think they are significant enough to call into question whether the initial retrieval of cloud optical depth can lead to an optimal solution.



Given this, and despite the extra processing required, we decided to remove the initial retrieval of cloud optical depth, and reprocess the AIRS data shown in the paper. In reflection of this, we have re-written Sec. 3.6, and modified the block diagram of Figure 2. The retrieval that produced the averaging kernels and row sums in Figures 4 and 5 of the original draft failed under this new processing, so we selected a new retrieval to produce the sample graphics. We thank the reviewer for bringing up this issue and prompting us to study it further.

Page 8, Line 23: "...nearly linear in vicinity of the solution" and a priori? Per the above comment about poor spectral fits after the first attempt, I wonder how well this linear assumption holds for all retrieval strategies employed. The authors should address the potential failure of this assumption and/or reasons why the assumption is valid in greater depth.

The assumption of linearity is perhaps most important in the error characterization. As the a priori covariances are largely ad hoc, we cautioned against the errors being reliable. We thought it best to leave validation of the error characterization, and the linearity assumptions that go with that, to a later study when AIRS-OE is more

refined. We have added a phrase noting that the assumption of linearity has not been tested. In the Summary and Discussion section, we mention adding linearity tests to future versions and point the reader to the section in Rodgers [2000] that expands on this.

Page 9, Line 33: "...on the spectrum" suggest to revise to "measurement spectrum."

Fixed. It now reads "...on the measured spectrum..."

Pages 9-10: How does the vertical resolution of the AIRS-OE algorithm compare to the operational AIRS V6 vertical resolution.

We've have not tested this specifically as it goes beyond the scope of this initial effort. Making an fair comparison between vertical resolution estimates between the two versions has to be performed and documented carefully as the averaging kernels of AIRS-OE and AIRS-V6 (the latter using singular value decomposition) are calculated quite differently.

Page 10, Line 12-13: Continuing the previous comment. "... depend on the amounts of trace gases present, the temperature lapse rate, the particulars of the cloud field" These statements are true for any retrieval strategy (cloud-clearing, cloudy, or clear) from AIRS or any other instrument.

True, but the section walks the reader through a sample averaging kernel for a simultaneous retrieval, and explains how its information can be extracted and used. We thought it helpful to tell the reader less familiar with averaging kernels that they're quite scene-specific.

The authors continue "...since the AIRS-OE retrievals are simultaneous and not sequential. . .," which seems to be a nod to the operational AIRS algorithm. Consider adding a reference (e.g., Maddy, E.S. and C.D. Barnet 2008. Vertical resolution estimates in Version 5 of AIRS operational retrievals. IEEE TGRS. v.46 Section 2A, p. 2377) to back up the first statement and provide contrast the second statement.

Section 3.7.1 now references the Maddy and Barnet paper, but we note that add that the averaging kernels from the AIRS operational retrievals do not work *across* species.

Page 10, Line 25: How is the noise error covariance defined? See above.

Section 3.3 has been expanded to describe this more fully, and we refer back to that section.

Page 10, Lines 29-30): "...spectral biases or other errors that are correlated across observations" calibration uncertainty, correlated instrument noise (if not included in measurement covariance . . . see previous comments).

The lines relate that we are not calculating systematic error in Eq. 11, and systematic error would be affected by calibration uncertainty. The correlated noise

would more affect random error, and we have noted that it's not there (but eventually will be) in the line below Eq. 10. We added a note that Strow et al. use tuning to fix some of the systematic error coming out of SARTA.

Page 12, Line 23-25: "...The morphology of the AIRS-OE retrieval fields are similar to the a priori, and the morphology of the averaging kernel fields are similar to each other. " This statement is confusing. Please revise for clarity..

We're re-written this to read: "Retrieved quantities have similar fields to their a priori. As indicated by the averaging kernels, sensitivity is enhanced for ice clouds, which is likely because of the higher thermal contrast with the surface."

Page 13, Line 26: "...for each of these..." Consider revising to "for each retrieval quantity"

Done. Thanks.

Page 13, Line 30: "...unphysically high values..." It's unclear whether the authors are referring to unphysically high differences (>100%) or unphysically super-saturated retrievals.

We've changed to read "...unphysically high RH values..."

Section 4.3: Spatial resolution is one of the main differences between AIRS-OE and AIRS V6. Have the authors performed a comparison between quality controlled AIRS-OE averaged onto the AIRS V6 effective footprint (i.e., 3x3 spatial average)? A short discussion of this type of comparison is suggested.

We agree that such a comparison would be interesting, but this is beyond the scope of the current study. Such an analysis would be useful after we ameliorate many of the retrieval's listed shortcomings and move to a more mature code.

Page 13, Line 33: "IGR" should be IQR.

Fixed. Thanks.

Page 14, Lines 11-16: Authors provide a qualitative description of the "a priori" sensitivity of the algorithm. A more detailed quantitative would be informative to the reader.

We're adding in a profile plot of the average row sums of the Tatm and H2O averaging kernels for retrievals using the ECMWF a priori. Plots using the climatology or neural-net a priori were very similar.

Page 14, Line 30: "...lower down. ." consider removing.

Removed.

Page 14, Line 33: "...because the water vapor retrieval was too high..." Too high relative to what? Consider revising for clarity. For instance, "because the reported retrieval water vapor was supersaturated."

We've changed the text to read, "We also found that average bias could often be significantly skewed by retrieval outliers – usually because the water vapor retrieval and relative humidity were unphysically high."

Figures 13, 14 and corresponding text: The AIRS-OE a priori statistics (bias and IQR) are generally more well behaved as compared to both the AIRS v6 and AIRS-OE retrievals and in some cases perform better than both physical retrieval algorithms (esp. near surface for temperature). I wonder how much do the cloud/other variable retrievals compensate for other variable/cloud errors. A more detailed analysis and/or discussion of an assessment of AIRS-OE retrieval increments of profile variables as well as surface temperature, and cloud optical properties relative to a priori as compared to Radiosonde and correlative CloudSat/Calipso retrievals should elucidate cross-talk between retrieval parameters is suggested.

We agree that more detailed analyses would be useful, but with the resources given this effort, we could only make limited comparisons against sondes and comparisons to AIRS-V6. As noted in the Summary and Discussion section, there are several non-trivial issues that we have identified against the algorithm that should be fixed for the upcoming versions. A more comprehensive validation and analysis would be warranted as the algorithm matures.

Page 15, line 30: "are low" – how low?

We've changed the line to read: "At NSA between about ~800 and ~220 mb, AIRS-OE temperature biases are low (less than ± 0.25 K) at altitudes above ~800 mb, as are water vapor relative biases (less than $\pm 6\%$)"

Page 16, line 9: ". . . has an information content analysis . . . operates both within and across different atmospheric parameters." It is unclear what is meant by "operates both within and across." Do you mean that the information content analysis provides diagnostic information regarding the temperature retrieval, water retrieval, etc. and interactions between temperature and water, temperature and cloud, water and cloud, etc. If so, please revise.

We changed the sentence to add examples: "As AIRS-OE rests on an optimal estimation framework, and includes simultaneous retrieval of profiles and scalar variables, it has an information content analysis that operates both within an atmospheric parameter (e.g., uncertainties in the temperature profile) and across different atmospheric parameters (e.g., uncertainties in water vapor due to uncertainties in temperature)."

Page 16, line 16-17: ". . . incorporating scattering by dust . . ." Is the radiative effect of an atmospheric dust signal in AIRS measurements large enough to cause significant degradation in a 1DVAR retrieval? Consider adding a sentence and or reference detailing why this is might be important (e.g., Maddy, E., DeSouza-Machado, S., Nalli, N., Barnet, C., Strow, L., Wolf, W., Xie, H., Gambacorta, A., King, T., Joseph, E., Morris, V., Hannon, S., and Chou, P.: On the effect of dust aerosols on AIRS and IASI operational level 2 products, GRL, 39, L10809, doi:10.1029/2012GL052070, 2012. 446)

Done. Thanks.

Page 16, line 17-18: Again, does the 4-stream RT algorithm limit the use of AIRS shortwave channels? Are there any other specific limitations with forward modeling that need to be addressed before the 4micron band could be utilized? Scattering tables, spectroscopy, etc.?

As noted above, these are non-trivial issues with SARTA (or most any other thermal-IR forward model). At this stage, we hesitate to speculate on exactly what changes are needed in SARTA, or SARTA-D4S to fix issues in the shortwave, although we note modifying the radiance error may help in the retrieval.

Page 16, line 32: “. . .compare well with operational AIRS-V6. . .” how well? how is it distinguished from AIRS v6? There are a number of characteristics (potential advantages and distinguishing features) of the AIRS-OE algorithm as compared to the AIRS v6 algorithm (higher spatial and vertical resolution, potentially better cloud characterization, comparable statistical results, etc.); however, these are scattered within the text and conclusions. It would be helpful to re-organize the conclusions such that these characteristics are emphasized.

We list what we think are the most useful and distinguishing features of AIRS-OE compared to AIRS-V6 in the first paragraph of Section 6, although we have cut the line about robust sensitivity as that needs more validation in more regions. Within this first effort, however, the quantitative comparisons of AIRS-OE and AIRS-V6 against radiosondes indicate that a qualitative phrase like “compares well” is reasonable. We thought we’d end the paper where the initial motivation for this effort began – a better horizontal resolution for temperature and water vapor, which when combined with cloud data, we expect to be useful in many kinds of targeted studies.

Additionally, we found a programming error that led to an undercount of successful AIRS-OE retrievals matched to sondes. This programming error has been corrected.

We thank the reviewer for perceptive comments and good questions.