Reviewer 1 - Christian Mätzler (our responses to the comments are given in red below)

General comments:
The objective of the study is not focussed on snow science, but on the search for a better use of microwave radiometer data from satellites in weather forecasting, especially in polar regions. Snow and ice surfaces produce the variable background of the atmospheric emission to be looked for. Different disciplines, dealing with atmosphere and cryosphere join here in a highly intricate way. Fortunately, the authors use up-to-date models (ARTS, SMRT) for simulating sensor signals at frequencies near 89, 118, 157, 183 and 243 GHz, i.e. at wavelengths roughly between 1 and 3 mm. Since snow-structure parameters cover a similar range, volume scattering by snow is dominant and can be highly variable. The selection of these frequencies, however, is not optimum for snow. It was based on the atmospheric properties to be sensed. The atmospheric window channels at 89, 157 and 243 GHz are highly sensitive to liquid-water clouds, to water vapour, and with increasing frequency also to ice clouds, the 118 GHz channels are used for temperature profiling around an O2 line, and the 183 GHz channels are used for humidity profiling around a strong line of water vapour. Only the wing channels around these lines played a role here. In a future study, window channels at lower frequencies that are optimal for snow should be used as well. When reading the manuscript, it appears that the focus was more on snow than on the overall aspects. No information was given on the atmospheric opacity range at the given channels. Because the airborne system was optimised for the atmosphere, it is not surprising that the results were only suboptimal. Imaging microwave radiometers optimised for the surface use conical scanning with constant incidence angle. The authors found a way out of this problem by limiting the study to radiance from nadir direction. However, this limitation was a trap for various raisons as shown below.

We thank Prof. Mätzler for the comprehensive review and thoughtful comments provided on our paper.

Atmospheric opacity ranges for the different frequencies are:

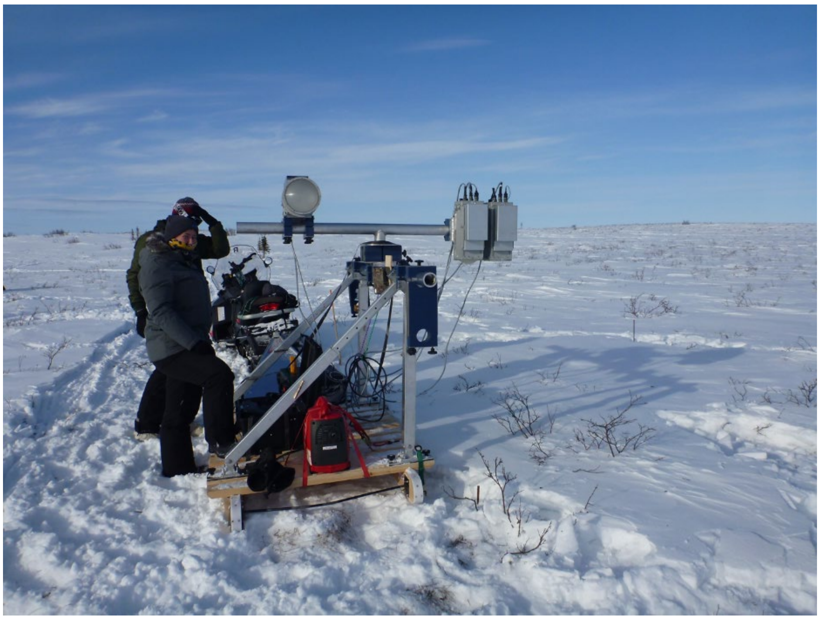
	C087	C090
89	0.092	0.093
157	0.162	0.156
183±7	0.745	0.714
118±5.0	0.316	0.327
243	0.335	0. 321

These are included for the discussion record. However, the general reader may find the downwelling brightness temperature measurements from the airborne instrument more useful. These are given in the Appendix, Figure A1 (this figure will be reordered according to frequency rather than instrument to be consistent with other figures). We disagree that these results are suboptimal: the challenges of point simulations vs areal observations will always exist and the purpose of this study was to demonstrate that we could explain the airborne observations through radiative transfer simulations, and that we can account for the surface effects in observations of the atmosphere, which we have done. This will be added to line 72-74 to make the reader more aware of that we are addressing the point vs areal challenge head on. The lower window frequencies that are commonly used for snow won't be very sensitive to the surface and wind-slab layers that have a big impact on the 183GHz, which is one of the key frequencies for atmospheric assimilation. For that, the 157 and 243GHz window channels are the best available. It would have been nice to have 50GHz channels to cover the key temperature sounding band, but then it would have been more critical to have e.g. 19 and 35 GHz for evaluation. To our knowledge there is no airborne system that covers the full frequency range 19-243 GHz. Conical scanning would be nice to give mapping and polarization, however, it also generally makes it very difficult/impossible to simultaneously measure the atmospheric downwelling Tb which is an important component of the validation (as demonstrated by this study). This study makes best use of a unique dataset which, to our knowledge, has no comparable equivalent elsewhere in time or location.

Discrepancies between surface and aircraft observations and changes in observations between different flights were interpreted by small-scale heterogeneity and by temporal variations. Radiometric data presented were taken at an unspecified aircraft altitude. Also, data taken with a surface-based instrument on a sledge at 89 GHz were transformed to flight altitude. I cannot understand how the setup measured the snow surface in nadir direction without distortion by shadowing of the sky and by its own radiation towards the footprint. It is much better to observe at a sufficiently large nadir angle (50° to 60° off nadir, as conically scanning satellite instruments do). Then the distortion by the instrument and its setup can be negligible, and additional information by the difference between vertical and horizontal polarisation can be obtained, e.g. to separate between specular and diffuse scattering. This would help in quantifying the effect of the ice lens. Transformation to nadir direction could still be done approximately by combining SMRT and ARTS. The surface-based instrument should also be used to measure the downwelling sky radiation (tipping curves for calibration, and to determine the zenith opacity of the atmosphere). I missed information on such measurements. Indeed, atmospheric and surface radiation are linked in many ways!

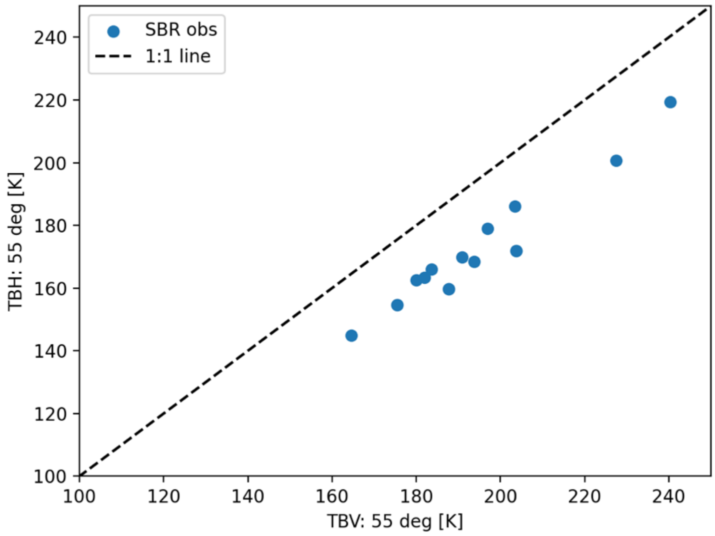
The aircraft altitude was approximately 500m – this will be added to the text. We have downwelling sky radiation from the aircraft measurements, as shown in Figure A1. While there may be differences between the atmospheric emission at the time of flights and the time of the ground-based measurements, these will be small at 89 GHz. The TB difference between flights was small and the variability within flights was of the order 5 K.

The radiometer setup is shown in the figure below:



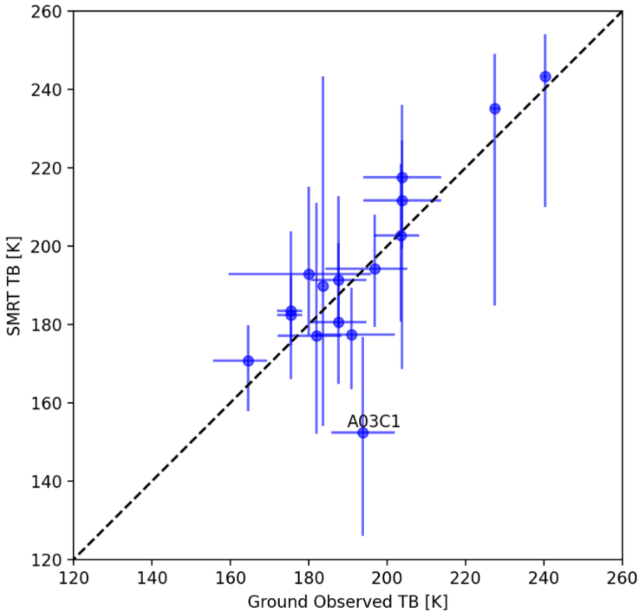
The use of a boom minimises the impact as much as possible. For the shadowing of the sky, there will be some specular component of reflection that will make the nadir measurement more sensitive to shadowing effects. However, the instrument is likely to be far from a black body and thus reflect the radiation far more than it will emit. Any metallic elements of the instrument will have reflectivities close to unity, so it may be appropriate to ignore the emission of the instrument and think of it as scattering the radiation. This means the TB measured will be larger than it would be without the instrument there, but the effect will be small. The results for the simulation of H and V shown below indicate that the error due to the radiometer set up is likely to be small.

We do have measurements at 55 deg and there are differences between H and V polarisation as shown below.



In terms of the simulations, we had initially looked at both sets of measurements (nadir and 55 deg) but decided to focus only on nadir measurements to simplify the paper. Nevertheless, the 55 deg simulations are interesting in themselves to look at diffuse vs specular reflection and/or the impact of ice lenses, as Prof. Mätzler indicates above.

A comparison between V-pol simulations and SBR observations at 55 deg is shown below:



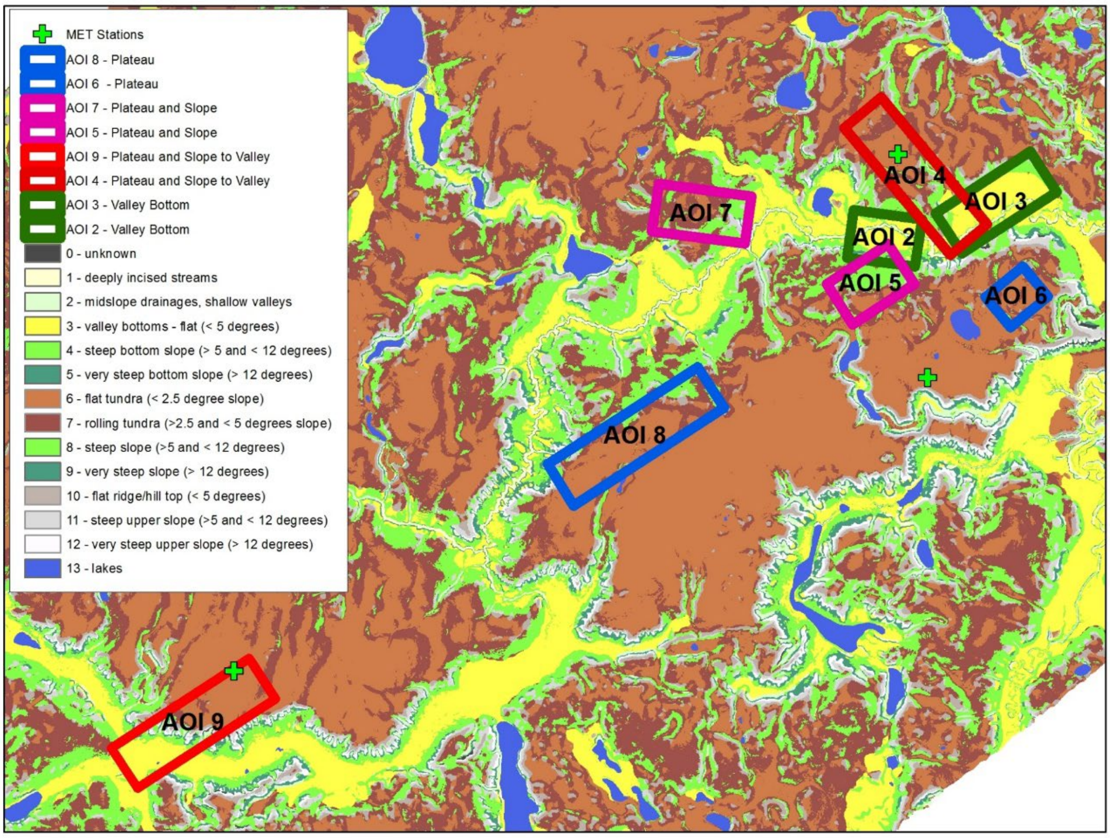
Pit A03C1 remains an outlier, but overall the simlations are not dissimilar to the nadir simulations in Figure 4, with the caveat that there are three additional pits with measurements at 55 deg: A02C1, A02E1 and A02W1. The mean error above is 0.4K and RMSE 12.9K, demonstrating better between agreement with observations than at nadir if all pits are included (compared with ME -7.1K and RMSE 16.6K). If pit A03C1 is excluded, the mean error is 3.2K and RMSE 8.0K, which is less good than the nadir observations excluding the outlier pits (compared with ME -0.03K and RMSE 7.5K). Emission from the instrument may be a source of error for the nadir observations but we do not think this is a large error.

The mean difference between H and V pol measurements at 55 degrees is 22.7K, whereas for simulations (Rayleigh-Jeans, at ground level) the mean difference is 19.0K. Simulations without ice lenses result in a TB difference of 11.7K. We consider SMRT performs reasonably well at both nadir and at 55 degrees with ice lenses present.

Special comments:
1) Figure 1: I miss geographic location and altitude range. It is unclear how rugged this terrain is, how steep the slopes and therefore how large the topographic effects are.

The details of the research basin will be added to the manuscript. Data were collected within the within the research basin of Trail Valley Creek (TVC), NWT, Canada (68◦44′ N, 133◦33′ W). The elevation range is 9 to 187 m.a.s.l and the topography is mostly gently rolling slopes with some deep valleys (Marsh et al., 2010 https://doi.org/10.1002/hyp.7786). For further details about the vegetation characteristics, see Grünberg et al., 2020 https://doi.org/10.5194/bg-17-4261-2020.

A more detailed classification of Figure 1 is shown below for discussion purposes. Slopes were generally less than 12 degrees, and although steeper slopes were present (dark green or white in the figure below), they formed a small proportion of the scene. We chose a simplified version of this figure for the manuscript in order to provide a more concise representation.



2) Line 132: How large is the difference between the Rayleigh-Jeans equivalent TB and the physical TB based on the Planck function (especially at the highest frequency used)? Give some typical examples. The difference between the two starts to diverge with increasing frequency and decreasing temperature.

At 89 GHz the difference is 2.1K, increasing to 5.8K at 243 GHz. This is based on simulation of the snowpits at an incidence angle of 55 deg. As discussed in lines 233-237, when we compared simulations to airborne observations we did an approximate conversion of the simulations to Rayleigh-Jeans Tb. The error of the approximation we used is less than 0.1K even for the minimum brightness temperatures observed and highest frequencies.

3) The use of SSA in Table 2 and elsewhere: It would be easier for the reader to get the correlation length in mm (Eq. 1) than SSA in kg/m^2 because the wavelength is in mm, too.

We have presented the data collected in Table 2 rather than the processed data. This is to allow the reader to use the data for their own purposes. We have given the necessary information in equation 1 and in the accompanying text to allow the reader to calculate the correlation length in the same way that we have done. We do not propose to change the paper, but include the same table here with correlation length for those interested to pick this information out from this discussion.

Pit	Date	Depth [m]	Density [kg m^{-3}]			l_{ex} [mm]			Topographic Index
			SS	WS	DH	SS	WS	DH	
A02C	15/03/2018	0.2	-	298	255	-	0.1	0.44	Valley
A02E	15/03/2018	0.2	-	328	282	-	0.07	0.27	Valley
A02W	14/03/2018	0.22	252	323	249	0.08	0.11	0.3	Valley
A03C1	17/03/2018	0.2	40	-	230	0.1	-	0.28	Valley
A03E	17/03/2018	0.42	159	-	264	0.06	-	0.37	Valley
A03W	17/03/2018	1.09	132	368	270	0.06	0.06	0.28	Slope
A04C	16/03/2018	0.31	-	314	226	-	0.09	0.33	Plateau
A04C1	16/03/2018	0.27	-	271	297	-	0.09	0.35	Plateau
A04N	16/03/2018	0.27	-	302	272	-	0.14	0.35	Plateau
A04N1	16/03/2018	0.24	-	232	265	-	0.07	0.2	Plateau
A04S	16/03/2018	0.38	-	332	257	-	0.08	0.28	Plateau
A04S1	16/03/2018	0.4	-	308	262	-	0.09	0.29	Plateau
MetS	22/03/2018	0.62	-	297	252	-	0.06	0.24	Plateau
A05C	21/03/2018	0.3	96	380	246	0.06	0.08	0.34	Slope
A05C1	20/03/2018	0.66	60	324	251	0.09	0.09	0.36	Plateau
A05E	20/03/2018	0.47	65	310	257	0.07	0.12	0.31	Plateau
A05N	21/03/2018	0.26	58	367	277	0.06	0.1	0.28	Slope
A05W	20/03/2018	0.94	75	336	202	0.07	0.07	0.34	Plateau
A06C	18/03/2018	0.24	158	310	244	0.07	-	0.36	Plateau
A06N	18/03/2018	0.19	52	222	216	0.06	0.05	0.3	Plateau
A06S1	18/03/2018	0.24	60	285	222	0.08	0.18	0.42	Plateau
A07C	21/03/2018	0.45	86	299	263	0.06	0.08	0.33	Slope
A07W	21/03/2018	0.32	76	336	269	0.06	0.06	0.24	Plateau
A08C	20/03/2018	0.31	90	287	238	0.06	0.12	0.32	Plateau
A08E	20/03/2018	0.36	73	421	283	0.06	0.06	0.31	Plateau
A08W	20/03/2018	0.18	94	250	196	0.06	0.11	0.48	Plateau
A08W1	20/03/2018	0.24	80	205	258	0.05	0.14	0.38	Plateau
A09E	20/03/2018	0.35	127	319	292	0.05	0.1	0.24	Plateau
A09W	20/03/2018	0.39	38	307	349	0.04	0.06	0.23	Valley

4) The identification of snow pits in Figures 2, 4, 5, 7, and Table 2 is cumbersome when changing between text parts, Tables and Figures. Please use simple numbers from 1 to 29. You still can add things like C2, such as 1-C2 for Pit 1. This is much clearer than A0C2 because all pits are now called A0...

Thank you, we will renumber pits as suggested.

5) Line 179-180: What is the thickness range of the observed ice lens? How does SMRT treat its effects? Do coherent reflections between the upper and lower boundary play a role?

The ice lens thickness ranged from 1mm to 1cm, with mean of 2mm. These are not treated explicitly in SMRT as coherent effects have not yet been implemented.

6) Figures 3, 8, 10: Text labels and numbers are too small.

We will increase the size of these.

7) Line 204 and 210 "Atmospheric correction": This term is irritating. Tb was adapted to flight height, not corrected. Only errors can be corrected in my understanding.

We will use the alternative phrase 'adjusted' in line 210 (and elsewhere) and relabel the section 'Adjusting for the atmosphere' in line 204.

8) Line 232 (and elsewhere, including the abstract) "Anisotropic atmosphere": What do you mean? An atmosphere that contains anisotropic particles, such as ice crystals? Or charged particles in a magnetic field? Or do you mean anisotropic radiance? It is clear, that radiance varies with direction, even in a plane-parallel atmosphere. Therefore, the tipping-curve method has been used for a long time. But this does not mean that the atmosphere is anisotropic.

We will use the terminology 'anisotropic atmospheric radiance'.

9) Line 240: "nadir ground-based TBs". See General comments, above. Measurements may be distorted (mostly enhanced) by the effects mentioned there.

We will include a discussion about these effects, as well as more detail on the limitations of this study. A note will be included around line 240 to state that the measurements themselves may be subject to error due to shadowing of the sky and emission from the radiometers.

10) Figure 7: I do not understand the box symbols. The key is incomplete.

The following will be added to the caption: the airborne data box extent shows the interquartile range, the internal line represents the median and box plot whiskers extend to +/- 1.5 times the interquartile range. Open circles are outliers in the airborne observations. SMRT simulations of the base case are shown by the blue spots

11) Figure 8: Why not without atmosphere, or with a time-constant atmosphere. It is not clear if the changes are due to the atmosphere or due to the surface.

There are minimal differences in the simulations: the same pit information is used in both sets of simulations. The only difference is the interpolation of temperature and the atmospheric emission. Within Figure 8, there are only noticeable changes in observations. The observed changes between flights can be represented by adding / removing the low density surface snow in the simulations i.e. by comparing Figure 8 and Figure A2.

12) Figure 9 is very helpful because it shows the weather history. Its information could be used to better interpret the radiometer data. The temperature remained below freezing. No changes are expected for the ice lenses. Temperature-gradient metamorphism with slow changes only. The figure also indicates that time series of radiometric measurements at the same temporal resolution might be valuable.

We agree that a time series of radiometric observations would be hugely valuable – regretfully these do not exist for this dataset. We will reflect these comments in the results section description of Figure 9 by adding '*No significant changes are expected in layer microstructure throughout the course of the field campaign as the temperature remained below freezing and only small changes in SSA expected over the days between flights'.* We did originally attempt to model changes in microstructure due to temperature-gradient metamorphism between flights and found these to be negligible.

13) Table 4: Text unclear. I don't see any "effect of thin...". I only see numbers. Please clarify. Why are they all negative?

The numbers are negative because including surface snow lowers the brightness temperature. We will add the following text to the caption: ‘Negative values indicate that inclusion of low-density surface snow reduces the brightness temperature’.

14) Line 330: "This suggests that emission from the atmosphere may dominate ... at 183 GHz". It appears to me that the author did not check the actual brightness temperatures & opacities involved.

We will add the following sentence: ‘This is consistent with the higher measured and simulated emission at 183 GHz shown in Appendix Figure A1.’

15) Figure 10 is very helpful. It is the only part where we clearly see the influence of the atmosphere. However, the analysis and description should be improved, e.g. on Line 333: "the atmosphere reduces the RMSE of the base simulation medians". I cannot see any RMSE in this figure. Do you mean the widths of the distributions shown? Please don't call this an error. And certainly not of the medians. Later, on Line 338, you mention something with respect to the distributions. I was unable do understand this text.

We will use Root Mean Square Difference rather than RMSE. We will add the following sentence ‘This means that the simulations form different distributions to the observations’

16) Line 345: Upper frequency limit of IBA: There is no fixed limit. The error of the approximation just increases with increasing frequency (and is larger for scattering in backward than in the forward hemisphere). "radius" should be defined, or else replaced by "correlation length".

We will replace the text with: ‘With an estimated limit of wavenumber k_0 ~ 1.5 x radius of spheres to keep the error of the approximation within reasonable limits, as specified by Picard et al., 2022, the IBA upper frequency for ….

17) Line 350: "Underlying topography": Do you really mean topography, here? Or dielectric properties of the underlying ground? The topography, in terms of slope steepness, orientation, and the solid angle of open sky above the surface point is relevant at all frequencies. The discussion that follows seems vague and not enough specific to the situations of the study.

The topography of snow surface would be relevant at all frequencies, but the accumulation of snow in depressions will mean the topography of the snow surface is smoothed compared with the surface underneath the snow. Both simulations and data show less differentiation between topography classifications at higher frequencies. There is no difference in dielectric properties of the underlying ground in the simulations, so from that perspective only the measured snowpack properties (driven in part by topography) changes between pits. We propose to adjust the paragraph to read:

Underlying topography is relevant at 89 GHz but becomes less relevant at higher frequencies. As the frequency increases, the penetration depth reduces and the sensor may only see the upper portion of snowpack. This is the dominant effect and results in smaller differentiation between TB classified by ground topography. However, structural changes and spatial variability in snowpack properties driven by topography may result in a topographical signal in the TB despite the signal not penetrating to the base of the snowpack. Small differences between topographical types persist even at 243 GHz in Figure 8.

18) The discussion on Lines 366 to 370 indicates that the selection of sensors used was not optimal. A mapping sensor with sufficient spatial resolution would have been more helpful, even if it is at a single frequency, such as 89 GHz. As an alternative a movable radiometer on a sledge would also give information on the spatial variability. Of course, this is no argument against the mentioned snow micropenetrometer. Both together would be excellent.

Rather than non-optimal selection of sensors, this discusses a limitation of the study. The sensors used are used or planned atmospheric frequencies and the question is whether we can use electromagnetic modelling to account for the snow emission and ultimately retrieve atmospheric information from these data. There will always be challenges in comparing point and areal measurements, and ideally we would have ground-based instruments at these frequencies too, used to make spatially distributed observations. We hope that such a field campaign may be possible in the future.

19) Finally, I am surprised about the large standard deviation of all simulated TB values. What is the reason? And how can you get more specific results that better focus on the actual situations?

The large standard deviation comes from the range of measured SSA and density within individual snowpits. They cover a plausible range of observations, with the 'base case' giving the best estimate. In order to get more specific results, we would need the 3D structure of the snow over the footprint of the sensor and solution of Maxwell's equations e.g. with NMM3D (e.g. Xu et al., 2010: https://doi.org/10.1109/JSTARS.2010.2053919), but we do not have this information.