Review AMT-2021-9:
Improving thermodynamic profile retrievals from microwave radiometers by including Radio Acoustic Sounding System (RASS) Observations

ANSWERS to REFEREE 1

General Comment:
The authors have replied to most of my comments and most questionable aspects. The new version where the discussion on NN retrievals is moved to a dedicated Appendix seems to be better suited due to the lack of bias-correction for the NN retrievals but still interesting results to show. I still recommend to answer a few questions about the MWR data bias correction. As biases of the data used in this study are quite large, I think it is important to clearly discuss and clarify the choices made by the author.
After answering these few points, I would recommend the publication of the manuscript to AMT.

We wish to thank very much both Referees for carefully reading our manuscript and for offering many comments towards its improvement. In revising the manuscript, we have considered all of their comments. Many changes have been applied to the current version of the manuscript and we believe that our manuscript benefited significantly from the constructive comments made by both the Referees. Please, find below our point-to-point answers in red.

Most of the figures are also of poor quality and difficult to read in the pdf version. The authors mention that it is only due to the pdf formatting but original figures are of good quality. I let the editor check this point in the final version.

In the revised version of the manuscript we have made sure to present figures clearly, both in the format and in the quantitative discussion of the results. Same for the equations. Nevertheless, in the final version of the manuscript we'll make sure to provide the Editorial office with the high-resolution version of the images in *.eps format and Equations from Microsoft's equation editor.

1) I am still a bit puzzled by the bias-correction calculation. This is an important aspect of the manuscript as the authors mention the large bias affecting MWR opaque channels of the MWR_NOAA unit that could make sub-optimal the analysis presented in this paper.
If I understand properly the bias has been computed from differences with the « prior » profiles and thus a climatological monthly mean. Biases in the most opaque channels are significantly affected by the accuracy of the boundary layer temperature profiles used in the simulation by night when low level inversions are present. I am not convinced that a monthly average could correctly infer the bias in the most opaque channels if such cases (low level temperature inversions) are taken into account in the statistics. Similarly, K-band and transparent V-band channels are sensitivity to the integrated water vapor which might be poorly represented by a monthly average too.

We have taken this comment from the Referee very seriously and have modified large parts of the manuscript to address these concerns. The Referee is correct that a dataset like XPIA has the advantages to have many radiosondes that could be used instead of the prior to reduce any observational systematic offsets in the MWR Tbs. While this is true for the XPIA dataset, for other field campaigns this approach might not be possible because radiosonde observations are not always available. For this reason, the Section regarding the bias correction of the MWR measured Tbs has been completely rewritten and renamed *"3.2 Bias-correction of MWR observations using radiosondes or climatology"*. In this section, two bias-correction methods are now presented, the first using radiosonde data (radiosonde BC) and the second using the prior (TROPoe BC). Through the revised manuscript, results using the two different bias-correction procedures are presented and analyzed. As expected, we find that the radiosonde BC method gives retrieved profiles closer to the radiosonde temperature profile than when using TROPoe BC.
Although the more accurate radiosonde BC method has been included in the revised version of the manuscript, the final goal of this study is not to assess the sensitivity to different bias-correction approaches, but to verify that the inclusion of RASS observations does improve retrieved temperature profiles. We show that this result is still valid independently of the bias-correction method used.

•Could you explain why clear-sky radiosondes profiles of the campaign could not be used to infer this bias correction?

Following the Referee's comments, as mentioned in the answer above, in the revised version of the manuscript two bias-correction methods are now presented, the first using radiosonde data (radiosonde BC) and the second using the prior (TROPoe BC). When using the radiosonde BC method, the procedure to identify the clear-sky days is now one-hour centered around the radiosonde times. On the other hand, when using the TROPoe BC, to identify clear-sky days we keep using the same procedure as before, to consider a field campaign scenario that does not benefit from so many radiosonde launches as the XPIA campaign.

• Did you investigate the sensitivity of the bias-correction to the database used to derive this correction (prior profiles versus radiosondes)?

Following the Referee's comments, as mentioned in the answers above, we did indeed investigate the sensitivity of the bias-correction to the database used to derive this correction (radiosonde BC versus TROPoe BC), including the results obtained from both approaches in the revised version of the manuscript. A direct comparison of the different biases is shown in Fig. 1 of the revised manuscript.

•From figure R2 of your answer: are you convinced that the bias is the same at zenith and oblique scans for this case study (I am sorry but this is very hard for me to read the numbers in the y-axis to check if the values look approximately the same at zenith and oblique scans)?

Following this Referee's comments and also the comments from the other Referee, we did measure the bias in the MWR observed Tbs (in all 22 MWR channels of the zenith scan, and in the four opaque channels of the oblique scans, i.e. the channels used by TROPoe), in comparison to radiosonde Tbs (radiosonde BC) and in comparison to the TROPoe Tbs (TROPoe BC). These results are now presented in Fig. 1 of the revised manuscript, which we hope it is easier to read. The following description of Fig. 1 has been included in the text: *"The biases from the two bias-correction schemes are within the uncertainties of each other for most of the channels except at the higher frequencies in the V-band. Biases in the most opaque channels are significantly affected by the accuracy of the boundary layer temperature profiles. When TROPoe BC is used, a monthly average prior temperature profile is used in the PR, and thus differences between this prior profile and the actual temperature profile can result in a spectral bias in the more opaque MWR channels. On the contrary, the radiosonde BC uses a direct measurement of the temperature profile (from the radiosonde), and thus is more accurate. It is also important to note that, in both approaches, the biases in the opaque channels for zenith and for oblique scans (for radiosonde BC these are red and blue, respectively; and for the TROPoe BC these are black and green, respectively) are very similar to each other.* ***This supports the assumption that the true bias is nearly independent of the scene, or that the sensitivity to the scene (e.g., zenith or off-zenith) is small."***

•Did you check that the calculated biases were approximately the same through the whole period at zenith and oblique scans as you mention line 420?

As shown in Fig.1 of the revised manuscript, the standard deviation over all clear-sky radiosonde profiles (red error bars for the zenith scan) are around 0.1-0.2 K in the opaque channels, which demonstrate that the biases do not change much during the duration of the XPIA campaign.

•Figure5: In your answer, you show that the oblique scan at 165° is probably affected by a cloud that is probably not yet detected at 15°. Could not that be a problem in the retrieval to mix two TB measurements: one-clear-sky and one cloudy-sky? How does the PR handle this? I think resolving elevated temperature inversions often observed during stratus clouds is already challenging for MWRs but probably even more if you mix two scans one in clear-sky and one in cloudy-sky.

This is a very good point. In the revised version of the manuscript we identify the clear-sky period looking at the zenith scan and the averaged oblique scans separately. Data from the zenith scan indicates more than 35 radiosonde profiles in clear-sky conditions, but the same evaluation of the oblique scans reduces this number to 18. We decided to use only these 18 radiosonde profiles so that we are sure that all scans (15°, 90°, 165°) observe clear-sky conditions.

•Line 601: you mention that the bias appears « in this case »: can you just clarify that it is probably affecting the whole time series and not only the case shown in figure2

According to the data in Fig. 1, the biases from the two bias-correction schemes are within the uncertainties of each other for most of the channels except at the higher frequencies in the V-band, where there are almost 1°C differences between the TROPoe biases and radiosonde biases for four opaque channels. The radiosonde data are the direct measurements, so as TROPoe Tb biases do not match radiosonde biases for opaque channels, we may confirm that TROPoe biases couldn't correct Tb data perfectly and therefore the cases like one shown in Fig. 5, are not unique but could affect the whole data set.

1) In figure R3 of your answer you clearly demonstrate that the problem of the MWR retrievals comes from the large bias in the MWR_CU unit that would be mostly solved with the MWR_NOAA unit. I would remove channel 57.2884 GHz from the analysis as this channel is likely affected by a hardware problem. But I think including this result in your manuscript would be very interesting to clearly demonstrate your hypothesis for the reader (like it is clearly demonstrated in your reply)

We decided to keep using the MWR CU units in the study only because the other unit was unavailable for almost the entire month of April. Nevertheless, we do believe that following the Referees' comments (i.e., including the radiosonde BC method to the study) largely demonstrated that these biases could be corrected (see Fig. 5a of the revised manuscript), and the assumed hypothesis, that the inclusion of the RASS observations to the retrievals improves the temperature profiles, could still be verified.

2) line 318: The uncertainty in MWR observations is evaluated as the standard deviation of Tb measurements during clear-sky measurements. Firstly, could you specify the time window for this uncertainty calculation and secondly could you discuss about forward model uncertainties? Are these taken into account in the observation error covariance matrix?

The uncertainties in MWR observations are calculated differently for two bias-correction methods but both methods use the time-series data in the 30 GHz channel to identify the clear-sky time period. For radiosonde BC, the 18 launch times are chosen based on the small (<0.4°C) standard deviation of Tb in one-hour data centered at the radiosonde launch time, from both scans, zenith and oblique. TROPoe BC uses several clear-sky days (four in our case). The uncertainty in the MWR Tb observations was set to the standard deviation from a detrended time-series analysis for each channel during these cloud-free periods.

Regarding the forward model uncertainties, we do mention in the manuscript that *"The bias of the retrieval depends on both the absolute accuracy of the forward model and on any observational systematic offset".* Additionally, since in the revised version of the manuscript we are now presenting the posterior covariance matrix, Sop, we use it to provide a measure of the uncertainty of the retrievals (Figs. 3 and 4, especially Fig. 4e).

3) Line 410: you mention that 5 RS are under rainy conditions. Could you confirm that these data have been discarded from the results in figures 6 to 9 (statistical analysis)?

Yes, we do confirm that the rainy conditions have been excluded from the results presented in the statistical analysis. This has now been clarified in the text: "*Moreover, while accurate RASS* *data can be collected during rain, MWR data could be potentially deteriorated due to water deposition on the radome. Therefore, six profiles (three for March 13, and one each on May 1, 3 and 4) were eliminated from the statistical evaluation. These restrictions lowered the number of total radiosonde launches used in this study to 52"*

Review AMT-2021-9:
Improving thermodynamic profile retrievals from microwave radiometers by including Radio Acoustic Sounding System (RASS) Observations

ANSWERS to REFEREE 2

This submission has great potential, but from my point it cannot be published like this. There are just too many basic things that still don't live up to a peer-reviewed scientific publication, starting from all sorts of formal issues over unclear & erroneous methods to non-scientific ways of argumentation.

We wish to thank very much both Referees for carefully reading our manuscript and for offering many comments towards its improvement. In revising the manuscript, we have considered all of their comments. Many changes have been applied to the current version of the manuscript and we believe that our manuscript benefited significantly from the constructive comments made by both the Referees. Please, find below our point-to-point answers in red.

General: Formal issues, equations, figures and figure descriptions don't fulfill the general standards; are there is still not enough quantitative discussion.

In the revised version of the manuscript we have made sure to present figures clearly, both in the format and in the quantitative discussion of the results. Same for the equations. Nevertheless, in the final version of the manuscript we'll make sure to provide the Editorial office with the high-resolution version of the images in *.eps format and Equations from Microsoft's equation editor.

Below please find my general major points, without going into the details of the Result section.

We have taken the Referees' comments under consideration and have modified the manuscript accordingly. The Results Section has also been modified as a consequence of the changes applied throughout the whole manuscript and we do believe that the flow of the manuscript, as well as the presentation of the results, is much improved as a consequence of these changes.

Lines 198-200: Still no reason (or reference) given why MWR off-zenith TBs improve the temperature profile, respectively are used.

Crewell and Löhnert (2007) demonstrated conclusively that the addition of elevation scans improves the accuracy of the retrieved temperature profiles. Furthermore, this improvement is because the elevation scans increase the information content in the retrievals, and is shown in other papers (e.g., Turner and Löhnert 2021). Most MWR vendors and groups routinely use elevation scans in their temperature retrievals as part of their usual operating procedure.

Section 3.1: (too) many formal issues

Section 3.1 and the equations wherein have been checked, hopefully fixing the "formal issues" mentioned by the Referee.

Line 384: Why are "Physical retrieval bias-correction" and "temperature profiles" included in one sub-section?

We do agree with the Referee that the previous organization of Sections 3 and 4 (and their Subsections) was not straightforward, so we did reorganize the manuscript completely in the revised version. We changed it to:

3. Physical retrievals
   - 3.1 Iterative retrieval technique
   - 3.2 Bias-correction of MWR observations using radiosondes or climatology
   - 3.3 Analysis of physical retrieval characteristics
4. Results
   - 4.1 Statistical analysis of the physical retrievals up to 3 km AGL
   - 4.2 Statistics for the profiles least close to the climatology
   - 4.3 Virtual temperature statistics

With this new organization we do believe that the flow of the manuscript should be improved.

Lines 384-389: The terms bias and uncertainty seem mixed up.

Commented [1]: Text on lines 384-389 was: "Fig.1. Bias for the four chosen clear-sky days (red-dashed lines) and their mean (red solid line) for the original observations in the top panel, and for the bias-corrected data in the bottom panel. Green lines are the uncertainty boundaries around the mean bias. Frequencies used in the PR algorithm are marked with black triangles in both panels."

Fig. 1 has been modified in the revised manuscript, as well as the text referring to it: "*Fig. 1 shows the Tb biases found for all 22 MWR channels from both bias-correction approaches. The biases calculated with the radiosonde BC scheme are shown for all channels used in our analysis: 22 channels of the zenith scan, in red, and four V-band opaque channels of the oblique scans, in blue. The black and green triangles represent the biases calculated using the TROPoe BC approach for zenith and for zenith+oblique scans, respectively.* ***All biases are presented with associated uncertainties****.*"

Line 351: "bias of the retrieval depends on both the sensitivity of the forward model..." Do you mean the absolute accuracy of the forward model or really the sensitivity? If the latter: sensitivity to what?

The Referee is correct and we have modified the mentioned sentence to: *"The bias of the retrieval depends on both the absolute accuracy of the forward model and on any observational systematic offset"*.

Line 351-370: The method described in this paragraph is scientifically not sound: I am puzzled about mixing up the terms systematic error reduction, uncertainty determination and clear & cloudy-sky determination.

We have taken the comments of both Referees very seriously on these points and have modified large parts of the manuscript to address their concerns. For this reason, the Section regarding the bias correction of the MWR measured Tbs has been completely rewritten and renamed *"3.2 Bias-correction of MWR observations using radiosondes or climatology"*. In this section, two bias-correction methods are now presented, the first using radiosonde data (radiosonde BC) and the second using the prior (TROPoe BC). In this updated Section we have now explained in detail what are the steps in each of the bias-correction approaches, starting with the clear-sky days identification, moving to the MWR Tbs biases calculation and the bias removal.

Lined 371-383: A bias calculation should not be carried out with respect to the prior. Also: need to show with measurements that bias in zenith and off-zenith TBs do not differ, a mere assumption here is not enough.

Following both Referees' comments, and as mentioned in the answer above, in the revised version of the manuscript two bias-correction methods are now presented, the first using radiosonde data (radiosonde BC) and the second using the prior (TROPoe BC).
Additionally, we did determine the bias in the MWR observed Tbs (in all 22 MWR channels of the zenith scan, and in the four opaque channels of the oblique scans, i.e. the channels used by TROPoe), in comparison to radiosonde Tbs (radiosonde BC) and in comparison to the TROPoe Tbs (TROPoe BC). These results are now presented in Fig. 1 of the revised manuscript. The following description of Fig. 1 has been included in the text: *"The biases from the two bias-correction schemes are within the standard deviation of each other for most of the channels except at higher frequencies. [...] It is also important to note that, in both approaches, the similarity between the biases in the opaque channels for zenith and for oblique scans [...] are very similar to each other.* ***This supports the assumption that the true bias is nearly independent of the scene, or that the sensitivity to the scene (e.g., zenith or off-zenith) is small."***

Lines 428-431: The averaging kernel must be applied to the radiosonde profiles, not to the retrievals

The Referee is of course right, and we corrected the error accordingly.

Fig. 2: not explained how why and how smoothed radiosonde profiles differ, this would involve

Commented [2]: Text on lines 351-370 was: "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. To this aim, we first looked for clear sky days (to reduce the degrees of freedom associated with clouds) during the period of the measurements. One method to identify clear-sky times is to use Tb observations in the 30 GHz liquid water sensitive channel. 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.) 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 2-3 hours around 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 both sets of uncertainties described above were extremely similar with the averaged difference less than 0.05 K.

Commented [3]: Text on line 371-383 was: "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 observations. We compute the bias in the bias-correction procedure only from the zenith scans, assuming that the same bias is suitable for other scans. Also, we assume that the true bias is an offset that is nearly independent of the scene, so that the sensitivity to the scene (e.g., clear or cloudy, zenith or off-zenith) is small. To investigate that, we eliminated the radiosondes launched during rainy periods (5 out of 58 cases) and found that the average temperature profiles were very little different than when all radiosonde profiles were used, with the maximum bias and RMSE absolute differences 0.12 K and 0.11 K respectively up to 5 km AGL. Fig. 1 shows the results of the bias-correction for the four chosen clear-sky days. The green lines on this figure indicate the MWR random errors; these are 0.3-0.4 K for K-band channels and 0.4-0.7 K for V-band channels."

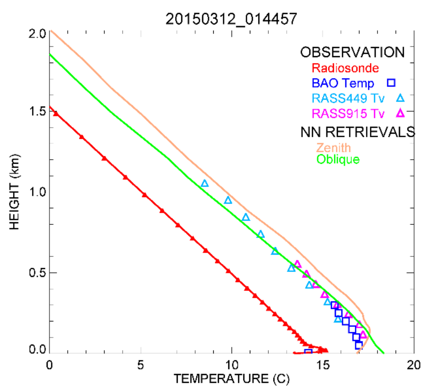
explaining how the Jacobians are built up and differ for each instrument combination (not given).

The Referee is correct in the fact that the radiosonde profiles smoothed using the four Akernels can be very different from each other. In the revised version of the manuscript we do note that *"while comparison of the retrievals to the relative Akernel-smoothed radiosonde profiles can be used to minimize the vertical representativeness effects due to the different vertical resolutions of these profiles, we note that a statistical comparison between the four configurations of the observational vector would not be fair if each of their retrieved profiles is compared to a different Akernel-smoothed radiosonde profile. Therefore, in the statistical analysis presented later in the manuscript (section 4.2), mean bias, root mean square error (RMSE), and Pearson correlation coefficients will be computed between the various MWR's retrieval configurations and the unsmoothed radiosonde profiles, just interpolated to the same vertical levels of the retrieved profiles."*

Lines 485-487: scientific reasoning missing

Commented [4]: Text on lines 485-487 was: "However, one radiosonde profile showed a large bias (> 5 oC) against all seven levels of BAO temperature measurements and against all PRs, therefore we decided to exclude this particular radiosonde profile from the statistical calculations."

We included the following text in the revised version of the manuscript: *"However, **one radiosonde profile showed a large bias (> 5 °C) against all seven levels of BAO temperature measurements and all available Tv measurements from the RASS 915 (eight measurements up to 600 m AGL) and from the RASS 449 (nine measurements up to 1100 m AGL), therefore this particular radiosonde profile was excluded from the statistical analysis** (Figure is included). Moreover, while the RASS data could be collected properly under rainy conditions, MWR data could be potentially deteriorated due to water present on the radome, therefore six profiles (three for March 13, and one for May 1, 3 and 4) were eliminated from the statistical evaluation. These restrictions lowered the number of total radiosonde launches used to 52."*



Conclusions: Argumentations are in general not comprehensible

The Conclusion Section has now been rewritten and we do hope that the summary we present is now clearly understandable.

Lines 737-741: "Of the PRs configurations tested, we find better statistical agreement with the radiosonde observations when the RASS 449 is used together with the surface observations and brightness temperature from only the zenith MWR observations and doubling the random radiometric uncertainty on the MWR observations (MWRz2sigma449) relative to the uncertainty calculated over the selected clear-sky days." I cannot figure out what this sentence is supposed to say.

As mentioned above, the Conclusion Section has been completely rewritten, hopefully being clearly understandable now. Also, the text mentioned by the Referee is now deleted because we decided to completely remove the analysis presented before where the random radiometric uncertainty on the MWR observations relative to the uncertainty calculated over the selected clear-sky days was doubled (MWRz2sigma449). We decided to remove this part because we realized the motivation was too specific to this dataset, but wouldn't be useful in general, on another dataset.

Lines 744-746: "The larger assumed radiometric uncertainty in the MWR Tb observations allows the retrieval to overcome both (a) 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 (Cimini et al. 2018)." How can there be "small systematic errors" between MWR TBs and RASS measurements? They are of different physical dimensions. What systematic errors exist in the forward radiative transfer model, how large are they and how do we know that the PR can overcome them?

As mentioned in the above answer, the text mentioned by the Referee is now deleted because we decided to completely remove the analysis presented before where the random radiometric uncertainty on the MWR observations relative to the uncertainty calculated over the selected clear-sky days was doubled (MWRz2sigma449).