Reviewer2:

The manuscript egusphere-2024-2884 by Shi et al presents the development of a fast radiative transfer model (called ARMS-gb) for simulating brightness temperatures observed by ground-based microwave radiometers (GMRs). The characteristic of the model are described, including peculiarities in training and profile interpolation. The resulting simulations are compared with simulations from a similar existing code and with real observations from two GMR instruments. Also, the use of ARMS-gb to compute Observation Minus Background (OMB) differences and monitor GMR calibration stability is shown.

As similar codes and analysis already exists, the degree of novelty is relatively small. But the manuscript does fit the scope of the journal, and I find it clear and easy to read. However, there are aspects that need to be clarified, which could possibly undermine some of the results and conclusions.

Therefore, I could recommend publication only after addressing the following comments.

Major comments:

1) The authors miss to explain what's special about the seven profiles added to the training set and how the authors determined that this addition improved the training. They mention "moist environment" but to my knowledge the 101-level ECMWF 83 profiles do provide very humid profiles (see also minor comment on line 96).

Answer: In the revised manuscript, we give a figure (Figure 1) to show the statistical comparisons of the water vapor profiles from the ECMWF 83-profile dataset and the 7 additional profiles from the UMBC 48-profile dataset. The humidity range of the additional profiles exceeds the mean values plus the standard deviation of the ECMWF 83-profile dataset, particularly in the lower levels of the troposphere. Furthermore, the upper bound for optical depth regression is extended. Corresponding description is added. See Lines 113-118 in diff.pdf.

To evaluate the impact of enriching the training dataset, we trained two sets of fitting coefficients: one using the ECMWF 83-profile dataset (hereafter referred to as

Coef_EC83) and the other using the new training dataset (hereafter referred to as Coef_New90). RT simulations based on these two coefficient are intercompared using the 101L ECMWF 83-profile and UMBC 48-profile dataset. A table (Table. 3 in diff.pdf) is added to show AVG, STD and RMS for each HATPRO channel. For the 101L ECMWF 83-profile dataset, the accuracy of the two fitting coefficients is comparable. For the 101L UMBC 48-profile dataset, results using Coef_New90 demonstrate significantly higher accuracy compared to those using Coef_EC83. Corresponding descriptions are added in the revised manuscript. See Lines 213-239 in diff.pdf.

The statistical metrics, including AVG, STD and RMS, for each MP3000A channel are provided in the following two tables.

Table B1. AVG, STD and RMS of each channel of MP3000A. RT simulations based on Coef_EC83 and Coef_New90 are performed under the 101L ECMWF 83-profile dataset.

Coefs	Channel	1	2	3	4	5	6	7	8
Coef_EC83	AVG (K)	0.0200	0.0171	0.0123	0.0136	0.0151	0.0172	0.0188	0.0193
Coef_New90	AVG (K)	0.0168	0.0136	0.0099	0.0093	0.0125	0.0152	0.0163	0.0196
Coef_EC83	STD (K)	0.0344	0.0348	0.0333	0.0274	0.0263	0.0260	0.0268	0.0298
Coef_New90	STD (K)	0.0362	0.0369	0.0349	0.0302	0.0275	0.0266	0.0278	0.0307
Coef_EC83	RMS (K)	0.0398	0.0387	0.0355	0.0306	0.0304	0.0312	0.0328	0.0355
Coef_New90	RMS (K)	0.0399	0.0393	0.0363	0.0316	0.0302	0.0307	0.0322	0.0364
Coefs	Channel	9	10	11	12	13	14	15	16
Coef_EC83	AVG (K)	0.0161	0.0151	0.0084	0.0100	0.0106	0.0047	0.0016	0.0001
Coef_New90	AVG (K)	0.0129	0.0037	0.0076	0.0090	0.0065	0.0062	0.0025	0.0013
Coef_EC83	STD (K)	0.1019	0.1004	0.0956	0.0829	0.0630	0.0397	0.0204	0.0110
Coef_New90	STD (K)	0.1096	0.1049	0.0984	0.0851	0.0634	0.0403	0.0205	0.0111
Coef_EC83	RMS (K)	0.1032	0.1016	0.0959	0.0835	0.0639	0.0400	0.0205	0.0110
Coef_New90	RMS (K)	0.1103	0.1050	0.0987	0.0855	0.0638	0.0408	0.0206	0.0112
Coefs	Channel	17	18	19	20	21	22		
Coef_EC83	AVG (K)	0.0005	0.0004	0.0004	0.0003	0.0003	0.0001		
Coef_New90	AVG (K)	0.0007	0.0006	0.0005	0.0003	0.0003	0.0002		
Coef_EC83	STD (K)	0.0072	0.0051	0.0037	0.0028	0.0023	0.0018		
Coef_New90	STD (K)	0.0072	0.0053	0.0037	0.0028	0.0023	0.0019		
Coef_EC83	RMS (K)	0.0072	0.0052	0.0037	0.0028	0.0023	0.0018		
Coef_New90	RMS (K)	0.0072	0.0053	0.0038	0.0029	0.0023	0.0019		

Table B2. Same as Table B1, but the intercomparison performs under the 101L UMBC 48-profile dataset.

Coefs	Channel	1	2	3	4	5	6	7	8
Coef_EC83	AVG (K)	0.0095	-2.3810	-0.0864	-0.1762	-0.4770	0.0130	0.0492	-1.0961
Coef_New90	AVG (K)	0.0439	0.0417	0.0366	0.0312	0.0292	0.0290	0.0282	0.0306
Coef_EC83	STD (K)	0.0992	2.3736	0.1746	0.3379	0.5818	0.3210	1.0516	2.0890
Coef_New90	STD (K)	0.0440	0.0409	0.0326	0.0228	0.0198	0.0206	0.0231	0.0264
Coef_EC83	RMS (K)	0.0997	3.3620	0.1948	0.3811	0.7523	0.3213	1.0527	2.3592
Coef_New90	RMS (K)	0.0621	0.0584	0.0490	0.0386	0.0353	0.0356	0.0365	0.0404
Coefs	Channel	9	10	11	12	13	14	15	16
Coef_EC83	AVG (K)	-22.0969	-12.0147	-45.9800	-9.1837	-3.5609	0.0266	-0.0359	-1.1075
Coef_New90	AVG (K)	0.0294	0.0236	0.0297	0.0258	0.0177	0.0112	0.0023	-0.0012
Coef_EC83	STD (K)	17.8324	21.8970	35.3179	8.5716	7.0411	0.0451	0.1228	1.3003
Coef_New90	STD (K)	0.1073	0.1058	0.0979	0.0916	0.0719	0.0510	0.0322	0.0243
Coef_EC83	RMS (K)	28.3948	24.9767	57.9786	12.5624	7.8903	0.0524	0.1279	1.7080
Coef_New90	RMS (K)	0.1113	0.1084	0.1023	0.0952	0.0741	0.0522	0.0323	0.0243
Coefs	Channel	17	18	19	20	21	22		
Coef_EC83	AVG (K)	-0.6983	0.0740	-0.2351	0.1186	0.0518	0.1917		
Coef_New90	AVG (K)	-0.0031	-0.0038	-0.0040	-0.0040	-0.0038	-0.0034		
Coef_EC83	STD (K)	0.8547	0.1105	0.3059	0.1576	0.0812	0.2153		
Coef_New90	STD (K)	0.0217	0.0201	0.0180	0.0161	0.0146	0.0131		
Coef_EC83	RMS (K)	1.1037	0.1330	0.3858	0.1973	0.0964	0.2883		
Coef_New90	RMS (K)	0.0219	0.0205	0.0184	0.0166	0.0151	0.0135		

Due to its unique geometries and atmospheric paths, observations of GMRs are more sensitive to temperature and humidty in low altitude than satellite observations. Therefore, it is necessary to enrich the training dataset for GMRs' observation operators.

2) Section 2.2 introduces a vertical interpolation method, which effects are then compared in Section 3.1 with those of another commonly used method.

However, it not clear if Section 3.1 compares these two methods (called mode 1 and 2) as both implemented within ARMS-gb or if it rather compares RTTOV-gb (implementing mode 1) with ARMS-gb (implementing mode 2).

The authors need to clarify this point, as the two situations would lead to different conclusions (see also minor comment on line 196).

Answer: We implement both interpolation modes within ARMS-gb first and perform comparisons across HATPRO channels. To isolate the impact of the interpolation modes and exclude differences related to the training process (e.g., LBL RTMs and the training dataset), only Coef_New90 is used. Corresponding descriptions are added in the revised manuscript. See Lines 254-255 and Lines 258-259 in diff.pdf.

3) Section 4 shows comparison between observations and simulations in clear sky, considering three cloud detection criteria. As written, criterium 3 does not seem correct, or at least does not correspond to the criterium given in the quoted references (Turner et al., 2007; Cimini et al., 2019). These two references identify cloudy conditions by setting thresholds on the standard deviation of observed BT at 31.4 GHz over a time period, while the authors states they use standard deviation of OMB. If the background simulation is constant over the 10-min period, then the results should be the same, but the authors need to clarify this point as it sounds like an unnecessary complication (see also minor comment on lines 251 and 267-268).

Answer: It is indeed a mistake. We correct it in the revised manuscript (See Line 308). The statistical results of ARMS-gb and RTTOV-gb don't noticeably change after the correction.

4) In a validation experiment, such the one described in Section 4.1, three sources are contributing: (i) the RT model, (ii) the input profiles, and (iii) the observing instrument. Figures 4 and 5 reports the results for two stations, representing relatively drier and moister environments. The two figures report very different results, although the input profiles and the simulations come from the same sources (ERA5 and ARMS-gb/RTTOV-gb, respectively). I understand the ERA5 and RT models may perform differently in different environments, but the one aspect that doesn't seem to be considered is the GMR instrument, which are of different type (Airda-HTG4 and YKW3) and independently calibrated.

Unless the absolute calibration can be validated properly at the two sites, using for example radiosonde profiles, a miscalibration of either instrument cannot be excluded. The stability of OMB does not suffice, as it only indicates a stable calibration, but does not say much about calibration accuracy (see also minor comment on lines 266-267 and 286-297).

Answer: According to the reviewer's suggestion, we also use radiosonde data as input for RT simulations, and the results from RTTOV-gb and ARMS-gb are compared. Scatterplots of simulated versus observed BTs are presented in Fig 5 and Fig.7 in diff.pdf. For the Karamay case, RTTOV-gb simulations align more closely with observations compared to ARMS-gb in the K-band channels. In the V-band channels, simulation accuracy of ARMS-gb and RTTOV-gb is comparable. For the Tanggu case, ARMS-gb provides more accurate results than RTTOV-gb, with higher correlation coefficient and smaller OMB median values and STDs in channels 1 and 2. For channels with central frequencies ranging from 54.5 GHz to 58.8 GHz, both RTTOV-gb and ARMS-gb accurately simulate observed BTs, with correlation coefffcients for both RTMs reaching up to 0.98. Corresponding sentences are added in the revised manuscript. See Lines 352-358 and Lines 372-380 in diff.pdf.

Minor comments:

- Line 9: either "also differ" or "are also different"

Answer: Corrected (use "are also different"). See Line 14 in diff.pdf.

- Line 18: "thermal" -> "thermodynamical"

Answer: Corrected. See Line 25 in diff.pdf.

- Line 20: "which extends" -> "which may extend"

Answer: Corrected. See Line 27 in diff.pdf.

- Line 43-45: "RTTOV-gb is trained using AMSUTRAN"; this is correct for RTTOV, but not for RTTOV-gb. Section 2.2 of Cimini et al., 2019 says: "Conversely, RTTOV-gb was trained using a later version of MPM, described by Rosenkranz (1998, hereafter R98), which is probably the most used among the ground-based microwave radiometry community. This model is continuously revised and freely available (Rosenkranz, 2017, hereafter R17), and its uncertainty has been carefully investigated (Cimini et al., 2018). Therefore, RTTOV-gb has been trained using the

R17 model also (version of 17 May 2017 available at http://cetemps.aquila.infn.it/mwrnet/lblmrt_ns.html, last access: 14 November 2018). Coefficients for both the R98 and R17 models are now available within RTTOV-gb v1.0."

The authors should modify the statement with a short summary of the above and remove the corresponding sentence at line 256 (as also suggested below) and in Table 3.

Answer: Thanks for the information. Sentences are changed to "While the coefficients for RTTOV are trained using AMSUTRAN (Turner et al., 2019), the coefficients for RTTOV-gb are trained using an updated version of the Millimeter-wave Propagation Model, as detailed by Rosenkranz (1998) (hereafter referred to as R98). A further updated version of R98 is introduced by Rosenkranz (2017) (hereafter referred to as R17), and its uncertainties are analyzed by Cimini et al. (2018). RTTOV-gb v1.0, now supports both coefficient trained using the R98 and R17.". See Lines 51-56 in diff.pdf.

Corresponding sentences in Section 4.1 and Table 3 are also revised. (See the answer to your comment about Line 256).

- Line 96: what's special about those seven profiles?

Answer: We give a figure (Figure 1 in diff.pdf) to show the statistical comparisons of the water vapor profiles from the ECMWF 83-profile dataset and the 7 additional profiles from the UMBC 48-profile dataset. The humidity range of the additional profiles exceeds the mean values plus the standard deviation of the ECMWF 83-profile dataset, particularly in the lower levels of the troposphere. Furthermore, the upper bound for optical depth regression is extended. Corresponding descriptions are added. See Lines 113-118 in diff.pdf.

- Line 102: what's "channel spectral V"? I guess it is channel bandwidth? Also, I guess Eq.4 is discretised as a sum; the authors should also state what spectral resolution they used to compute the sum.

Answer: We use "transmittance across the channel bandwidth V" to replace "transmittance in spectral channel V". See Line 124 in diff.pdf.

In practice, the channel bandwidth V is divided into 256 intervals and the integral in Eq. (4) is approximated by a discrete sum. Corresponding descriptions are added in the revised manuscript. See Lines 127-128 in diff.pdf.

- Line 120: "dense" -> "denser"

Answer: Corrected. See Line 143 in diff.pdf.

- Line 172: N was previously used to indicate the number of channels. I'd suggest to change letter to avoid confusion.

Answer: The letter is changed to "N_S". See Lines 200-203 in diff.pdf.

- Line 175: As above: the integral is computed as a sum, and the adopted spectral resolution should be stated.

Answer: Similar to Eq. (4), the integral calculation is also discretised as a sum, with the channel bandwidth V divided into 256 intervals prior to summation. Corresponding descriptions are added in the revised manuscript. See Lines 208-209 in diff.pdf.

- Line 191: either "two vertical interpolations are required" or "vertical interpolation is required twice"

Answer: Corrected (use "two vertical interpolations are required"). See Lines 241-242 in diff.pdf.

- Line 196: it is not clear if the two modes are applied both the ARMS-gb or rather mode 1 is used with RTTOV-gb and mode 2 with ARMS-gb. The two situations would lead to different conclusions.

Answer: We implement both interpolation modes within ARMS-gb first and perform comparisons across HATPRO channels. To isolate the impact of the interpolation modes and exclude differences related to the training process (e.g., LBL RTMs and the training dataset), only Coef_New90 is used. Corresponding descriptions are added in the revised manuscript. See Lines 254-255 and Lines 258-259 in diff.pdf.

- Line 251: not sure if there is a typo, but otherwise criterium (3) does not correspond to that used by Turner et al., 2007 or Cimini et al., 2019. It's the 10-min std of observed Tb at 31 GHz to be checked against the 0.2 K threshold, not the OMB.

Answer: It is indeed a mistake. We correct it in the revised manuscript (See Line 308 in diff.pdf). The statistical results of ARMS-gb and RTTOV-gb don't noticeably change after the correction.

- Line 253: As stated at line 71, ARMS-gb is limited to clear-sky simulations. As such, it is not clear why the threshold for cloud water content is set to 100g/m^2 and not to 0 g/m^2. Is cloud water provided in input at all to either ARMS-gb or RTTOV-gb?

Answer: Total column cloud liquid water content and ice water content from the ERA5 reanalysis dataset are used as index only for cloud clearing to make sure input profiles are under clear-sky scene. The threshold is set to 100 g/m^2 according to Moradi et al. (2020). We also evaluated OMB statistics under different thresholds (e.g., 10 g/m^2, 1 g/m^2) and results don't noticeably change. Corresponding descriptions are added in the revised manuscript. See Lines 310-313 in diff.pdf.

Reference

Moradi, I., Goldberg, M., Brath, M., Ferraro, R., Buehler, S. A., Saunders, R., and Sun, N.: Performance of Radiative Transfer Models in the Microwave Region, Journal of Geophysical Research: Atmospheres, 125, e2019JD031831, 620, 2020.

- Line 256: Please, remove "It accounts gaseous absorption by ODPS which is trained by AMSUTRAN (Turner et al., 2019)" and refer to Cimini et al., 2019 for the absorption model. Same in Table 3.

Answer: The corresponding sentence is changed to "It accounts gaseous absorption by ODPS which is trained by R98 (Rosenkranz, 1998) or R17 (Rosenkranz, 2017).".

Cimini et al., 2019 is used for referring to RTTOV-gb. Corresponding information in the Table is also corrected. See Lines 316-317 and Table. 4 in diff.pdf.

- Lines 261-264: The altitude above sea level and the surrounding orography of the two sites should also be reported, as these may have an effect on the simulated BTs (e.g., if orography is complex, bilinear interpolation may be misleading).

Answer: The altitudes above sea level and STD of surface pressures from the four nearest ERA5 grid points, which reflects the situation of surrounding orography, of Karamay, Tanggu and Minfeng are reported in the revised manuscript. See Lines 324-326 and Lines 394-395 in diff.pdf.

- Lines 266-267: The stability of OMB indicates that the calibration may be stable, but does not say much about calibration accuracy.

Answer: The sentence is changed to "Due to the stability of the OMB trend during this period, it is assumed that the quality of the calibration may be stable.". See Lines 330-333 in diff.pdf.

- Lines 267-268: This seems to hint that std of simulated BTs are used for cloud detection (see previous comment to line 251), which I think is wrong or at least does not correspond to the screening used by Turner et al., 2007 and Cimini et al., 2019.

Answer: It is indeed a mistake. We correct it in the revised manuscript (See Line 308 in diff.pdf). The statistical results of ARMS-gb and RTTOV-gb don't noticeably change after the correction.

- Lines 286-297: The results in Figure 5 are very different from those in Figure 4. If the analysis is correct, one would expect similar results, for example at channels 1, 2 and 3 of HATPRO (or Airda-HTG4), which are very close to channels 1, 3 and 4 of MP3000A (or YKW3). This may be due to uncertainties in ERA5, as the authors seem to suggest, but also to GMR instrument miscalibration. This cannot be excluded, unless a proper calibration evaluation can be performed using, e.g., radiosonde profiles.

Answer: According to the reviewer's suggestion, we also use radiosonde data as input for RT simulations, and the results from RTTOV-gb and ARMS-gb are compared. Scatterplots of simulated versus observed BTs are presented in Fig 5 and Fig.7 in diff.pdf.

- Line 347: "HATRPO" is mispelled.

Answer: Corrected. See Line 440 in diff.pdf.

- Lines 371-372: Accuracy may be improved also updating spectroscopy to the newest developments. This is likely the case at 50-54 GHz, i.e. HATPRO channels 8-9-10, as shown in Figures 7-8 of Larosa et al., 2024 (https://doi.org/10.5194/gmd-17-2053-2024). To my knowledge, those spectroscopy improvements are not implemented in MonoRTM.

Answer: Thanks for the information. An intercomparison among different microwave LBL RTMs is necessary to construct a reliable transmittance dataset for the ODPS training process. The reference is added in the revised manuscript. Corresponding sentences are changed to "Selecting a reliable and accurate LBL model for training is also essential for improving the accuracy of RT simulations. For example, Larosa et al. (2024) incorporates the latest advancements in absorption spectroscopy to improve RT simulation accuracy in the 50-54 GHz frequency range. An intercomparison among different microwave LBL RTMs is necessary to construct a reliable transmittance dataset for the ODPS training process. ". See Lines 483-486 in diff.pdf.

We appreciate the Reviewer 2 very much for the constructive comments.