Replies to comments of Review #1

We would like to thank the reviewer for their very positive feedback and appreciate their time for reading and commenting on the submitted manuscript. The replies to the referees' comments are given in this document and structured as follows:

Referee's comments in italic – line numbers according to initially submitted manuscript
Authors' responses in roman – line numbers according to adjusted manuscript.
Citations from the initial and the adjusted manuscript are given in bold.

<u>Please note that we also implemented some minor text revisions in addition to the changes made in response to the reviewer's comments. For these changes, we refer to the tracked changes file.</u>

The manuscript presents a thorough and well-executed study of biases in satellite-derived estimates on cloud radiative effect over Arctic sea ice. The extensive research-aircraft campaign provides a unique and valuable dataset that enables a robust quantitative assessment of these biases. The analyses are comprehensive and make full use of the available satellite and aircraft observations. The care with which the analyses have been conducted is evident throughout the manuscript, giving the reader confidence in the results and their interpretation.

I have a single major comment and several minor suggestions for improvement. The issues raised should be straightforward to address. Overall, I consider this to be a high-quality contribution, and I believe the manuscript is well suited for publication in ACP following minor revisions.

<u>*Major comment*</u>

The authors should make it clear how the findings of this extensive study advance scientific understanding in the field. At present, the manuscript jumps directly from the final Results subsection (Section 4.3) to the Conclusions. The Conclusions section contains no references to earlier studies, making it difficult to assess the novelty and significance of the reported findings. The manuscript would be strengthened by either adding a dedicated Discussion section that highlights the key advances beyond existing knowledge or by incorporating a concise discussion of these advances, with appropriate references, within the Conclusions section. This would help readers better understand the study's contribution to the field and its scientific impact.

We agree that an extended discussion, which also puts our findings into context of existing literature, will strengthen the conclusions of the manuscript. Although the Conclusions section already contains some discussion points, they might be rather hidden and, thus, not get clear to the reader. Therefore, we revised the Conclusions section, which is now entitled "**Discussion and Conclusions**", by stating our important findings and directly providing a dedicated discussion with each finding. The updated version reads as follows:

"The major finding of this paper is that the total CRE simulated with the satellite observations suggests a cooling effect, although the clouds almost exclusively warm the surface according to the collocated aircraft measurements. The median total CRE switched from –23 W m^{-2} to +20 W m^{-2} during the first (May/June) and from –10 W m^{-2} to +5 W m^{-2} during the second (July/August) phase of ARCSIX. This discrepancy is related to a systematic low bias in the surface albedo used for the satellite retrieval, which amplifies the solar cooling effect of an identical cloud, such that optically thinner clouds are sufficient to turn the total

CRE into a cooling effect (Shupe and Intrieri, 2004). At a first glance, the satellite-based negative CRE might not seem contrary to prior knowledge, because a cloud cooling effect during summer is, in fact, a typical feature of the seasonal cycle of total CRE in the Arctic (Boeke and Taylor, 2016; Kay and L'Ecuyer, 2013). However, the values reported in satellite- or model-based studies usually account for an Arctic-wide average comprising both sea ice and open ocean surfaces, whereas the latter surface type was not sampled during ARCSIX. Consequently, the aircraft measurements appear rather consistent with observations over permanently snow-covered surfaces, which found that the summertime warming effect of clouds is either small (Intrieri et al, 2002) or does not exist at all (Miller et al., 2015). Satellite-based CRE calculations dedicated to sea ice were not evaluated before.

The identified low bias in the sea ice albedo resulted from an overrated melting assumption in the satellite retrieval, particularly during the first phase of ARCSIX. Representing the sea ice albedo by spatially invariant, season-dependent statistical values, similar to the approach applied in Kay and L'Ecuyer (2013), to parameterize sea ice melt does not account for regionally distinct effects or specific synoptic conditions that differ interannually. Although an underestimation of sea ice albedo is not unique to this study and was similarly noticed in CERES observations and different climate models (Huang et al., 2022; Kim and Taylor, 2026), the present study quantifies the radiative impact of the surface albedo uncertainty, which has implications on the accuracy of the surface albedo feedback and, thus, the strength of Arctic amplification. Additionally, satellite retrievals of cloud properties suffer from the surface albedo misrepresentation. These implications demand an improvement of the satellite-derived albedo over snow-covered areas to accurately represent spatial and temporal surface albedo variability. Jiao et al. (2019) developed a dedicated snow kernel to complement kernel-based retrievals of spectral surface albedo. Although their approach is not yet operational, it yields improved albedo values over snow surfaces as compared to current retrieval methods (Ding et al., 2019). In this sense, the ARCSIX field observations have immense potential for improving the representation of snow metamorphism, sea ice melt, or melt pond formation for more accurate estimates and predictions of sea ice properties.

The systematic surface albedo misrepresentation revealed in this study does not generally question previous literature results of Arctic-wide CRE estimates, because the cloud cooling effect over open ocean still dominates and the strength of surface-albedo-induced CRE biases may differ in various regions. However, our findings probably imply that the real cooling effect of clouds is lower than expected and that the transition from warming to cooling effect happens later in the season.

The cloud height from the MODIS (Collection 6.1) and VIIRS retrievals exposes a drastic overestimation on the order of 1 km compared to the ARCSIX aircraft observations, which is more severe for the VIIRS than for the MODIS retrieval algorithm. A similar overestimation in satellite-derived height of low marine clouds was found by Holz et al. (2008) for data from MODIS Collection 5 and attributed to algorithm inaccuracies in the case of temperature inversions. Although the algorithm has improved between the MODIS collections, an accurate ancillary thermodynamic profile is still crucial for a correct CTH retrieval, especially in the case of temperature inversions. Several reanalyses have shown a high-biased temperature in the Arctic boundary layer (Graham et al., 2019). Therefore, testing the performance of cloud height retrievals for real measured temperature profiles in future work will be beneficial to disentangle the impact of retrieval algorithm and temperature input on the cloud height uncertainty.

The thermodynamic analysis data from GEOS-FP used in this study revealed a similar warm bias below 500 m, especially during the first phase of ARCSIX, leading to an underestimated strength of the capping inversion. While the effects of cloud height and temperature bias partly compensate for CBT and, thus, $F^{\downarrow}_{\mathrm{TIR}}$, the cloud height bias alone decreases $F^{\downarrow}_{\mathrm{TIR}}$ by 5–9 W m^{-2}. This shows the need for consistent temperature profiles for cloud height retrieval and irradiance calculations. However, the impact of cloud height and temperature bias on the CRE is low compared to the surface albedo bias. The solar radiation is weakly affected by both cloud height and thermodynamic profile differences.

Correlations between satellite- and aircraft-derived COT were weak, which is probably related to the spatial resolution difference as well as measurement and retrieval uncertainties. Although the low-biased surface albedo used in the satellite retrievals is expected to yield a too high COT, individual over- and underestimations occurred frequently during both campaign phases. <u>Statistically, the satellite-derived COT overestimated the aircraft-measured COT during the first phase, while an underestimation was observed during the second phase. However, only for the first phase increased this misestimation the TIR CRE. During the second phase of ARCSIX, the TIR CRE was largely unaffected due to the dominance of optically thick, opaque clouds, where emissivity is not sensitive to COT (Shupe and Intrieri, 2004).</u> Due to the weak sensitivities in the TIR and compensating effects between solar and TIR range, the impact of the COT discrepancy on the total CRE is small, amounting to 2 W m^{-2} and –11 W m^{-2} during the first and second campaign phase, respectively.

In contrast to the results reported by Chen et al. (2021), clouds undetected by the MODIS and VIIRS retrievals were rarely encountered in the cases analyzed here, mainly because the flight plans were based on clouds clearly visible in satellite images. Nevertheless, the cloud detection problem must not be ignored as a potential source of uncertainty for satellite-based SRB estimates and should be addressed in upcoming research.”

<u>Added literature:</u>

- Boeke and Taylor (2016): https://doi.org/10.1002/2016JD025099
- Kay and L’Ecuyer (2013): https://doi.org/10.1002/jgrd.50489
- Intrieri et al. (2002): https://doi.org/10.1029/2000JC000439
- Jiao et al. (2019): https://doi.org/10.1016/j.rse.2018.11.001
- Ding et al. (2019): https://doi.org/10.3390/rs11131611
- Holz et al. (2008): https://doi.org/10.1029/2008JD009837
- Graham et al. (2019): https://doi.org/10.1029/2019GL082781

Minor comments

Line 2: Clouds have a significant impact

We considered this suggestion (line 2).

Lines 10-11: The phrase "between May and August" is ambiguous.

We see the ambiguity regarding the inclusion of May and August in the denoted period. To be more precise, we specified the formulation to "**from late May to mid-August**" (line 10).

Line 18: specify the period, as the number is very sensitive to it.

We agree and the studies cited here indeed investigated different periods. To take these differences into account, we modified the text to **"... which is accelerated by a factor of two (since ~1960) to four (since ~1980) compared to the rest of the globe ...**" (line 18).

Lines 33-34: Include wind forcing on ice dynamics as a factor controlling sea ice evolution. Also, in winter the origin of coldest air masses is often in northern Siberia and Canada.

To account for the second part of your second comment, we skipped the airmass origin. Furthermore, we considered the reviewer's addition by revising the sentence as follows: "**Besides dynamical forcing of wind and ocean currents, the sea ice evolution is largely controlled by thermodynamic advection of warm or cold airmasses and the local surface energy budget components ...**" (lines 34–35).

Line 63: perhaps "continuous ground-based observations and temporary campaign measurements"

To avoid the double occurrence of "**observations**" within the sentence concerned, we decided to keep our original version.

Lines 100-101: Do you mean "Moreover, passive satellite imagers do not provide direct information on cloud-base properties. As a result, cloud-base height (CBH) can only be inferred ..."

We considered the reviewer's suggestion (lines 103–104).

Lines 154 and 197: Porpoise is a small marine mammal. Perhaps you mean a purposive leg and purposive sampling.

The verb "to porpoise" has the meaning of "to move like a dolphin", which resembles the shape of the leg we want to describe by this term. However, we understand that this term causes confusion. Therefore, we revised the regarding sentence to improve the explanation: "**To obtain information on the vertical structure of the cloud along the wall, an additional leg with**

alternating ascents and descents through the cloud layer, which is referred to as saw tooth leg, was added." (lines 157–158).

Subsequently, we replaced all occurrences of "**porpoise**" by the probably more intuitive term "**saw tooth**".

Line 191: It is assumed that

We removed the comma (line 194).

Lines 270-276: Do multiple reflections between the cloud base and snow/ice surface play a role here?

Multiple reflections between surface and cloud are always present and partly determine the albedo difference between cloudy and cloud-free conditions. However, the main reason for this difference is the changed distribution of direct and diffuse illumination, which affects the single scattering geometries at the surface. While we neglected these changes for the reasons mentioned in the manuscript, 1D effects of multiple reflections are accounted for by the radiative transfer simulations performed in this study. However, 3D effects due to horizontal inhomogeneities are not considered and their impact is assumed negligible on average. Regarding the thin clouds below the aircraft, the albedo was rather modified by the reflectance of the cloud itself than by multiple reflections.

Line 294: What do you exactly mean by aggregation for all cloud walls?

We mean that we processed the data as described in Sect. 2.2 for all cloud walls and collected them afterwards to build statistics. Our revised sentence reads: "**The quantities derived according to Sect. 2.2 from the aircraft and satellite observations along all cloud walls are accumulated and analyzed statistically.**" (lines 304–305).

Additionally, we replaced "**an aggregation**" by "**a collection**" in line 126.

Figure 3. Explain the meaning of the grey bars?

The meaning of these bars is already explained in the text. However, we added the following entries to the figure legend to be independent of the text:

- light gray: "**wavelength range not covered by SSFR**"
- dark gray: "**water vapor absorption bands**"

Lines 323-332: Most of the paragraph is about weak correlations but the text ends with very positive remarks, which I found confusing.

The beginning of this paragraph compares the cloud optical thickness derived from aircraft and satellite observations, which indeed correlate weakly. In contrast, the last three sentences address the success of cloud detection and COT retrieval in general, regardless of the agreement of the retrieved COT with aircraft observations. In earlier studies (Chen et al., 2021), cloud detection failures occurred frequently, especially for optically very thin clouds. However, such failures were rare in our cases, although the retrieved COT values vary largely. By revising the transition between these two parts as follows, we aimed to achieve a clearer separation between them: **"Despite the frequent mismatch between aircraft- and satellite-derived COT and unlike Chen et al. (2021), …"** (lines 342–343)

Line 367: exhibits a behaviour similar to that in the first phase

We considered the reviewer's suggestion (line 379).

Line 401: What do you mean by current assumptions for sea ice melt?

We mean the assumption of scaling the statistical sea ice albedo by a factor of 0.8 during the summer months that is currently applied to account for sea ice melt. At least for our cases, this scaling yields too low sea ice albedo values, even during summer. We tried to clarify this by revising the sentence concerned as follows: "**Omitting the albedo scaling would increase the median cloud-retrieval broadband albedo to 0.75, indicating that the current scaling assumption overrates the impact of sea ice melt even in summer, …**" (lines 412–414). Additionally, we replaced the phrase "**melt parameterization**" by "**scaling employed to parameterize sea ice melt**" two sentences earlier to strengthen the comprehension with the use of the coherent term "scaling".

Line 412-414: One more challenge is representation of the the surface scattering layer (SSL) that develops through snow metamorphism and the ice-to-snow transformation process.

Absolutely! Such processes modify, i. a., snow grain size and, thus, the scattering characteristics of the surface, which affect surface albedo. To keep concise, we just added snow metamorphism to our list of weakly represented processes: "**…, i. a., sea ice melt, melt pond formation, and snow metamorphism processes, …**" (lines 425–426).

Line 518: This assumption seems unrealistic. Were there any alternatives for such an assumption?

We agree that assuming equality of the thermodynamic profiles in cloudy and cloud-free conditions is unrealistic, but it is one of the common approaches to define the cloud radiative effect. Other, more complicated, approaches that somehow account for thermodynamic changes and yield different CRE estimates exist (see, e. g., Stapf et al., 2021). However, such approaches depend on the time that has passed since the disappearance of the cloud as the thermodynamic profiles transition smoothly. This means that there is no unique solution for cloud-free conditions. Furthermore, the removal of the cloud is hypothetic and the thermodynamic transition might depend on the moment when the cloud is removed. Given these caveats, we decided to use the approach with unchanged thermodynamic properties for CRE definition in our study.

To make that clear, we added the following sentences already at lines 525–527: "**Note that the thermodynamic profiles used for the cloud-free simulations are unchanged compared to the cloudy cases because accounting for their transition after cloud removal is challenging. This approach is common for CRE determination.**"

The actual sentence is then revised to “**Due to the unchanged temperature profile between cloudy and cloud-free conditions, the surface skin temperature and, thus, $F_{\mathrm{TIR}}^{\uparrow}$ are identical in both cases. Statistics of the latter quantity are illustrated in Figs. 9e and 9f.**” (lines 533–534)

Lines 545-551: This subsection is very short bu the strong impacts of surface albedo become very evident in the following sub-section. Consider merging or re-naming the subsections.

It is true that this subsection is short despite discussing the effects of the surface albedo difference as the most impactful driver for CRE uncertainty. However, we believe that the relationship between albedo and irradiances is systematic, straightforward to understand and does not require much interpretation. The message from Fig. 9c and the clear statements within the subsection speak for themselves. Mainly for reasons of consistency with subsection 4.2.1, which focusses on microphysical cloud properties as the other important driver, we decided to keep the structure as is.

Lines 638-641: Clarify the text.

These sentences changed while revising our conclusions section. Please find the updated version of this part as underlined text in the reply regarding the major comment (lines 693–696 in the updated manuscript).

Line 649: "... and should be addressed in upcoming research."

We considered the reviewer’s suggestion (line 701). Additionally, we replaced the earlier occurrence of “**should not**” within the same sentence by “**must not**”.

Line 672: "... albedo values assumed ... do not well represent..."

This phrase no longer occurs in the revised conclusions section.