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

General comments : The authors discuss about the improved accuracy of temperature profile retrievals obtained by combining active Radio Acoustic Sounding System (RASS) measurements and passive microwave radiometer (MWR) observations. This topic is of high interest for the scientific community as synergetic combinations of passive and active remote sensing instruments should help alleviate limitations when using each instrument alone. The structure of the manuscript is fine. However I found that the quality of English should be improved and the authors should be more precised in the presentation of their results. Several times, some details are missing to fully understand what they are discussing (the MWR configuration especially in terms of channels and elevation angles). I also think that several major scientific issues should also be addressed by the authors before publication to AMT. These issues concern the interpretation of the results using MWRs oblique measurements and some conclusions that, in my opinion, are not enough balanced due to the problems with the instrument used in this study. I also have some doubts for some interpretations for which additional explanations and potential new figures would be necessary to clarify the results. Thus I suggest the manuscript for publication after some major corrections in the results interpretation and minor english spelling errors which I list below.

Specific comments :

<u>Major comment 1</u> : 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).

<u>Major comment linked with biases in MWR oblique scans :</u>
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 intepretation 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)
→ 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).
→ 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.

<u>Second major comment about NN retrievals:</u>
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 bias-correction 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).

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 observ**ed** 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».

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.

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..

Introduction line 125 : The litterature 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.

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

Section 2, line 172 : change included into including.

Section 2, line 196 : change manufacturing into manufacturer.

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

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

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

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

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

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) ?

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 jacobians ?

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 .).

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

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 ?

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.

Section 3.1, line 313 : please correct the sentence into : « its dimensi**on** increas**es** ».

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

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?)

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.

Section 3.2, line 345 : can you at least mention that NN biases could be improved by applying a bias-correction ?.

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 ?

Section 3.2 line 366 : Modify the sentence into « demonstrate **a** better agreement ».

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

Section 3.3, line 425 : please correct vs into 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 ?

Section 3.3, line 437 : change dash lines into dash**ed** lines.

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

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).

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

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

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

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

Section 4.2, line 566 : change « above and below 1.5 km » into « by up to 5 km AGL »
.
Section 4.2, line 567 : change statistical measures into statistical scores.

Section 4.2, line 567 : I do not understand this sentence which is in contradiction with the previous one. Line 566 you mentin 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.

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

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.

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.

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 ?

Section 4.3, line 120 : what do you mean by « weighted average over the 42 vertical heights » ?

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.

Conclusion line 703 : I honnestly 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

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 parenthesis).

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

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.