Reviewer two, answers.

We thank the reviewer for reading our manuscript and for offering many useful comments towards its improvement. In the revised manuscript, we included modifications addressing almost all of these comments.

The submitted manuscript takes up on the ground-based remote sensing synergy approach of combining microwave radiometers (MWR) and RASS by applying a state-of-the-art physical retrieval approach. This is important, since MWR are known to show very accurate performance in temperature profiling in the lowest 500 m, whereas RASS are able to adequately capture the typical temperature inversion at the top of the atmospheric boundary layer (ABL) and thus, in theory, the synergy of both could lead to an improved temperature profile throughout the whole ABL.

Major points

1.) The way to showing the latter point above, however, is obviously severely hampered by the quality of the MWR data, most probably in terms of a TB bias. While the authors do show a bias correction applied to the MWR TBs, it is unclear whether this was done only for zenith observations or also at 15° elevation (Fig. 1). Here a detailed analysis is missing. If this manuscript is to be accepted for publication using real TBdata, the reason for the biases shown in Figs. 6 c and f (black a grey lines) must be identified, discussed and corrected for.
We thank the reviewer for this specific comment. Some parts of the bias-correction description were indeed missing. Additional text has been included in Section 3.2: "We compute the bias in the bias-correction procedure only from the zenith scans assuming that the same bias is suitable for the oblique scans. Also, we use the assumption that the true bias is an offset that is independent of the scene, so that the sensitivity to the scene (e.g., clear or cloudy, zenith or off-zenith) is small. To investigate this, we eliminated the radiosondes launched during rainy periods (5 out of 58 cases) and found that the averaged temperature profiles were very little different than when all radiosonde profiles."

More detailed discussion of the temperature biases shown in Fig. 6, especially near the surface layer, will be included in Section 4.2 in the final version of the manuscript.

2.) The paper shows hardly any quantitative discussion, which is necessary for a sound scientific analysis. Except for just a few passages, discussions of the figures are carried out only in a qualitative, rather unspecific manner. With respect to this, specifically the sections 3 and 4 should be thoroughly rewritten. E.g., avoid using "This might..", "We believe…", "seemingly", "Differences", "better" or "improve" etc. without referring to adequate statistical measures. A lot of the data is there in the XPIA data set und you can use to confirm, deny or to quantify your assumptions, respectively results.
We have tried to avoid purely qualitative descriptions, and to provide quantitative details in the indicated sections for the new version of the manuscript, thank you.
For ex., in Section 3, especially 3.2, we included the detailed descriptions of how the clear-sky days were chosen and how the uncertainty and the bias for each MWR channel were calculated.

3.) Because the authors write they could not apply any bias correction to the NN approaches, I strongly suggest omitting them from the paper. The comparisons are thus "unfair" and I do not see what benefit the reader has from including the NN retrievals when the actual goal is evaluating the MWR/RSS synergy potential that can be achieved with the PR. Instead, in all the corresponding figures, I would like to see the results of RSS-only PR, i.e. without including the MWR so the reader has an impression what these systems are capable of in a stand-alone manner.

We thank the referee for this particular comment. Following your recommendation and also the opinion of Reviewer #1 on this matter, we decided to move all comparisons of PR and NN profiles to Appendix A, while also making note of the fact that without mentioning NN retrievals our analysis would be incomplete for the community of MWR end-users. We think that the possibility to do the bias correction in the PR is just one of the advantages the PR has. The NN retrievals are provided by the manufacturer and have the disadvantage that no bias correction is performed. They are nevertheless used by most end-users. We believe that the comparison between PR and NN is still very important and should be included in some way in the manuscript, while noting the unequal basis for the NN and bias-corrected PR comparison. These issues are now addressed in the Appendix.

Regarding a RASS-only PR, we do not see the value of this because RASS without MWR in MonoRTM will be used as RASS + prior, so we should get mostly the profile of the prior because the RASS covers only a small portion of the 17 km temperature profile. On the other hand, the RASS measurements are included in the figures, especially in Fig. 10 in the manuscript, showing what these instruments can provide in a stand-alone manner.

4.) How did you deal with clouds, what about precipitation? Did you retrieve LWP simultaneously to temperature and water vapor? What influence do clouds have on the retrieval? I find no information about this throughout the manuscript.
Most of the radiosondes were launched during clear-sky time. See also answers to Referee #1 about this.

We included several new paragraphs in Section 3.2, e.g.:

"we found that from 58 radiosonde launches used in our statistical analysis, 41 belong to the clear-sky category, 12 - to cloudy but non-precipitating conditions and 5 - to rainy periods".

A discussion of the impacts of clouds on the retrieval is mentioned in comment 1) above and will be included in the manuscript.

5.) The sections describing microwave radiometry need more background and scientific accuracy.
We had already included many references in order to avoid a detailed description of the basic principles of microwave radiometry.

Further specific points and questions to be addressed

1.) Abstract, last paragraph: It is not clear if the improvements described refer to the PR compared to the NN or the MWR+RASS combination compared to the MWR-only retrieval.
As we moved the discussion of PR and NN profiles comparison in Appendix A, this paragraph has been changed to highlight the purpose of this paper as:

"Having the possibility to combine the information provided by the MWR and RASS systems, in this study the physical-iterative approach is tested with different observational inputs: first using data from surface sensors and the MWR in different configurations, and then including data from the RASS. These temperature retrievals are assessed against 58 co-located radiosonde profiles. Results show that the combination of the MWR and RASS observations in the physical-iterative approach allows for a more accurate characterization of low-level temperature inversions compared to the physical retrievals of the MWR passive measurements, and that these retrieved temperature profiles match the radiosonde observations better than the temperature profiles retrieved from the MWR in the atmospheric layer between the surface and 5 km AGL. Specifically, in this layer of the atmosphere, both root mean square errors and standard deviations of the difference between radiosonde and retrievals that combine MWR

and RASS are improved by ~0.5 K compared to the difference between radiosonde and MWR retrievals. Pearson correlation coefficients are also improved.
We provide the comparison of the temperature physical retrievals to the neural network retrievals in Appendix A."

2.) Introduction: A description of the physical principle that allows temperature (& humidity) profiling (and LWP retrieval) from passive MWR observations is missing. When doing so, please consider reformulating the advantages and disadvantages of the MWR retrieval methodology, because they are currently not scientifically sound.
Be sure to differentiate how the frequency dependence and elevation angle dependence of TB can both lead to resolving the temperature profile in the vertical.
We are not sure what the Reviewer is suggesting here. There are many articles describing the MWR temperature and humidity retrievals as well as physical principles of such retrievals, and we had already included many of these references in the manuscript in order to avoid a detailed description of the basic principles of microwave radiometry.

Nevertheless, we include a description of the temperature retrieval frequencies in Section 2.1:

"V-band frequencies or channels also could be divided in two categories: the opaque channels, 56.66 GHz and higher, which are more informative in the low layer of the atmosphere from the surface to ~1 km above the ground and the transparent channels, 51-56 GHz, which are more informative above 1 km in the temperature profile".

3.) Line 109: MWR don't "apply radiative transfer equations and neural network retrievals…" – please reformulate.
This paragraph is reformulated to: "Radiative transfer equations are commonly used to train statistical retrievals or as forward models within physical retrieval methods".

4.) Line 115: Please make clear what you mean with "deep layer of the atmosphere".
Changed to: "the layer of the whole troposphere ".

5.) Section 2.1, lines 203-204: The purpose of using observations at 15° elevation is not to "average out small scale horizontal inhomogeneities of the atmosphere" but to obtain TB observations at different optical depths.
This paragraph has been modified according to your suggestion:

"In this study we make use of the NN zenith and of the NN oblique measurements, where the latter can obtain TB observations at different optical depths."

6.) Section 3.1, lines 280-286: Why does the Y vector and the error covariance matrix contain both "zenith" and "zenith+oblique" components. If I understand correctly, you can choose to use only zenith observations and add the off-zenith (=oblique) TBsto improve the retrieval? So then should it not be "zenith" and "oblique"? Please clarify.
Table 1 as well as observational vectors Y2, Y3 and Y4 and matrix Sε are modified:

	T_{sfc}	Q_{sfc}	Tb_{zenith}	$Tb_{oblique\_avrg}$	$Tv_{RASS915}$	$Tv_{RASS449}$

Y₁ *= MWRz*	*X*	*X*	*X*			
Y₂ *= MWRzo*	*X*	*X*	*X*	*X*		
Y₃ *= MWRzo915*	*X*	*X*	*X*	*X*	*X*	
Y₄ *= MWRzo449*	*X*	*X*	*X*	*X*		*X*

$$\mathbf{Y_1} = \begin{bmatrix} T_{sfc} \\ Q_{sfc} \\ \boldsymbol{Tb_{zenith}} \end{bmatrix} \qquad \mathbf{Y_2} = \begin{bmatrix} T_{sfc} \\ Q_{sfc} \\ \boldsymbol{Tb_{zenith+oblique\ avrg}} \end{bmatrix}$$

$$\mathbf{Y_3} = \begin{bmatrix} T_{sfc} \\ Q_{sfc} \\ \boldsymbol{Tb_{zenith+oblique\ avrg}} \\ \boldsymbol{Tv_{RASS915}} \end{bmatrix} \qquad \mathbf{Y_4} = \begin{bmatrix} T_{sfc} \\ Q_{sfc} \\ \boldsymbol{Tb_{zenith+oblique\ avrg}} \\ \boldsymbol{Tv_{RASS449}} \end{bmatrix}$$

$$\boldsymbol{S_\varepsilon} = \begin{bmatrix} \sigma^2_{Tsfc} & 0 & & 0 & 0 \\ 0 & \sigma^2_{Qsfc} & & 0 & 0 \\ 0 & 0 & \boldsymbol{\sigma^2_{Tb_{zenith}}} \text{ (1)} & \boldsymbol{or\ \sigma^2_{Tb_{zenith+oblique\ avrg}}} \text{ (2)} & 0 \\ 0 & 0 & & 0 & \boldsymbol{\sigma^2_{Tv_{RASS915(449)}}} \text{ (3 or 4)} \end{bmatrix}$$

7.) Line 310: Do you mean the covariance between the uncertainties of the measurements?
This part of the manuscript is reformulated:

"The uncertainty in the MWR Tb observations was set to the standard deviation from a detrended time-series analysis for each channel during cloud-free periods. The derived uncertainties ranged from 0.3 to 0.5 K in the 22 to 30 GHz channels, and 0.5 to 1.0 K in the 52 to 60 GHz channels. We assumed that there was no correlated error between the different MWR channels.
For the RASS, collocated RASS and radiosonde profiles were compared and the standard deviation of the differences in Tv were determined as a function of the radar's signal-to-noise ratio (SNR). This relationship resulted in uncertainties that ranged from 0.8 K at high SNR values to 1.5K at low SNR values. Again, we assumed that there was no correlated error between different RASS heights. Following all these assumptions, the covariance matrix $\mathbf{S_\varepsilon}$ is diagonal."

8.) Section 3.2: There seems to be a non-consistent use of terminology. Please use "uncertainty" only in the sense of random uncertainty and distinguish it clearly from systematic offset (=bias).
We have made certain to consistently refer to the random uncertainty of Tb as the uncertainty, and the systematic offset as the bias.

9.) Lines 323-324: erroneous, please reformulate in a consistent manner
The text is changed as:
"While the bias of the retrieval depends on both the sensitivity of the forward model and the observational systematic offset, we can try to eliminate, or at least to reduce, the systematic error in the MWR observations."

10.) Line 327: The 30 GHz channel is not predominantly water vapor, but liquid water sensitive.
Changed.

11.) Lines 328-330: "The random uncertainty in brightness temperature was calculated as its standard deviation during clear sky times and for this channel is approximately 0.3 K": Why is this calculated standard deviation related to the TB uncertainty? Over what time window did you average? What about water variability in the atmosphere during the calculation time? Why actually did you calculate this standard deviation and where do you use it in the course of your study?
Thank you for this comment. We included a much more detailed description of the uncertainty calculation in the text in Section 3.2:

"A threshold value of 0.3 K has been used for the uncertainty calculation. The random uncertainty in Tb is calculated as an average of the Tb standard deviation in a one hour sliding window through all data points of a day. It also could be computed as the standard deviation of the difference between Tb and the smoothed Tb to eliminate daily temperature variability. Finally, there is a "standard" set of uncertainties used as the high boundaries for Tb uncertainty per MWR channels calculated empirically in the previous experiments. Four clear-sky days have been chosen using a criterion of 0.3 K uncertainty in the 30 GHz channel: March 10 and 30, and April 13 and 29, 2015.
During periods with liquid-bearing clouds overhead, this criterion is markedly higher (more than 0.7 K) and much higher for the rainy periods (> 4 K). While those calculations were applied on a daily basis, it is important to mention that the days are not uniform in terms of cloudiness or rain. Therefore, we used the data for the 2-3 hours bracketing the time of radiosonde launches to determine to which category a particular radiosonde profile belongs, clear-sky, cloudy or rain. In this way, we found that from 58 radiosonde launches used in our statistical analysis, 41 belong to the clear-sky category, 12 - to cloudy but non-precipitating conditions and 5 - to rainy periods. For the four chosen clear-sky days not only were the daily uncertainties of 30 GHz Tb below 0.3 K, but all three sets of uncertainties described above were extremely similar with the averaged difference less than 0.05 K."

12.) Lines 332-333: How were the clear-sky days selected?
Please, see above.

13.) Lines 333-334: How did you calculate the bias?
From the modified text in Section 3.2:
"The bias was computed for each of the 22 channels as the averaged difference between the observed Tb from the MWR zenith observations, and the forward model calculation applied to the prior, over these selected clear-sky days, and then subsequently removed from all of the MWR observations."

14.) Before line 358: a description and a quantitative discussion of the Sa and Se matrices applied needs to be given before going on describing retrieval results.
Sa and Se matrices are described in Section 3.1 and retrieval results are discussed in 3.2.

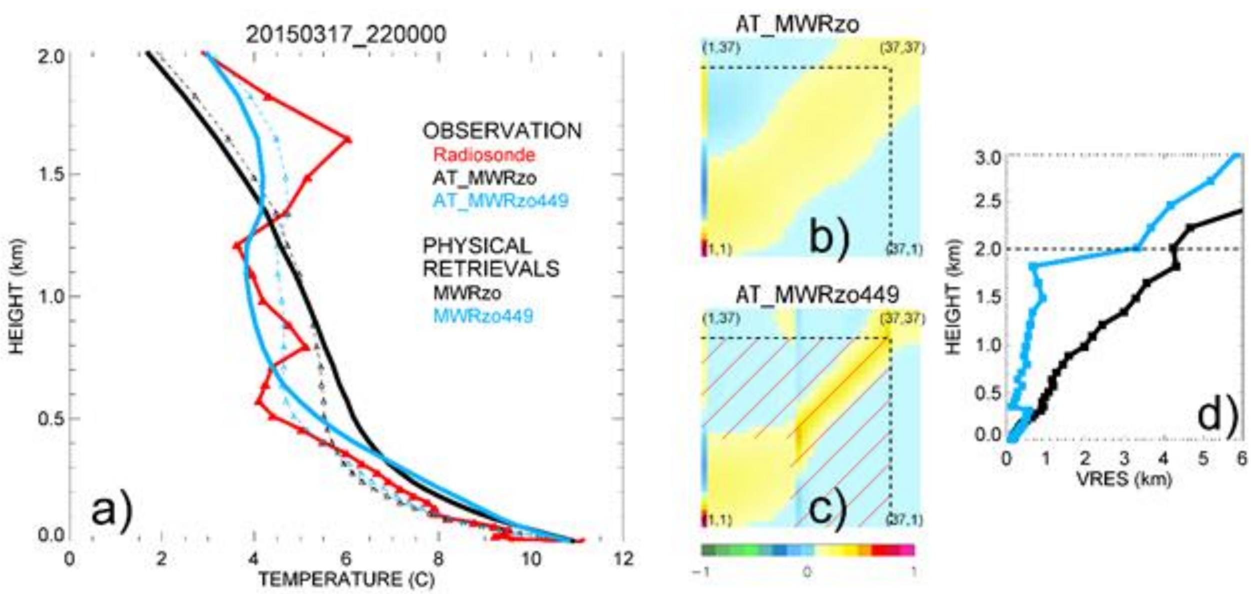
15.) Lines 425 and following, referring to Fig. 3: quantitative argumentation missing and VRES "jumps" in Fig. 3 are not discussed
We thank the Reviewer for this comment very much because it prompted us to reconsider the method used to calculate the vertical resolution.

There are two ways to compute the vertical resolution from the averaging kernel. First, we applied a method that Tim Hewison published (TGRS 2007, reference below) that uses only the diagonal data of the averaged kernel. It works well when the retrieval uses only the input from the passive observations, like the MWR, but is not very suitable for the passive/active combination of inputs, as is seen in Fig. 3d in the manuscript (with the creation of the "jumps"). So, we returned to the method (that we actually erroneously mentioned in the paper) that computes the vertical resolution as the full-width half-maximum (FWHM, Maddy and Barnet, TGRS, 2008, reference below) value of the averaging kernel at each height.

T. J. Hewison, "1D-VAR Retrieval of Temperature and Humidity Profiles From a Ground-Based Microwave Radiometer," in IEEE Transactions on Geoscience and Remote Sensing, vol. 45, no. 7, pp. 2163-2168, July 2007, doi: 10.1109/TGRS.2007.898091.

Maddy, E. S. and C. D. Barnet, 2008: Vertical Resolution Estimates in Version 5 of AIRS Operational Retrievals. *IEEE TGRS*, VOL. **46**, NO. 8, AUGUST 2008, doi:10.1109/TGRS.2008.917498

Using the FWHM method, Fig.3 is changed to the one below, where the "jumps" in panel d are significantly reduced:



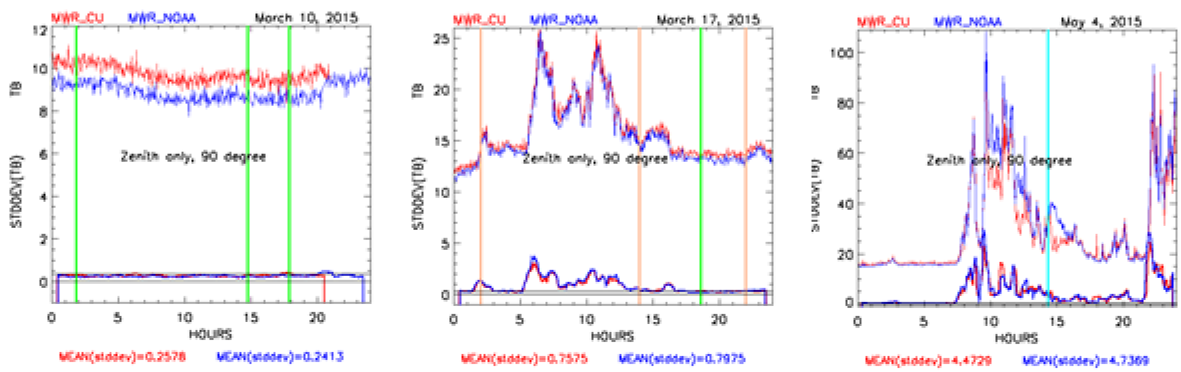
16.) Section 4.1, lines 469-471: unspecific sentence, please reformulate

This sentence is deleted because soon after the similar text is followed:

"MWRzo449 has the best statistical measures compared to the other PRs, particularly below 2 km AGL, where RASS 449 measurements are available".

17.) Fig. 5: How many cases are used for the statistics, how many are clear-sky, how many are cloudy sky? How did you deal with cloudy cases in general?

Statistical results are shown in Figs. 4, and 6-10, not in Fig. 5 of the manuscript (where a single case profile - 18 March, 2015 at 0200UTC is presented). For the statistical analysis, from 58 valid radiosonde profiles 41 have been launched in clear-sky periods, 12 - in cloudy but non-precipitating conditions and 5 - in rainy time. This information is now included in the manuscript, Section 3.2. We defined those categories using the 30 GHz channel Tv as in these figures:



Zenith Tb from a 30 GHz channel for a clear-sky day (left panel), cloudy day (middle panel) and rainy day (right panel) from the CU radiometer in red and NOAA radiometer in blue. STDDEV(Tb-SMOOTH(tb,11)) is shown at the bottom in each panel with its average values printed under the panels in corresponding colors. Vertical (green – for clear-sky, beige – for clouds and cyan – for rain) lines show the time of radiosonde launches.

18.) Fig. 8: Can you derive meaningful statistical measures such as RMSE from only 15 cases?
This is a valid comment. We are interested in describing the "worst case" most extreme events, when the radiosonde temperature profiles are most different from the prior profile, and so, by definition the number of cases needs to be limited, otherwise they are no longer extreme. On the other hand, some level of statistical significance is desired. Given that we have 58 radiosondes, 15 events are already nearly 25% of the total. We felt that this was a reasonable compromise given the limitations of the data set.

19.) Fig. 9: The MWRz2sigma449 performs best compared to the other retrievals. This retrieval relies on an increase in the MWR uncertainty, which was chosen in an arbitrary manner. This choice should be thoroughly justified and set into context with the performance of the 449-only retrievals which I would like to see (see "Major points" above).

The choice of double MWR uncertainty for MWRz2sigma449 is not arbitrary, but the reviewer is absolutely right, it is not qualitatively justified in the manuscript. It was chosen based on the "worst" XPIA temperature profile on March 18, 2015, 02:00 UTC showing in Fig.5 in the manuscript. This particular case is not only the worst in the XPIA experiment in terms of temperature inversions (three of them in one profile, with one near the surface), but with other complications. We found that the MWR Tb from the opaque channels of both zenith and obliques scans, have biases (to the forward model calculation of radiosonde Tb) of around 1 K. We wanted to check our hypothesis about too little freedom of the PR approach in the layer between surface and RASS measurements. As is mentioned in the text, "After several trials", we indeed made many additional runs, but we wanted to keep our recommendations general, and not be very specific about this particular case.

20.) Section 4.4, lines 683-686: This sentence is formulated in a general, rather nonspecific way and could be given without any of the studies conducted here.
This paragraph is removed.

Technical comments

1.) Figures are given in rather low resolution, a higher one would have been nice to be able to better interpret the results.
All figures are in tiff format that has a high resolution. The deterioration of the images comes from the conversion to PDF. Original tiff format files will be provided to the editorial office when requested.

2.) Equation fonts appear in a non-standard, unorganized way.
Equation font is changed to be the same throughout the paper.

3.) In general: please write K or °C, but not °K.
Checked and fixed.

4.) Section 3.1, lines 280-286: Numerate all equations, be consistent with equation fonts and text fonts, be consistent with variables (i.e. L, LWP), explain all variables (and indices) in the text. Please be neater.
Lines 280-286 consist of only one equation, Eq. (1), which is numbered, and the descriptions of all its terms. We changed the text to have consistency in fonts and text fonts, and we consistently used LWP in the revised text.

5.) Line 348 and following: use a new sub-section, the paragraphs are not related to "Bias-correction" anymore
We renamed the Section **3.2 PR's bias-correction** to **3.2 PR's bias-correction and PR's temperature profiles.**

6.) Section 3.3, lines 415-421: move text to Fig. 3 caption
The text on these lines reformulates the Fig. 3 description in a more explanatory way.