Review of "Comparison of optical-equivalent snow grain size estimates under Arctic low Sun conditions during PARMARCMiP 2018" by Jäkel et al.

General comment

The authors apply different methods to retrieve the optical-equivalent snow grain size over an Arctic area (part of Greenland and adjacent sea ice) from airborne and satellite albedo observations. The dataset is extremely interesting (because of the paucity of observations in that region) and challenging (because of the high solar zenith angles and the heterogeneity of the surface). However, the applied methodology has serious deficiencies. The main problems are a misinterpretation of the role of roughness on albedo, and the fact that shadows and snow-free young ice are not excluded from the retrieval and therefore cause artefacts and biases in the results (see detailed comments to Sect. 5.3) making the comparison of grain size retrievals from the different satellite scenes meaningless. I therefore recommend the editor to not accept the paper until a major revision is done and these methodological deficiencies are solved. Also, the sections describing the applied models should be clearer, and in general the paper requires a linguistic revision, as in many sections the logic of the syntax is poor.

Detailed comments

p.3, line 4-5: as you clarify here that the term "snow grain size" will refer to "optical-equivalent snow grain size" throughout the paper, please be coherent and replace in all following sections (including the figures) "SSA" and "optical-equivalent snow grain size" with "snow grain size".

p.4, line 30: remove ""covered the sea ice" as it is repeated immediately after.

p.4, Section 2.1: please add a description of the broadband albedo data collected at Villum (instrument used, measurement uncertainty), which are particularly important as they will be used in the analysis.

p.5, line 2: please explain what is the vertical resolution of the DEMs generated from the laser scanner observations.

p.5, line 22: " " Recent founding revealed that the accuracy of the Ice Cube measurements is lower than this 10%, even for the reported snow density of 230±30 kg m^{-3}. See for instance Calonne, N., B. Richter, H. Löwe, C. Cetti, J. ter Schure, A. Van Herwijnen, C. Fierz, M. Jaggi and M. Schneebeli (2020). "The RHOSSA campaign: multi-resolution monitoring of the seasonal evolution of the structure and mechanical stability of an alpine snowpack." The Cryosphere 14(6): 1829-1848. This study reported that SSA measured with the IceCube was systematically higher than values computed on tomographic images, approximately by a factor of 1.3. A definitive explanation for this disagreement has still to be found. Anyhow, as also reported in the introduction, optically-based SSA retrievals (either from albedo or from the single wavelength reflectance measurement applied in the IceCube) suffer from a ±20% uncertainty (i.e. 40% uncertainty range!) when the shape of the grains is not taken into account. Thus, in this section, I would include these arguments when presenting the total uncertainty associated to SSA measurements obtained from the IceCube.

p.6, line 28: please remove "However,"

p.7, line 1: Replace "Visual camera data of the surface taken aircraft-based" with "Aircraft-based photos of the surface"

p.7, line 2: please remove "However," as you are not introducing a contraposed sentence.

p8. lines 5-7: "The temperature profile on 26 March shows the strongest inversion of all three flights with surface temperatures of -30 C and -10 C around 880 hPa corresponding to an altitude of 1.1 km. The weakest inversion of about 5K was measured on 27 March." Please consider re-phrasing the first sentence (e.g.:...three flights, with -20 C difference between the surface temperature of -30 C and the temperature at the inversion height located around 880 hPa, corresponding to an altitude of 1.1 km"), and use the same unit for temperatures here and throughout the manuscript.

p.8, line 11: here an introductory paragraph describing the modelling strategy and the role of each of the applied models would be needed.

p. 8, line 13: "To simulate the snow surface albedo, the open-source Two-streAm Radiative TransfEr in Snow model (TARTES) was used (Libois et al., 2013)." In reality, you used several radiative transfer models, either in direct mode (to simulate spectral albedo from snow grain size) or in inverse mode (to retrieve snow grain size from measured albedo). As this section is dedicated to Tartes only, it would help a lot if you clarify for what purposes Tartes is applied (it seems to me that it was used in inverted mode to retrieve snow grain size from aircraft albedo measurements, in direct mode to calculate the albedo given as input to atmospheric radiative transfer modelling, to assess the impact of clouds on snow albedo and the impact of the assumed snow grain shape on the retrieved grain size...)

p.9-10, section 3.3 and 3.4: please clarify here that ICON-ART and Flanner and Zender scheme are applied only over land, to simulate the snow grain metamorphism in the Villum's region. I would place both models in the same section ("e.g. "modelling of the snow grain size metamorphism"?), as they will be used only for this purpose.

p. 9, line 30: "The acceleration of growth due to rain was additionally added and the reduction by snowfall was adjusted." Could you please be more specific? If this is the first time that this parameterization is applied, it has to be detailed in such a way that reader can reproduce it.

p.10, line 4: replace "temporal" with "temperature controlled" (moving it before "SSA evolution")

p.10, line 21: "Snow grain size and snow particle shape are then obtained by minimizing the surface directional reflectances at two wavelengths": I believe that this statement is wrong. Maybe you wanted to write something like "Snow grain size and snow particle shape are then obtained by minimizing the differences between modelled and observed surface directional reflectances at two wavelengths"?

p.10, section 4.2: here the applied terminology for various types of albedo is very unclear. If I interpreted correctly, I recommend using:

- "Black-sky albedo, corresponding to directional-hemispherical reflectance in case of only direct illumination" instead of "plane surface albedo, corresponding to the hemispherical reflectance and assuming only direct illumination"
- "White-sky albedo, corresponding to bi-hemispherical reflectance" instead of "spherical albedo"
- "Blue-sky albedo, corresponding to the actual albedo, i.e. a linear combination of black-sky and white-sky albedo" instead of "weighted albedo" (as it appear in Fig 2).

p.11, line 8: replace "as approximated by (Kokhanovsky, 2003)" with "as approximated by Kokhanovsky (2003)"

p.12, section 4.3.1: this section is carelessly written and extremely unclear, in both the logic and the language. Are you trying to describe two retrieval methods, one using TARTES and the other called "modified SGSP"? If so, state it, and explain the difference between the two methods. Here are some examples of unclear and inaccurate expressions (there are more, so please rewrite this section completely, with a much cleared logical structure and with precise statments):

- "the albedo ratio (R) of the SMART measurements at Lambda1 = 1280 nm and Lambda2 = 1100 nm wavelength" maybe it means "the albedo ratio (R), which is the ratio between the SMART albedo measurements at..."??

-"to minimize the retrieval uncertainty, which is affected by the measurement uncertainty of the spectral albedo" You cannot expect that the readers go to read Carlsen et al. (2017) to understand the meaning of this sentence. Please explain.

-" These aircraft measurements were performed over the Antarctic Plateau under clean atmospheric conditions" Do you mean that "Eq. 9 was applied by Carlsen et al. (2017) to retrieve snow grain size from aircraft observations over the Antarctic Plateau and, thus, at high elevation and in very dry air conditions"? High elevated and dry air, not clean… that is the point, to avoid the need of atmospheric correction.

- "…for the conditions during PAMARCMiP…" which conditions?

-"Different to Carlsen et al. (2017), in this study TARTES simulations were performed together with libRadtran calculations to generate LUTs accounting for the specific atmospheric conditions during the PAMARCMiP aircraft observations." In which way it is done differently from Carlsen et al. (2017)? Is it necessary here to know? What these LUT tables include? albedo versus snow grain size??

p. 13, line 5: "The albedo parametrization used in Sec. 4.3.1 is valid for observations at the surface". First of all the parameterizations are two, and they are not used, but "described" in Sec 4.3.1. Secondly, the statement is not true, as 1) SGSP was used for aircraft-based observations (Carlsen et al. 2017) and 2) the derived TARTES LUT tables fit to the atmospheric conditions observed during aircraft observations. I understand that you mean that, in the observed Arctic conditions, you need to apply the atmospheric correction for aircraft observations. Well then write it clearly, being rigorous and correct in your statements.

p.14, line 3-4: remove "at both wavelengths".

p.14, Fig 3: use a different line for simulated albedo at 3000m altitude and surface albedo with ropt=60micrometers. In the caption note, replace "2000" with "200".

p.14, line 14: replace "wavelength at 1700nm" with "1700 nm wavelength"

p.15, line 7: "i.e. decreasing optical depth of the snow layer." It should be "increasing" optical depth, but I would remove this part of the sentence, you have already explained the concept and this addition would not clarify anything.

p.15 Section 4.3.4: Also this section is particularly badly written, it needs to be reformulated. Especially the logic of the syntax is poor.

p.15, line 16: remove ",respectively" at the end of the sentence.

p.15, line 17: replace "form" with "from"

p.15, line 19: replace "is affected by" with "propagates from"

p.15, line 22-24: "Figure 4a compares the true snow grain size (without variation) and the uncertainty range of the retrieved snow grain size including variation in both directions ($\pm\alpha$)." The content of this sentence is expressed quite poorly: you probably mean that "Figure 4a shows the snow grain sizes calculated from the measured albedo (using Tartes) and from R (using the modified SGSP method) as well as the associated grain size uncertainties due to the errors in the measured albedo and R."? Please rewrite. Figure 4a is quite unclear:

- The use of "R" and "1700nm" to indicated when SGPS and Tartes methods are applied to retrieve *r* is confusing: it would be preferable to directly use Tartes and SGPS acronyms (e.g. r_{Tartes} and r_{SGSP})
- Do the two methods provide exactly the same *r*? If so, you must have applied some tuning, playing e.g. with the grain shape. You need to clarify this, and report how the match was achieved.
- Due to the unclarity of Sect 4.3.1, I could only guess which scheme you are applying for which wavelength (Tartes for α at 1700 nm and SGSP for R?) The logic between sections is broken, and the reader cannot really follow which scheme is applied where.
- The concept of "true optical-equivalent grain size" in x-axis is very ambiguous, please use a more correct terminology, something like "Retrieved optical-equivalent grain size from the measured albedo applying Tartes (r_{Tartes}) and SGSP (r_{SGSP})"
- In the y axis, I would write something "Retrieved optical-equivalent grain size from the upper and lower boundaries of the albedo uncertainty interval applying Tartes and SGSP".

p.15, line 26: "…with a larger slope of the LUT (Fig. 2)." Only now the reader starts to guess what this LUT may include. Here you need to be rigorous with the syntax: you don't really mean the slope of the LUT, but the slope of the curve representing the r_{opt}-albedo relationship in the LUT. Right?

p.15, line 26-27: "The relation between snow grain size and surface albedo for 1700 nm wavelength shows a stronger decrease than using R." Again there is a syntax problem: please be rigorous in the logic of your sentences, here and throughout the manuscript (I will not point to all the mistakes!). Maybe here you mean that "the r_{opt} decrease with the increase of albedo at 1700 nm is steeper that the r_{opt} decrease with the increase of R."

p.15, line 27-28: "This leads to a higher sensitivity on the SMART measurement uncertainty applying R, and an absolute error, which is about three to five times higher than using (1700 nm) for the snow grain size retrieval." From your previous sentence, it seems to me that the opposite should be concluded. Please replace "applying R" with "applying SGSP", and "using 1700nm" with "using Tartes" (also in sentence below, in line 31): isn't it so that you are here applying two different methods, Tartes and SGSP??

p.17, line 13: "Fig 5 displays the range…" Please explain what causes this range (uncertainty range in some input parameter?)

p.18, lines 1-27: again a paragraph that is very painful to read. Please rewrite it with rigorous logic (below are just few suggestions, the whole section need to be reformulated).

p.18, line 6-9: "The sensitivity of tau (T) and kappa (T) to temperatures in the range between -20 °C and -50 °C (see Tab. 4 in Flanner and Zender, 2006), differs from the parametrization for the range between -20 °C and 0 °C, resulting in this lower fitting quality. This likely is caused by the suppression of the temperature gradient effect at -50°C, which is strongly affecting the parametrization between -20 and -50°C." Did you skip some logical steps? The meaning of this paragraph is totally obscure. Should "sensitivity" replace "parameterization"?

p.18, line 14: "Radiative transfer simulations…": using which model? TARTES?

p.18, line 16: "While period I (22 March) was characterized by a SZA of 83 °": This cannot be, the period was overcast! This means, that the effective SZA is of the order of 52-55°. You should re-calculate the albedo with the correct SZA for overcast conditions.

p.18, line 17-18: "The simulations revealed differences of 0.09 between period I (α = 0.96) and period II (α = 0.87), which are higher than observed from the pyranometer measurements (0.07)." What is the observed albedo (by pyranometers in Villum) in these two cases, and what is its uncertainty? In the best case (with high standard and freshly calibrated sensors), the measured albedo has a 1% uncertainty in overcast conditions and 5% uncertainty in clear sky (due to the low SZA), so I consider a 0.02 difference in overcast-clear albedo between model and observations insignificant. However, it seems to me that the modelled albedo is much higher than the measured albedo. Can you hypnotize an explanation for this?

p.18, line 23: "period II was re-simulated using the spectral surface albedo from TARTES, which assumed only a diffuse component (spherical albedo) due to the impact of the cloud layer from period I". Very poorly written sentence: you mean "spectral albedo in Period II was re-calculated with TARTES assuming diffuse incoming irradiance to simulate overcast conditions"?

p.18, line 30: replace "…using satellite data" with "…using MODIS and Sentinel-3 data, respectively,"

p-19, lines 2-4: "Note, that the ICON-ART simulations were restricted to areas where the model-based land mask indicated land surfaces only, and are not included in the analysis here." Please remove this sentence, it is totally out of place here, especially if you clarify from the beginning that ICON-ART and Flanner & Zender are only applied over land.

p.19, fig 6: it would very much helpful if the coastline contour of Greenland would be marked in the maps. Currently, there is black contour that is not explained, and it seems that does not correspond to the coastline (what is it?). Latitude and Longitude are inverted in x and y-labels of the plots.

p.21, line 10-11: "The best agreement to IceCube measurements (ropt = 53 µm) on this day were derived from the 16:45 UTC MODIS overpass (ropt = 54 µm)". When you add the uncertainty in the IceCube measurement (see also my comment below related to Fig. 8) you may conclude that 2-3 ropt estimations fall into the uncertainty range of the in situ observations.
p.21, line 20: replace "…leads to a large diversity, compared to the SGSP and SMART results" with "…leads to a large difference between SGSP and SMART results" (if I well understood the meaning…).

p.22, fig 8: please replace "IRQ" with "IQR" in subplot b. For the discussion, it would be important to include the uncertainty of the in situ optical effective radius measured with the IceCube (a minimum of ±20% uncertainty, but also keep in mind the systematic bias observed by Calonne et al. 2020).

p.23, line 12:"The obvious larger variability of…" Why obvious? I recommend removing "obvious"

p.23. line 13: "The retrieval methods are based on radiative transfer theory assuming." Uncomplete sentence, please correct.

p.23, line 14: "Ice drift in the Fram strait caused an increase in surface roughness" Increase? With respect to what, the rest of the Arctic?

p.23, line 20-24: please use the letters a), b) c) and d) to label the subplots of Fig 9. When referring to them, please indicate the correct labelling of the plot ("Fig 9a, point 1" instead of [1], etc.)

p.24, line 5: I recommend replacing "limited" with something like "not suitable".

p.24, line 17: "As shown in Fig. 9, also regions with young ice contributions, where snow cover is greatly diminished,…" I would specify that you are referring to Fig 9c and d, and I would replace "greatly diminished" with "thinner than in the surrounding areas".8

p.24, line 27-28: "Therefore, the identification of shadows from the camera images is well suites to quantify the surface roughness affecting the surface radiative properties." I don't see the point here: if you have the DEM generated from the laser scanner data, you are perfectly able to estimate the extension and orientation of the shadows once you know lat, lon, and time. You don't need the camera images for this.

p.24, line 33: please add "surfaces" after "Significant percentages of light rough and rough snow"

p.25, line 2: please replace "when" with "where"

p.25, line 3-4: "It clearly demonstrates that small scale variations in snow coverage, for example, due to refrozen leads, needs to be excluded from the snow grain size evaluation." That is certainly true, although the concept is badly expressed: instead of "small scale variations in snow coverage" you should use a more precise expression, such as "areas without snow or with thin snow layers that are not optically thick". However, as you have recognized this problem (artifact) in your effective snow grain size retrieval, you should correct for it, either excluding the thin ice areas from your analysis, or estimating the bias that this problem causes in your results, and, thus, present the results without this bias. Otherwise, you will include physically meaningless results.

Anyhow, the main problem of this section is a misinterpretation on the role of roughness on albedo. Although fraction of shadowed areas and surface roughness are correlated (because

shadows are caused by roughness) it does not mean that roughness and shadows cause the same effect on albedo. Roughness decreases the albedo (by few percent) because photons reach the surface features that receive most of the irradiance with lower zenith angle than the surface slopes that receive less irradiance. This effect should affect the measure albedo independently on the platform from which albedo is measured, so I don't see how it could explain the seen variability of snow grain size retrieved by the different methods and platforms. Shadows, on the contrary, do not decrease the surface albedo. They are areas that look darker than the surrounding because they receive less illumination, but their actual albedo is higher than the surrounding, because they only receive diffuse light, which is richer in the visible wavelengths. This means that, to estimate the snow albedo from satellite images, the shaded areas should be masked out, as they represent an artefact, negatively biasing the albedo. Since the fraction of shadows in a defined area depends on the time of the observations, it is not at all surprising that albedo (and, consequently, retrieved snow grain size) exhibits significant variations among the different satellite images when the shadow artifact is not corrected. This problem is particularly serious at high SZA, as the authors noticed. For instance, I suspect that the difference between the frequency distribution of retrieved snow grain size from Aqua scene at 10:35 and Terra scene at 11:50 (Fig 7) is due to the fact that shadows cover a larger area at 10:35 and therefore the uncorrected albedo is lower and the retrieved grain size is larger than at 11:50. Since the shadow artifact is not corrected, the comparison of grain size retrievals from the different satellite scenes is meaningless.

p.25, line 16-17: "the effect of the unknown snow grain particle shape may lead to uncertainties in the range of 35% in extreme cases when using LUTs based on calculations for a mixed shape particle type, which is significantly higher than reported by Picard et al. (2009) with uncertainties of ± 20 %." Your results are perfectly in line with Picard et al. (2009), as a range of 35% is inside the 40% range (±20%) of uncertainty obtained by Picard et al. (2009). Please correct.

p.26, line 6: after "…the relative ~~retrieval~~ biases" please add "between the 1700nm and ratio R retrieval methods"

p.26, line 9-10: please correct the poorly written text, for example as "Both retrieval approaches show similar results for all snow shapes, ~~.The differences for~~ with the R-based retrieval being ~~are~~ only slightly higher than the 1700 nm retrievals. ~~, what indicates that in both retrievals the assumption of snow grain shape is crucial and may lead to systematic biases.~~ The assumption on the grain shape has a much more critical impact on the retrieval".