Reply to Referee#1 on Manuscript # acp-2021-1064 in ACPD: "Evaluating seasonal and regional distribution of snowfall in regional climate model simulations in the Arctic" by Lerber et al.

We would like to express our sincere appreciation to the reviewer for the valuable comments and corrections on our manuscript and the opportunity to revise and improve the manuscript.

We have included the reviewer comments below in bold and italic, and responded to them individually following the numbering. The individual comments and responses are followed by the revised text, with changes highlighted with colors: deletions in red and additions in blue.

1. The major comment for this manuscript is that the authors could consider some objective metrics in the evaluation, e.g., spatial correlation, Taylor skill score.

Thank you for your comment, but we have to take the poor sampling of CloudSat into account, which limits a throughout quantitative evaluation. Therefore, we believe that our presented approach to quantitatively evaluate HIRHAM with CloudSat is solid and the best one could do. Saying this, we present the spatial distribution (Figs. 3, 6, 8), means (Figs. 4b, 5), and frequency distribution (Fig.4a).

Here is our rationale why we came to concise to the present version. As stated, the manuscript studies the differences with two approaches, i.e. the assessment of surface snowfall rate (observation-to-model) and the radar reflectivity factor profiles (model-to-observation).

In the part of surface snowfall rates, the results are mostly compared to the study of Edel et al. (2020), where the analysis was performed by looking at the distribution of snowfall rates, distributed yearly or seasonally, and studying differences of mean snowfall rates as we presented them in this current version of the manuscript. We want to report our findings comparable to theirs to be able to build the link between these two studies. Additionally, with surface snowfall rates we compare the model and observations, how well model qualitatively reproduces the CWTs.

Due to CloudSat's long revisiting time, daily skill scores are not providing meaningful comparison as it was shown in Souverijns et al. (2018). For the monthly and yearly means, the spatial differences and frequency distribution are a well-suited evaluation. According to your comment, we include now additionally the RMSE:

In lines 341-343: "Though model and observations show similar spatial distributions, distinct spatial differences occur (Fig. 3c), and e.g. root-mean square error in the yearly surface snowfall rates is high with 148 mmyr^{-1} between HIRHAM5 and CloudSat, and 175 mmyr^{-1} between ERA-Interim and CloudSat."

In the part, where we investigate the differences in radar reflectivity factor profiles and CFTDs, the obviously largest differences occurred due to the small reflectivity portion of the too small ice particles and how, mainly due to this, in general, model and observations have highest reflectivity quantities in different altitudes, although clearly by regionally and spatially in the Arctic, the model reproduced the snowfall well as seen e.g. in Figure 8. Spatial correlation scores gave unjustified poor scores due to this above-mentioned difference. Therefore, we stay with these shown differences.

2. Lines 105-106: Why was 70N selected as the boundary of two rings? Please justify it

Many literature takes the 70°N latitude as a measure of the central Arctic or "mean Arctic". Examples are: Wang et al. (2012, doi:10.1155/2012/505613), Screen et al. (2011, doi:10.1029/2011JD015847), Koenigk et al. (2014,10.1007/s00382-013-1821-x), Bintanja (2018, doi: 10.1038/s41598-018-34450-3).

The additional description has been added:
"The two rings are separated to clarify the different characteristics of the southern and northern regions, where the 70°N defines the central Arctic boundary and also coarsely separates the Arctic Sea regions from the Arctic continental regions."

3. Lines 128-130: I might not agree with this statement. Simulation uncertainty comes from many aspects, and microphysics parameterization is only one of them. Boundary layer parameterization can also significantly influence the model dynamics and then influence the snowfall simulation. If the authors did not conduct the sensitivity test on model physics schemes, it is not suitable to give this statement.

We suggest a modification to the text:

"Therefore, it is assumed that the differences between the modeled snowfall and observations are in lesser degree related to the simulated large-scale flow but mostly caused by the ECHAM5 boundary layer and microphysical parameterization employed in HIRHAM5 and observational uncertainties."

4. Section 2.3: Please provide the quantitative uncertainties of the two CloudSat products.

The 2B-GEOPROF-product output is reflectivity factor profile. The sources of measurement uncertainty include uncertainty in the absolute radiometric calibration and measurement noise. The noise characteristics of the CPR vary with signal strength. It is estimated in Wood et al. that the resulting uncertainties range from 3 dBZ for a reflectivity of -30 dBZ to about 0.1 dBZ for reflectivities above -10 dBZ. Calibration errors, which would result in a bias in the measured reflectivities, are expected to be less than 2 dB based on a prelaunch calibration error budget (Tanelli et al., 2008), but the value of this bias is basically unknown and typically not considered. We added in line 168:

"The minimum detectable reflectivity is dependent on, e.g. cloud cover, seasonal changes in temperature, surface type, and atmospheric attenuation, typically varying by ~1 dB over the globe in the range from -30.9 to -29.9 dBZ (Tanelli et al., 2008) and the measurement uncertainties related to noise range from 3 dBZ for a reflectivity of -30 dBZ to about 0.1 dBZ for reflectivities above -10 dBZ (Wood and L'Ecuyer, 2018).

The uncertainty of snow profile product, 2C-SNOW-PROFILE, was expressed in Edel et al. 2020 as a relative uncertainty, which is the ratio between the mean single surface snowfall rate uncertainty and the surface snowfall rate, ranging from 1.5 to 2.5, with higher values associated to complex topography and high frequency of mixed phase precipitation. There are a few studies, which have estimated the accuracy of the product respect to weather radar estimated snowfall rate (Cao et al. 2014, Norin et al. 2105) with very similar results. The product has a good detectability of light snow (snow water equivalent less than 1 mm h^{-1}), however limited ability to retrieve at the higher end of snowfall intensity distribution (> 1 mm h^{-1}). We added in line 193

"Thus, we are confident to use the output of 2C-SNOW-PROFILE product as the ground truth, though acknowledging the relevant unreliability stemming from the uncertainties in observed reflectivities, the used retrieval parameters and it's a priori assumptions (Edel et al., 2020). The product has shown a good detectability of light snow (snow water equivalent less than 1 mm h^{-1}), however limited ability to retrieve at the higher end of snowfall intensity distribution (> 1 mm h^{-1}) when compared to weather radar estimated surface snowfall rate (Cao et al. 2014, Norin et al. 2015). The relative uncertainty of the product increases with complex topography and higher frequency of mixed phase precipitation (Edel et al. 2020)."

5. Line 144: "Sect. A" -> "Appendix A"

Corrected.

6. Line 280: "East regions" -> "East region"

Corrected.

7. Line 294: "mmyr−1" -> "mm yr−1", please correct others throughout the manuscript.

Corrected.

8. Section 4.1: It is better to show the locations of Greenland, Barents, Kara Seas, etc.

The names of the Seas are now written in Figure 1 to ease the reading of the Section 4.1.

9. Lines 400-401: It would be better to show the results in other seasons in appendix or supplement.

The CFTDs images for other seasons are added to Appendix with other additional images. And text is added in the line 401

"For investigating the differences in the vertical reflectivity structure between the different regions we focus on the winter season (DJF) which cover snowfall rates of approximately 30% over all seasons. Furthermore, we reduce problems related to mixed-phase conditions as temperatures are generally low. The other seasons are shown in the Appendix B."

10. Section 5.1: Please discuss the differences more quantitatively.

We have added quantified occurrences and other values to the section in lines 417- 426.

Due to the lower occurrence (< 0.1%) of cold temperature reflectivities, reflectivities at warmer temperatures are relatively more frequent in HIRHAM5 than in CloudSat observation, with occurrences of > 0.8 % for HIRHAM5 and with occurrences between 0.6 - 0.8 % for CloudSat. However, HIRHAM5 is able to reproduce regional differences seen by Cloudsat correctly. Enhanced reflectivity related to the snow mode (- 10 and 5 dBZ) occurs at the warmest temperature in the North Atlantic (around -10°C) in both observations and model, similar at slightly warmer temperature in the Kara Sea regions. In the Chukchi Sea, occurrences (0.4 - 0.8 %) are confined to a narrow temperature range between -20 and -35°C, while in the Laptev Sea the distribution broadens to colder temperature again in both observations and simulations. In the Chukchi Sea, HIRHAM5 can also reproduce the increased reflectivity occurrence (0.6%) around -20°C in the lower latitude region compared to the higher latitude region. The strongest difference between the observed and simulated CFTDs is visible for Greenland where the simulations show reflectivities at much warmer temperatures (-20 to -10°C) and higher reflectivities (0-10 dBZ) consistent with the overestimation in snowfall rate by HIRHAM5 discussed before.

11. Figures 4 and 9: Please conduct significance test of difference for (b).

The student t-test is performed to the differences with random samples (10% of the total amount) of the defined difference distributions. The results show that the shown median difference is statistically robust. The tables for the results are shown below for you, and the text is added to captions of Figure 4b and Figure 9b:

Figure 4b:
"The significance of the median difference for both HIRHAM5 and ERA-Interim compared to CloudSat observations is shown to be statistically robust for all seasons performing the student t-test with random samples (10% of the total amount) of the observed difference distributions."

Figure 9b:
"The significance of the median difference for both original Tompkins and Sundqvist schemes to modified Tompkins scheme is shown to be statistically robust for all seasons performing the student t-test with random samples (10\% of the total amount) of the modeled difference distributions.

In the tables, the t-value quantifies the difference between the population means. Here, the other population is the 10% random samples of the difference distribution, and the other is the total distribution. The p-value is the probability of obtaining a t-value.

For boxplot 4b:

	HIRHAM5		ERA-I	
	t	p	t	p
DJF	0.621	0.535	-0.507	0.612
MAM	0.332	0.740	-0.563	0.574
JJA	0.205	0.838	-0.092	0.926
SON	0.336	0.737	-0.507	0.612

and for the boxplot 9b

	Original Tompkins		Sundqvist	
	t	p	t	p
DJF	0.401	0.689	-0.444	0.657
MAM	0.338	0.735	-0.577	0.564
JJA	0.032	0.974	0.388	0.698
SON	-0.452	0.651	0.392	0.695

12. Figure 7: "CFTD" -> "CFTDs", What's the meaning of "86500 - 6.6·10^6" in the caption?

The sentence is modified to make it clearer:

"the normalization is done by the sum of total hits, which varies from region and season, but typically the number of hits is ranging between 86500 - 6.6·10^6."

13. Lines 456-457: Please show the locations of the North Atlantic region, the East Siberian Sea and the Beaufort Sea in Figure 9.

The regions with approximate longitude degrees are now written to the text in in lines 456-457. The modified text is:
"The highest reflectivity values due to rain particles are concentrated in the North Atlantic region (20°W - 10°E) and some higher values are also modeled in the East Siberian Sea (150°E - 180°E) and the Beaufort Sea (150°W - 130°W)."

Reference:

Norin, L., Devasthale, A., L'Ecuyer, T. S., Wood, N. B., and Smalley, M.: Intercomparison of snowfall estimates derived from the CloudSat Cloud Profiling Radar and the ground-based weather radar network over Sweden, Atmos. Meas. Tech., 8, 5009–5021, https://doi.org/10.5194/amt-8-5009-2015, 2015.

Qing Cao, Yang Hong, Sheng Chen, Jonathan J. Gourley, Jian Zhang, and Pierre E. Kirstetter, "Snowfall Detectability of NASA's CloudSat : the First Cross-Investigation of ITS 2c-Snow-Profile Product and National Multi-Sensor Mosaic QPE (NMQ) Snowfall Data," Progress In Electromagnetics Research, Vol. 148, 55-61, 2014,doi:10.2528/PIER14030405