Interactive comment on “Evaluation of Biases in mid-to-high latitudes Surface Snowfall and Cloud Phase in ERA5 and CMIP6 using Satellite Observations” Hellmuth et al.

General Comments

In the following response to the reviewer we keep the reviewers comments (black) and our answers from the previous review (dark blue). The responses to the current review phase are in light blue, additionally newly added text to the manuscript is in italic-bold.

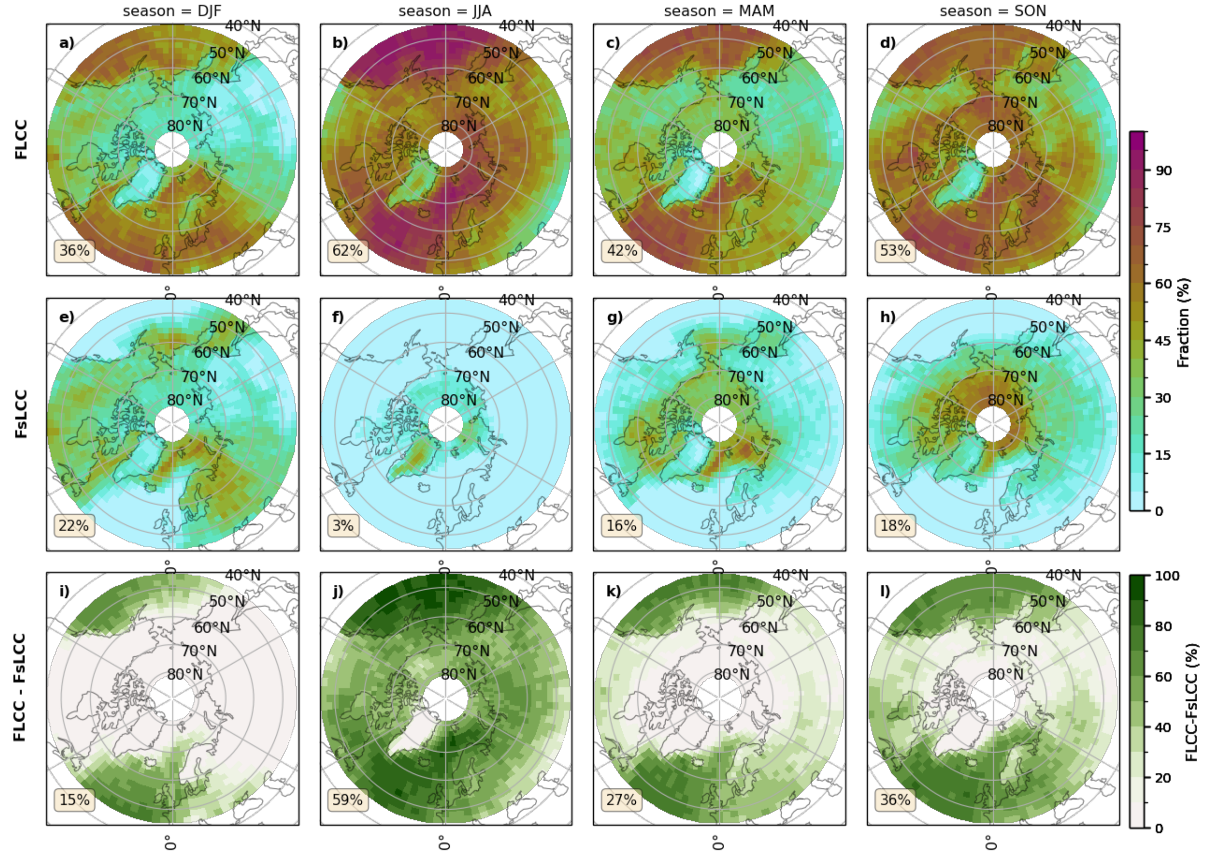
The authors have addressed in detail all my comments and concerns from the initial submission. I commend their efforts and appreciate the additional clarity of the revised manuscript.

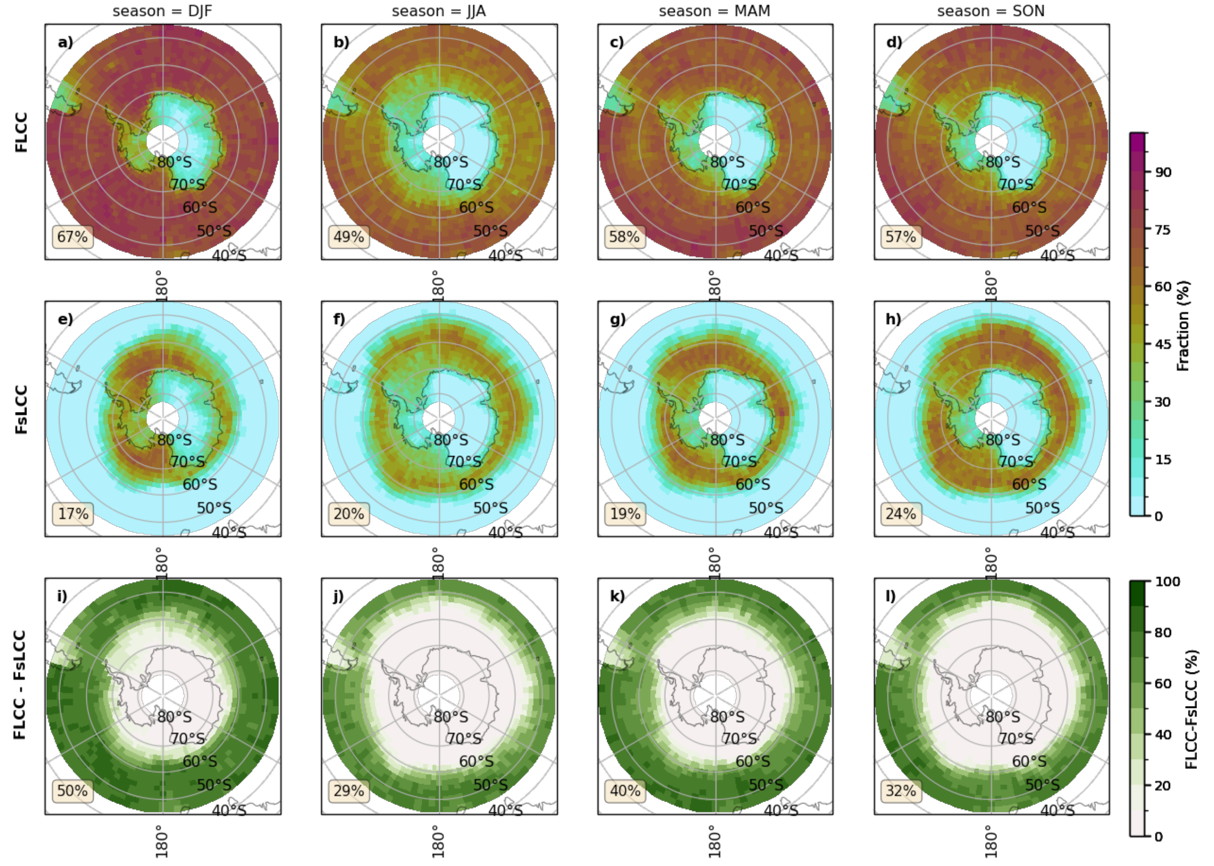
My main concern from the initial submission was in regard to the generalizations/conclusions presented. The authors have added more precise language to better limit the scope of their conclusions based on their results.

I am satisfied with the authors’ responses to all my minor suggestions and most of my main suggestions. Below are follow-up comments on one of my main suggestions. I have left the original suggestion and authors’ response for completeness, and indicated my follow-up comment with ***.

(5) Lines 451-453: “*In contrast, McIlhattan et al. (2017) showed that the CESM-LE underestimates the fLCC by ∼ 17% and overestimates the fsnow by ∼ 57% in the Arctic. However, since we utilize a different metric (sLCC instead of LCC), there is no reason to expect the model biases to be identical.*” From my understanding of the two metrics, fLCC and fsLCC wouldn’t be likely to produce biases with the opposite sign in the high-latitudes (outside of perhaps summer) if the models contained similar cloud systems. It seems more likely that the difference between this study’s biases and those found by and McIlhattan et al. (2017) arise from differences in the models’ cloud systems. The same metrics from McIlhattan et al. (2017) were used to evaluate LCCs in CESM2 (a CMIP6 generation model; McIlhattan, E. A., et al. (2020). Arctic clouds and precipitation in the Community Earth System Model version 2. Journal of Geophysical Research: Atmospheres, 125, e2020JD032521. https://doi.org/10.1029/2020JD032521), and those findings appear to be more similar to those presented here. I strongly encourage the authors include the issue of model generation when comparing their results to earlier studies.

The following figures show the fLCC (a-d) as defined by McIllhattan et al. (2017) and the fsLCC (e-h) as defined in our study. The lower panel in the figures (i-l) show the difference between the two metrics. fLCC and fsLCC are especially different over the ocean.


season = DJF
season = JJA
season = MAM
season = SON
FLCC
FsLCC
FLCC - FsLCC
a)
b)
c)
d)
e)
f)
g)
h)
i)
j)
k)
l)
40°N
50°N
60°N
70°N
80°N
0°
36%
62%
42%
53%
22%
3%
16%
18%
15%
59%
27%
36%
Fraction (%)
0
15
30
45
60
75
90
FLCC-FsLCC (%)
0
20
40
60
80
100


season = DJF
season = JJA
season = MAM
season = SON
FLCC
FsLCC
FLCC - FsLCC
a)
b)
c)
d)
e)
f)
g)
h)
i)
j)
k)
l)
80°S
70°S
60°S
50°S
40°S
180°
67%
49%
58%
57%
17%
20%
19%
24%
50%
29%
40%
32%
Fraction (%)
0
15
30
45
60
75
90
FLCC-FsLCC (%)
0
20
40
60
80
100

We included the additional reference given by the reviewer and added some sentences to the discussion on model versions. Specifically, we added the following sentences in **Lines 492-509**:

"*However, since we utilize a different metric (sLCC instead of LCC), there is no reason to expect the model biases to be identical. Nevertheless, in a further study by McIlhattan et al. (2020) it was shown that the LCC frequency in the newer CESM version is more aligned with the satellite observations except for in the summer months, where it overestimates the LCC frequency.*

Similarly, we generally observe that ERA5 and the CMIP6 mean overestimate the fsLCC to various extents during all seasons. These overestimations are likely linked to the microphysical parametrizations of cloud processes that govern cloud phase. This finding aligns with McIlhattan et al. (2020), indicating that while newer model versions have advanced in representing LCC frequencies more accurately, they can still overestimate cloud occurrences, particularly in specific seasons. Precipitation in the new CESM version is more frequent but lighter overall compared to the previous version, which is similar to our findings indicating that the models' sLCCs produce continuous snowfall, analogous to the "perpetual drizzle" problem (Mülmenstädt et al., 2020; Lavers et al., 2021).

Furthermore, while McIlhattan et al. (2017, 2020) considered a single ESM, our study considers an ensemble of CMIP6 models. Nevertheless, the insights from McIlhattan et al. (2017, 2020) provide relevant context for our finding that ERA5 and the CMIP6 model mean produce sLCCs more frequently than observed in the NH and SH mid-to-high latitudes, especially over the sea ice and land depending on the season (Figs. C1 and C2). In these regions, not only is the frequency of occurrence of sLCCs too high, but the sLCCs are too efficient at producing snowfall in ERA5 and CMIP6 models (Figs. 5 and 6). The latter finding is consistent with the findings of McIlhattan et al. (2017) that LCCs produce snow too frequently in the CESM-LE model."

****I appreciate these clarifications and additions. The two plots are very helpful in understanding the regional differences in the two metrics. I had not fully thought about the implications of the authors' requirement of FsLCCs having temperatures below freezing at the surface: it essentially removes from the analysis all clouds from most open ocean grid-boxes. This effect would be particularly strong in summer but important in other seasons as well.*

Is it correct to reason that slightly colder model/reanalysis surface temperatures relative to CloudSat could result in FsLCC overestimations relative to observations – independent from differences in cloud representation? For example, if a given region had liquid cloud properties/frequency that were identical in ERA5 and CloudSat, but ERA5 had surface temperatures slightly below 0C and CloudSat slightly above, then would ERA5 have higher FsLCC for that region? If this is a correct interpretation, I would suggest modifying the language of the paper to avoid making strong conclusions about sLCCs over the open ocean. (e.g. the abstract notes "Specifically, we find that the ERA5 reanalysis and ten CMIP6 models consistently overestimate the frequency of sLCCs and snowfall frequencies from *sLCCs compared to CloudSat-CALIPSO satellite observations, especially over open ocean regions.")*

Thank you for raising this point, that differences in sLCC between ERA5 and CMIP6 models and CloudSat could also come from temperature differences. While our additional analyses (presented below, Figs. Rev. 1, Rev. 2, and Rev. 3) shows that this is generally not the case, a notable exception is the CMIP6 ensemble mean in the central Arctic in summer (Fig. 1), we thank the reviewer for requesting the additional analyses that made this evident. We have now removed "especially over open ocean regions" from the abstract, and have added a sentence in the results and a short discussion about this in the manuscript.

Lines 305 - 307: "*It is reasonable to assume that the temperature in the ECMWF-AUX product used in CloudSat-CALIPSO is quite similar to the ERA5 daily mean* ***as indicated by the seasonal mean 0°C isotherm in Figs. 1 and 2****.* ***However, ERA5 shows a slight variation in the 0°C isotherm line over Central Europe during DJF compared to ECMWF-AUX (Fig. 1 e). Furthermore, a*** *comparison of the 2m temperature between ECMWF-AUX and ERA5 shows a latitudinal average difference of 0.24K ± 0.22K (Fig. D1).*"

Lines 540 - 548: "***For the most part, the 0°C isotherm shown in Figs. 1 and 2 support our argument that the primary issue with the ERA5 and CMIP6 datasets lies not with the simulated temperature itself, but with the representation of cloud properties and microphysics. This distinction highlights that the observed deviations in fsLCC and fsnow are driven more by inaccuracies in cloud simulation than by temperature discrepancies. Exceptions occur over Central Europe during DJF between ECMWF-AUX, ERA5 and the CMIP6 ensemble mean (Figs. 1 a, e, i). However, more notably is the difference in the CMIP6 ensemble mean over the central Arctic during summer (Fig. 1 k), where simulated temperatures appear to be too cold. In this specific case, the cloud bias could stem from a temperature bias, suggesting a potential link between temperature inaccuracies and cloud simulation in the CMIP6 ensemble mean for this region and season.***"

As for the text highlighted by the reviewer (*"Specifically, we find that the ERA5 reanalysis and ten CMIP6 models consistently overestimate the frequency of sLCCs and snowfall frequencies from sLCCs compared to CloudSat-CALIPSO satellite observations, especially over open ocean regions."*), this simply states the facts, without attributing the bias to any particular cause such as cloud microphysics or temperature. In the previous review we introduced Figure D1, which displays the zonal mean 2m temperature of ECMWF-AUX (solid black line) and ERA5 (dashed orange line) for the mid-to-high latitudes. We noted that the temperature errors are less likely than those associated with microphysical variables. Further strengthening this argument, we observe minimal differences in grid cell 2m temperature correlation between ECMWF-AUX and ERA5 (shown in Figure Rev. 1), which both come from the same weather forecast centre. The correlation coefficients for these datasets are high (R^2=0.81°C for the Northern Hemisphere and 0.97°C for the Southern Hemisphere) when looking at the 0°C±2°C grid cell 2m temperature range. This suggests that discrepancies are likely due to cloud occurrence and not the applied temperature threshold. Moreover, the seasonal averages of fLCC for both hemispheres (Figs. Rev. 2 and Rev. 3) show that ERA5 and CMIP6 ensemble mean tend to overestimate the liquid containing cloud occurrence, further supporting that the 2m temperature values are not the primary issue, as no temperature threshold was applied to obtain fLCC.

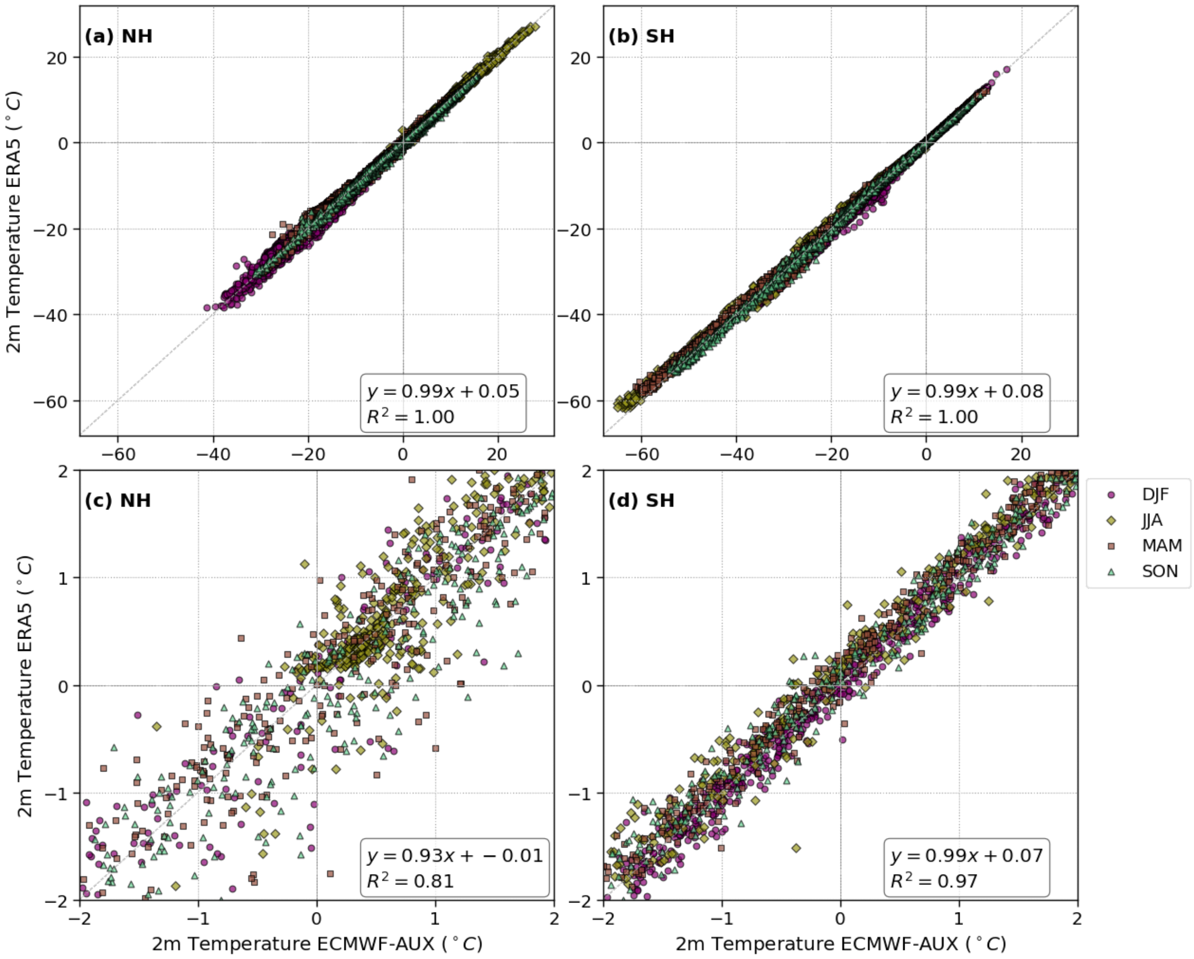


Figure Rev. 1: Correlation between grid cell 2m temperature from ECMWF-AUX and ERA5 for the Northern Hemisphere (NH, panels a and c) and Southern Hemisphere (SH, panels b and d), with colours indicating individual seasons. Panels (a) and (b) display correlation across the full temperature range for all grid cells, whereas panels (c) and (d) restrict the view to temperatures close to 0°C, excluding grid cells with temperatures above 2°C or below -2°C. The lower-right box in each plot shows the linear regression between ECMWF-AUX and ERA5 grid cell 2m temperatures and provides the correlation coefficient (R^2) for 2007-2010, independent of the season, highlighting the strength of alignment between the two datasets.

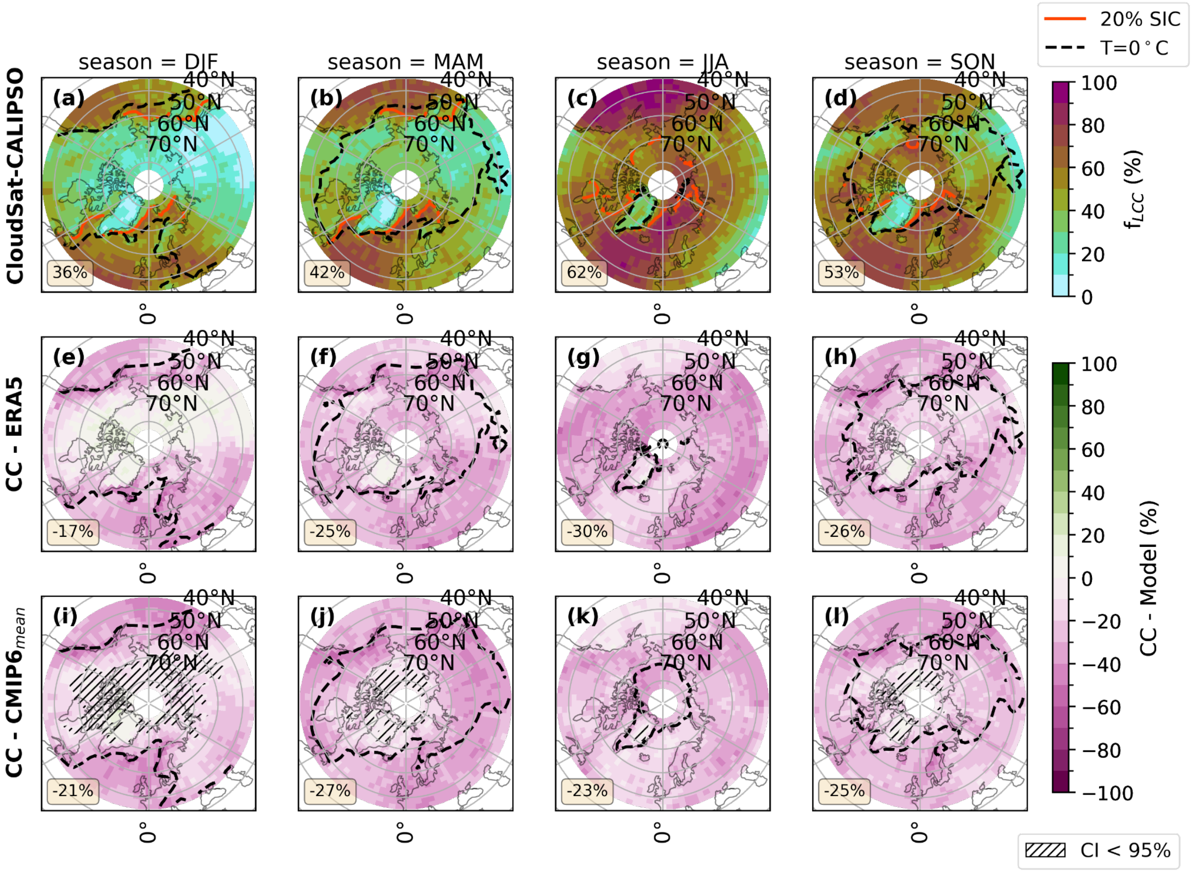


Figure Rev. 2: Seasonal averages of fLCC in the NH mid-to-high latitudes between 2007 and 2010. Combined CloudSat and CALIPSO observations are shown in the first row (a-d). The last two rows are the difference plot. They are CloudSat-CALIPSO (CC) observations minus ERA5 (e-h) or CMIP6 model mean (i-l) where valid data occurs, with green (pink) values showing underestimation (overestimation) in ERA5 and the CMIP6 model mean concerning the satellite observations. Areas where the difference between CloudSat-CALIPSO and CMIP6 model mean is not significant (< 95%) are marked with hatches. The area-weighted averages for the study area where CloudSat-CALIPSO has observations are displayed in the lower-left corner of each map. The black dashed line represents the seasonal mean 2m temperature 0°C isotherm for each individual product. The red line (in a-d) shows the average sea ice edge of 20% sea ice concentration (SIC) between 2007 and 2010, for the given season.

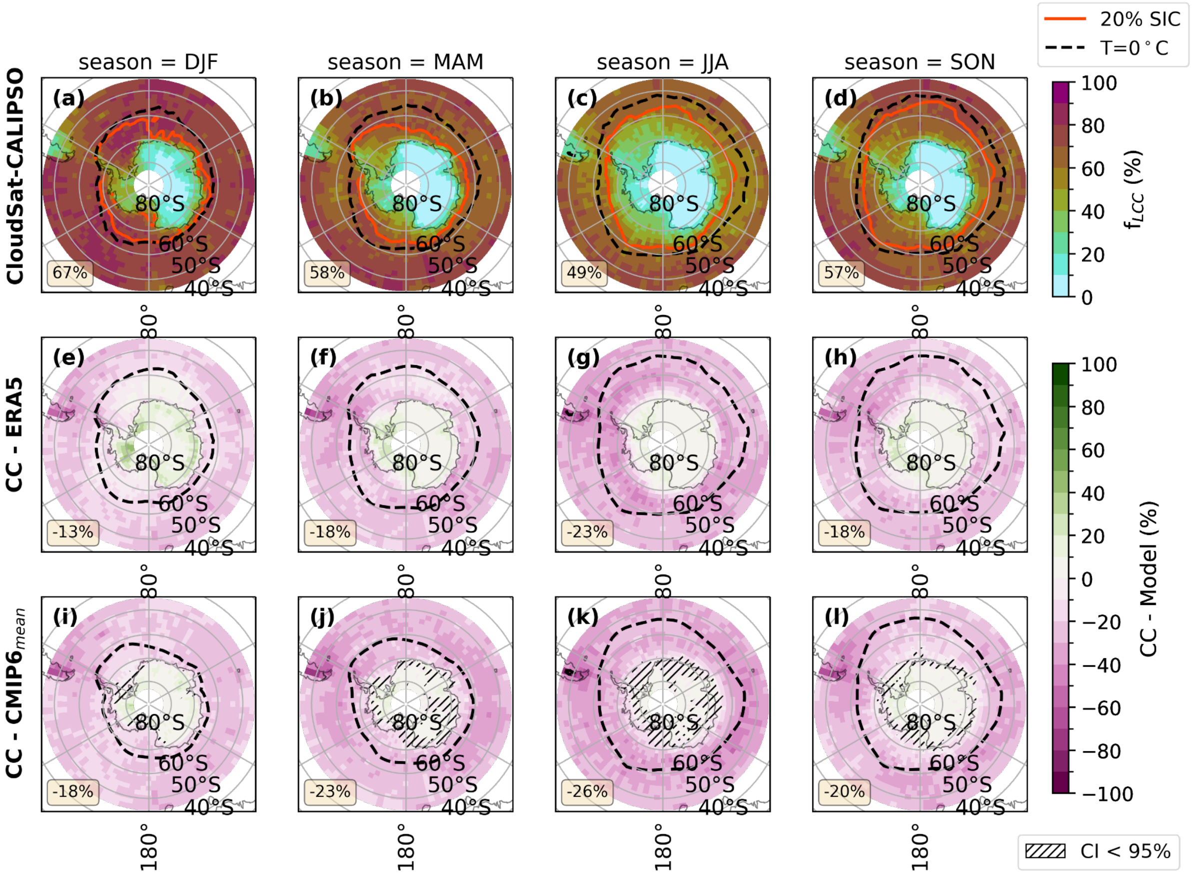


Figure Rev. 3: Seasonal averages of fLCC in the SH mid-to-high latitudes. Layout and differences are identical to Figure Rev. 2.

It would also be very helpful to include mean spatial areas where T2m < 0C for the subplots in Figures 1&2 and/or the charts in Figure 3. Ideally these areas would be similar across the obs/reanalysis/model datasets. If the areas are much larger in the reanalysis and models, that would be an important contributing factor to include in the discussion of the area weighted mean FsLCC values in Figures 1&2&3.

In light of this and the previous comment, we have now revised Figures 1, 2, 5, and 6 in the manuscript, and added the seasonal mean 2m temperature 0°C isotherm on these Figures to show the areas with 2m temperatures below 0°C.

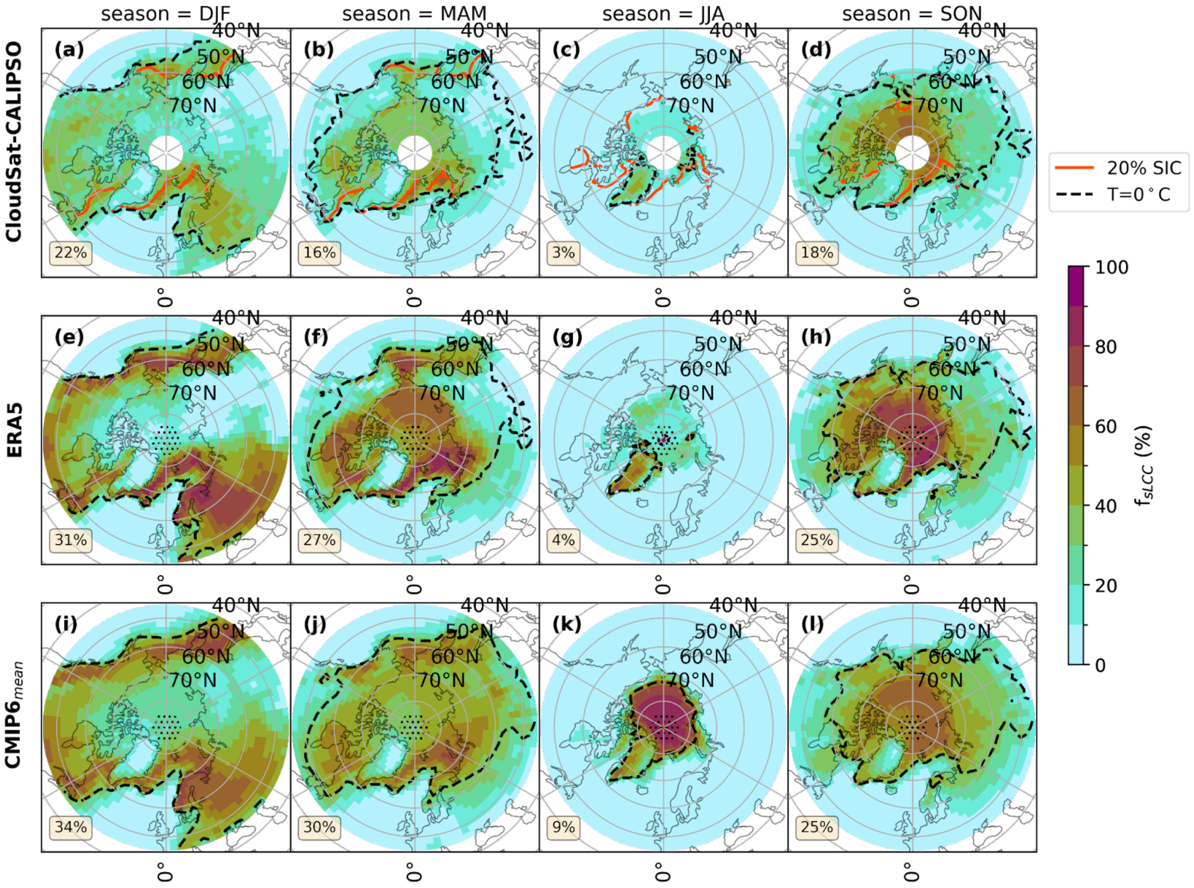


Figure 1. The NH mid-to-high latitude seasonal averages of fsLCC. The first row (a-d) for CloudSat-CALIPSO, the second row (e-h) displays ERA5 data, and the third row (i-l) shows the CMIP6 model mean. Each map includes an area-weighted average for the study area (lower left corner). These averages are calculated for areas where CloudSat-CALIPSO have valid observations (between 45°N − 82°N) and exclude the dotted area (in e-l). ***The black dashed line represents the seasonal mean 2m temperature 0°C isotherm for each individual product.*** The red line (in a-d) shows the average sea ice edge of 20% sea ice concentration (SIC) between 2007 and 2010, for the given season.

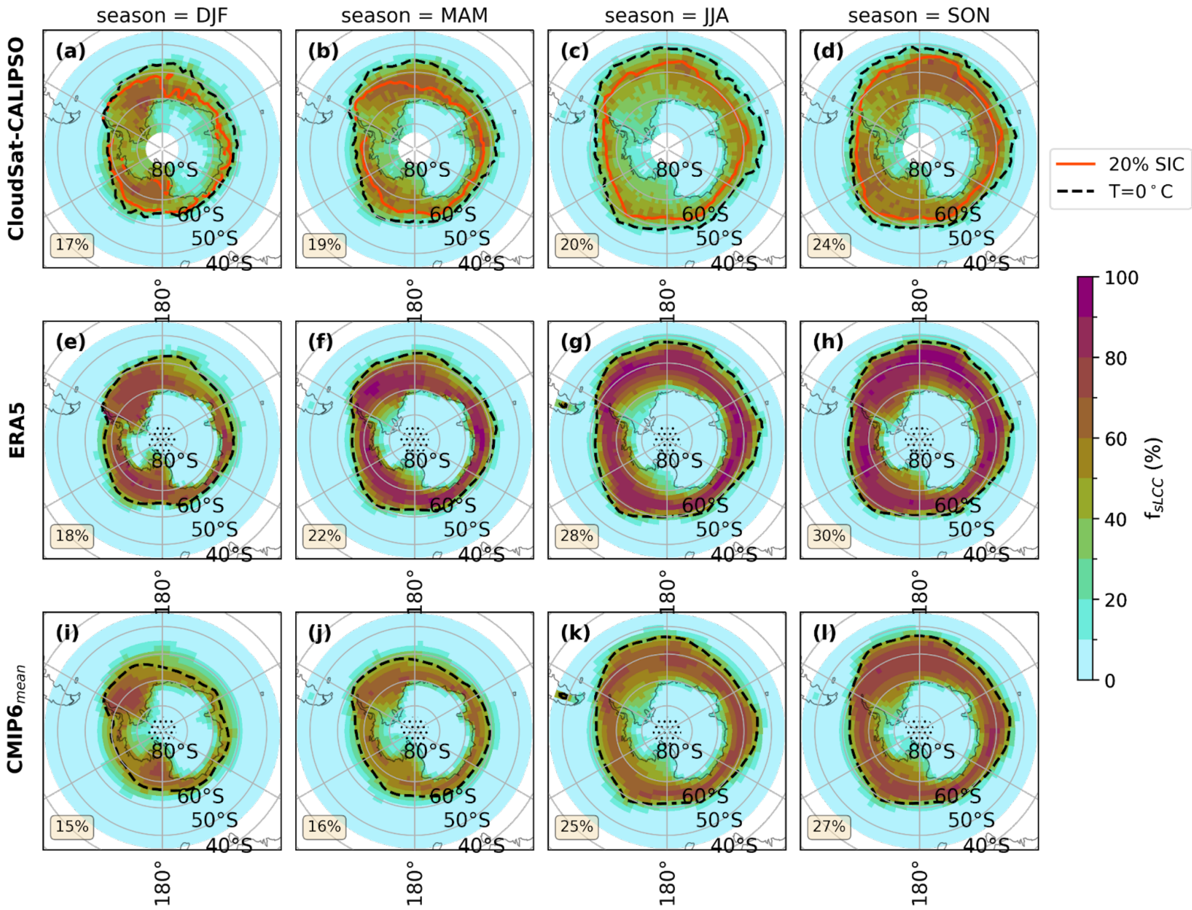


Figure 2. The SH mid-to-high latitude seasonal averages of fsLCC. The first row (a-d) for CloudSat-CALIPSO, the second row (e-h) displays ERA5 data, and the third row (i-l) shows the CMIP6 model mean. Each map includes an area-weighted average for the study area (lower left corner). These averages are calculated for areas where CloudSat-CALIPSO have valid observations (between 45°S − 82°S) and exclude the dotted area (in e-l). ***The black dashed line represents the seasonal mean 2m temperature 0°C isotherm for each individual product.*** The red line (in a-d) shows the average sea ice edge of 20% sea ice concentration (SIC) between 2007 and 2010, for the given season.

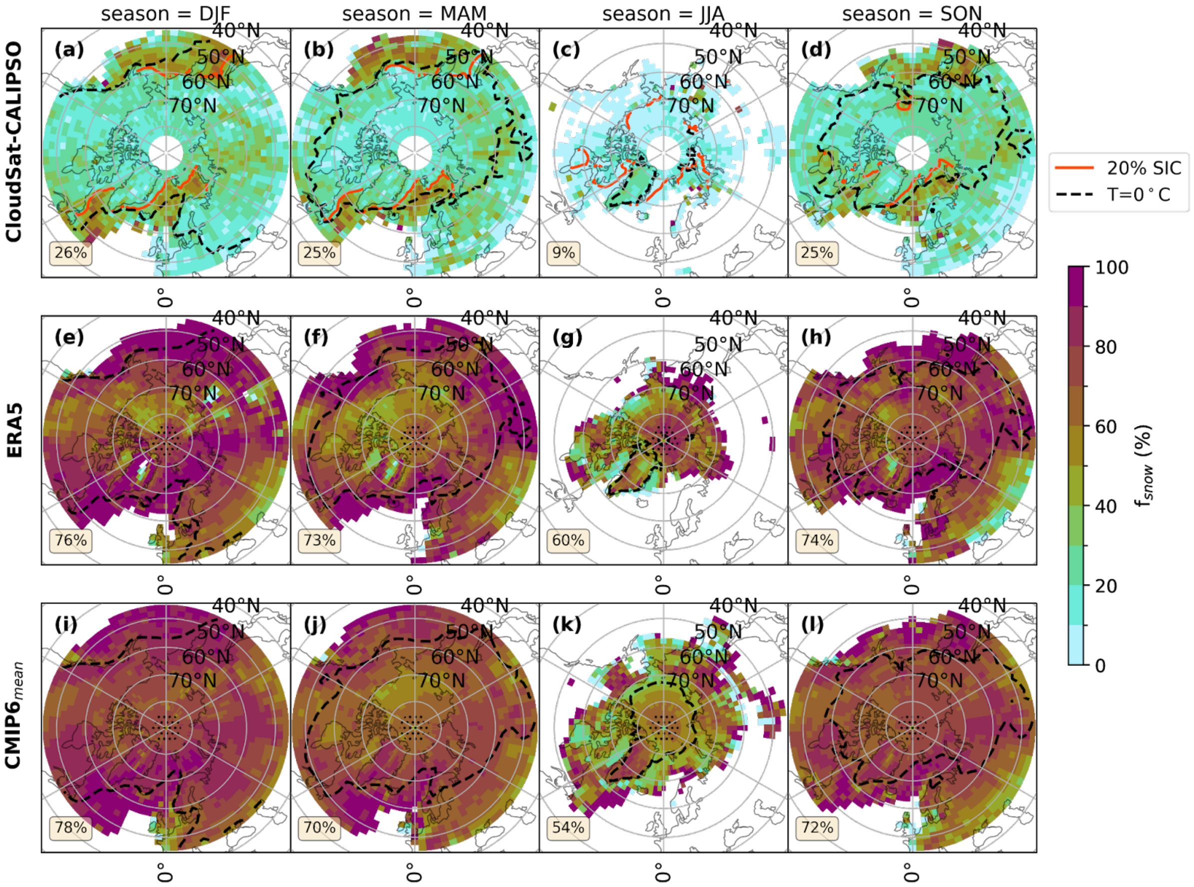


Figure 5. The figure presents the seasonal averages of fsnow in the NH mid-to-high latitudes. The layout and area-weighted averages are calculated the same as those shown in Figure 1.

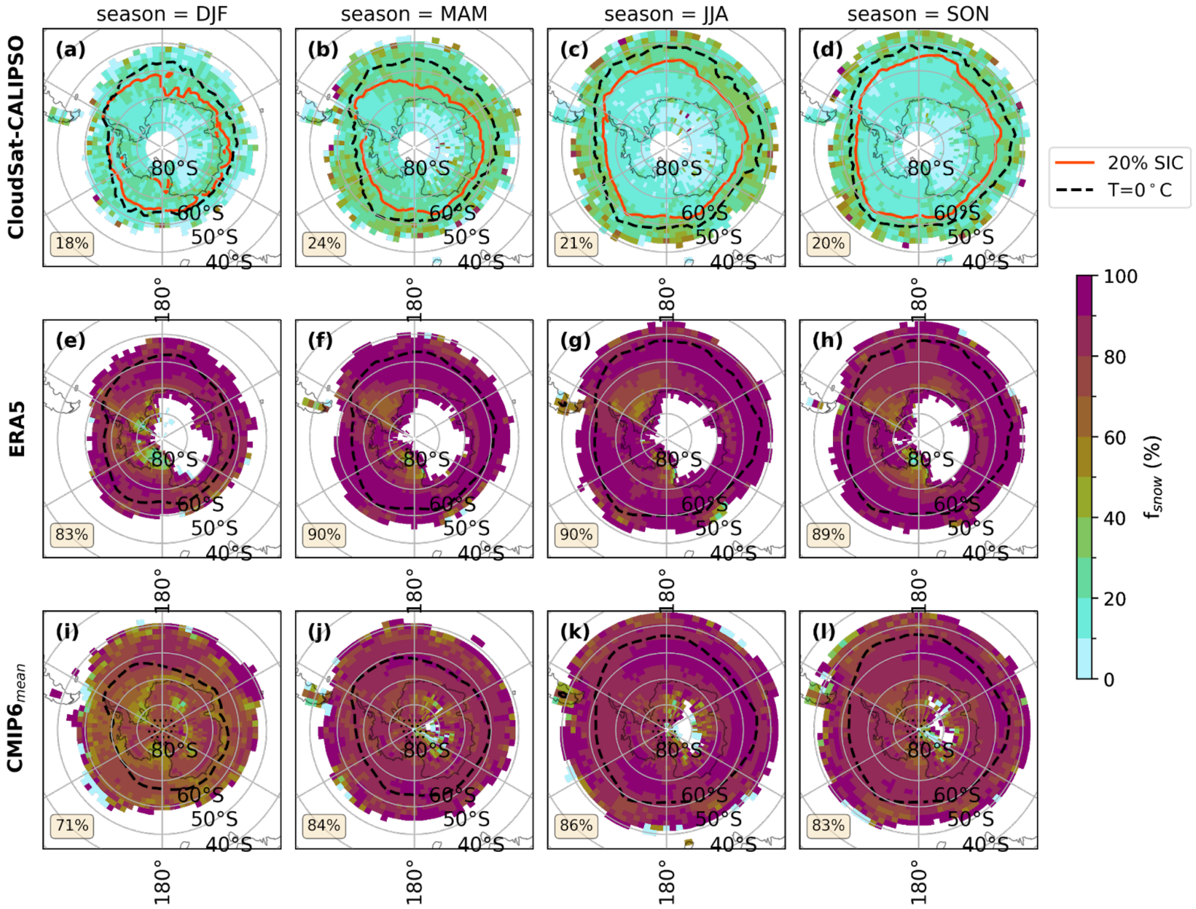


Figure 6. The figure presents the seasonal averages of the fsnow in the SH study region. The layout and area-weighted averages are calculated the same as those shown in Figure 1.