Reviewer one, answers.

We wish to thank the reviewer very much for carefully reading our manuscript and for offering many comments towards its improvement. In revising the manuscript, we have taken into account almost all these comments.

Major comment 1 : It should be clarified which MWR channels are used for each configuration : oblique versus zenith as well as for PR versus NN retrievals. In fact using transparent channels for lower elevation angles is often avoided as the homogeneity assumption is violated (especially if there is cloud or rain in one direction and not in the other direction when the two elevation scans are averaged). Thus, all my interpretation assumed that transparent channels are not used at low elevation angles for the manuscript. If this is not the case and transparent channels have also been used at low elevation angles, the authors should explicit which quality control has been used to identify inhomogenous scenes when they average the two microwave radiometer scans (refer to Cimini et al 2006).

Two identical radiometers (Radiometrics MP-3000A) were used during the XPIA experiment. Both MWRs have 35-channels spanning a range of frequencies, with 21 channels in the lower (22-30 GHz) K-frequency band, from which 8 channels were used during XPIA: 22.234, 22.5, 23.034, 23.834, 25, 26.234, 28 and 30 GHz, and 14 channels in the higher (51-59 GHz) V-frequency band, all used in XPIA: 51.248, 51.76, 52.28, 52.804, 53.336, 53.848, 54.4, 54.94, 55.5, 56.02, 56.66, 57.288, 57.964 and 58.8 GHz, with elevation angles of 90 degrees (zenith) and 15 & 165 degrees (obliques). Section 2.1 has been modified to include these additions.

The Reviewer is correct in assuming that only the opaque channels are used from the oblique scans, when these are used in the Physical Retrieval approach. More specifically, the Physical Retrieval has two options for radiometer measurement inputs: using only the zenith scan, or using the zenith plus oblique averaged scans. From the zenith scan, Tbs from all 22 channels are used in both configurations, while for the oblique scans, when they are included, only the opaque channels (56.66, 57.288, 57.964 and 58.8 GHz) are used. Additional RASS active instrument measurements are used together with the second option, while 2-m in-situ observations of temperature and humidity are used in all configurations. So, Table 1 in manuscript has been revised to be:

	T_{sfc}	Q_{sfc}	Tb_{zenith}	$Tb_{oblique\_avrg}$	$Tv_{RASS915}$	$Tv_{RASS449}$
$\boldsymbol{Y_1}$ = *MWRz*	X	X	X			
$\boldsymbol{Y_2}$ = *MWRzo*	X	X	X	X		
$\boldsymbol{Y_3}$ = *MWRzo915*	X	X	X	X	X	
$\boldsymbol{Y_4}$ = *MWRzo449*	X	X	X	X		X

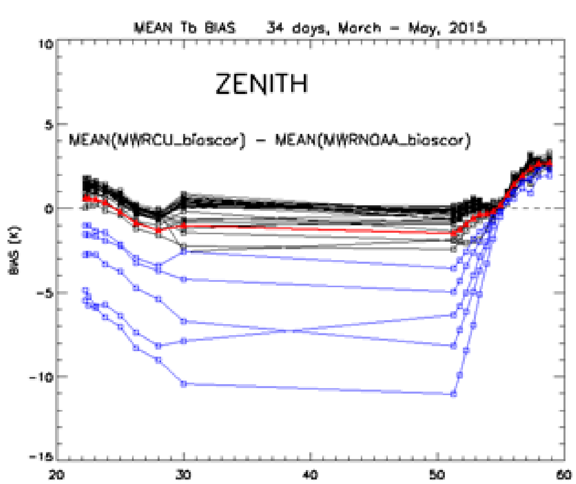
The text has been modified in Section 3.1:

"The MWR provides **Tb** measurements from 22 channels from the zenith scan for the zenith only configuration (**Y_1,** which also includes the 2-m in-situ observations of temperature and humidity), while when using the zenith plus oblique Tb inputs (**Y_2**, **Y_3**, and **Y_4**, also including the 2-m in-situ observations of temperature and humidity) the same 22 channels were used from the zenith scan but only the four opaque channels (56.66, 57.288, 57.964 and 58.8 GHz) from the oblique scans."

The second major comment that should be addressed is about the interpretation of figure 5 where the degradation of the temperature profiles with MWRzo below 200m is attributed to biases in the MWR oblique scans. I think it is important to be rigorous there because nowadays many MWRs dedicated to temperature profiling use low elevation angles down to 5.4° to improve temperature retrievals. Thus, all your interpretation in the manuscript of the improvement brought by RASS measurements is sub-optimal if oblique scans cannot be used (at least for your conclusions below 2 km altitude where RASS brings most of the information). First of all, I think this needs to be addressed in the paper and clearly explained and discussed in the conclusion. Secondly, I also found that the hypothesis provided in line 507 that your degraded results below 200 m with MWRzo comes from a bias is not convincing for several reasons :

→ line 207 : you mention that the two MWR units have a very good agreement in the temperature profiles in the overlapping dates both in terms of bias and correlation. Thus, when you conclude later that the MWR unit used in the paper presents a bias in the oblique measurements it means the two units were in fact biased and not well calibrated, which I found surprising (we can imagine a problem in one calibration but for two calibrations it seems that there is a problem in the deployment)

The NN retrieved temperature profiles from the two MWRs indeed have a good agreement with statistically low bias (0.5 K) and high correlation (0.994). While in line 207 we refer to these statistical measures, line 507 refers to the particular case of March 18, 02:00 UTC which is certainly "a worst case scenario" in the XPIA experiment (certainly a difficult one to retrieve accurately from passive instruments because of the many temperature inversions, three in one profile including one at the surface!)



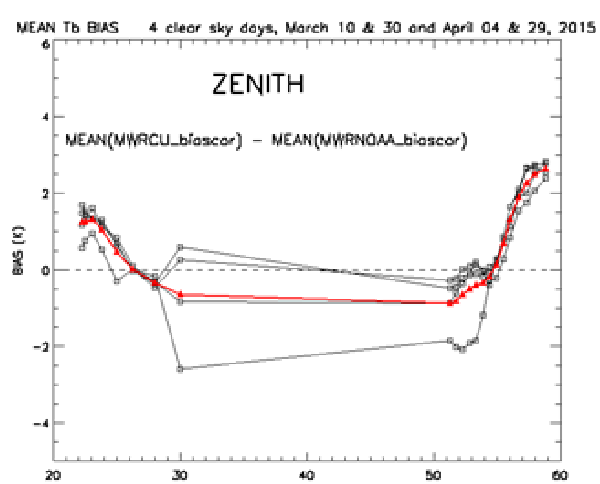


Fig. R1. *Difference between bias-corrected Tb of MWR_CU and MWR_NOAA shown for all 34 days when both MWRs were available (left panel) and for the chosen four clear-sky days only (right panel). The averaged difference is shown in red. Blue lines Tb differences correspond to the rainy days.*

Previously, in Bianco et al., 2017: “we compared the brightness temperatures (Tb) of the two MWRs for each retrieved channel **for 1 day**, finding almost all channels in good agreement (with differences of ~ 2 K for channels 51.248, 51.760, 52.280, 56.020, 57.288, 57.964, and 58.800 GHz; differences of ~ 1 K for channels 22.500, 26.234, 30.000, 52.804, 53.848, and 56.660 GHz; while the remaining channels did not show appreciable differences)”. Using **all available data** from both radiometers we still found that the differences of their daily averaged Tb from the zenith scans, even bias-corrected, were 2-2.5 K for the opaque channels.

→ biases are in general very low for opaque channels that are the most informative below 1 km altitude (and even more below 200m where the degradation is observed). Liquid nitrogen calibration does not change so much the calibration for these channels as they are in general well calibrated by the hot load calibration (every 5 minutes but I do not know if this is the case for the Radiometrics).

Yes, the opaque channels have been changed less than the transparent ones by applying the bias-correction, so the initial difference between the two radiometers for these channels almost did not change. Additionally, the MWR_NOAA Tb had a problem with measurements in the 57.288 GHz channel (with an initial ~ 5 K difference with the MWR_CU Tb) but that difference was reduced in half with the bias-correction.

The Tb difference between the CU and NOAA radiometers for all available dates (34 days) after applying the bias-correction for the opaque channels (>56.5 GHz) still shows a difference around 2-2.5 K, while for the transparent channels (52-56 GHz) the differences are mostly improved with the final biases < 1.5 K. Finally, the K-band channel biases were not extremely large even without bias-correction, and after bias-correction they are less than a degree K different.

→ In figure 6, we can also see that NN retrievals using oblique measurements manage to improve NN with zenith only below 700 m, the degradation appears above 1 km when transparent channels are used and are more subject to large biases. Thus, the use of opaque channels below 700 m does not seem to degrade NN retrievals as much as shown for PR below 200m in figure 5. We observe the same thing in figure 2 : if we look at the NN retrievals, there is a significant modification of the profile below 250m when including oblique measurements that we do not observe with the physical retrievals.

In order to confirm your hypothesis, could you check the biases for oblique measurements as it is done in figure 1 ? If you compared to simulated TB from radiosondes and assuming homogeneity in an area around ~ 1km from the instrument, could you re-use the RS to investigate more in depth the biases at low elevation angles (as it is done in figure 1) to confirm this hypothesis ? Alternatively, you could also use model data (analysis or very short-term forecasts) during clear-sky conditions similarly to the paper of De Angelis et al 2017. I think this check is very important to confirm your conclusions lines 507 and 521. Depending on your answer about the channels used at oblique measurements, did you try to restrict MWRzo to only the most opaque channels (very close to 58 GHz)? It would be interesting to identify if the supposed bias occurs for all V-band channels and/or only the most transparent ones

Following the reviewer's suggestion, we compared the Tb measurements from the opaque channels for the same time shown in Fig.5 of the manuscript with Tb calculated by the forward model applied on the radiosonde data, Fig.R2 below:

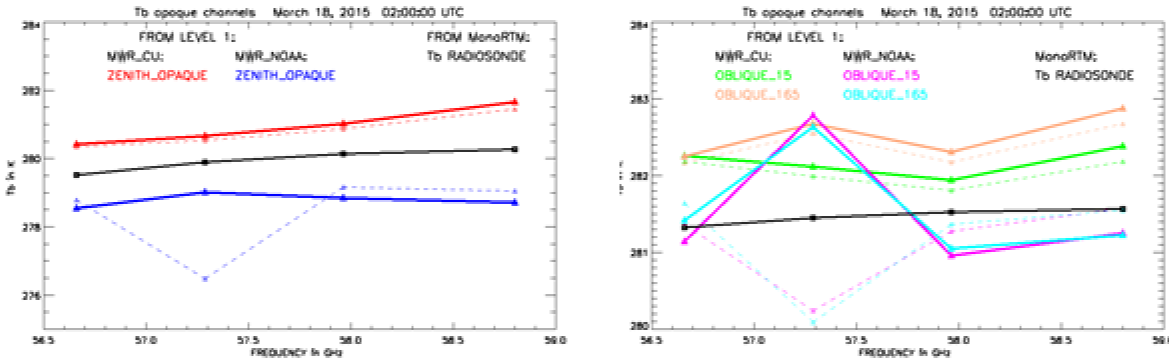


Fig.R2. *March 18, 2015, 02:00 UTC: the Tb of the opaque channels, 56.66, 57.288, 57.964 and 58.8 GHz, from the zenith scans (left panel) of the MWR_CU in red and MWR_NOAA in blue and from both oblique scans (right panel) in colors, and the Tb from the MonoRTM forward model using the corresponding radiosonde profile in black. Dashed color lines mark the original Tb data and solid color lines – bias-corrected Tb.*

For this time period the bias-correction does not improve the Tb observed by the MWR compared to those derived from the radiosonde, the bias-corrected Tbs are further from the radiosonde Tbs compared to the uncorrected Tbs, except for the 57.288 GHz channel of the MWR_NOAA that shows measurement problems before bias correcting it. We have to admit that radiosonde Tb data cannot be claimed as the "true" because these data are the output of the forward model that has its own uncertainty.

The bias-correction in general improves the temperature profiles for most of the test time reducing the bias in the 1-2 km AGL layer by 0.5 K for all PR averaged profiles.

Additional text included in Section 3.2 (with some editing): "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."

Fig.R2 shows the bias between the opaque channels' Tb and radiosonde-derived Tb. While these differences are similar in absolute values, but of opposite sign, for the zenith scans, the oblique channels show a noticeable difference between MWR_CU and MWR_NOAA Tbs compared to radiosonde Tbs (Fig.R2, right). These differences resulted in very different PR profiles from the two radiometers, as shown in Fig. R3:

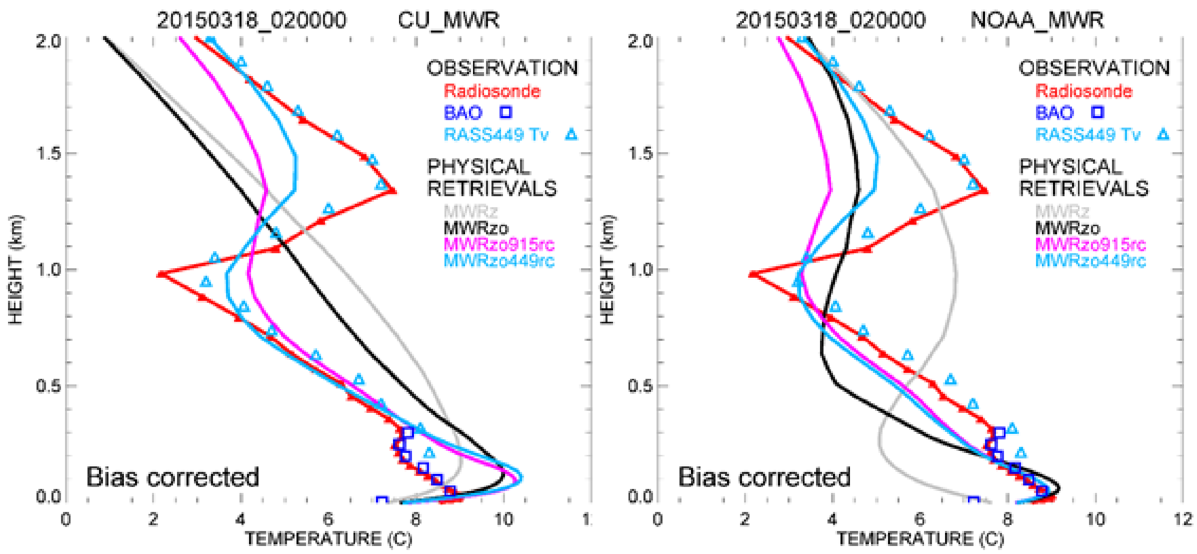


Fig.R3. *March 18, 2015, 02:00 UTC case. Observations from radiosonde are in red, and from BAO seven levels – in blue squares. The four PR profiles are in gray (MWRz), black (MWRzo), magenta (MWRzo915) and light-blue (MWRzo449). PRs from the MWR_CU are on the left and from the MWR_NOAA - on the right.*

According to the right panel of Fig.R2, the MWR_CU has a bigger Tb bias for the opaque channels from the oblique scans compared to the MWR_NOAA measurements for this case, that resulted in MWR_NOAA temperature profile to be closer to the radiosonde profile in the layer of 0-300 m above the ground, shown in Fig.R3.

Still, with the measurement problem in the 57.288 GHz opaque channel of the MWR_NOAA instrument, and because of its limited time availability during the XPIA campaign, we decided to limit our analysis to the MWR_CU data only.

Second major comment about NN retrievals: Line 347 you mention that you cannot un-bias the BT from neural network. I can understand especially if you did not train the neural network by yourself but I think this is a major concern in all your evaluation of the next sections. We can see that NN retrievals have a degraded accuracy due to an increase bias above 1 km altitude which is probably due to the large V-band bias for transparent channels. However, after this small remark line 347, you never discuss this issue again. I think it is not fair when you compare with the PR which takes into-account a bias-correction which is very large for transparent V-band channels. At minimum, the authors should always remind this limitation to the reader : the problem might not be due to the NN approach itself but to a biascorrection that needs to be applied to NN retrievals similarly to PRs (you should also cite Martinet et al, Tellus, 2015 which shows how NN bias can be decreased after bias correction).

I am also wondering if, through the manufacturer software, you could re-process the NN retrievals by modifying the binary of TB files including the bias that you provided in figure1. This should be feasible and at least would give some ideas if the NNs are improved when using the same BT as for PRs (but keeping in mind that your bias correction for NN would not be perfect as probably a different RTM has been used to deduce the bias and train the NNs).

I also only understood at the end of the paper that the green line for the NN oblique measurements never use zenith observations. Thus, I assume NN with oblique measurements only does not use transparent channels as this would violate the homogeneity assumption. So, it is totally normal that the bias of NN with oblique measurements is degraded above 1 km altitude…If NN with oblique measurements only use opaque channels at low elevation angles, all your results to compare with NN retrievals should combine the two temperature profiles that you obtain: the one from zenith only mainly above 1 km altitude and the one obtain from oblique measurements below 1 km altitude. This has to be done if you want to compare with the configuration MWRzo which uses both zenith and oblique measurements. If I also understood correctly that zenith observations are not used for NN retrievals I think that figure 8 should stop at 1 km above ground maximum and not 5 km. Either you want to go up to 5 km altitude and you need to create a composite temperature profiles from the NN retrievals and make again your statistics with this new profile. Or you should limit your averaging of the bias and RMSE up to 1 km altitude because you cannot take into account statistics from the NN which are biased because they do not use observations informative of higher altitudes (or observations which are not bias corrected like the PRs).

We thank the referee for this particular comment. Following the very insistent recommendation of Reviewer #2 and your questionable opinion about the temperature comparison of bias-corrected input for the PR data and uncorrected NN data, we decided to move all comparisons of PR and NN profiles to Appendix A. The suggestion about NN bias-correction using the Tb biases from PRs looks interesting, but we decided not to mention it because of the artificial mix of two approaches. Instead, we included (in Appendix A) the comparison of PR profiles with separate NN profiles from zenith and from oblique averaged scans and with NN profiles calculated from the combination of the scans using NN oblique scans up to 1 km and NN zenith scans above.

Technical corrections:

Introduction, line 109 : I think the sentence is a bit too long and complex to follow. The radiative transfer equations are in general used to train the neural network retrievals or used directly inside physically-based retrievals whereas from the sentence it seems not connected. I think the sentence would be more rigourous rephrased that way :
« in order to estimate profiles of temperature and humidity from observed brightness temperatures, they apply regressions, neural network retrievals or physical retrieval methodologies which include more information about the atmospheric state in the retrieval process. Radiative transfer equations are commonly used to train statistical retrievals or as forward models inside physical methods».
Rephrased as suggested.

Introduction line 116 : I do not agree with the argument that MWRs have a limited accuracy due to the fact that they do not actively measure temperature and humidity profile. We can of course improve their retrievals but it is hard to find sensors with accuracy better than 0.5 to 1.5 K during all conditions for temperature. I agree with the other drawbacks (lower accuracy during rain, coarse vertical resolution especially) but not with that one or you should give more arguments.
We deleted the comment on the accuracy of temperature and humidity measurements.

Introduction line 121 : site specific climatology is only a disadvantage for regressions or neural networks. This is not the case when using 1D-Var retrievals combined with model outputs. I think it would worth mentioning a few reference papers using 1D-Var approaches combined with NWP model : Hewison 2007, Cimini 2011, Martinet et al 2020 etc..

We have now added the following text in the Introduction, together with the mentioned References:

“Some studies have used analyses from NWP models as an additional constraint in these variational retrievals (e.g., Hewison 2007, Cimini 2011, Martinet et al. 2020); however, we have elected not to include model data in this study because we wanted to evaluate the impact of the RASS profiles on the retrievals from a purely observational perspective “

Introduction line 125 : The literature refers more to low accuracy of MWR LWP retrievals for values below 20 g/m², 50g/m² seems a bit overestimated please modify or provide a reference for this statement.

Changed from 50 to 20.

Introduction line 142 : add an « s » to lowest several kms.

Included.

Section 2, line 172 : change included into including.

“Included” is right.

Section 2, line 196 : change manufacturing into manufacturer.

Changed.

Section 2.1, line 203 : Please correct into : « NN zenith and of the NN oblique measurements. »

Included.

Section 2.1, line 205 : can you mention the date of the last calibration with liquid nitrogen for the data used in the paper ?

Prior to the experiment, both MWRs were calibrated using an external liquid nitrogen target and an internal ambient target and thoroughly serviced (sensor cleaning, radome replacement, etc.). The MWR used in this study was serviced and calibrated on 2/27/2015. This text was included in the manuscript.

Section 2.2, line 221 : can you mention in which conditions RS were launched (how many clear-sky or cloudy-sky?)

Of 58 valid radiosonde profiles, 41 were launched in clear-sky periods, 12 - in cloudy periods, and 5 during rain. We defined those categories using Tb in the 30 GHz channel, as shown in the figure below:

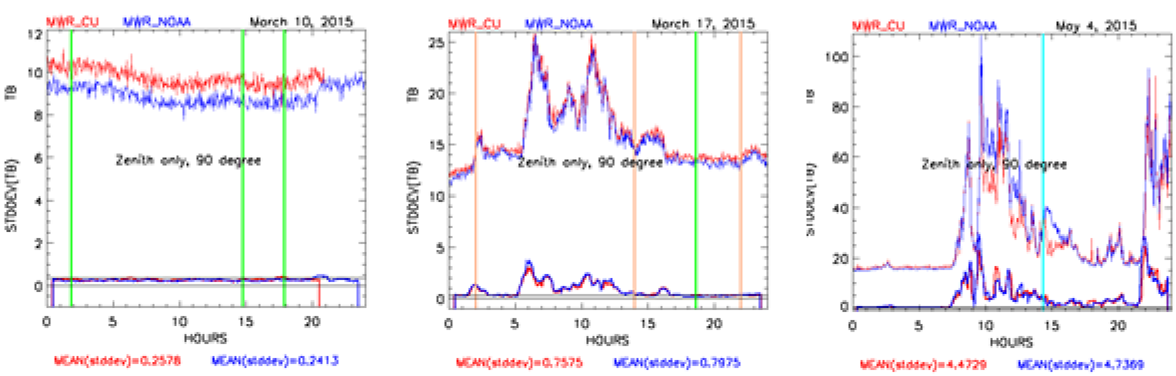


Fig.R4. *Zenith Tb from the 30 GHz channel for a clear-sky day (left panel), cloudy day (middle panel) and rainy day (right panel) from the CU radiometer in red and NOAA radiometer in blue. STDDEV(Tb-*

SMOOTH(tb,11)) is shown at the bottom of each panel with its average values printed under the panels in corresponding colors. Vertical lines (green – for clear-sky, beige – for clouds and cyan – for rain) show the time of radiosonde launches.

We also included the following text in the manuscript:

"Four clear-sky periods have been chosen using a criterion of less than 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."

Section 2.3, line 225 : Please correct same location as the MWR.

Corrected.

Section 3.1, line 270 : please specify : integrated content of liquid water

Included.

Section 3.1, line 282 : could you add some spaces between the Sa matrice and the specification of the Jacobian Kij ? Could you also specify in this notation what is i and j ? (I assume channel and vertical level). Could you be consistent with the definition of Xa line 267 (always use L for LWP or only LWP everywhere) ?

Xa and Sa are changed (from L to LWP). Jacobian is moved to form the straight-line definition. Notations of "i" and "j" are included.

Section 3.1, line 294 : can you say a word on how the Sa matrix has been computed ?Section 3.1, line 296 : can you mention the perturbation size that you used to compute your Jacobeans ?

We included the additional description of the Sa matrix in the text in Section 3.1: "Using 3,000 radiosonde launched by the NWS in Denver, we interpolated each profile to the vertical grid used in the retrieval, after which we computed the covariance of temperature and temperature, temperature and humidity, and humidity and humidity for different levels."

Section 3.1, line 300 : could you please mention which MWR channels are used in the retrievals for zenith only and for oblique measurement ? (all of them or just a sub-sample .).

As mentioned earlier, 22 channels were used from zenith measurement and 4 channels (opaque) – from oblique (included in Section 3.1).

Section 3.1, line 312 : could you mention the uncertainty values used in the Se matrix ?

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 K 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 1.5 K at low SNR values. Again, we assumed that there was no correlated error between different RASS heights.
These additions are also included in Section 3.1.

Section 3.1 and table 1 : Does Tbzenith-oblique means both TB measured at zenith and at oblique elevation angles ? If this is the case, why there is a cross at the column indexed « Tbzenith » too ? It is a bit confusing as it seems that Tbzenith is used twice in the retrievals which I assume is not the case. Could you clarify this point in table 1 but also line 283 in the Se matrix ?
Table 1 as well as observational vectors Y2, Y3 and Y4 and matrix Sε have been modified.

Table 1 with its modifications has already been shown above. Vectors Y2, Y3 and Y4 and matrix Sε are modified as follows:

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

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

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

Section 3.1, line 278 : the sentence is confusing. It seems equation (1) is here to show how the Y vector is estimated from the state vector X whereas equation (1) shows the new atmospheric state updated at each iteration of the minimization depending on the previous state, the different matrices (Sa, K, Se) and the forward model. Please correct the sentence accordingly so that it makes more sense.
Corrected in Section 3.1: "The MonoRTM model **F** is used as the forward model from the current state vector **X**, Eq. (1), and is then compared to the observation vector **Y,** iterating until the difference between **F(X)** and **Y** is small within a specified uncertainty."

Section 3.1, line 313 : please correct the sentence into : « its dimension increases ».
Done.

Section 3.2, line 319 : please correct into « will contribute to a bias in the retrievals ».
Done.

Section 3.2, line 328 : could you mention what thresholds and criteria you used from the 30 GHz Tb to identify clear-sky periods ? (standard deviation over which time period and which threshold?)
This text was added to Section 3.2 (with some editing):
“A threshold value of 0.3 K has been used for the uncertainty calculation. Fig. R5 (see below) shows one of the clear-sky days, March 10, 2015. The final uncertainty equals the average of the Tb standard deviation in a one-hour window sliding through all data points of a day. It also could be computed as the standard deviation of the difference between Tb and 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 previous experiments.”
“For the four chosen clear-sky days not only the daily uncertainties of **30 GHz Tb** were below 0.3 K, but all three sets of uncertainties described above were extremely similar with the averaged difference less than 0.05 K.”

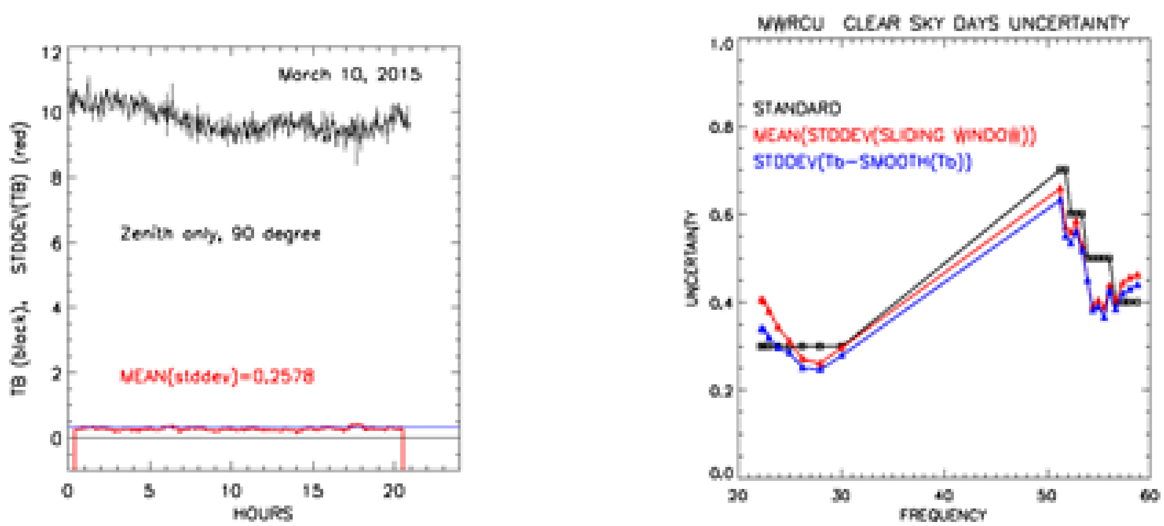


Fig. R5*. Left: Tb from MWR_CU 30 GHz channel for March 10, 2015, one of the chosen clear-sky days. The standard deviation (at the bottom, in red) is calculated as the averaged standard deviation of Tb in a one-hour window sliding through all data points of the day. Right: MWR_CU uncertainty, computed as an average over four clear-sky days using a sliding window (in red), smooth function (in blue), and the before mentioned “standard” values (in black) for all 22 channels.*

Section 3.2, line 333 : How the bias is computed ? Is it a difference with simulated BT from radiosondes ? Can you please clarify this in the manuscript.
From the modified text in Section 3.2:

“The bias was computed for each of the 22 channels as the averaged difference between the 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.”

Section 3.2, line 345 : can you at least mention that NN biases could be improved by applying a bias-correction ?.
We moved the NN and PR comparison in Appendix A and mentioned this possibility.

Section 3.2, figure 2 : Can you specify if it is a clear-sky day or a cloudy day ? I suspect that this is a cloudy day with elevated inversion which often causes trouble to MWRs. If possible, a comparison with a clear-sky day by night with a sharp temperature inversion close to the surface could be interesting too. Could you say a word in the manuscript why you are have 0.5 to 1K difference between the RS measurements and the BAO tower measurements which are used a the « truth » for validation ?

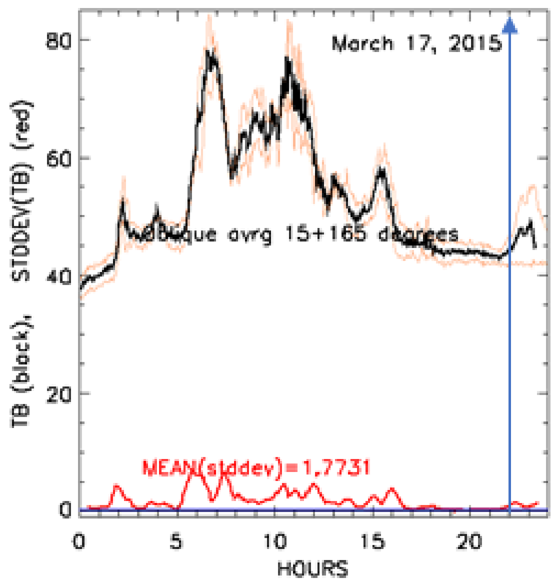


Fig.R6. *Oblique channels Tb from 30 GHz channel, March 17, 2015. Blue arrow marks the time of the day, 22:00 UTC, for the radiosonde case shown in Fig. 2 of the manuscript.*

This is the 30 GHz channel Tb from the oblique scans. A difference between the two scans of 15 and 165 degrees at 22:00 (time of the radiosonde launch) just started to grow that may indicate the cloudiness in the view of one of the obliques.

Fig. 5 in the manuscript shows exactly one of the difficult cases you are mentioning: evening hours with sharp temperature inversions, one of them close to the surface.

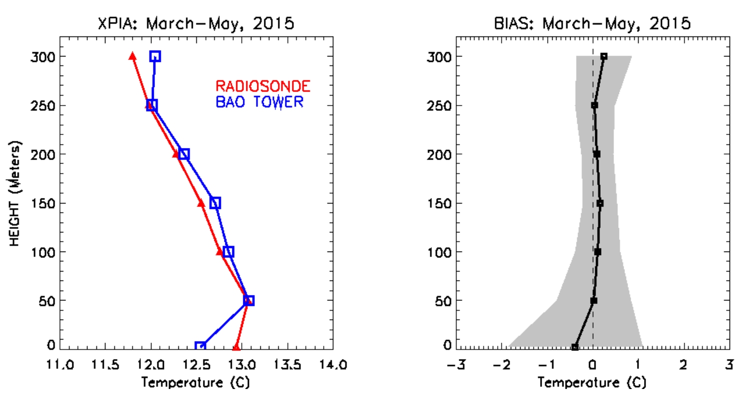


Fig. R7. *Averaged temperature at BAO tower heights from radiosonde (red) and from BAO levels (blue) in the left panel and their biases at each level with shaded image of standard deviation over 58 radiosonde launches in the right panel.*

We indeed use BAO measurements as the “truth” having very close agreement between the radiosonde and BAO measurements. The special case of Fig. 2 in the manuscript has larger differences between the radiosonde and BAO, which on average were less than 0.5 K, which is within the expected accuracy of the radiosondes.

Section 3.2 line 366 : Modify the sentence into « demonstrate a better agreement ».
In the text now: “the MWRzo449 profile (in light-blue) demonstrates a better agreement”

Section 3.3, line 388 : please rephrase into « Akernal provides useful information ».
Done.

Section 3.3, line 425 : please correct vs into versus.
Included “versus”.

Section 3.3, figure 3 : As it is, Panel a) does not sound really relevant to me as it is the same as figure 2. However, in this section we would expect to see the smoothed RS profiles for the two configurations selected (MWRzo and MWRzo449). Could it be added to panel a) ?
Can you also explain why you get a strange vertical line in the Atkernel on the left part of the figure ?
The smoothed Radiosonde profiles from MWRzo and MWRzo449 are included in panel R8a).
First left vertical lines in panels R8b-c) indicate surface data (see the definitions of observational vectors Y). To confirm this, we repeated those runs without including surface temperature and humidity data in the observational vector. This indeed caused the disappearance of the vertical lines in Fig.R8b, c (not shown).

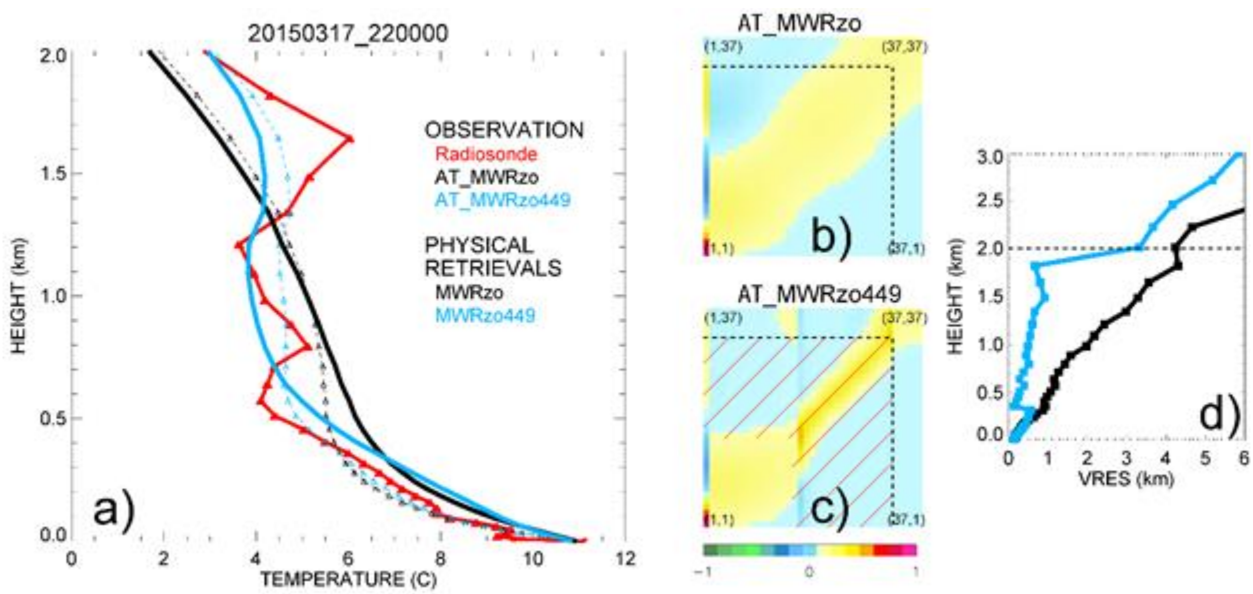


Fig. R8. *The same as Fig.3 in the manuscript with two changes: T radiosonde profiles smoothed by AT_Kernel in MWRzo (dashed black) and MWRzo449 (dashed light-blue) are included in panel a), panel d) shows Vertical resolution calculated by FWHM method. These changes are included in the manuscript.*

We also change the panel d) in this Figure by changing 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. This method 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 was 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, 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

Section 3.3, line 437 : change dash lines into dashed lines.
Changed.

Section 4.1, line 468 : to be consistent add a space to 1km => 1 km
Added.

Section 3.3, figure4 : can you explain why MWRzo915 does not make any improvement of the vertical resolution above ~600m compared to the MWRzo ? From panel c) it seems the spread around the diagonal is significantly reduced compared to MWRzo. However, the black and purple lines are almost on top of each other in panel e).
The reason why the vertical resolution of the MWRzo915 is very similar to that of the MWRzo above ~750m is explained by the fact that above this height much fewer RASS measurements are available (as in fact presented in Fig. 10), therefore the positive impact brought by the inclusion of RASS measurements is greatly reduced above that height.

Section 4.1, line 479 : change dash lines into dashed lines.
Changed.

Section 4.1, line 531 : add a space to « 5 km » to be consistent through the manuscript.
Added.

Section 4.1, line 535 : change as good as that during XPIA into « as good as during XPIA ».
Deleted "that".

Section 4.2, line 544 : please changed into « smoothed radiosonde using the averaging kernel matrix ».
Changed.

Section 4.2, line 566 : change « above and below 1.5 km » into « by up to 5 km AGL »
We think it is important to refer to the 1.5 km height because this is the maximum height reached by most of the RASS measurements.

Section 4.2, line 567 : change statistical measures into statistical scores.
Changed.

Section 4.2, line 567 : I do not understand this sentence which is in contradiction with the previous one. Line 566 you mention that statistical scores are very different for all PRs but then line 567 that above 1.5 km AGL they are similar. What do you mean? Please correct the text accordingly.

There is no contradiction in these lines. We use a separation level 1.5 km to highlight the different behavior of the scores: all profiles are more smoothed and uniform above 1.5 km (with MWRzo449 having the best RMSE and BIAS) but less so closer to the surface.

Section 4.2, line 570 : Please change « NN retrievals are very variable » into « the accuracy of NN retrievals is very variable ».
Changed.

Section 4.2, line 571 : Your conclusion is only true above 1 km altitude, below 1 km altitude, NN retrievals perform better than MWRz and MWRzo and even the two configurations with RASS measurements. The degradation of NN retrievals above 1 km is mainly due to a large bias which might be due to the fact that you do not apply the bias correction to MWR measurements for NNs whereas you apply it to the PRs. This needs to be justified and clearly stated here. Linked to my previous comment, I do not understand how NN retrievals can be improved below 1 km with oblique measurements whereas you concluded in section 4.1 that oblique measurements present a large bias. Additionally, the MWRz using only zenith measurements also present a large bias (above 1 K) below 1 km altitude which seems to conclude that probably opaque channels are biased both at zenith and oblique measurements. Could you also comment on the degradation of the accuracy of MWRzo915 between ~ 200 m and 1 km ? In figure 5 you showed an example were the RASS 915 measurements were able to improve temperature retrievals of MWRz and MWRzo above 200m but averaged over all the profiles it is not the case any more. It seems to come from a bias in your retrieval that we do not observe with MWRzo 449.
Comparison to NN profiles are moved to Appendix A where the reviewer's questions have been addressed.
The degradation of MWRzo915 above 200m is also seen in Fig. 10 of the manuscript. While the availability of RASS 449 data is almost constant from 300m to 1.6 km, RASS 915 data availability faded quickly in height with its reduction from 100% availability at 300 m to almost 10% at 1km.
Fig. 5 shows the most complicated temperature profile during XPIA, it is also a very interesting case in terms of all possible active measurements' availability, from both RASS 499 and RASS 915.

Section 4.2, figure 6 : I think the vertical blue and red lines to identify a correlation of « 1 », perfect RMSE of 0 and bias of zero are confusing for me. The figure being already crowded, I would remove these additional coloured lines for only a vertical black dashed line for panels c and f only.
Vertical lines in Fig. 6 will be changed to black in the new version.

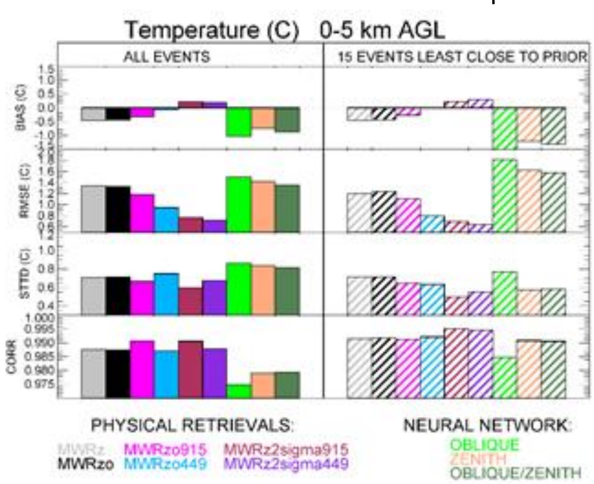
Section 4.2, figure 7 : I am surprised that you use oblique measurements from the MWR for humidity retrievals: can you comment on the fact that this probably violates the homogeneity assumption necessary to use low elevation angles? In general, only opaque channels are used at low elevation angles and they are not sensitive to water vapor. If you used low elevation angles, did you apply a quality-control to detect inhomogenous cloudy scenes ?

Tb obliques data are used as an average from two scans, 15 and 165. We note that most of the radiosonde launches were made in periods without liquid clouds, so the oblique scans should be similar. Also, the K-band channels from the oblique scans are not used in the retrieval, thus spatial variability in water vapor is not an issue. We only use the more opaque V-band channels for the oblique scans. Therefore we believe that our calculation of humidity retrievals is valid.

Section 4.3, line 620 : what do you mean by « weighted average over the 42 vertical heights » ?
The following text is included in the text:
"The vertical resolution of the Physical Retrievals is not uniform, with more frequent levels closer to the surface. If the data from all levels are used as the simple average, the near-surface layer will be weighted more compared to the upper levels of the retrievals. To avoid this, a vertical averaging in 0-5 km profiles is performed with separate weights at each vertical level calculated by the distance between the levels."

This is a very common validation procedure over some slice of the model with uneven vertical resolution.

Section 4.3, figure 8 : Could you comment about the potential modifications to your figure if you had calculated the statistics up to 1 km or 2 km AGL instead of 5 km ?
As you do not apply a bias correction to the NN retrievals where V-band transparent channels have a strong bias I am wondering if the conclusions are not wrongly biased for the evaluation of NN retrievals with this averaging up to 5 km. As already mentionned previously, it is not fair to compare two retrievals not applied on the same dataset (one with bias correction, another one without). At least this should be again commented when discussing the results of this section.
We made the statistical evaluation of temperature profiles up to 1, 2 and 5 km heights (see Fig. R9).



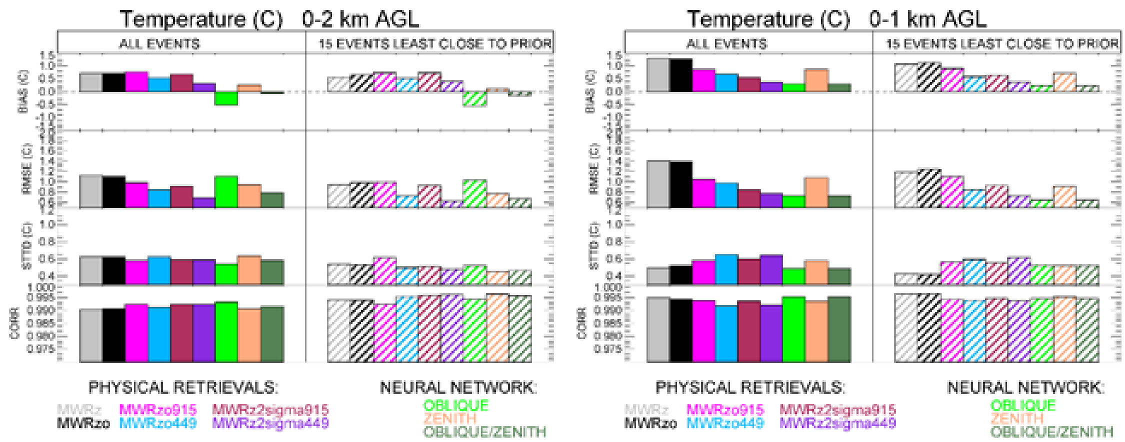


Fig. R9. *On each double-panel plot from top to bottom: biases (retrievals minus ATkernel radiosonde), RMSEs, standard deviations of the difference between retrievals and ATkernel radiosonde, and Pearson correlations for the six PR configurations and three NN retrievals, oblique, zenith and their combination, averaged from the surface to 5 km AGL (top), to 2 km AGL (bottom left) and to 1 km AGL (bottom right), and averaged over the 15 events furthest from the priors (hatched boxes).*

Statistical analysis shows similar behavior for the PR configurations in terms of RMSE for all three vertical layers. For NN statistics, we included a third type of comparison against the radiosonde measurements, the combination of the oblique scan temperature profiles up to 1 km AGL and the zenith scan temperature profiles above 1km AGL. This combined NN has the lowest RMSE compared to the other two NN scans considered separately. Also, these combined NN profiles have the lowest RMSE in the lower layer of 0-1 km compared to all PR profiles, but larger RMSE in wider atmospheric layers such as 0-2 or 0-5 km. All three NN retrievals (oblique only, zenith only, oblique and zenith combined) have the highest RMSE compared to all PR configurations in the layer of the atmosphere up to 5 km. From the PR temperature profiles, the RMSE decreases from the passive instrument configurations (MWRz, MWRzo) to the configurations with active RASS measurements in very similar ways over the 0-5, 0-2, and 0-1 km atmospheric layers, especially when comparing the 0-1 and 0-5 km layers of the atmosphere. Bias also improves from MWRz/MWRzo to the configurations that include RASS. The setting of MWRz2sigma449 shows the best statistics in terms of bias and RMSE compared to all other PR retrievals, and better to all three NN retrievals in 0-2 and 0-5 km layers. In general, almost all PR profiles with RASS have RMSE below 1 K in all three vertical layers.

Conclusion line 703 : I honestly did not have understand that NN retrievals with oblique measurements do not use the zenith observation. This has to be more explicit directly in section 2.1, line189 to 201. This explanation arrives too late in the manuscript

Comparison to NN profiles are moved to Appendix A, where we clarified the difference between the NN configurations.

Conclusion : line 718 when MonoRTM is mentioned is redundant with line 719. I suggest modifying lines 717 to 719 into :

the small systematic errors that exist between the MWR observed Tb values and the RASS measurements and (b) the systematic errors that exist in forward microwave radiative models.

(I would thus remove all the text between parentheses).

Modified.

Conclusion, line 722 : please correct the most difficult to retrieve and the most important to forecast.
Corrected.

Conclusion, line 728: this sentence should be mitigated : the study proves that active sensors can improve MWR passive observations with zenith observations only but due to the weird results you obtain with lower elevation angles which are expected to improve the retrievals in the same area as the RASS measurements I think you should mention that the results could be different with MWRs with elevation angles usable down to 5° above the ground. In fact, with new MWR instruments using both zenith and low elevation angles we can expect RMSE between 0.5 and 1.5 K in the first 2 km (1.5 K for cloudy-scene when there is a temperature inversion in the upper layers). Thus, we cannot be sure that the improvement brought by RASS measurements would be as much informative in the first 1 km with a MWR unit for which oblique measurements could be optimally used. I think you should mention this in your conclusion.
The text in the manuscript is modified as:
“Even for this subset of selected cases we find that MWRz2sigma449 produces better statistics, proving that the inclusion of active sensor observations in MWR passive observations would be beneficial for improving the accuracy of the retrieved temperature profiles also in the upper layer of the atmosphere where RASS measurements are not available (at least up to 5 km AGL). However, we note that this result may be dependent on the fact that our oblique measurements were taken at a 15 degree elevation angle, and that MWRs in locations with unobstructed views allowing for scans down to 5 degrees may provide similar improved accuracy to the temperature profiles (reference below) in 0-1 or even 0-2 km AGL layers.”

Crewell, S., U. Löhnert, 2007: Accuracy of Boundary Layer Temperature Profiles Retrieved With Multifrequency Multiangle Microwave Radiometry, IEEE TGRS, VOL. 45, NO. 7, JULY 2007, **DOI:** 10.1109/TGRS.2006.888434