Reviewer 2- Alan Geer (our responses to the comments are given in red below)

Possibly for the first time, this paper demonstrates good agreement between snow radiative transfer simulations (driven by snow pit measurements) and downlooking microwave observations at higher frequencies, i.e. 89 GHz to 243 GHz. This illustrates a path towards using snow radiative transfer, driven by multi-layer snow models, in the assimilation of microwave observations for both weather forecasting and for the inference of surface snow properties. The coupling of ARTS and SMRT models (and the demonstration of why this is important) is also an important step. The paper also illustrates some of the remaining difficulties to be solved. One of these is the occasionally large mismatches between point snow brightness temperature (TB) measurements and airborne TB measurements with fields of view up to 100m. Another is the drop in TB of up to around 20 K observed from one flight to the next, which was attributed to fresh snow on the surface, and illustrates strong temporal variability.

Overall, this paper is an important step forward, it will be of great interest to many scientists in the field, and it is well presented. However, there are a few areas where the methodology or results could be better explained, there are some possible uncertainties that might deserve more consideration, and it is important that the abstract and conclusions should clearly indicate the scope and limitations of the work.

We thank Dr. Geer for the positive overview of this paper and will bring out the scope and limitations of the work more clearly.

Main points

1) As described in the paper's abstract, coupling ARTS and SMRT is a major developmental step. However, for such an important part of the paper there is very little detail. For example it is not clear how the downwelling atmospheric radiation field is represented by ARTS and then coupled into SMRT (presumably as radiances at the quadrature angles of the discrete ordinates solver used in SMRT, but this is not stated). Assuming ARTS and SMRT are not "fully" coupled, by which I mean a discrete ordinates problem is solved simultaneously in the snow and atmosphere, I imagine that ARTS is called first to simulate the downwelling radiance field, the upwelling radiance at the observation angle, and the surface-to-aircraft transmittance. Then presumably SMRT is called and its output corrected to aircraft level with the paper's equation 2. These issues should be clearly discussed in the paper in section 2.4. It would also be good to have details of the ARTS radiative transfer solver method, mainly just to exclude the unlikely scenario that atmospheric scattering is being represented too (which could need "full" coupling of the solvers).

We appreciate and agree this needs further detail. It is correct that the two are not fully coupled and we will include a flowchart in section 2.4 to describe the steps taken. This will then sit alongside the publicly available code. We will also add the following text: 'The ARTS Clear Sky (non-scattering) solver is used for a 1D atmosphere. The sensor is represented using a "top-hat" channel response in each of the two sidebands, with a frequency resolution of 0.1GHz.'

2) Some of the descriptions of how Arctic microwave observations are used at NWP centres (with ECMWF as the main example) could be made more precise. Microwave sounding radiances are used over snow and sea-ice surfaces if the surface contribution is small enough. For example the 183+/-3 GHz channels are usually assimilated over sea-ice and snow, whereas 183+/-7 GHz channels are not. Also, one of the main problems with ice and snow surfaces, from an NWP point of view, is that a constant surface emissivity cannot be assumed. Over non-snow land surfaces, the dynamic emissivity retrieval technique typically assumes that an emissivity retrieval can be extrapolated using

a constant in frequency approximation (for example an 89 GHz retrieval is used as the surface emissivity for 183 GHz assimilation over non-snow surfaces). A few more detailed points illustrating these issues:

line 28-29: “data over Arctic regions” could more precisely be “surface-sensitive data over Arctic regions” and the reason for the data exclusion is usually the possible presence of snow and ice.

We will make this substitution

line 30: “potential benefits of .. microwave data over Arctic regions” - but some Arctic microwave data is already being assimilated operationally, particularly in summer, as illustrated in the Lawrence et al. (2019) studies, and as is described clearly on lines 35-38.

We will remove the word ‘potential’

line 46: Baordo and Geer (2016) describe the assimilation of only snow-free land surface data, for the SSMIS instrument, and they eliminated surface-sensitive observations at latitudes greater than 60 degrees or for surface temperatures less than 278 K. Hence the point about using atlas in these possible-snow areas is not so relevant. Within Geer et al. (2014) there is a description of subsequent work that extended SSMIS usage over snow and sea ice surfaces following the above-described template. This actively assimilates 183+/-3 GHz and higher peaking channels. The dynamic emissivity retrieval is made at 150 GHz and then assumed to be valid also at 183 GHz, making sure the extrapolation in frequency is as small as possible (but even this small extrapolation induces errors that are too large to permit assimilation of channels that have stronger surface sensitivity, like 183+/-7 GHz). This snow and ice dynamical emissivity retrieval approach started at ECMWF even earlier with clear-sky MHS assimilation following the work of Di Tomaso et al. (2013: Assimilation of ATOVS radiances at ECMWF: third year EUMETSAT fellowship report. EUMETSAT/ECMWF Fellowship Programme Research Report No. 29, available from http://www.ecmwf.int.)

We will adapt the text to reflect this point and include the additional information on dynamic retrievals for narrow band channels, and associated references.

line 48: “microwave emissivity is highly spatially variable” - this could be a place to mention that it is also highly variable in frequency.

We will adapt the text to read ‘microwave emissivity is highly spatially variable, highly dependent on frequency and has high uncertainty due to its sensitivity to the microstructure…’

Just a discussion point, but these dynamic surface approaches are continuing to be improved for NWP, and in particular we are starting to improve representations of the frequency dependence of surface emissivity. Compared to the more physical approach of the paper under review, the dynamic approach has the advantage of being able to adapt the surface to match what is in the sensor’s the field of view, thus dealing with the time and space heterogeneity issues that are well illustrated in the paper, and hence they may continue to provide strong competition for the fully physical approach for some time to come.

This is a very welcome discussion point and incoming improvements to NWP. We hope that the physical approaches will underpin the representations of frequency dependence of surface emissivity and will help drive improvements in the land surface model representation also.

3) It would have been good to discuss the surface characteristics of the Trail Valley Creek site and how they relate to possible uncertainties in the surface radiative transfer. In particular, vegetation, since it appears the surface is being modelled as bare soil. The satellite pictures seem to show trees in the valleys and the possibility of grass or shrubs on the plateaus. Could vegetation have impact on the radiative transfer, particularly if it contains some liquid water, and particularly as the snow cover is not deep, e.g. 20 - 100 cm (lines 109-110)?

This is a very good point and could certainly impact the quality of the simulations. We will include this in the discussion. It's worth noting that the dominant land surface is tussocks (37%) followed by dwarf shrubs (24%), whereas trees only constitute 2% Grünberg et al., 2020 https://doi.org/10.5194/bg-17-4261-2020. Ideally we would have a radiative transfer model that simulates the effects of the vegetation, but this is not yet possible with SMRT. Vegetation was noted in many but not all pits, as shown in the Table below. We propose adding this table to the Appendix. Emission from the vegetation not accounted for could potentially contribute to an underestimation in simulated brightness temperature. However, the contributions from twigs and grasses are likely to be small. The change in snow structure due to vegetation in pit 4-3C1 (A03C1) – 'very loose snow towards bottom, blocked by vegetation' could be a contributing factor in the discrepancy between observations and simulations and will be included in the discussion.

Pit	Vegetation notes
1-2C	Tussocks and a few shrub twigs
2-2E	Tussocks and dwarf shrubs
3-2W	Grass tussocks
4-3C1	Grass (very loose snow towards bottom, blocked by vegetation)
5-3E	Lots of shrubs to 60cm
6-3W	-
7-4C	Tussocks and twigs
8-4C1	Tufts of grass
9-4N	Tussocks and twigs
10-4N1	-
11-4S	Tussocks and twigs
12-4S1	Lichen
13-MetS	-
14-5C	-
15-5C1	Lichen. Trees around pit
16-5E	Further from the trees than the other. Lichen
17-5N	-
18-5W	Lichen, shrubs, trees around snowpit
19-6C	Shrub, lichen, vegetation 7cm tall in pit
20-6N	Grass and lichen
21-6S1	Lichen, small bushes
22-7C	2m shrub in area, 30cm shrub in pit
23-7W	-
24-8C	Lichen
25-8E	-
26-8W	-
27-8W1	Grass and moss
28-9E	-
29-9W	-

4) There could be some more investigation of the way the temperature profile is determined, and whether this has any bearing on the radiative transfer uncertainties. Lines 168-170 describe a linear extrapolation from the air temperature (ultimately from dropsondes?) through the snowpack to a stable lower layer temperature. Is this sufficient to represent the relative complex dependence of the snow temperature profile on the surface air temperature, particularly its insulation properties and speed of heat transfer? For example, when trying to explain the drop in brightness temperature between flights C087 and C090, could this be relevant? Looking at Figure 9, at the time of the C090 flight, could the snow still be cold after a night that dropped below -25 degrees C, and hence has not yet caught up with the rapid rise in the air temperature?

All flights were around 2pm, so residual cold from nocturnal cooling is unlikely to have persisted. The interpolation is taken from the ground station measurement, not the dropsondes (i.e. from met data in Figure 9), and we are assuming the value measured at the Met Station is representative over the whole TVC site. This is a simplification, and a better method would be e.g. snowpack modelling. However, there were only minor differences between using the measured pit temperatures and interpolated temperature estimates, so full snowpack temperature profile modelling was deemed overkill.

5) It could be worth specifying also the type of seasonal snow in the abstract and conclusions. Currently on line 406 the work is described as relating to "an Arctic tundra snow environment" but that could more precisely be "an Arctic tundra snow environment in late winter". In order to use satellite observations for weather forecasting globally and in all seasons over snow and sea-ice, we will need to be able to simulate many other snow types such as wet snow and including diurnal cycles of freeze and thaw.

This is a good point and we will make the suggested change. We will also include this in a discussion paragraph describing the limitations of this study and future research identified as a result.

Minor points

line 40 - 19, 37 and 89 GHz channels are extensively used for water vapour, cloud and precipitation assimilation, but the statement that "window frequencies around 19, 37 and 89 GHz are used to obtain information about the surface (e.g. snow)" could be misread to exclude this and to imply that these frequencies are not useful for the atmosphere.

We will replace 'used to obtain' with 'typically chosen for applications requiring'

line 44-45 - "forecast and analysis"? rather than "forecast analysis" which is confusing.

We will make this change

line 115 - if the sled measurements are nadir to a snow surface that may be sloping, is any adjustment made when sled measurements are mapped to true nadir aircraft measurements?

No adjustments were needed – sled measurements were made over near horizontal surfaces line 196 - “representing the layer density and SSA by the largest and smallest observed values” - it’s not clear whether this means within a single pit, or across all pits.

This is within each layer per single pit, and will be clarified in the text.

line 197 - the “full range of plateau airborne observations” deserves some explanation, as at this stage it’s really not clear that (presumably) this means across the two flights and incorporating all plateau measurements in the relevant area illustrated in figure 6 and following the comparison strategy described in lines 262-265. It might be worth re-ordering some of this information (e.g. to put it in the section on aircraft data?)

This will be changed to ‘all airborne observations from the C087 flight over areas within A04 classified as plateau’

line 221-227 - in the adjustment of the background atmospheric profile to fit aircraft-measured downwelling radiances, can the dropsonde profile below the aircraft be modified to fit observations? It’s not clearly excluded in the text. And how representative is the lowermost dropsonde air temperature of the snow temperature? (See main point 4)

As suggested in the comment, because no downwelling observations were available below the aircraft, it is not possible to modify the profile below aircraft height. We can update the text to make this clearer, e.g. “Temperature and water vapour profiles used as input for ARTS were retrieved for each AOI in each flight. Background profiles were taken from a combination of dropsonde profiles, from sondes released before the low-level AOI runs, and profiles from the Met Office operational global NWP model (above sonde height). The retrieval adjusts these background profiles to match aircraft-level downwelling observations in the vicinity of each AOI at 183±1, ±3 and ±7 GHz. Because downwelling observations are only available above the aircraft, the profile below aircraft height (~590 m in the AOIs) is not adjusted in the retrieval.“

For additional information, the sondes were dropped when the aircraft was at a higher altitude before the AOI runs (~7500 m in C087, and ~1800 m in C090), meaning the sonde profiles start above the altitude at which downwelling observations were made in the AOIs (~590 m). A portion of the dropsonde profile is therefore modified during the retrieval, just nothing below the height of the aircraft in the AOIs. The section of profile below the aircraft is relatively small compared to the total atmospheric profile (max altitude 80,000 m), therefore any uncertainty resulting from this is also expected to be relatively small.

The dropsonde temperatures were not used in the interpolation of snow temperatures: this information comes from the meteorological station.

Figure 4 caption (figure 5 similarly) - the significance of the square could be explained in words in the caption (the lines indicating that it is a zoom are faint and easy to miss), The caption should also explain the meaning of the error bars

In the caption we will explain that the zoom is used to provide space to label these specific outlier pits, and will make the zoom box more obvious. We will also include text to indicate the SMRT error bars arise from the variability of the simulations i.e. using the maximum and minimum density and SSA within each layer, and the observed error bars arise from the maximum and minimum of three adjacent radiometric observations.

line 241-242 - linked to Figure 4 and the need to state clearly what the error bars mean, it's not clear how the "range of simulations" mentioned here is being generated.

Please see previous response

Figure 7 and 8 captions - need a careful description of the meaning of the various boxes, whiskers and spots.

As indicated in the response to Prof. Mätzler's comment, we will include the following text: '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'

Line 383 - this RMSE calculation is a headline result from the paper, quoted in the abstract, so it should be clear how it is obtained. For me the description "RMSE of the base simulation medians by frequency and flight" is not quite clear enough. For example whether this really is the RMS of the SMRT base simulation median minus the observation median and, if I understand correctly, that means the sample over which the RMS is computed is of size ten, e.g. "across 5 frequencies and 2 flights"? Could this be more precisely described in the abstract too, noting specifically the use of medians in the calculation? Because if it is based on medians, then we might expect the RMSE comparing the errors of individual pits to individual surface categories and AOIs to be somewhat higher. That would also be a useful figure to calculate.

We will use Root Mean Square Difference rather than RMSE to describe this. This is the difference between medians so it is correct that n=10. We will clarify this in the abstract and include the individual pit RMSD for flight C087 also. For individual surface categories within AOIs against pits, the RMSD is 35.7 K excluding the atmosphere and 18.4K with the atmosphere included for flight C087 (n=145). For flight C090 the RMSD without atmosphere is 29.2K and with the atmosphere is 21.7K. These will be included in the revised manuscript, but will not be presented as a combined figure for both flights because of the difference for C090 likely caused by the thin low density snow layer.

Line 396-397 - main point 2 again: "In current numerical weather prediction models, microwave emissivity is assumed to be constant over snow-covered surfaces or derived from a monthly climatology, with errors too large to be able to use satellite observations in the Arctic": Dynamic emissivity retrievals have been used over snow and sea-ice at ECMWF to allow assimilation of 183+/-3 GHz channels (and higher peaking channels) since the work described in Di Tomaso et al. (2013) and Geer et al. (2014). Hence the snow emissivity does not usually come from atlas and it is not assumed constant in time or space (but it is assumed constant with frequency from 150 to 183 GHz). Nonetheless, the dynamic emissivity retrievals are not yet good enough to permit assimilation of strongly surface sensitive channels (e.g. 183+/-7 GHz) over snow so there still is plenty that can be improved by physical modelling as described in the paper under review.

This is a valuable discussion point – we will include these references and briefly discuss dynamic emissivity retrievals.

Line 422-423 - "the addition of fresh, low density precipitation and a later wind event that removed it over the space of a few days caused differences in observed brightness temperatures." - is the attribution of these changes in TB to the fresh snow event certain enough to be able to say for definite it "caused" it here in the conclusion, rather than to say "likely caused", for example?

While is it consistent with the simulations and could not otherwise be explained, we accept that it is not possible to attribute this conclusively and will replace ‘caused’ with ‘likely caused’ as suggested.