We thank the anonymous reviewers for their careful reading of our manuscript and their many insightful comments and suggestions. Below we respond to the comments of each reviewer in detail, with reviewer comments in bold. We are also providing a revised manuscript that reflects their suggestions and comments. We feel that this has resulted in a stronger manuscript.

Note that below, the "Discussion" paper is the paper as published in ACPD, and the "revised" paper is the revised manuscript submitted in response to the reviewer's comments. Note also that, due to the addition of figures in the revised paper, the numbering of some figures has changed from the Discussion paper. In our response below, we use the Discussion paper numbering, and label the new figures added to the revised paper N1, N2, etc.

Anonymous Referee #1

The paper describes an analysis of spectral residuals from a set of IASI spectra from two different versions of the LBLRTM in order to establish the impact of modifications between the two versions both on spectral residuals and on retrieved profiles.

LBLRTM is an important tool in the interpretation of IASI data, and this paper attempts to address scientifically useful questions. Unfortunately, with the approach used, I do not believe the conclusions are robust enough to be publishable.

The ideal way to approach such a task is to use the 'true' atmospheric profile for each of the IASI scenes, run the LBLRTM, and compare the results with the IASI measurements. However, the authors argue that the best estimates of this 'true' profile, from ECMWF and climatologies, are insufficiently accurate for these purposes and instead modify this profile using a retrieval based on the LBLRTM itself. This has two undesirable effects: it can mask spectroscopic errors which are simply absorbed into the retrieved profiles, and also leads to complications in propagation of, particularly, temperature and water vapour errors into residuals of other species.

As detailed in our response to the reviewer's specific comments below, we have substantially revised our paper to address the reviewer's concerns, by (a) showing the average spectral residuals for both model versions when the a priori profiles are used and (b) comparing the retrieved temperature and H_2O profiles to the a priori profiles. We feel that this has resulted in a stronger manuscript without fundamentally altering the conclusions of our study, and thank the reviewer for this helpful suggestion.

In summary, our perspective is that, for the reasons discussed below, retrieving an optimal estimate of the atmospheric state allows the systematic issues in the spectroscopy to be more easily discerned, and thus our analysis focuses on the a posteriori (post-retrieval) residuals. However, we are not claiming that our retrieved states are necessarily closer to the true state than the a priori profiles in all cases. Thus we have performed a similar analysis of the a priori residuals to test the robustness of our conclusions. We find that the analysis of the a priori residuals (detailed below and in the revised paper) results in similar conclusions to the a posteriori residual analysis, although the results are less clear due to the combination of spectroscopic and atmospheric state errors in the a priori residuals. Since other readers may

share this reviewer's perspective and value the a priori analysis, we have added figures showing the a priori residuals and comparing the retrieved and a priori profiles to our revised paper.

We agree with the reviewer that the ideal way to approach an evaluation of spectroscopy with satellite data would be to use the true profile of temperature, water vapor, and all other trace gases, and if we had this true profile available, we would of course use it. However, as we noted in our Discussion paper (page 99, lines 5-12), errors in the specification of the atmospheric state are unavoidable. As discussed in detail in Shephard et al. (2009), the a priori profile, whether from in situ observations (e.g., radiosondes) or atmospheric models, is unlikely to precisely match the near-instantaneous sampling of the atmosphere within the field of view of the satellite instrument. For example, the comparisons of radiosondes launched from the ARM Southern Great Plains site an hour apart showed significant variability in temperature (0.5-1 K over 1 km averaged layers) and water vapor (25% for 2 km thick layers, Tobin et al., 2006). This collocation error is in addition to any systematic biases that might be present in the atmospheric model or in situ observations. Thus for a closure study like this one we must use some reasonable estimate of the true state, with the understanding that this is merely an estimate, and thus there is no way to completely separate spectroscopic errors from errors in the specification of the atmospheric state.

We thus chose to use retrievals to adjust (and hopefully improve) our specification of the atmospheric state prior to analysis of the residuals, as done in the previous ACP paper of Shephard et al. (2009) and in several other studies noted therein. Using the observed spectra to provide an optimal estimate of the atmospheric state can potentially remove the impact of collocation errors and systematic in situ or model biases on the residuals, allowing the systematic issues in the spectroscopy to be more easily discerned. In addition, one of the primary goals of our study was to assess the consistency of the spectroscopy between different spectral regions, both within the same band (i.e., the P- and R-branches of the H_2O ν_2 and O_3 ν_3 bands) and between bands (i.e., the CO_2 ν_2 and ν_3 bands). Using the a priori profiles as our best estimate of the atmospheric state makes it difficult to assess this consistency, as both spectral regions can have substantial residuals due to the incorrect specification of the atmospheric state (see our discussion of the a priori residuals below). In contrast, using retrieved profiles, as in our study, can make clear where substantial inconsistencies remain between the spectral regions. We have added some of this discussion to our revised paper (see page 8, lines 8-20).

We do agree with the reviewer's comment below that our procedure of using the retrieved profiles to specify our atmospheric state does complicate the analysis somewhat, in that both the retrieved profiles and the spectral residuals must be analyzed. We are also not claiming that our retrieved states are necessarily closer to the true state than the a priori profiles in all cases. Our analysis of the CH_4 band in Section 5.5 is instructive here, as we recognize that the CH_4 spectroscopy in both model versions is poor, impacting not only to the spectral residuals remaining after retrieval, but also leading to the retrieval of unphysically large tropospheric mixing ratios of CH_4. However, we feel the advantages in terms of reducing the impact of collocation errors and any systematic in situ or model biases and allowing the assessment of the consistency of the spectroscopy across bands outweighs the complication of the analysis.

We feel that we can address the reviewer's concerns by presenting the residuals from the use of both the a posteriori (retrieved) profiles and the a priori profiles in our revised paper and

providing further analysis of the retrieved temperature and water vapor profiles, as discussed below.

Tobin, D. C., Revercomb, H. E., Knuteson, R. O., Lesht, B. M., Strow, L. L. Hannon, S. E., Feltz, W. F., Moy, L. A., Fetzer, E. J., and Cress, T. S.: Atmospheric Radiation Measurement site atmospheric state best estimates for Atmospheric Infrared Sounder temperature and water vapor retrieval validation, J. Geophys. Res., 111, D09S14, doi:10.1029/2005JD006103, 2006.

SPECIFIC COMMENTS

My main criticism, then, is that results are not presented on the basis of residuals using the a priori profile. This would have provided a simple, and direct, comparison of the performance of the two models. If there are uncertainties in the true profile, these should be characterised and propagated through the forward model as uncertainties associated with the residuals. If there are clear deficiencies in the a priori, then these should be carefully identified and adjusted, and examined a posteriori to verify that the adjustments are reasonable.

Here we do not see what impact the retrievals have had on the a priori estimate, or if the resulting profiles are physically reasonable (not, for example, something with wide oscillations or unphysical surface-lower atmosphere temperature discontinuities). And, for the analysis of residuals, there should still be a single true profile for each scene, not one which depends on the model. However, by using the two different model versions, two different versions of the 'true profile' are created which appear to differ significantly from each other (but I could see no example of how they differ from the a priori). So, the spectral residuals analysed here are some complicated combination of spectroscopic errors, forward model errors and retrieval artefacts. The paper itself becomes vastly more complicated seems necessary, just from attempting to untangle these effects.

To address these concerns, we have added the mean a priori residuals for both model versions to Figures 2, 3, 7, 12, 22, 23, and 25 (using the numbering in the Discussion paper) in our revised paper. The figures are reproduced here for convenience, along with a brief discussion of how these a priori residuals relate to our conclusions. We have also examined how the retrieved temperature and water vapor profiles compare with the a priori profiles in new figures (labeled Figure N1 and Figure N3 below, and added to the revised paper as Figures 7 and 20) discussed below. In addition, we have added an examination of the a priori and a posteriori residuals in the H_2O band versus brightness temperature in a new figure in the revised paper (labeled Figure N2 below, and added to the revised paper as Figure 17), which helps to highlight the improvements in the P-branch residuals.

Full Spectrum
Figure 2 below shows that the RMS of the a priori residuals over the entire IASI spectral range is lower in v12.1 than in v9.4+, consistent with the a posteriori residuals (as noted on page 11, line 24 of the revised paper).

CO_2 bands
Figures 3 and 7 show that in both the CO_2 ν_2 and ν_3 bands the RMS of the a priori residuals is

lower in v12.1, consistent with our conclusion that the CO_2 spectroscopy is improved in v12.1. (Note the different scales used for the a priori and a posteriori residuals.) Thus, we conclude that, regardless of whether one uses the a posteriori or a priori specification of the atmospheric state, it is clear that the updated CO_2 spectroscopy in v12.1 is an improvement. We also note that it would be very difficult to evaluate the consistency of the spectroscopy in these two regions using the a priori residuals in these figures, whereas the improved consistency is obvious in the a posteriori residuals.

We also compared the retrieved temperature profiles to the a priori profiles, as suggested by the reviewer. Figure N1(a) below shows the difference between the a posteriori (after all retrieval steps, and thus corresponding to the final residuals in Figures 3 and 7) and a priori temperature profiles for LBLRTM v12.1. The thin red, green, and blue lines are the differences for the individual scans, and correspond to a PWV of 0-2, 2-4, and 4-6 cm, respectively. The thick black line shows the mean difference from the a priori over all 120 scans, with the black dashed lines showing the 25th and 75th percentiles. Figure N1(b) shows the same comparison for LBLRTM v9.4+, while N1(c) shows the RMS differences from the a priori at the retrieval levels for both v12.1 (black) and v9.4+ (red). The RMS differences from the a priori temperature profiles are indeed lower in v12.1 than in v9.4+ (with the slight exception of the 16.16 and 14.68 hPa levels), as are the RMS differences of the surface (skin) temperature (0.61 K in v12.1 versus 0.69 K in v9.4+). The average of the retrieved temperature profiles with the improved CO_2 spectroscopy in LBLRTM v12.1 is thus more consistent with the forecast average, which draws on information from a wide range of sources, and which is expected to be a reasonably accurate representation of the mean atmospheric state over the spatial and temporal range of the data.

In our revised paper, we have extensively revised Section 5.1 to discuss the a priori residuals and the comparison of the retrieved and a priori profiles.

H_2O band

Figure 12 shows that the RMS of the a priori residuals in the H_2O ν_2 band is larger in v12.1 than in v9.4+ for both the P- and R-branch. However, in the P-branch the standard deviation about the mean for the a priori residuals is reduced from 0.349 K for v9.4+ to 0.326 K for v12.1. This suggests that the residuals within the P-branch are more consistent with each other in v12.1. In addition, Figure N2 shows the residuals in the H_2O band versus the measured brightness temperature for our moderate PWV case, with the left column showing the a priori residuals and the right the a posteriori. Figure N2a and N2b show the results for the P-branch using v12.1, while Figure N2c and N2d show the P-branch residuals using v9.4+. It is clear that there is much less spread in the P-branch channels with brightness temperatures less than ~255 K (i.e., for the stronger H_2O lines) in v12.1 than there was in v9.4+, whether you use the a priori or a posteriori results. This is consistent with the work of Coudert et al. (2008), which generally increased the intensity of the strongest lines (in both the P- and R-branches) by up to 7%. In addition, the a posteriori P-branch residuals using v12.1 (Fig. N2b) show no correlation with the brightness temperature of the channel and a slope of ~0 (within the uncertainty), suggesting the slope with brightness temperature seen in the a priori residuals in Figure N2a can be removed by the retrieval procedure. Thus, we stand by our original conclusion that the a posteriori residuals in the P-branch have improved due to a combination of the CO_2 and H_2O spectroscopy improvements between v9.4+ and v12.1.

In the R-branch, the standard deviation about the mean for the a priori residuals increases from 0.473 K in v9.4+ to 0.507 K in v12.1. However, as noted in the Discussion paper, the updated spectroscopy does reduce some of the systematic residual features seen in the R-branch using v9.4+. This can be seen by comparing the a priori residuals versus brightness temperature from v12.1 (Fig. N2e) with those from v9.4+ (Fig. N2g). While the difference is not as dramatic as in the P-branch, some of the outliers at brightness temperatures less than ~255 K show reduced a priori residuals in v12.1 compared to v9.4+. Again, this is consistent with the work of Coudert et al. (2008). The a posteriori R-branch residuals in the right column are a little more "flat" (i.e., uncorrelated with brightness temperature) in v12.1 than in v9.4+, but not dramatically so. Thus we stand by our conclusion that the improvement in the R-branch a posteriori residuals is more modest than in the P-branch, and appears more heavily influenced by the improvements in the temperature retrieval rather than improvements in the H_2O spectroscopy alone.

We also compared the retrieved H_2O profiles to the a priori profiles, as suggested by the reviewer. Figure N3(a) below shows the ratio (as a percent) of the a posteriori (after all retrieval steps) and a priori H_2O profiles for LBLRTM v12.1. Figure N3(b) shows the same results for v9.4+, and Figure N3(c) shows the RMS differences between the retrieved and a priori H_2O profiles for both models. At altitudes below the 600 hPa surface, LBLRTM v12.1 has a much smaller RMS difference from the a priori than v9.4+, while the two models have similar RMS values at higher altitudes. The retrieved H_2O profiles with the improved CO_2 and H_2O spectroscopy in LBLRTM v12.1 are thus more consistent with the forecast.

In our revised paper, we have extensively revised Section 5.2 to discuss the a priori residuals and the comparison of the retrieved and a priori profiles.

Other bands

Figure 22 below shows the a priori residuals for the O_3 band. As we noted in the Discussion paper, the O_3 spectroscopy in this region is substantially the same for both model versions. However, both the H_2O and CO_2 spectroscopy have changed in this region, leading to slight differences in the a priori residuals between v12.1 and v9.4+. We note that the inconsistency of the spectroscopy between the P- and R-branches of the band is evident in both the a priori and a posteriori residuals. We thus conclude that there is no substantial difference between the model versions in this region, with both showing inconsistencies in the spectroscopy across the O_3 band, consistent with the conclusions of our Discussion paper.

Figure 23 below shows the a priori residuals for the CO band. Here, the dangers of using the a priori profiles as the “true” profiles are evident. Based on the RMS of the a priori residuals, one would be tempted to conclude that LBLRTM v12.1 performs better in this region. However, after the retrievals are performed, it becomes obvious that the spectroscopy in this region is worse in v12.1 than in v9.4+ due to the increase in the water vapor self continuum in this region, as concluded in our Discussion paper. Another way of saying the same thing is that the water vapor profile retrieved using v12.1 is inconsistent with the water vapor self continuum in the CO band, whereas the two were fairly consistent in v9.4+.

Finally, Figure 25 below shows the a priori residuals for the CH_4 band. Consistent with the a posteriori residuals, the RMS of the a priori residuals is higher in LBLRTM v12.1 than in v9.4+, and we are thus unable to conclude that the updated CH_4 spectroscopy is a clear improvement. Analysis of the CH_4 profiles retrieved using both models show unphysically high values, as noted in the Discussion paper (see Page 104, line 29). However, this positive bias was lessened when LBLRTM v12.1 was used, suggesting the updated CH_4 spectroscopy could potentially improve operational CH_4 retrievals, as noted in our Discussion paper. In fact, we have done some preliminary work with the Tropospheric Emission Spectrometer (TES) retrievals of CH_4 that show that the new CH_4 spectroscopy in LBLRTM v12.1 reduces the positive bias in TES CH_4 retrievals compared with aircraft observations from +40 ppb to +23 ppb while generally improving the spectral residuals between 1190 and 1320 cm^{-1} (Alvarado et al., 2013).

In our revised paper, we have revised Sections 5.3-5.5 to discuss the a priori residuals.

Alvarado, M. J., K. E. Cady-Pereira, V. H. Payne, S. S. Kulawik, J. R. Worden, and K. J. Wecht (2013), Evaluation of Recent Spectroscopic Updates and Their Impacts on TES CH_4 and CO_2 Retrievals, presented at the TES Science Team Meeting, Cambridge, MA, USA, 18-19 March

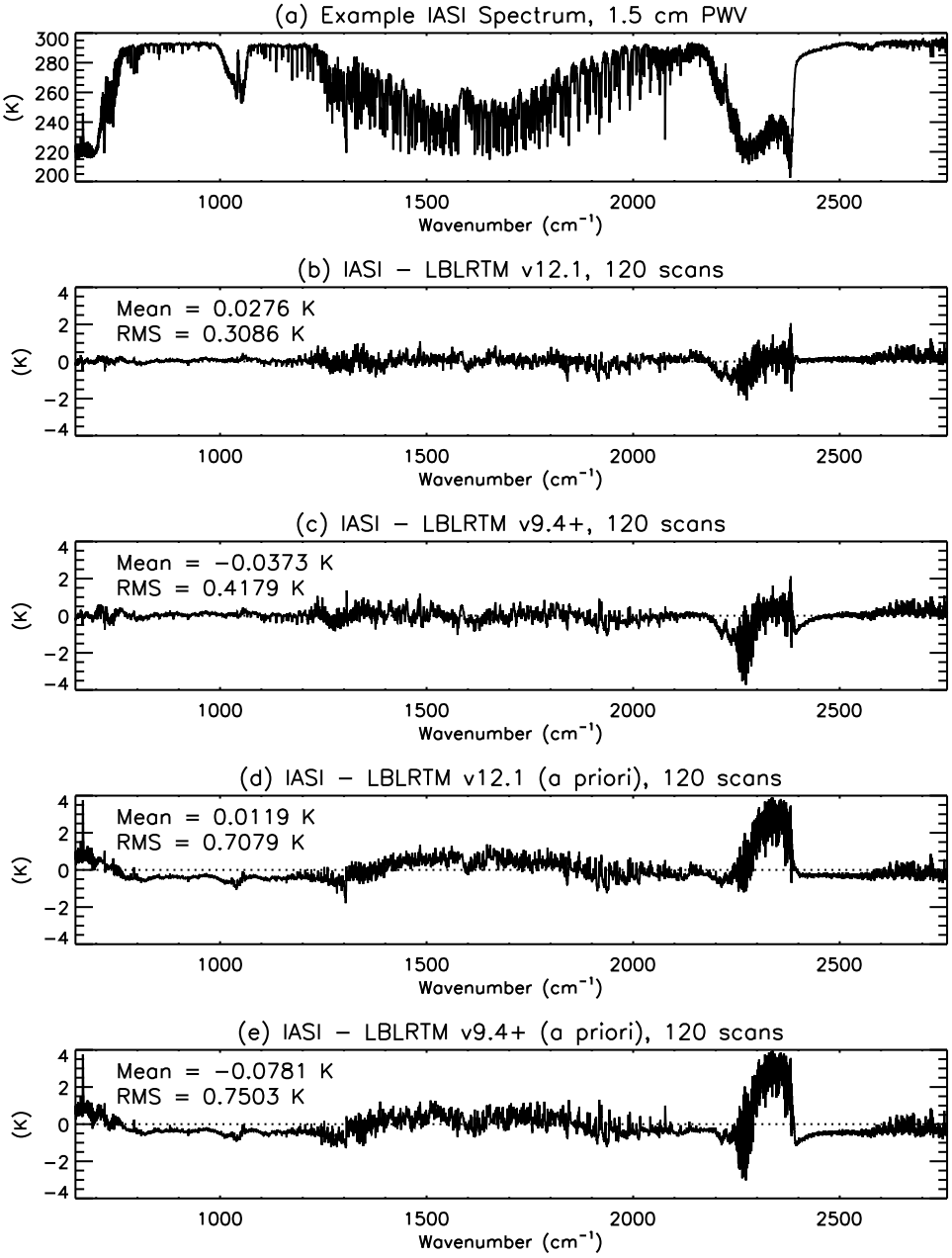


Figure 2. (a) IASI observed brightness temperature spectrum for an example profile with 1.5 cm PWV. (b) Mean of the final brightness temperature residuals for 120 spectra using LBLRTM v12.1. (c) Mean of the final brightness temperature residuals for 120 spectra using LBLRTM v9.4+. (d) Mean of the a priori brightness temperature residuals for 120 spectra using LBLRTM v12.1. (e) Mean of the final brightness temperature residuals for 120 spectra using LBLRTM v9.4+. Note that the RMS due to instrument noise alone is 0.17 K.

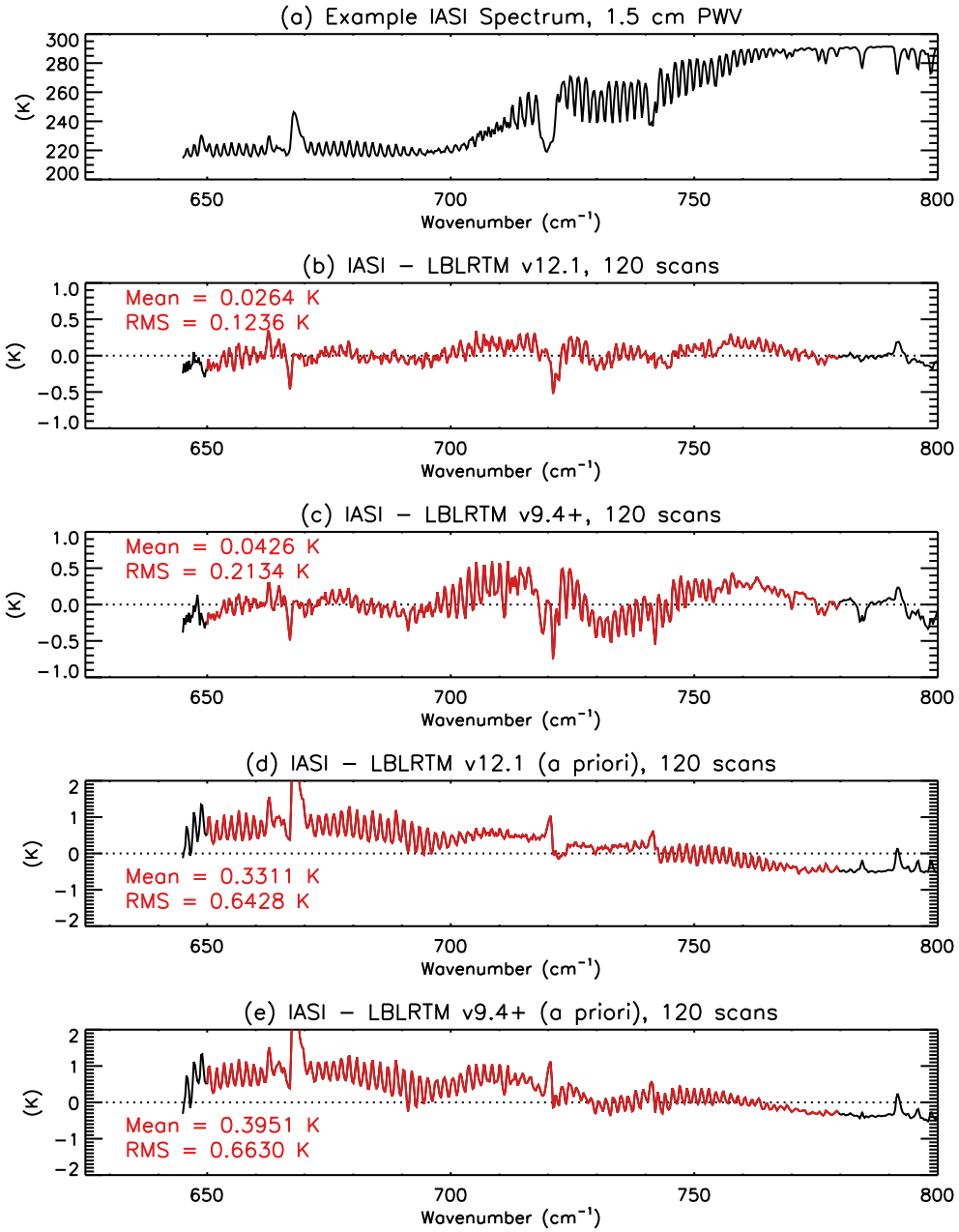


Figure 3. (a) IASI observed brightness temperature spectrum in the ν_2 band of CO_2 for an example profile with 1.5 cm PWV. (b) Mean of the final (a posteriori) brightness temperature residuals for 120 spectra using LBLRTM v12.1. (c) Mean of the final brightness temperature residuals for 120 spectra using LBLRTM v9.4+. (d) Mean of the a priori brightness temperature residuals for 120 spectra using LBLRTM v12.1. (e) Mean of the a priori brightness temperature residuals for 120 spectra using LBLRTM v9.4+. A priori residuals in the 667 cm^{-1} Q-branch peak at approximately 4 K. The ν_2 retrieval window (which was used to retrieve T_{atm} and T_{srf}) is highlighted in red. Note that the RMS due to instrument noise alone in this retrieval window is 0.045 K.

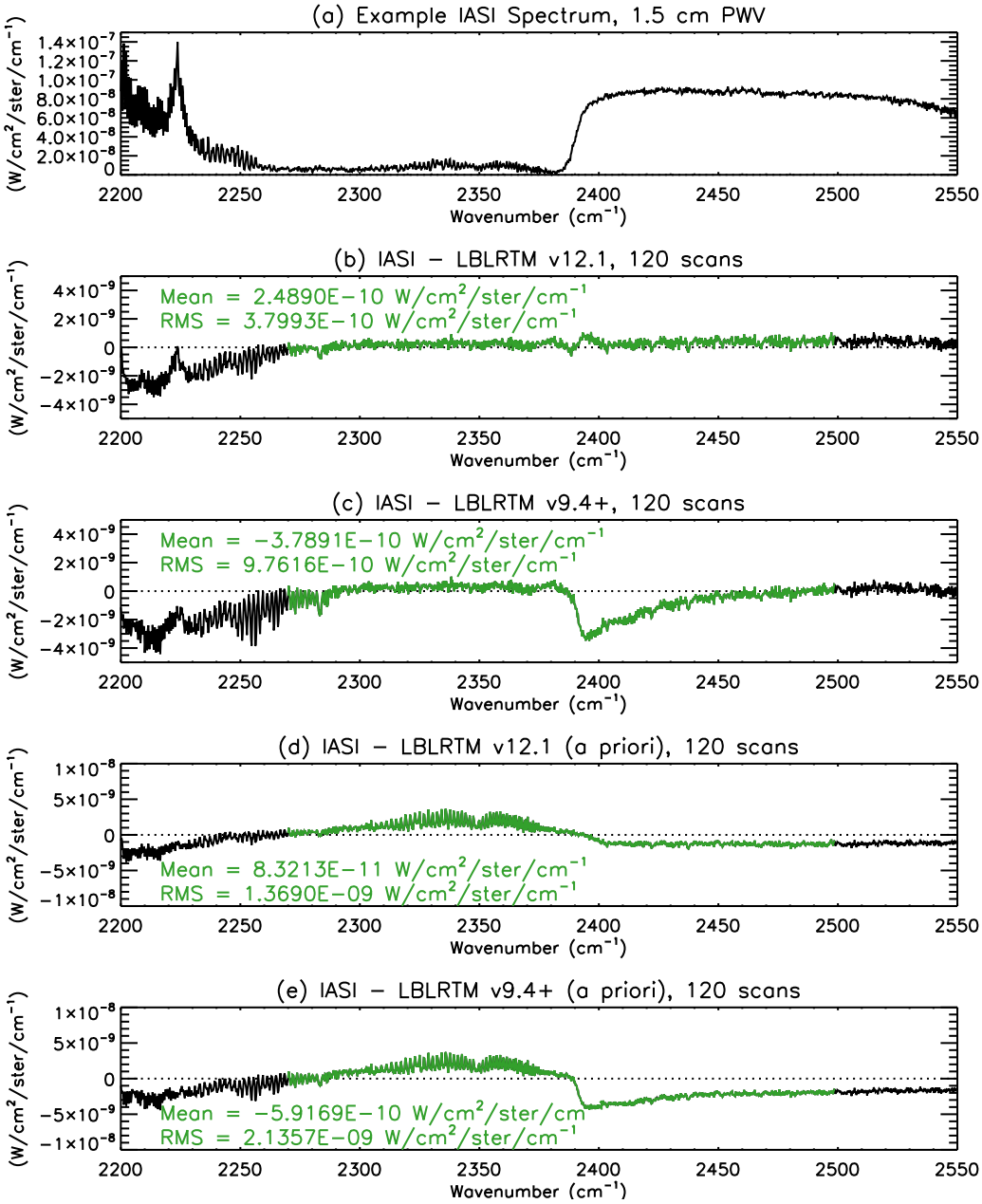


Figure 7. (a) IASI observed radiance spectrum in the ν_3 band of CO_2 for an example profile with 1.5 cm PWV. (b) Mean of the final (a posteriori) radiance residuals for 120 spectra using LBLRTM v12.1. (c) Mean of the final radiance residuals for 120 spectra using LBLRTM v9.4+. (d) Mean of the a priori radiance residuals for 120 spectra using LBLRTM v12.1. (e) Mean of the a priori radiance residuals for 120 spectra using LBLRTM v9.4+. The CO_2 ν_3 band, which was not used in the temperature retrievals, is highlighted in green. Note that the RMS due to instrument noise alone in this band is 2.4×10^{-10} W/cm^2/ster/cm^{-1}.

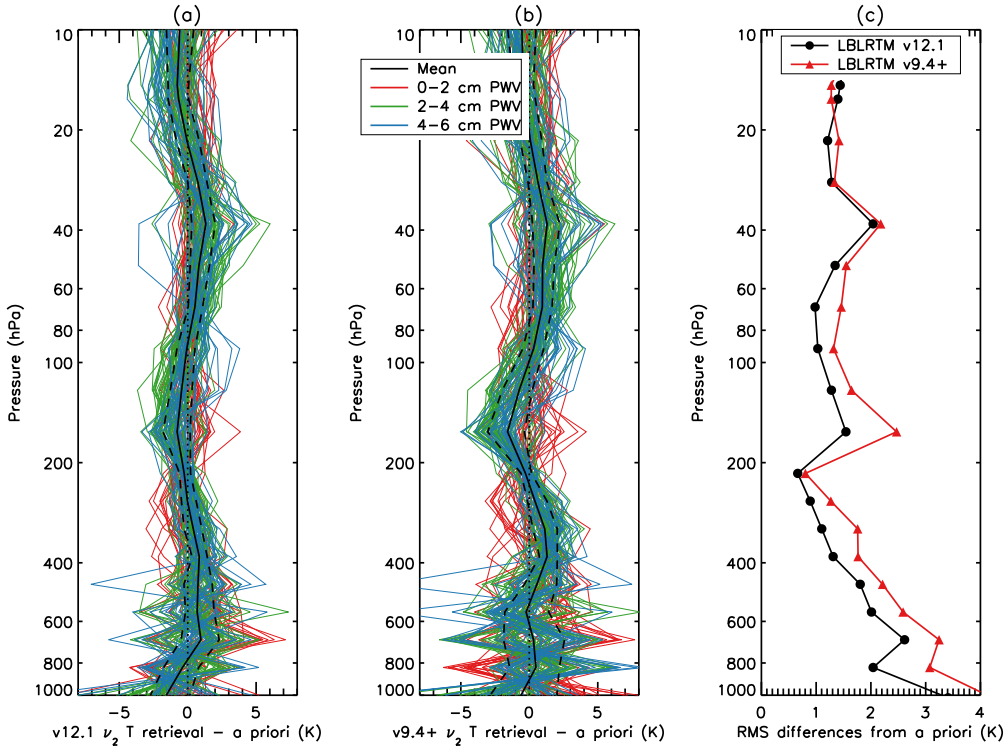


Figure N1: (a) Differences between the final (a posteriori) temperature profile retrieved using LBLRTM v12.1 and the a priori temperature profiles. The thin red, green, and blue lines are the differences for the individual scans, and correspond to a PWV of 0-2, 2-4, and 4-6 cm, respectively. The thick black line shows the mean difference from the a priori over all 120 scans, with the black dashed lines showing the 25th and 75th percentiles. (b) As in (a), but for LBLRTM v9.4+. (c) RMS of the differences between the a posteriori and a priori temperature profiles for v12.1 (black circles) and v9.4+ (red triangles). The RMS difference of the surface (skin) temperature from the a priori is 0.61K for v12.1 and 0.69 K for v9.4+.

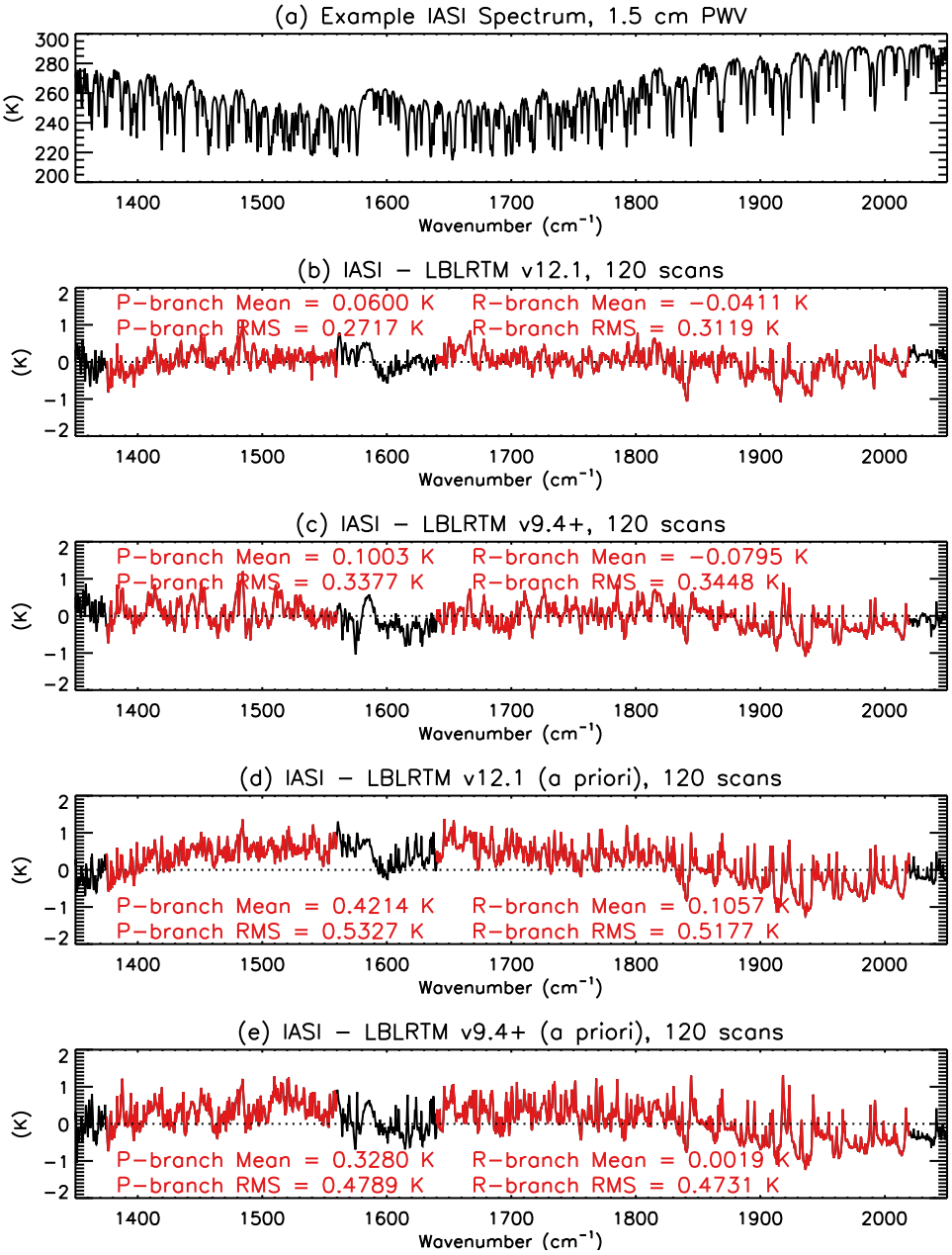


Figure 12. (a) IASI observed brightness temperature spectrum in the ν_2 band of H_2O for an example profile with 1.5 cm PWV. (b) Mean of the final (a posteriori) brightness temperature residuals for 120 spectra using LBLRTM v12.1. (c) Mean of the final brightness temperature residuals for 120 spectra using LBLRTM v9.4+. (d) Mean of the a priori brightness temperature residuals for 120 spectra using LBLRTM v12.1. (e) Mean of the a priori brightness temperature residuals for 120 spectra using LBLRTM v9.4+. The P- and R-branch H_2O retrieval windows are highlighted in red. Note that the RMS due to instrument noise alone in the P-branch is 0.041 K and in the R-branch is 0.088 K.

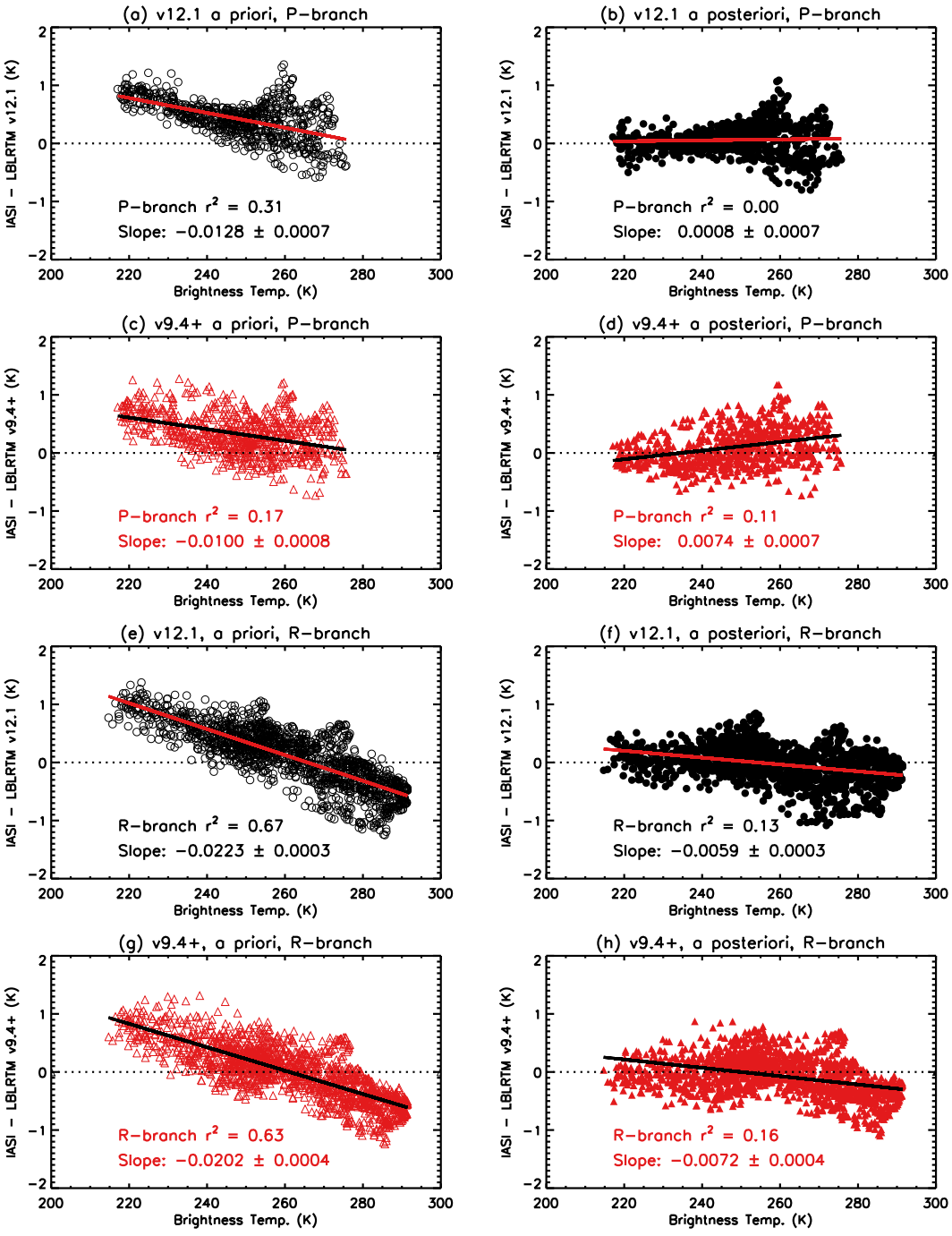


Figure N2. A priori (left column, open symbols) and a posteriori (right column, closed symbols) mean residuals of 120 scans in the P- and R-branches of the H_2O ν_2 band versus observed brightness temperature for an example spectrum with 1.5 cm PWV. Black circles are for residuals from v12.1, while red triangles are for residuals from v9.4+.

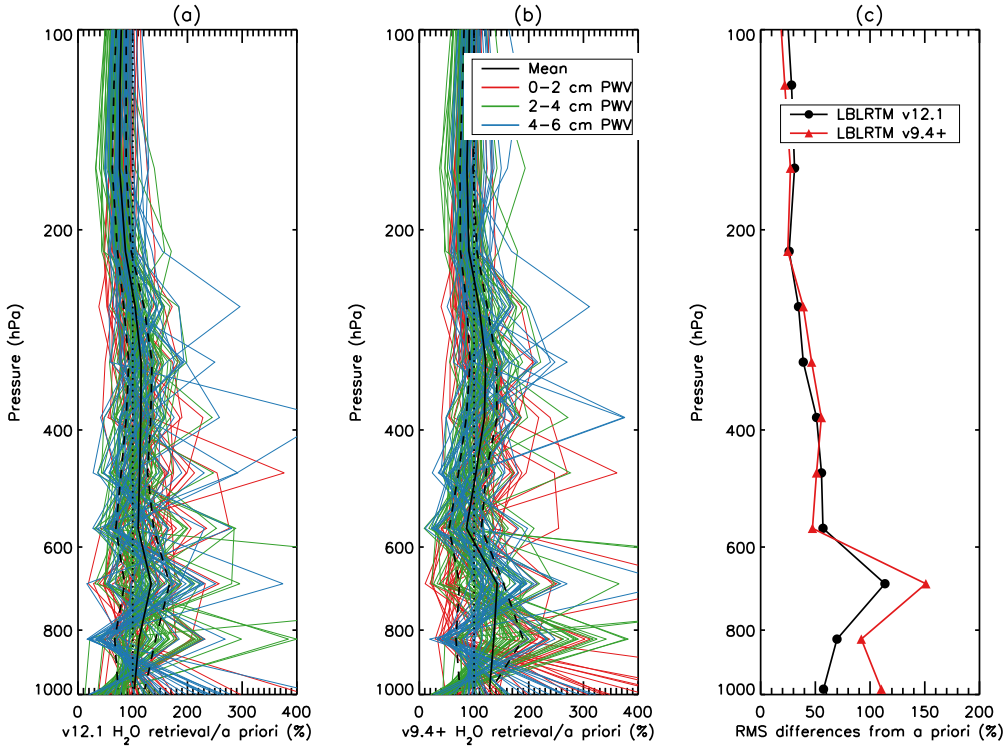


Figure N3: (a) Ratio of the final a posteriori (retrieved) H_2O profile using LBLRTM v12.1 to the a priori temperature profiles, as a percent. The thin red, green, and blue lines are the differences for the individual scans, and correspond to a PWV of 0-2, 2-4, and 4-6 cm, respectively. The thick black line shows the mean ratio to the a priori over all 120 scans, with the black dashed lines showing the 25th and 75th percentiles. (b) As in (a), but for LBLRTM v9.4+. (c) RMS of the percentage differences between the a posteriori and a priori H_2O profiles for v12.1 (black circles) and v9.4+ (red triangles).

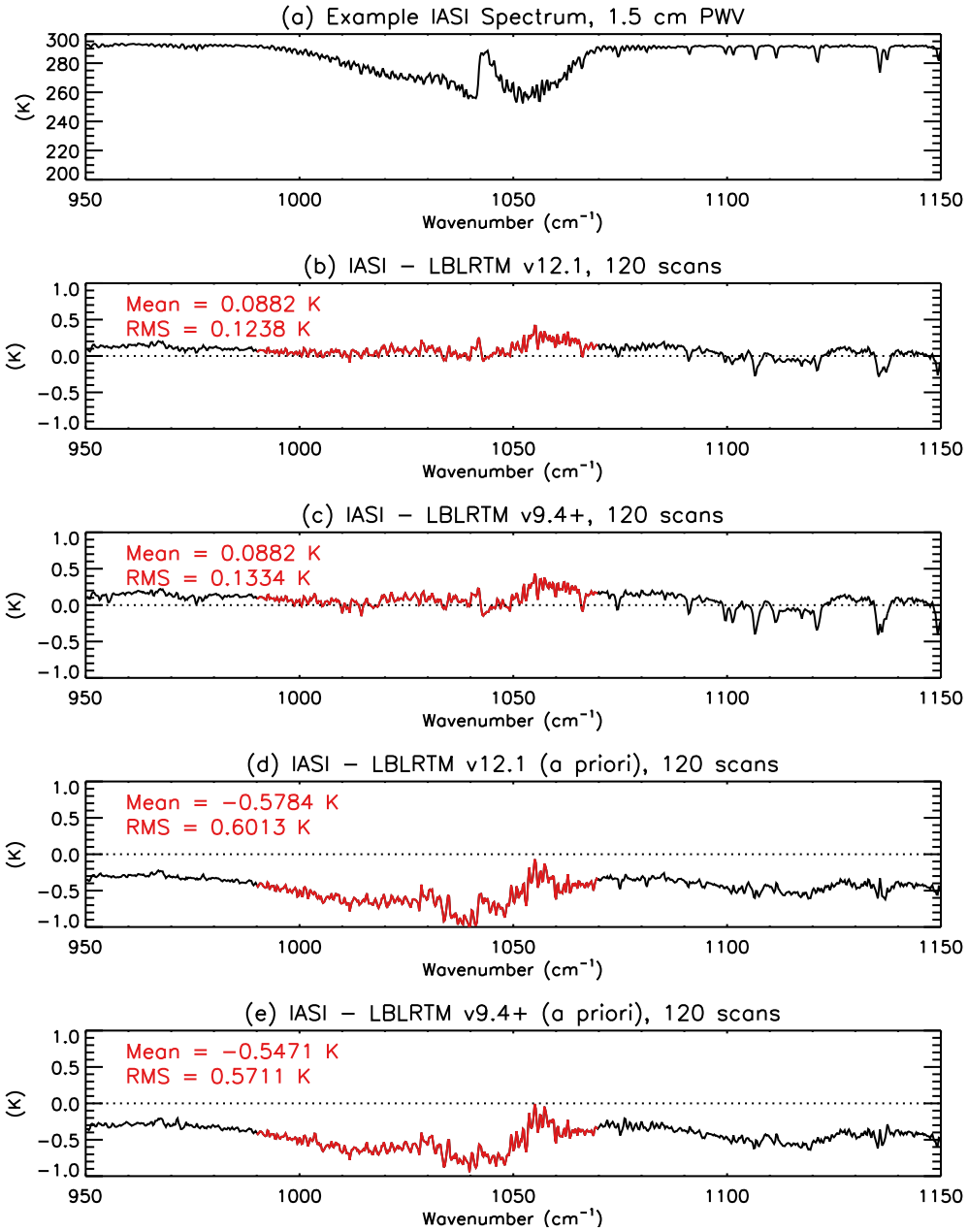


Figure 22. (a) IASI observed brightness temperature spectrum in the ν_3 band of O_3 for an example profile with 1.5 cm PWV. (b) Mean of the final brightness temperature residuals for 120 spectra using LBLRTM v12.1. (c) Mean of the final brightness temperature residuals for 120 spectra using LBLRTM v9.4+. (d) Mean of the a priori brightness temperature residuals for 120 spectra using LBLRTM v12.1. (e) Mean of the a priori brightness temperature residuals for 120 spectra using LBLRTM v9.4+. The O_3 retrieval window is highlighted in red. Note that the RMS due to instrument noise alone in this retrieval window is 0.027 K.

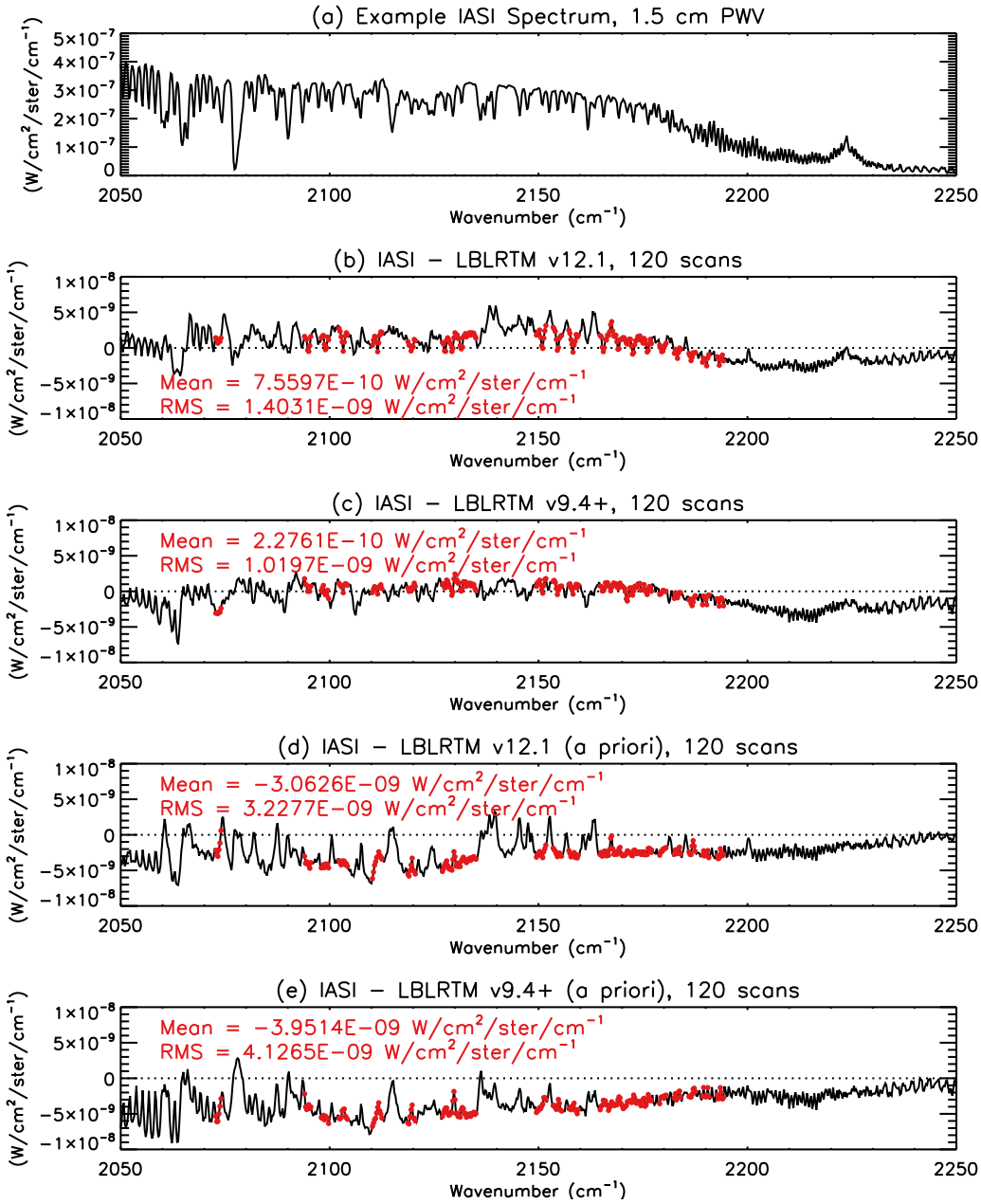


Figure 23. (a) IASI observed radiance spectrum in the fundamental vibrational band of CO for an example profile with 1.5 cm PWV. (b) Mean of the final (a posteriori) radiance residuals for 120 spectra using LBLRTM v12.1. (c) Mean of the final radiance residuals for 120 spectra using LBLRTM v9.4+. (d) Mean of the a priori radiance residuals for 120 spectra using LBLRTM v12.1. (e) Mean of the a priori radiance residuals for 120 spectra using LBLRTM v9.4+. The CO retrieval windows are highlighted in red. Note that the RMS due to instrument noise alone in these windows is 1.8×10^{-10} W/cm^2/ster/cm^{-1}.

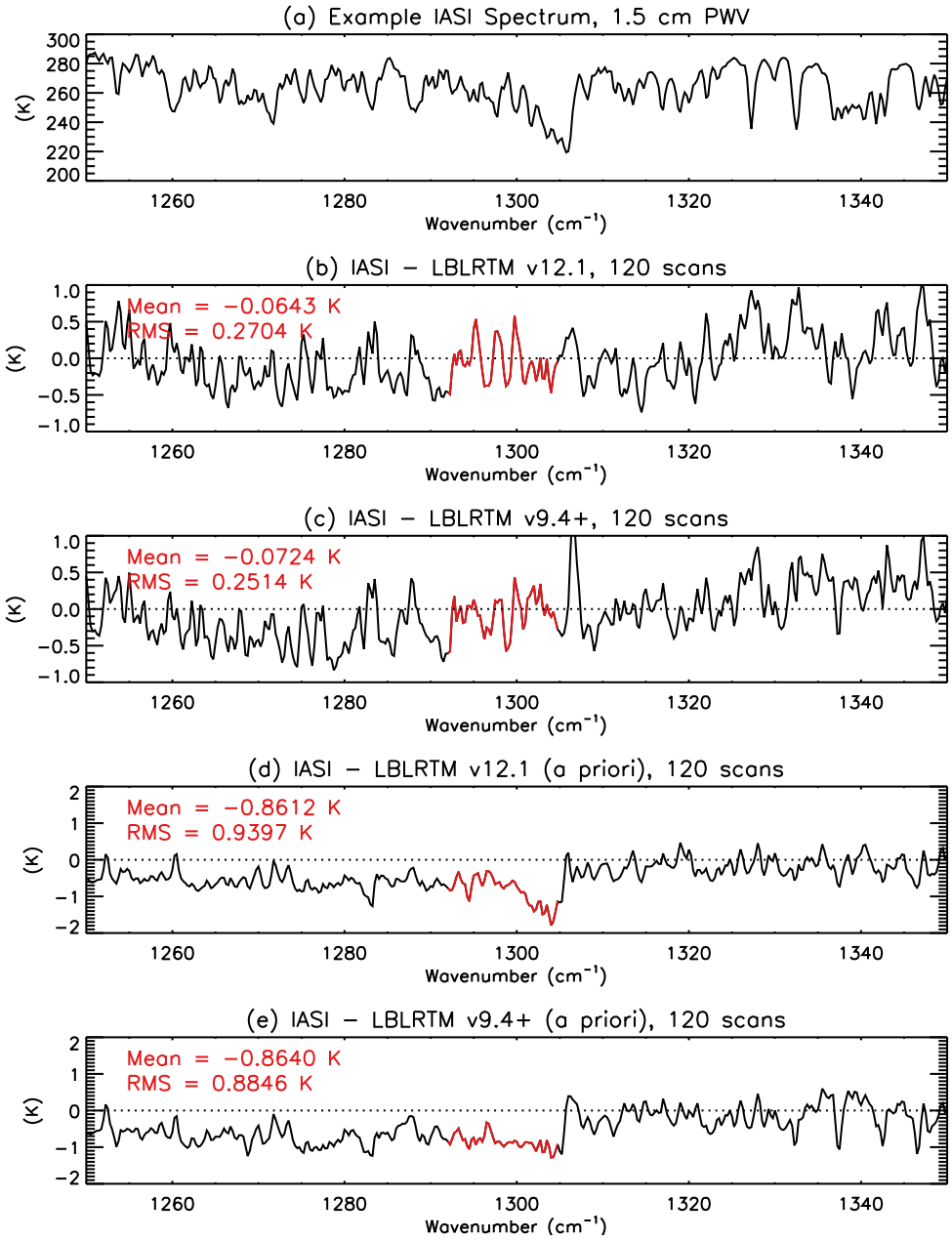


Figure 25. (a) IASI observed brightness temperature spectrum in the ν_4 band of CH_4 for an example profile with 1.5 cm PWV. (b) Mean of the final (a posteriori) brightness temperature residuals for 120 spectra using LBLRTM v12.1. (c) Mean of the final brightness temperature residuals for 120 spectra using LBLRTM v9.4+. (d) Mean of the a priori brightness temperature residuals for 120 spectra using LBLRTM v12.1. (e) Mean of the a priori brightness temperature residuals for 120 spectra using LBLRTM v9.4+. The CH_4 retrieval window is highlighted in red. Note that the RMS due to instrument noise alone in this retrieval window is 0.016 K.

If the resulting residuals are clearly (ie visually) improved, then one might overlook this. But in most cases the improvements can only be identified by examining the RMS or bias (incidentally, unclear which is preferred, and why).

We feel the RMS is generally the correct metric to look at as (a) the RMS includes contributions from both the mean bias and the standard deviation about that mean and (b) the retrievals explicitly minimize the mean bias. The exceptions are when different spectral regions used in the same retrieval are compared (as in the comparison of the H_2O P- and R-branches, or different regions within the R-branch in Section 5.2), as differences in the mean bias between these sub-regions can be a useful indicator of inconsistent spectroscopy within a band. We have added this discussion to our revised paper (see page 11, lines 15-21) and thank the reviewer for noting the oversight.

So, of the various conclusions, I am inclined to believe that the CO2 modelling in v12.1 is better than v9.4 - although even here it is not clear whether this is due to the underlying spectroscopic data or the inclusion of P,R line-mixing in v12.1. However, nothing else seems to point unambiguously to v12.1 being more accurate than v9.4.

We agree with the reviewer that there are many spectral regions where the spectroscopy in LBLRTM v12.1 is not a clear improvement over v9.4+. We have corrected our discussion of Figure 2 in the Discussion paper, which may have misled readers into believing the spectroscopy in v12.1 is improved throughout the IASI spectrum (see page 11 lines 22-24 in the revised paper). We think it is important to note that our Discussion paper only concludes that the spectroscopy in v12.1 is clearly more accurate than v9.4+ in three cases, i.e. the CO_2 spectroscopy in the CO_2 ν_2 band, the spectroscopy in the CO_2 ν_3 band region (including the bandhead), and the H_2O spectroscopy in the P-branch of the H_2O ν_2 band (see page 105, lines 12-19 and page 106, lines 7-9 of the Discussion paper). We state that the improvement in the a posteriori residuals in the P-branch of the H_2O ν_2 band are from a combination of improvements in the temperature retrieval and the improvements in the H_2O spectroscopy (page 106, lines 7-10). This conclusion is reinforced by our further analysis of the a priori and a posteriori residuals in Figure N2 above, which shows clear improvements for strong lines in the P-branch, consistent with the use of the Coudert et al. (2008) H_2O line intensities in v12.1. In contrast, we state in the Discussion paper that that the improvements are more modest in the R-branch, and are more heavily influenced by the improvements in the temperature retrieval rather than improvements in the H_2O spectroscopy (page 106, lines 10-13). We state that v9.4+ and v12.1 are nearly identical in the O_3 band, with significant average residuals remaining in the R-branch, showing the need for further work on the O_3 spectroscopy in this region (page 107, lines 4-6). We state that the changes in the water vapor self continuum have degraded the model performance in the CO fundamental band, which could lead to significant errors if LBLRTM v12.1 is used of CO retrievals (page 107, lines 7-11). Finally, we state that significant residuals remain in the CH_4 band even with the updated spectroscopy, leading to unphysically large retrieved values (page 107, lines 13-16). However, we also note that this positive bias in retrieved CH_4 is significantly lower in v12.1 than v9.4+, suggesting that the CH_4 spectroscopy, while still poor, may be a moderate improvement, and deserves further investigation. We feel that these conclusions, as stated in the Discussion paper, are thus a fair reflection of the results of our study, and that we have clearly noted the areas

where the updated spectroscopy in v12.1 is improved from, is worse than, or is indistinguishable from v9.4+.

Given that the reviewer is inclined to agree with us on the improvement of the CO_2 spectroscopy, our only disagreement seems to be about the H_2O spectroscopy in the P-branch of the H_2O ν_2 band. We hope Figure N2 and the accompanying discussion above have convinced the reviewer that the H_2O P-branch spectroscopy is a clear improvement over v9.4+.

We discuss the impact of CO_2 P- and R-branch line mixing further below in response to another comment.

I would also be wary of assuming statements on the impact of these changes on temperature and water vapour retrievals are generally applicable. Most assimilation or operational retrieval schemes do not use the complete spectral range as used here but just a small selection of channels. For these, the retrieval characteristics, hence impact of LBLRTM differences, may be quite different.

We agree with the reviewer that the precise impacts of the updated spectroscopy on operational assimilation and retrieval schemes will depend on the other parameters included in the retrieval, including but not limited to differences in channel selection, a priori profiles and constraint matrices, and convergence criteria. Since evaluating the impacts on all current operational assimilation schemes and retrievals is beyond the scope of this paper, we have instead opted to evaluate the impact on retrievals that use large spectral ranges. This follows from our emphasis on evaluating the consistency of the spectroscopy within and between bands. We recognize, however, that our results can only be suggestive of the impacts on operational schemes, and that the impact on any specific retrieval or assimilation method would have to be carefully evaluated on its own. We have added this caveat to our revised paper (see page 10, lines 2-5) and thank the reviewer for pointing out its importance.

Other comments:

HDO Depletion The argument that HDO depletion is the cause of some of the spectral residual differences seems entirely plausible but, following the rest of this paper, I would have expected the argument to be made by demonstrating that a reduced HDO concentration results in a reduced RMS. Instead, it is argued that the HDO depletion causes a bias, but since I assume the retrieval attempts to minimise the RMS of the spectral residual it is not obvious to me that the result would be a bias rather than an enhanced RMS - it is, after all, just equivalent to another spectroscopic error and the argument elsewhere is that these increase the RMS difference.

We first note that using an improved a priori HDO profile does reduce the RMS within the H_2O P-branch very slightly, from 0.362 K to 0.361 K, as shown in Figure 15 of the Discussion paper. The small effect on the total RMS for the P-branch is expected as only a small number of channels are sensitive to the HDO profile (i.e., have HDO optical depths > 1.0).

The reviewer is also correct that the retrieval attempts to minimize the RMS and mean bias *across*

all channels included in the retrieval. However, that does not mean that *subsets of those channels* cannot show biases relative to each other. What Figure 15 shows is that channels that are significantly impacted by HDO have higher residuals on average than channels that are not impacted by HDO – in other words, the HDO impacted channels are "biased" relative to the other channels in the P-branch. Using an improved, vertically varying HDO profile appears to remove this discrepancy between the HDO-sensitive channels and the insensitive channels, as seen in Figure 15. Also, the use of the improved HDO profile reduces the positive mean bias of the P-branch *relative to the R-branch* (see Figure 12). We have revised our discussion of Figure 15 in our revised paper (see page 18, lines 20-26) to make this clearer.

H2O R-band If HITRAN data produces a larger bias while keeping the RMS unchanged, doesn't this mean that the scatter (SD) about the mean must also have been reduced with HITRAN? Otherwise you would expect RMSˆ2 = SDˆ2 + Biasˆ2 In fact, coupled with the results of Table 4 I would suggest that the HITRAN data for the R branch is better than using AER, ie that v9.4 is better here than v12.1.

Referring to Figure 16 of the Discussion paper, the printed RMS is only for the "region of disagreement," that is between 1750 and 2020 cm^{-1}. We agree with the reviewer that the similar RMS coupled with the higher bias for HITRAN 2008 in this region implies that the HITRAN H_2O spectroscopy data is more self-consistent than AER v3.1 within the range 1750-2020 cm^{-1}. However, our concern is with the discontinuity between the regions on either side of 1750 cm^{-1} in HITRAN 2008. This can be seen by examining difference in the mean bias in different subregions of the band. For example, the difference in mean bias between the P- and R-branches is higher in HITRAN 2008 than in AER v3.1: the differences are 0.02, 0.12, and 0.10 K for the low, medium, and high water vapor cases (respectively) when AER v3.1 is used, and 0.15, 0.25, and 0.24 K when HITRAN 2008 is used (see Table 3 in the Discussion paper), suggesting that the P- and R-branches are less consistent in HITRAN 2008 than in AER v3.1. Furthermore, when the mean bias in the R-branch as a whole is compared with the "region of disagreement," AER v3.1 is again more consistent, with differences in mean bias of -0.03, 0.06, and 0.08 K versus 0.01, 0.10, and 0.12 K for HITRAN 2008. Thus, the H_2O spectroscopy in AER v3.1 is more consistent across the R-branch and between the P- and R-branches than HITRAN 2008, with the caveat that HITRAN 2008 is more self-consistent within the 1750-2020 cm^{-1} subregion. We have modified the revised text (see page 19, lines 15-26) to make this argument clearer.

TECHNICAL CORRECTIONS AND MINOR COMMENTS
p80 l14: "below 10hPa" is ambiguous. Presumably "altitudes below the 10hPa surface" is what is meant.

We agree with the reviewer and have adjusted our revised text (see page 1, line 26 of the revised document).

p81 l26: "MeteoSat" should presumably be "Eumetsat"

We agree with the reviewer and have adjusted our revised text as suggested (see page 2, line 31 of the revised document).

p83 l3: suggest adding RTTOV to the list of fast RTMs that are trained on LBLRTM spectra.

We agree with the reviewer. We did not intend to exclude RTTOV from the list of models trained on LBLRTM. We have added it to the revised text (see page 4, lines 1-2 of the revised document) and thank the reviewer for bringing this omission to our attention.

p83 l19: the heavy molecule cross-section data supplied with HITRAN is generally both temperature *and* pressure-dependent.

The reviewer is correct in noting that HITRAN provides both temperature and pressure-dependent cross-sectional data for heavy molecules. However, LBLRTM v12.1 (and previous versions) do not use the pressure-dependent data. Instead, the input cross-section data for all temperatures is at the lowest pressure available in HITRAN and the pressure scaling of the cross-section is then performed internally by performing a convolution of the cross section spectrum with an appropriate Lorentz function. We have rewritten the text in the revised document to make this clearer (see page 4, lines 14-17).

p87 l12: "3.1 p/m 3.3 x 10^{-7}" is presumably a typographical error

We have double-checked, and the stated value in our paper is the value stated in the original reference, and values for this metric are shown there to occasionally be negative, so the standard deviation makes sense.

p88 l23/24 "below 10hPa" and "above 0.1hPa" also ambiguous.

We agree with the reviewer and have adjusted the text as suggested (see page 8 line 30 of the revised document).

p89 l19: "a priori spectral surface emissivity" implies that the surface emissivity was also a retrieved parameter. I assume it wasn't since there is no other mention of it.

We agree with the reviewer that "a priori" can be confusing when used for parameters that are not retrieved, as is the case for ocean surface emissivity. We have adjusted the text in the revised document to refer to this as a "fixed, best estimate" of surface emissivity (see page 9, line 19 and other locations in the paper).

p89 l21: there is no mention of the number or location of the retrieval levels in each profile.

As stated on page 89, lines 26-28 of the Discussion paper, the pressure levels included in the state vector of the retrievals are presented in the Supplement for the interested reader. This notice is also in the revised document (see page 9, lines 26-28).

p89 l24: was surface pressure retrieved?

No, the surface pressure was not retrieved, and instead came from the ECMWF model output. We have revised the text to clear up this confusion (see page 8, line 29 and page 9, line 24 of the

revised document.)

p89 l12: was any a priori correlation assumed between T_srf and T_atm for the lowest temperature profile level.

No, the T_{srf} and T_{atm} were assumed to be uncorrelated, although both parameters were retrieved simultaneously. The RMS differences of T_{atm} from the a priori are shown in Figure N1 above, and the RMS differences of T_{srf} are given in the caption. We have revised the text to make this clear (see page 10, lines 14-15 of the revised document).

p91 l11: any particular reason that the CO correlation was 2 km while 1 km seems to have been assumed for everything else?

In general, we attempted to follow the procedure of Shephard et al. (2009) in our retrieval procedures, including in the specification of the correlation lengths. However, there is no particular scientific reason for the specification of 2 km for the correlation length of CO, but this specification has little impact on our results for the CO fundamental band, where the spectroscopy in LBLRTM v12.1 is clearly worse than in v9.4+ due to changes in the self-continuum for H_2O, as stated in the text.

p91 Eq1: it would be useful to have an approximate figure for how much of this RMS total can be ascribed to instrument noise.

We agree with the reviewer, and thus have added the RMS attributable for noise in each spectral region analyzed to the captions of Figures 2 (full spectrum, 0.17K), 3 (CO_2 ν_2, 0.045 K), 7 (CO_2 ν_3, $2.4x10^{-10}$ W/cm^2/ster/cm^{-1}), 12 (H_2O ν_2, 0.041 K in P-branch, 0.088 K in R-branch), 22 (O_3, 0.027 K), 23 (CO, $1.8x10^{-10}$ W/cm^2/ster/cm^{-1}) and 25 (CH_4, 0.016 K) (using the Discussion paper numbering). In all of these regions, the RMS attributable to noise is less than the observed RMS with either version of LBLRTM, suggesting that the remaining residuals are primarily due to errors in the spectroscopy and/or errors in the profiles rather than noise. We than the reviewer for the suggestion.

p92 l12: it is claimed that the reduction in residual is due to the inclusion of P and R branch line mixing, but how do you know that it is this and not simply the difference in the basic CO2 line parameters between the two versions?

Our claim that the reduction in the residuals in the CO_2 ν_2 band are mainly due to the previous work of Shephard et al. (2009), which showed substantial differences in the modeled spectrum on either side of the 720 cm^{-1} Q-branch (see Fig. 15 of Shephard et al., 2009). This matches the area and magnitude of the major differences between the v9.4+ and v12.1 residuals in Figure 3 of our Discussion paper. However, the reviewer is correct that we did not present any estimates of the relative importance of P- and R-branch line coupling compared with the difference in the basic line parameters. To rectify this, we have created Figure N4 below (and added it to the revised paper as Figure 4), which compares the modeled spectrum between LBLRTM simulations using the CO_2 line parameters from LBLRTM v9.4+, v11.1 (HITRAN 2004 with P-,Q-, and R-branch line coupling from Niro et al., 2005), the Lamouroux et al. (2010) line coupling database (HITRAN

2008 with P-,Q-, and R-branch line coupling, but without the CDSD position and strength changes), and v12.1 for our moderate water vapor case. It is obvious that the biggest changes between the versions came with the introduction of P- and R-branch line coupling, with relatively smaller impacts from the use of the CDSD positions and strengths, and the differences between Niro et al. (2005) and Lamouroux et al. (2010) negligible except for a few key regions. We have added this plot and discussion to the revised paper at page 12, lines 17-26.

Niro, F., Jucks, K., and Hartmann, J.-M.: Spectral calculations in central and wing regions of CO2 IR bands. IV: Software and database for the computation of atmospheric spectra, J. Quant. Spectrosc. Radiat. Transfer., 95, 469-481, 2005.

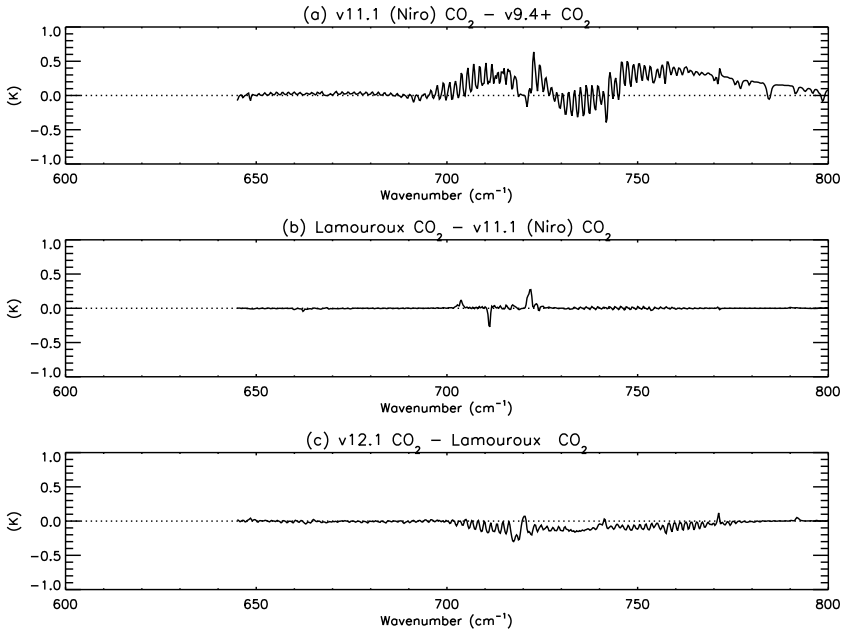


Figure N4. Differences of the modeled brightness temperature spectrum (moderate water vapor case) in the CO_2 ν_2 band due to changes in CO_2 spectroscopy. All results are for the moderate water vapor case. (a) LBLRTM v11.1 CO_2 (which includes the Niro et al. (2005) P-, Q-, and R-branch line coupling parameters) minus v9.4+. (b) Lamouroux et al. (2010) line coupling parameters (based on HITRAN 2008) minus Niro et al. (2005) (based on HITRAN 2004). (c) v12.1 CO_2 parameters (with CDSD position and strengths) minus Lamouroux et al. (2010).

p92 l24: it seems highly unlikely to me that these residuals are affected by water vapour -

more likely the atmospheric temperature structure.

We agree with the reviewer that the residuals in the 720 cm^{-1} Q-branch are more likely affected by the atmospheric temperature structure – this is what we were trying to communicate in page 92, lines 24-27 of the Discussion paper. We have altered the text to make this clearer (see page 13, lines 1-6 of the revised document).

p93 l20: could it be that the residual at 667cm-1 represents poor vertical resolution of the stratopause in your retrieval?

That is certainly a valid hypothesis. Our upper atmospheric retrieval levels are at 4.64, 2.61, 1.62, 1.00, 0.38, 0.1, and 0.01 hPa, as shown in the Supplement. It is possible that this is insufficient to resolve the stratopause, thereby leading to residuals in the 667 cm^{-1} Q-branch. Since we have not fully explored this possibility, our statement on remaining systematic spectroscopic errors in this region in the Discussion paper is probably too strong. We have revised the text to point out this alternate hypothesis for the remaining residuals (see page 13, lines 23-25 in the revised document.). We thank the reviewer for this astute observation.

p93 l13: convert this figure to brightness temperature.

We converted the figure to brightness temperature (0.07 K) and added it to the revised manuscript (see page 13, line 20).

p93 l24: there is an implication in this statement that the oscillations in Fig.6 come from the v9.4 retrievals. However, the real test would be to compare the retrievals against the a priori profiles to see which one has the best RMS difference - but that's not shown.

As shown in Figure N1 above, the RMS differences from the a priori temperature profiles are indeed lower in v12.1 than in v9.4+, showing the updated CO_2 spectroscopy in LBLRTM v12.1 gives temperature retrievals that are more consistent with the forecast model than those from LBLRTM v9.4+. We have added this figure and discussion to our revised paper (see Page 13, lines 26-32, Page 14, lines 1-5, and Figure 4) and thank the reviewer for the suggestion.

p94 l10: Fig 1 suggests that a version of the T retrieval *is* performed using the v3 band, but here it is stated that residuals plotted are based on the v2 band T retrieval.
So what was the purpose of the T retrieval using the v3 band? This is actually explained later (p95) but I was confused at this point.

We agree with the reviewer that this can be confusing, because we are performing two separate tests of the consistency of the CO_2 ν_2 and ν_3 bands: (1) using the retrieved ν_2 temperature to simulate the ν_3 band and (2) comparing the temperature profiles retrieved separately with the ν_2 and ν_3 bands.

To make this clear as early in the document as possible, we have rewritten the paragraph as follows (see page 14, lines 23-28 of the revised document):

"In our study, we assessed the consistency of the spectroscopy between the CO_2 ν_2 and ν_3

bands in two ways. First, we used the temperature profile retrieved using the CO_2 ν_2 band to simulate the radiances in the ν_3 band. (We discuss our second assessment method, comparing the temperature profiles retrieved using each band, further below.) Figure 9 shows the mean of the final and a priori residuals for both model versions in the CO_2 ν_3 band for the 120 IASI spectra. The retrieval window for the ν_3 atmospheric temperature retrieval is shown in green."

p121 Fig3 (and elsewhere) it seems odd to describe a spectral region of more than 100cm-1 width as a 'microwindow' (which are more typically _1cm-1 width).

We agree with the reviewer and have thus changed the term to "retrieval window" throughout the text.

p125 Figs7-9 y-axis should be in Kelvin for consistency with other plots?

For presenting results in IASI band 3 (2000 cm^{-1} and higher), we have chosen to present the results in radiance units ($W/cm^2/ster/cm^{-1}$), as the low radiance values in this region (from the fall off of the Planck function) make a small change in radiance appear as a large change in brightness temperature. This is the same procedure that was followed for the analysis of IASI residuals in this region in Shephard *et al*., ACP, 2009, and we feel strongly that this is the most constructive way to analyze the results in this region. However, as reviewer #2 points out, we did not make this reasoning explicit in the Discussion paper. We have thus added this discussion to the revised document (see pages 14, lines 28-32) and also referred the interested reader to Figure 2, where the mean residuals in all spectral regions are plotted in units of brightness temperature.

p142 Fig23: radiance units rather than BT?

See above – this region is also in IASI band 3.

p142 Fig24: optical depth for convolved spectra is an ambiguous quantity I assume this represents -ln(convolved transmittance) rather than convolved (-ln(monochromatic transmittance)).

The values are -ln(convolved transmittance) for LBLRTM runs with only a single absorbing species. We have made this distinction clearer in the captions of Figure 24 and 26 (using Discussion paper numbering).

Anonymous Referee #2

General Comments:I have read with interest the paper by Alvarado et al. that deals with a thorough investigation of the performance of the LBLRTM forward model. The paper covers aspects that are very relevant for the users of satellite data. I think it is a good paper based on sound science although no new concepts or ideas are introduced (with the possible exception of the investigation of the mis-specification of the water vapour isotopic ratios). My main criticism is that sometimes (probably due to the vast amount of material presented in the paper) the discussion of the results is lacking in detail. Possible alternative interpretations of the results are overlooked or discarded altogether. The number and the quality of the figures is adequate. To summarise, I think the paper should be accepted for publication with minor modifications. I invite the authors to take my comments constructively and pursue the publication of the paper.

Specific Comments:

Abstract

First paragraph: I do not entirely agree with the use of "critical" in the following statement "Reducing the uncertainties in our knowledge of spectroscopic line parameters and continuum absorption is thus critical to improving the application of satellite data to weather forecasting". Instead of "critical" I would suggest the use of, for instance, "an important factor". Forward modelling is a key element in the assimilation of satellite data but it is not the only factor controlling the impact of satellite data on weather forecasts. To my knowledge, it has never been demonstrated that the use of an improved spectroscopy has critically improved the quality of weather forecasts.

We agree with the reviewer, and have changed “critical” to “important” in the revised document (see, for example, page 1, line 17).

Introduction
Page 81, Second paragraph: strictly speaking, NWP centres (or at least the vast majority of them) do not retrieve atmospheric profiles. They assimilate satellite radiance data using variational data assimilation techniques. Rather then "retrieved", the authors should speak of "analysed" profiles.

We agree with the reviewer, and have changed “retrieved” to “analyzed” in the revised document (see, page 2, line 19).

Page 81, line 16: again, I do not agree with the use of "critical".

We agree with the reviewer, and have changed “critical to” to “an important part of” in the revised document (see page 2, lines 23).

Section 2.2
Page 86, line 18: what is the justification for the use of empirical scaling factors in the CO2

continuum model?

These corrections and their justification are discussed in Mlawer et al., 2012, as referenced in the Discussion paper, and for convenience we repeat that discussion here. These scaling factors were applied to the CO_2 continuum calculated using the "impact approximation" with line coupling. The scaling factors themselves were derived using the "dry" observations (i.e., of profiles with less than 0-2 cm PWV) of the IASI and AERI spectra analyzed in Mlawer et al., 2012. The theoretical justification for such empirical corrections to the CO_2 continuum is that, strictly speaking, the impact approximation should only apply near the line center, and so deviations from the CO_2 continuum calculated under the impact approximation should be expected. A second approach would be to apply a "chi" function to account for the sub-Lorentzian behavior of the far wings due to the finite duration of collisions, as in Souza-Machado et al. (1999). However, this approach requires the empirical fitting of the chi function parameters (e.g., average collision duration), and so is still an empirical correction.

De Souza-Machado, S., Strow, L. L., Tobin, D. C., and Hannon, S. E. : Improved atmospheric radiance calculations using CO2 P/R-branch line mixing, Proc. SPIE 3867, Satellite Remote Sensing of Clouds and the Atmosphere IV, 188 (December 8, 1999); doi:10.1117/12.373071, 1999.

Page 86, line 20: this sentence can be misleading. Do you actually mean the temperature dependence of CO2 continuum? More in general, could you please clarify the nature of this correction? What about the CO2 continuum in the nu2 band? Could you please give any detail?

The reviewer is correct, we are referring to the temperature dependence of the CO_2 continuum due to the temperature dependence of the line coupling coefficients of CO_2, as discussed in Mlawer et al., 2012. This temperature dependence could in principle apply in other regions of the CO_2 continuum as well, but we presently do not have any evidence from satellite or surface observations that this temperature dependence is an important effect outside of the bandhead region, and thus a temperature-dependent correction is not currently applied elsewhere. We have modified our text (page 6, lines 30-31 of the revised paper) to make this clearer.

Section 3
Page 87, line 7: "1/e" should be defined explicitly in the text.

We have added a more thorough description to our revised text (see page 7, lines 11-12).

Page 87, line 15: Could you please clarify what criteria did you use to assess the cloudiness of a scene? How confident you are the spectra are not affected be residual cloud contamination?

As in Matricardi (2009), only spectra that are detected as clear by the ECMWF cloud detection algorithm were used in this study. This algorithm (described in McNally and Watts, 2003) works by taking the first guess departures (i.e., the difference between the observed BT and BT calculated from a good estimate of the atmospheric state - typically a forecast from an NWP model) and looking for the signature of opacity that is not included in the clear-sky calculation (i.e.

cloud or aerosol). The goal is to eliminate channels (or, in our case, entire spectra) where residual cloud contamination leads to departures in brightness temperature greater than ~0.2 K. We have added this discussion and reference to our revised paper (see page 7, lines 19-21).

McNally, A.P. and P.D. Watts, 2003. A cloud detection algorithm for high-spectral-resolution infrared sounders, *Q J Roy Meteorol Soc*, **129**, 3411-3423.

Page 87, lines 15 to 20: why do you think 9 profiles failed to converged and one profile showed biased residuals?

Of the nine profiles that failed to converge, only two failed to converge for both versions of LBLRTM, while the other 7 converged with v12.1 but not v9.4+, suggesting spectroscopic errors in v9.4+ may have played a role. Many of these non-convergence cases were oscillatory, i.e. the retrieval bounced between two states that were sufficiently different that the convergence criteria were not met. Finally, we removed another case on the basis of a careful review of the a posteriori residuals from the 121 remaining cases. While we can only speculate as to the physical reason behind the failure to converge and the biased residuals (poor spectroscopy in v9.4+, overly strict convergence criteria, small but significant cloud contamination, etc.) we removed these cases from our analysis to ensure that these cases would not contaminate our analysis of the spectral residuals and retrieved profiles.

Section 4
Page 88, line 23: I wonder why the authors have chosen to use a priori profiles obtained from an older version of the ECMWF model. I think it is very likely that a more recent version of the model would have produced more accurate a priori profiles. Is this choice in any way related to the Matricardi (2009) paper often quoted by the authors? If this is the case, could you please clarify why?

As mentioned in the Discussion paper, we use a subset of the profiles used in Matricardi (2009) for our analysis. These profiles were provided to us by directly by Marco Matricardi, as mentioned in the acknowledgements of the Discussion paper. While a more recent version of the model might indeed provide more accurate a priori profiles, this is the model output we had available at the start of this project, and given our retrieval procedure the exact value of the a priori profiles should have relatively little impact on our results.

Could you please also clarify what do you mean by ECMWF model output? Have you used analysed (i.e. retrieved) profiles or background (i.e. forecast) profiles?. I am asking this because it could have implications for the accuracy of the profiles.

According to Matricardi (2009), these profiles are from the short-range operational forecasts of temperature, water vapor, and ozone. As stated above, we feel our retrieval procedure should adequately account for minor errors in the a priori profiles.

Page 89, line 3: the authors quote that water vapour profiles are taken from the ECMWF model as discussed in Matricardi (2009). I doubt this can be the case. If I am not mistaken, Matricardi (2009) uses very short range (3 hour) forecast profiles

obtained inside the 12 hour 4D-Var assimilation window and, to my knowledge, this profiles are (usually) not publicly available. The use of 3 hour forecast profiles has the obvious implication that these profiles can be closely matched in time to IASI observations, an aspect that is of particular relevance for humidity fields. Could you please provide any clarification?

As mentioned above, we did indeed use the profiles used in Matricardi (2009), which were provided to us by directly by Marco Matricardi.

Page 89, line 20: the authors state that spectral emissivities have been computed at zero viewing angle. Does this mean that they have used only IASI spectra at Nadir (i.e a Zenith angle of zero degrees at surface)? Please clarify.

We only used near-nadir IASI spectra, with a maximum angle off-nadir of 17.75 degrees. We have added this clarification to the revised document (see page 7, line 18).

Page 90: could you please clarify what is the basis for the choice of the error covariance values used in the retrievals? How do you justify that choice?

In general, our a priori covariance values were chosen to be consistent with those used in the analysis of Shephard et al. (2009). While other choices could change the numerical results, they would be unlikely to alter our major conclusions. H_2O, O_3, and CO were assumed to have a covariance of 20% to reflect their variability in the atmosphere, which might not be fully captured by the a priori profiles, while the well-mixed greenhouse gas CH_4 was assumed to have a smaller covariance (5%), reflecting the greater certainty in the a priori profile for this gas. For temperature, we assumed that the ECMWF surface temperature was fairly accurate (std. dev. of 1 K), as is the ECMWF temperature profile at altitudes below 200 hPa (std. dev. of 2 K). We then relaxed the temperature constraint at higher altitudes, as our corrected ECMWF temperature profile in these regions should be uncertain to roughly the size of the correction applied.

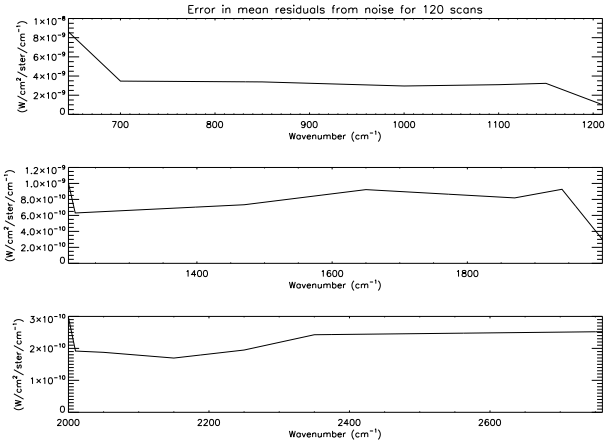
Section 5.1
Page 92: lines 19 to 25: the authors provide evidence of a dependency of the residuals on PWV (at least in some spectral regions). Why do they think residuals should increase in warmer and wetter regions? What mechanism should be responsible for that? Is there any inherent weakness of the LBL model that could be responsible for that?

We do present the dependence of residuals in the CO_2 and H_2O band on PWV, as stated by the reviewer. The primary purpose of this analysis is to identify regions where the H_2O continuum might be incorrect, as it was in the CO_2 ν_3 band in LBLRTM v9.4+ (see Figure 9 of the Discussion paper). In our revised paper, we separate our discussion of the PWV dependence of the residuals in the 720 cm^{-1} Q-branch and the residuals between 755 and 770 cm^{-1}. In the 720 Q-branch, the apparent PWV dependence in the CO_2 ν_2 band is more likely due to the differences in atmospheric temperature profiles in the different PWV groupings. Potential sources of temperature-related error in the CO_2 ν_2 band include (a) errors the temperature dependence of the CO_2 half-widths and line coupling parameters and (b) the lack of temperature dependence in the MT_CKD calculated CO_2 continuum in this region, but further work is necessary to definitively identify the source of the error. We have added a brief discussion of these potential sources of error to our

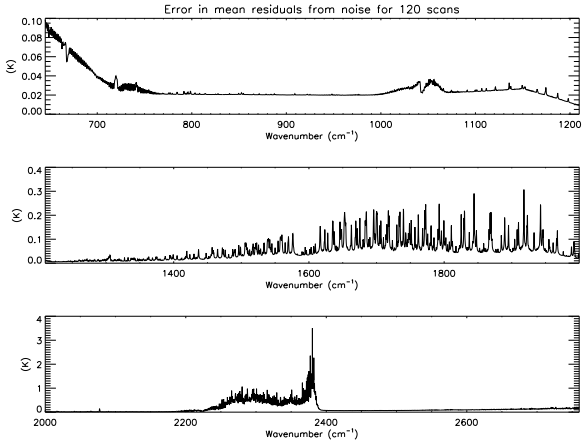
revised paper (see page 13, lines 3-6). As the channels in the 755 to 770 cm^{-1} region are more sensitive to the lower troposphere, and the optical depths of CO_2 and H_2O are comparable in this region, it is possible that either CO_2 or H_2O spectroscopic errors could be responsible for the residuals in this region, as noted in our revised paper (see page 13, lines 6-9).

Page 93, line 7: the authors mention Fig.12 of Masiello et al. (2011). I think it would be very informative if , following Masiello et al, the authors could also plot the error in mean residual due to noise.

We generally do not plot the error in the mean residuals due to noise in our residual plots, as with 120 cases, the error in the mean from noise (i.e., the noise for a channel in a single scan divided by the square root of 120) is very small compared to the mean residual features, and is generally indistinguishable from the zero line on the scale needed to plot the mean residuals. However, to address the reviewer's concern, we have added plots of the error in the 120 case mean from noise across the IASI spectrum (in brightness temperature and in radiance units) to the revised Supplement, and added a reference to these plots in the revised document (see page 12, lines 6-9). These plots are reproduced below. We also note the expected RMS from noise for each spectral region analyzed in the captions of the residual plots in the revised document, as suggested by Reviewer #1.



Supplementary Figure 7: Error in mean residuals (averaged over 120 scans) from noise in radiance units.



Supplementary Figure 8: Error in mean residuals (averaged over 120 scans) from noise in brightness temperature units (calculated by adding the radiance error to the observed radiances of the moderate water vapor case).

Page 94, line 16: in contrast to the previous Figures, in Figure 7 the residuals are expressed in terms of radiances rather than brightness temperatures. I fully understand the reason for that (i.e. the strong non linearity of the Planck function in the short wave) and had like to have a sentence in the text explaining it. This said, the authors should notice that it is very difficult to check the consistency of the results in the long wave and short wave regions if results are expressed using different units.

As we noted in response to Reviewer #1 above, we have chosen to present the results in IASI band 3 (2000 cm^{-1} and higher) in radiance units ($W/cm^2/ster/cm^{-1}$), as the low radiance values in this region (from the fall off of the Planck function) make a small change in radiance appear as a large change in brightness temperature. This is the same procedure that was followed for the analysis of IASI residuals in this region in Shephard *et al*., ACP, 2009, and we feel strongly that this is the most constructive way to analyze the results in this region. However, as Reviewer #2 points out, we did not make this reasoning explicit in the Discussion paper. We have thus added this discussion to the revised document (see page 14, lines 28-32) and also referred the interested reader to Figure 2, where the mean residuals in all spectral regions are plotted in units of brightness temperature.

Page 94, lines 20 to 25: I must admit I am not fully convinced by the N2O argument. Does the supposed day to day stratospheric N2O variability account for such a large signal? After all, the weighting functions of many of these channels peak in the middle to lower troposphere. Why have the authors totally discarded any issue related to the CO2 spectroscopy in that

region?

The reviewer is correct in pointing out that we did not present any evidence that the residuals in the 2200-2270 cm^{-1} region are substantially due to errors in our N_2O profile in the stratosphere. We were merely attempting to note that we cannot definitively say the errors are solely due to poor N_2O spectroscopy, as other potential errors (such as the N_2O profile in the stratosphere, or the CO_2 spectroscopy as noted by the reviewer) might also be present. We now recognize that the way this section was written in the Discussion paper might give the reader the incorrect impression that we think the N_2O and CO_2 spectroscopy is correct in this region. We have thus rewritten this discussion as follows (see page 15, lines 3-14 in the revised paper):

> *In addition, the large negative residuals between 2200–2270 cm^{-1} in both Figure 7(b) and 7(c) suggest that the optical depth in this region is still largely underestimated in LBLRTM v12.1. This is likely due to errors in the N_2O, CO_2, and/or H_2O spectroscopy in this spectral region, and we note that the updated CO_2, N_2O, and H_2O spectroscopy in LBLRTM v12.1 has reduced the residuals in this region. However, there may also be a small contribution from errors in our N_2O profiles. While our N_2O profile should be fairly accurate in the troposphere (where N_2O is a well-mixed gas), this spectral region is sensitive to N_2O in the stratosphere, which can have a significant day-to-day variability (Randel et al., 1994). However, our calculations suggest that a 33% increase in our assumed stratospheric N_2O would be required to remove a -2x10^{-9} W/cm^2/ster/cm^{-1} (~0.5 K) residual in this region. While it is possible for midlatitude stratospheric N_2O to vary by this amount, it is unlikely to have had a consistent bias of 33% across all 120 cases, and thus errors in the stratospheric N_2O profile are unlikely to account for a large portion of the residuals seen in this region.*

Page 94, line 26: it seems to me that the residuals in the 2200 cm-1 to 2250 cm-1 region have a strong dependence on PWV in both versions of LBLRTM. Yet, there is no mention of this in the text. Why have the authors chosen to ignore this aspect of the results?

We thank the reviewer for pointing this out and giving us a chance to correct this oversight. The residuals do indeed show a dependence with PWV, suggesting that not only does LBLRTM have too little optical depth in this region, but that this discrepancy is worse for wetter, warmer, tropical profiles than for drier, colder, higher latitude profiles. We think it is unlikely that this is due to issues with the MT_CKD water vapor continuum, as this explanation, when combined with our CO band results in Figure 23 of the discussion paper, would require that the MT_CKD water vapor continuum be too low from 2200 to 2250 cm^{-1} but too high from 2050 to 2200 cm^{-1}, which is unlikely. In addition, the standard deviation of the residuals about the mean is relatively high in this region, and the histograms of the residuals for the 0-2 cm PWV cases overlap significantly with the histograms for the 4-6 cm PWV cases, suggesting that the apparent dependence on PWV is not statistically significant. We have added this discussion to our revised paper (see page 15, lines 30-32 and page 16, lines 1-4).

Page 95: ljnes 1 to 11: what the about the contribution of foreign water vapour continuum in this region? It has been suggested in the literature that models like MT_CKD can underestimate its contribution by an order of magnitude.

First, we note that our analysis in this paper (see Figures 7 and 8) doesn't leave much room for a substantial increase in the water vapor foreign continuum in the CO_2 ν_3 region unless there are corresponding decreases in the other sources of optical depth in this region (H_2O self continuum, CO_2 continuum, N_2 collision-induced continuum, etc.). We assume the reviewer is referring to a recent paper by Ptashnik et al. (2012), which suggested that the water vapor foreign continuum in MT_CKD v2.5 was too low (see their Figure 5). We are skeptical of their result for two reasons. First, using the same experimental technique this team suggested that the water vapor self continuum in MT_CKD was also too low, but, as noted on Page 95, lines 1-11, this would negatively affect the residuals for the IASI cases examined here. Thus we suspect that their experimental system may be overestimating the absorption in some manner. Second, the comparison to MT_CKD v2.5 in Figure 5 of Ptashnik et al. (2012) did not include the additional N_2 absorption induced by the addition of H_2O to the experimental cell that is accounted for in MT_CKD v2.5, and this may account for some of the discrepancy in their figure. Given these concerns with the Ptashnik et al. (2012) comparison, and the fact that MT_CKD v2.5 is able to reproduce the observed radiance in the CO_2 ν_3 band very well, we do not feel an increase in the MT_CKD water vapor foreign continuum is justified at this time. We have added a mention of the Ptashnik paper to the revised paper at page 15, lines 23-25.

Ptashnik, I. V., McPheat, R. A., Shine, K. P., Smith, K. M., & Williams, R. G.: Water vapour foreign-continuum absorption in near-infrared windows from laboratory measurements, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 370, 2557-2577, 2012.

Page 96, lines 1 to 4: I agree that profiles retrieved with LBLRTM v12.1 are more consistent. However, I cannot fail to notice that in the middle troposphere differences between profiles retrieved using the nu2 and n3 bands can still reach 2K. Do the authors have any opinion regarding the possible source of this discrepancy?

We thank the reviewer for pointing this out. We believe the discrepancy is due to the fact that, when the CO_2 ν_3 band is used to retrieve atmospheric temperature, the only spectral region in that band sensitive to the middle troposphere is the ν_3 bandhead. Our evaluation of the CO_2 ν_3 residuals using the CO_2 ν_2 retrieved temperature profile (Fig. 7 and page 94, lines 8-25 in the Discussion paper) found small, systematic residuals in the bandhead region, which is consistent with the differences seen in the retrieved temperature profiles in the middle troposphere. We have added this discussion to our revised paper (see page 16, lines 23-27).

Section 5.2: To what extent the residuals can be affected by errors in the (foreign?) water vapour continuum? This aspect of the discussion is conspicuously absent in the paper. If the authors think that water vapour continuum is not relevant in this spectral region, then they should state it explicitly in the paper.

The reviewer is of course correct that the water vapor continuum is relevant in this spectral region. However, the foreign and self water vapor continua in this region did not change between LBLRTM v9.4+ and v12.1, and so are not discussed as part of the differences in residuals between the model versions. It is possible that some of the remaining residual features are due to errors in the H_2O continua rather than errors in the H_2O line parameters. We have thus added a discussion

of this point to our revised document (see page 17, lines 29-30).

Section 5.4: How do the authors intend to modify the foreign continuum?

We believe the H_2O self continuum in the CO fundamental band is what needs to be corrected, not the foreign. The fitting procedure used to develop the scaling factors in the MT_CKD v2.5 H_2O self continuum is discussed in detail in Mlawer et al., 2012. This fitting procedure will be repeated, but with the water vapor self continuum in the CO fundamental band constrained at the values from MT_CKD v2.4 (i.e., the scale factors applied to the MT_CKD v1.0 continuum values in the spectral range 2050-2200 cm^{-1} will be reset to 1.) We have added a reference to this procedure to our revised document (see page 23, line 23-24).

Section 5.5: Are the authors suggesting that an improved spectroscopy is potentially degrading the residuals?

We apologize for our muddled language in this section of our Discussion paper. What we were trying to convey is that, in evaluating the CH_4 spectroscopy, it is important to look both at the after-retrieval residuals and the retrieved CH_4 profiles. While the RMS of the residuals is slightly worse in LBLRTM v12.1, the retrieved CH_4 profiles with LBLRTM v9.4+ are more realistic, with a substantially lower positive bias. Thus, while we cannot conclude on the basis of this study that the CH_4 spectroscopy in LBLRTM v12.1 is a clear improvement over that in v9.4+, we do suggest that the new spectroscopy might help to reduce the positive bias seen in operational CH_4 retrievals that use this spectral region. In fact, we have done some preliminary work with the Tropospheric Emission Spectrometer (TES) retrievals of CH_4 that show that the new CH_4 spectroscopy in LBLRTM v12.1 reduces the positive bias in TES CH_4 retrievals compared with aircraft observations from +40 ppb to +23 ppb while generally improving the spectral residuals between 1190 and 1320 cm^{-1} (Alvarado, M. J., K. E. Cady-Pereira, V. H. Payne, S. S. Kulawik, J. R. Worden, and K. J. Wecht (2013), Evaluation of Recent Spectroscopic Updates and Their Impacts on TES CH_4 and CO_2 Retrievals, presented at the TES Science Team Meeting, Cambridge, MA, USA, 18-19 March). We have rewritten Section 5.5 in our revised draft to better convey our meaning.

Conclusions: this section should be updated based on the response to the comments above.

We have revised both the abstract and conclusions to be consistent with our other revisions.