Response to the interactive comment on "Characterization of downwelling radiance measured from the ground-based microwave radiometer using the theoretical reference data"

First of all, the comments given by two anonymous reviewers are really appreciated. Thanks to the reviewers' comments, the overall quality and part of conclusion of the paper has been improved. Please note that author responses to the reviewer's comments are given in blue italic. Furthermore, the improvements and revisions mentioned in the author responses are reflected in the revised manuscript in red characters. A few additional references added for the revision is also identified with the red characters.

Before providing our responses to the individual reviewers, it should be stated that a mistake that is not noticed either by the authors or the reviewers before is found and corrections are made with the revision process. Although the overall conclusion and important findings, except the argument on the frequency shift, are not significantly affected by the mistake, It would be appropriate to identify the mistake along with the revisions made due to the mistake, before giving our item-by-item responses to the reviewers' comments.

The mistake was in the definition of the top of the atmosphere for the radiative transfer model (RTM) simulation. As the T & q profiles retrieved from the microwave radiometer have the top altitude of 10 km, we use, without giving a much thought, 10 km as the top of atmosphere for the RTM simulation. Thus, the simulated Tb actually represented downwelling radiances from the atmospheric layer in between the ground and 10 km (we may call it as "tropospheric radiance"), rather than a whole column of the atmosphere. Thus, the simulated Tb used for the current version of the manuscript is smaller than the actual Tb by the "missing radiance", i.e the downwelling radiance coming from the atmosphere located at above 10 km. To rectify the mistake, we re-run the RTM simulation with full scale of vertical profiles of ECMWF (up to 40km), KLAPS (up to 20km) and radiosonde (up to the available height) and then update (or revise) the all relevant tables, figures, and manuscript (the clearest impact is the significant improvement in the comparison between the measured Tb and simulated Tb).

Thus, even with the new simulations, overall conclusion and findings are not significantly affected except the issues related to the frequency uncertainty of the radiometer. The new differences between the measured and simulated Tbs show a significant improvement, especially at the lower frequencies, as shown in the revised Table 3. For example, the previous bias at 51.26 GHz was 4.6 K with a 2.6 K of variability, while it becomes -0.6 K with the variability of 2.3 K. With the new simulation, the only significant difference is evident in the 52.28 and 53.84 GHz channels. Thus, the discussion on the frequency shift is mainly focused on these two channels. Further investigation of the frequency shift also directed authors to accept the main cause of the bias is probably due to other causes such as uncertainty in the spectroscopy data or uncertainty in the calibration. The revised manuscript includes all of the new arguments.

The background for such a change is mainly improvement in the RTM simulation by correctly including the "missing radiance" . Especially, the "missing radiance" produces differential effects on the simulated Tb depending on the frequency, as shown in Figure R1. For example, at the wings of the O2 absorption band, the simulated Tbs with the top altitude of 40 km (which is the top altitude that ECMWF analysis data extends) are much warmer than those with the top altitude of 10 km, the difference is as large as 5.6 K at 51.25 GHz. However, the differences are getting smaller toward the center of the absorption band, and become 0.0 at 58.00 GHz.

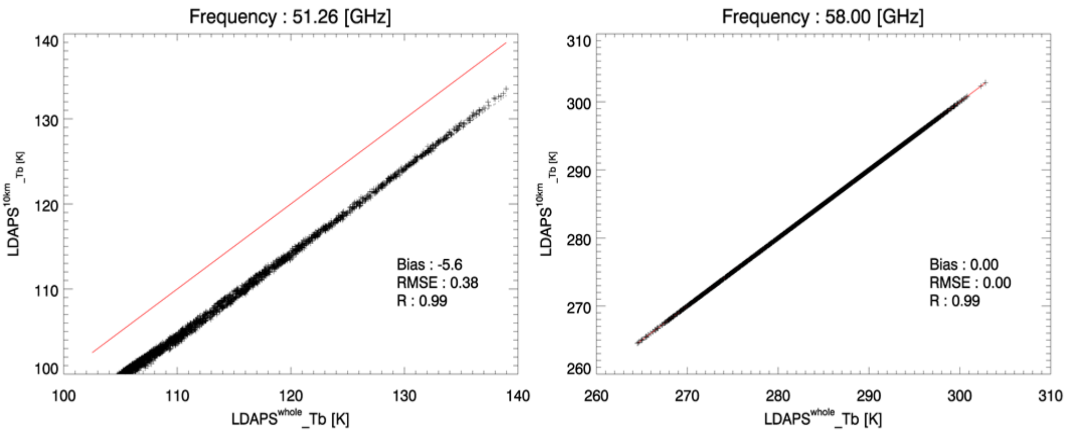

Figure R1. Comparison between the simulated Tb with the top altitude of the atmosphere at 10 km (Y axis) and at 40 km (X axis) for the 51.26 GHz (left) and 58.00 GHz (right) frequency. The bias value in between the two frequencies decreases with the increasing frequency.

The reason for the differential response in the simulated Tb with the frequency is due to the relative magnitude of the "missing radiance" to the "tropospheric radiance". As the absolute magnitude of the "tropospheric radiance" at the wing channels are much smaller than at the center channels (the measured Tb is much cooler at these channels), the relative effect due to the "missing radiance" is shown to be much larger at the wing channels (this effect also produces a relatively large variability in the Tb difference at the lower Tb values at the 51.26 GHz). Thus, the comparison results with the measured Tb also depend on the channel, i.e. a relatively large change in the lower frequency channels.

Anonymous Referee #1

Major comment:

Although I understand the expedient of adjusting the center frequency of the observed data to an "effective center frequency" that minimizes the differences between ECMWF and observed data I fundamentally disagree with the approach outlined by the authors to use the ECMWF data to "calibrate" the MWR data. In my opinion such expedient does not provide a real estimate of the center frequency shift. It merely ensures that when using the data in conjunction with the ECMWF model the resulting estimator is unbiased.

One of the points of strength of microwave radiometers is the fact that they provide an independent set of measurements to be used to evaluate radiosondes and model performances. As a matter of fact well-calibrated radiometers are used to correct for radiosondes biases and have been successfully used to refine radiative transfer models, not the other way around. When model data (or even radiosondes) are used to calibrate the radiometers the dataset won't be independent anymore, but it will be biased to whatever dataset was used to calibrate them (in this case ECMWF or KLAPS) with the addition of uncertainty in the radiative transfer.

This is shown in Table 4 where differences between the ECMWF and KLAPS-adjusted center frequencies vary from 10 MHz to 60 MHz. As a reference for the authors I note here that if the center frequency is properly selected with a local oscillator the accuracy is expected to be of the order of 100 KHz (0.1 MHz). I understand that the data collected can't be changed, however I think the authors should discuss the limitations and drawback of the technique that they are using as a warning to potential users of the data. For future data, my suggestion would be to fix the radiometer (if this wasn't done already) so that it can be reliably and independently calibrated.

- ⇨ *Well, we first need to acknowledge the reviewer's persistent argument and the whole points given by the reviewer is reflected with the revisied manuscript, i.e., the argument for the frequency shift is dropped based on the analysis of the Tb difference and the variability of the necessary value of the frequency shift. A few additional explanation is given below.*
- ⇨ *With the correction of the mistake in the RTM run, the systematic biases are reduced significantly. As shown in the revised manuscript, most of V-band channels are shown to be in line with the known and/or previous comparison results, except the 52.28 GHz and 53.86 GHz channels.*
- ⇨ *At the same time, we checked the characteristics of the Tb difference (difference between the measured Tb and simulated Tb) for the original and shifted frequency (the necessary shift value is obtained by the same process as described in the original manuscript). As the error that could be introduced by the frequency shift is characterized with the dependence of the Tb difference on the measured Tb (because of the temperature sensitivity of Planck function), the original data clearly indicated that direction. However, when we checked the Tb difference after application of the new frequency to the simulated Tb, there is no improvement in the dependence, i.e slope between the Tb difference and the measured Tb is the same as shown in Fig. 6 of the revisied manuscript (also shown below). Furthermore, the uncertainty associated with the necessary frequency shift value to best match the measured and simulated Tb is way too high compared to the value that suggested by the review. Thus, we dropped the frequency shift proposition in the revisied mansuscript, including the arguments given in our review response.*

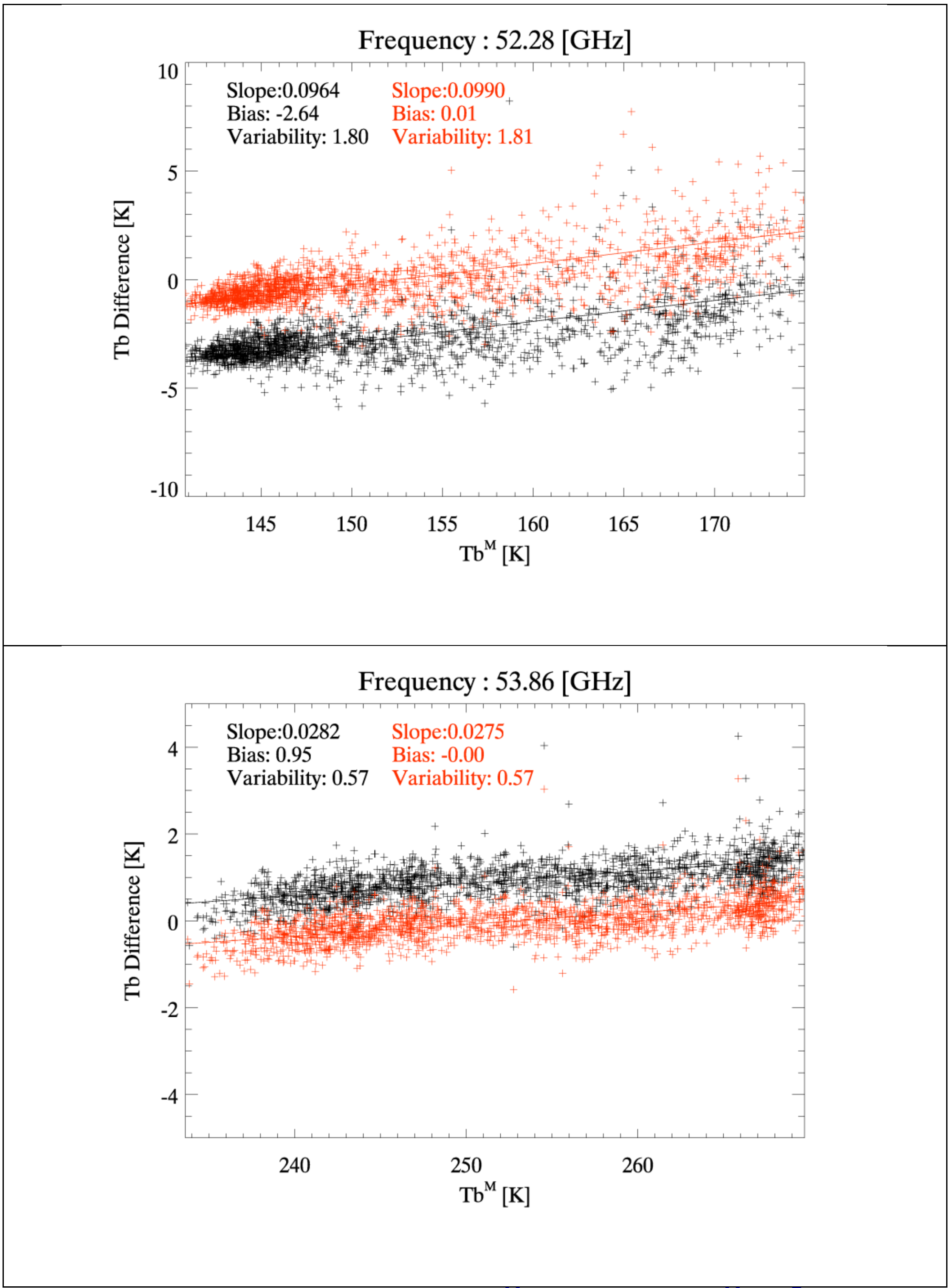

Figure. The Tb difference as a function of Tb^M (black crosses: Tb^M - Tb^E, red crosses: Tb^M - T^{new}) for the data set obtained after removing data with the faulty calibration and contaminated by clouds. The solid lines are the best linear fit.

Minor points

The explanation of Table 3 seems mismatched with the table content.

⇨ *Thank you for your comments and sorry for the mistake. We replace the table contents.*

Anonymous Referee #2

SPECIFIC COMMENTS

1) Generally I think the authors should rework some figures and tables thoroughly. For example:
Fig.1: If a bias is plotted, the authors should let the reader know, which differences are calculated. In the subfigures is noted 'sonde&KLAPS' and 'sonde&ECMWF'. Does it mean 'sonde minus model'? Probably not, as in the captions is mentioned "the bias of the temperature profiles of the NWP data compared to the radiosonde data". Does it mean 'model minus radiosonde'? But finally, in the text is formulated (Page 4353, line 22) that Fig.1 shows a comparison between available radiosonde and NWP data'. That would mean 'sonde minus model' after all.

- ⇨ *Thanks a lot for the careful readings. We corrected the mistakes as follows; agree to your comments and the paper was revised. Concerned with the data number, you mentioned, we unify the number of used data (ECMWF, KLAPS, Radiosonde, and radiometer) and describe the number of data set for the clarifications.*

Fig.4: Maybe the authors had in mind to plot different ranges of Y-axis as noted in the captions. Unfortunately they didn't. It should be done as nothing can be recognized from the last three subfigures

- ⇨ *We change the subfigures in Fig.4 to recognize the different range of Y-axis.*

Tables: I miss units for several columns.

- ⇨ *We add units for the tables.*

Table 1: Typo – negative variability (SD)

- ⇨ *Corrected.*

Table 2: What is the basis for the number of samples for KLAPS data (37230)? Three years of hourly data result in about 26300.

- ⇨ *Actually, 26,304 KLAPS data for 3 years are available. But only 24,532 pairs of KLAPS and radiometer are compared in Table 1 because radiometer data during the rain conditions are excluded. With the same reason, 4,508 ECMWF and 115 radiosonde data also compared in this study as the original state (all data are 4384 and 117, respectively).*

Table 3: Captions and column headings don't agree.

- ⇨ *Table 3 is corrected.*

Table 2 and 3: In both tables 'Bias, Variability and R' for the data set referred to as 'Original' are listed and have identical values. Why then the numbers of cases differ (4384 and 3972)?

- ⇨ *We correct the number of data for Table 2 and Table 3 "original".*

Table 4: Typo – frequencies

- ⇨ *Corrected in the revision. Thanks.*

2) Section 2.2.2 and 3.1:
A total of 117 radiosonde data are used to compare at first radiosondes with two model data sets (local model KLAPS and ECMWF) and secondly, simulated brightness temperatures. The numbers of samples taken into account for the analysis of the two models differ (117 and 67, respectively) and therefore different data sets are compared. This is ignored by the authors and not discussed. Not any information is given whether model forecast values or data of the numerical analysis were used. I assume that the two data sets are composed differently due to the different temporal resolution.

➪ *Yes. That is correct. As the ECMWF data are available every 6 hours, while the KLPAS data are available every hour, there is a large discrepancy in the available number of data.*

Possibly, deviations between the models result from their variable pre-treatment. The question arises whether KLAPS data are available for the subset of 67 cases used for the ECMWF comparisons? If so, add bias/SD calculations as well as Tb computations on the basis of this ‘really comparable’ data set. Further, how many cloudless cases are contained in both data sets? Clarifications are needed.

➪ *Yes. Indeed the “really comparable” subset of data set could be used to make the comparison more equatable. The reason that we used the different number of data points is quite a simple, to keep as many as available data to increase the reliability of the error statistics.*

➪ *On the other hand, as suggested by the reviewer, our conclusion could be double-checked by following the reviewer’s suggestion. The results are shown in Table R1. As shown in the table, even with only 66 available data points of the all three types of data set at the same time, error statics of the three different sources of T & q profiles show quite a similar result. The magnitude of bias, variability, and correlation coefficient among radiosonde, ECMWF, and KLAPS are similar.*

➪ *For the purpose of checking the relative compatibility, we do not remove the cloud contaminated data points for the Table R1 simply to keep the number of data points that are large enough to provide the statistically meaningful error statistics. We could easily speculate the large variability shown in the lower frequencies (and decreasing with increasing frequency) are due to mainly cloud contamination.*

Table R1. Error statistics of the radiometer Tb~~Tb~~ (Tb^M) compared to the simulated Tb (Tb^R, TB^E, and TB^K)~~Tb~~ using the limited number of radiosonde data, ECMWF and KLAPS data at the same time (total number of data is 66)

Frequency (GHz)	Radiosonde			ECMWF			KLAPS		
	bias	variability	R	bias	variability	R	bias	variability	R
51.26	7.56~~13.4~~	21.00~~19.1~~	0.68~~0.73~~	6.47~~9.3~~	21.19~~24.9~~	0.65~~0.65~~	6.92~~8.4~~	20.80~~13.1~~	0.69~~0.74~~
52.28	3.85~~10.6~~	16.15~~14.8~~	0.71~~0.75~~	2.90~~7.3~~	16.27~~19.3~~	0.69~~0.68~~	3.22~~6.6~~	16.00~~10.2~~	0.71~~0.77~~
53.86	2.35~~7.7~~	3.69~~3.4~~	0.94~~0.94~~	1.88~~7.0~~	3.72~~4.9~~	0.94~~0.92~~	1.95~~6.8~~	3.73~~2.4~~	0.94~~0.97~~
54.94	0.65	0.63~~0.7~~	1.00~~0.99~~	0.30~~1.6~~	0.75~~1.2~~	0.99~~0.99~~	0.40~~1.7~~	0.72~~0.7~~	0.99~~0.99~~
56.66	-0.06~~-0.4~~	0.66~~0.6~~	1.00~~0.99~~	-0.46~~-0.3~~	0.90~~1.0~~	0.99~~0.99~~	-0.39~~-0.1~~	0.63~~1.0~~	0.99~~0.99~~
57.30	-0.16~~-0.5~~	0.71~~0.6~~	1.00~~0.99~~	-0.57~~-0.3~~	0.93~~1.0~~	0.99~~0.99~~	-0.54~~-0.2~~	0.67~~1.1~~	0.99~~0.99~~
58.00	-0.18~~-0.6~~	0.77~~0.6~~	1.00	-0.58~~-0.4~~	0.99~~1.0~~	0.99~~0.99~~	-0.59~~-0.1~~	0.71~~1.2~~	0.99~~0.99~~

3) In the paper is shown that a screening of data is necessary to perform frequency adjustments and that clouds have the largest impact on the simulated Tb. It is entirely reasonable to show it as presented in Fig.3. But is it also appropriate to compare simulated and measured Tb without consideration of the cloud conditions (Table 2)?

- ⇨ *Sorry about the confusions. Table 2 and Figure 3 show the comparison results between the measured and simulated data without any screening process (only data affected by rain are excluded). The results after the screening process (cloud effect and the faulty calibration) are presented in the separated table, i.e. Table 3, which is also corrected with the revision. Hope, this clarifies the confusions introduced by the incorrect labeling of Table 3.*

In Section 2.3 (p. 4354) is stated that Tb simulations are done with the clear sky assumption. According to the authors (Table 3) about 40% of all data points are sorted out due to cloudiness. It means that the assumption is not valid for 40% of the Tb simulations which are widely discussed in Section 3. The conclusion that for an accurate assessment cloudless data points are needed (page 4348, line 2) is trivial as it was the general assumption for the Tb calculations. From my point of view assessments of model differences as discussed in Section 3.2 and offered in Table 3 should focus more on screened data then on data which are mostly incorrect by definition.

- ⇨ *Again, sorry for the confusion. The major conclusions with the sensor characterization are coming from the comparison results with the clear sky data (the second column of the Table 3.). The main point that we included the "original data" which is affected both by cloudy and the faulty calibration is the emphasize the importance of the cloud screening and monitoring of the calibration characteristics. Furthermore, the last uncertainty, i.e. frequency shift, is discussed with the only data that passed the screening process for the rain, cloudy, and the faulty calibration.*

MINOR COMMENTS

Probably due to different contributions by the co-authors the paper appears inhomogeneous. It concerns both the language and the text flow. For example, measured Tb is used three times before TbR is introduced for it (Page 4351, line 24). In the following TbR (5x), measured Tb (9x), measured TbR (3x) and radiometer Tb (2x) are listed alternately. Please, try to homogenize the text.

- ⇨ *Thanks for the comment. Careful revisions are made for such an annoyance. For example, we use the abbreviation Tb^E, Tb^K, Tb^R, and Tb^M for simulated brightness temperature using ECMWF, KLPAS, and Radiosonde data and measured brightness temperature by radiometer respectively.*