

Response to comments

Note: All line numbers refer to the revised manuscript (with track changes hidden).

5 Reviewer 1 (RC1)

Review of Wojciechowska et al., 2025.

10 This paper tests the performance of a simplified calculation of all-sky albedo vs CERES satellite data. The simplified calculation consists of a function or look-up table (LUT) using the MODIS observed 1x1 degree daily cloud fraction (CF), droplet number (N_d) and Liquid Water Path (LWP) values as inputs. It is constructed using a kernel approach that (presumably – there needs to be more detail on this in the paper) uses several other MODIS variables as inputs along with the CF, N_d and LWP. Using a single (time and global) mean LUT leads to large errors that exhibit a spatial pattern and a dependence on
15 Estimated Inversion Strength (EIS) and Solar Zenith Angle (SZA). Attempts were made to improve the LUT. Correcting for the SZA bias using the single global mean LUT had only a small impact. However, moving to using a separate time-averaged LUTs for each 1x1 gridbox led to significant improvements leaving only small errors. This suggests that regional information (in addition to the CF, N_d and LWP values) is needed for an accurate estimate of albedo.

20 The paper describes a potentially very useful simplified way to calculate albedo quickly and easily based on only 3 cloud variables. I recommend its publication after the changes below are made – mostly clarifications of the methods and some extra description.

25 General and line-by-line comments

It is interesting that there is no need for separate seasonal/time-varying LUTs – one time-mean LUT for the whole data period for each location seems sufficient to get low errors. It might be worth commenting on this a little more.

30 Reply: We have introduced two additional modifications of the method related to the temporal aggregation used in the kernel construction: modifications IX and X. In the original method, the kernel is constructed using CF, N_d , and LWP values from the full annual time series. In modification IX, the kernel is instead constructed from monthly-averaged CF, N_d , and LWP, while in modification X it is constructed from daily-averaged values of CF, N_d , and LWP. When evaluating the difference between
35 estimated and observed albedo, the reconstructed albedo is compared against observations corresponding to the same month (modification IX) or the same day (modification X), respectively.

Both temporal aggregations lead to an improvement in the method (Table 1 in the manuscript and Table R1 below). The ratio of correct albedo reconstructions increased from 40.9% for no temporal aggregation (original method without modifications) to 57.7% when the kernel is constructed from the
40 monthly-averaged CF, N_d , and LWP and 60.6% when the kernel is constructed from the daily-averaged CF, N_d , and LWP.

The revised text in the manuscript now reads as follows:

45 Lines 232–237: “In all of the above modifications, the kernel was constructed using CF, N_d , and LWP values from the full annual time series. The last two modifications were related to the temporal aggregation used in the kernel construction. In modification IX, the kernel is instead constructed from monthly-averaged CF, N_d , and LWP, while in modification X it is constructed from daily-averaged values of CF, N_d , and LWP. When evaluating the difference between estimated and observed albedo, the reconstructed albedo is compared against observations corresponding to the same month (modification IX) or the same day (modification X), respectively.”

50 Lines 267–273: “When applying modifications IX and X, an improvement in the method is observed, with the ratio of correct albedo reconstructions increasing from 40.9% for no temporal aggregation (original method without modifications) to 57.7% when the kernel is constructed from the monthly-

55 averaged CF, N_d , and LWP and to 60.6% when it is constructed from the daily-averaged CF, N_d , and LWP (Tab. 1). However, it should be noted that the daily-averaged version (modification X) may also be subject to bin undersampling, as the number of datapoints available on individual days is considerably smaller than over longer averaging periods. Therefore, the daily results should be interpreted with similar caution as the finest spatial resolution case.”

This response and Table R1 are also included in response to a similar comment from the other reviewer.

60 Table R1. Ratio of underestimated ($\Delta\alpha < -0.02$) and overestimated ($\Delta\alpha > 0.02$) cases of albedo for methodological modifications no. IX and X. Values in brackets indicate the difference with respect to the original method.

Methodological modification	Ratio of cases (%) with:			
	$\Delta\alpha < -0.02$	$\Delta\alpha > 0.02$	$ \Delta\alpha > 0.02$	$ \Delta\alpha \leq 0.02$
Original method	27.7	31.4	59.1	40.9
IX: monthly-averaged	20.0 (-7.7)	22.3 (-9.1)	42.3 (-16.8)	57.7 (+16.8)
X: daily-averaged	19.2 (-8.5)	20.2 (-11.2)	39.4 (-19.7)	60.6 (+19.7)

65 Is there an advantage to using the LUTs vs using the kernel method directly? Especially if there is a need to have a separate LUT for each grid-box requiring a fairly large array to be stored? If we knew what is required by the kernel method then this might be clearer to the reader – presumably it requires lots of extra variables (see comments later)?

70 Reply: The kernel we use is a 3D joint histogram constructed from global observational data, where mean albedo is calculated for discrete bins of CF, LWP, and N_d . Once constructed, this kernel functions as a reference distribution that maps combinations of these three cloud properties to expected albedo values, without requiring additional variables beyond CF, LWP, and N_d . This differs from radiative kernels such as that in Duran et al. (2025), which diagnose the TOA shortwave radiative response rather than albedo. Our approach is purely diagnostic and conceptually closer to the cloud-feedback kernels of Zelinka et al. (2012), as it is derived from observational relationships without requiring model perturbations or radiative transfer calculations.

75 The advantage of using regionally defined kernels is that the CF–LWP– N_d –albedo relationship varies geographically, so a single global kernel would not adequately represent this spatial variability.

We have clarified the kernel methodology in the manuscript:

80 Lines 111–118: “The main method of the study is a 3D joint histogram (kernel) constructed from global observational data, where mean albedo is calculated for discrete bins of CF, LWP, and N_d . Once constructed, this kernel functions as a reference distribution that maps combinations of these three cloud properties to expected albedo values, without requiring additional variables beyond CF, LWP, and N_d . This differs from radiative kernels such as that in Duran et al. (2025), which diagnose the TOA shortwave radiative response rather than albedo. Our approach is purely diagnostic and conceptually closer to the cloud-feedback kernels of Zelinka et al. (2012), as it is derived from observational relationships without requiring model perturbations or radiative transfer calculations. The kernel was
85 constructed as follows.”

Line 94 – How much does the filtering by ice cloud fraction restrict the altitude of the clouds studied? It would be useful to show this somewhere. E.g., are you just looking at low-altitude clouds after the filtering?

90 Reply: Filtering by ice cloud fraction indeed restricts the analysis primarily to low-altitude clouds (Figure R1). We have added the following sentence to the manuscript:

Lines 108–109: “, which restricted the analysis primarily to low-altitude clouds (with cloud top pressure usually above ~680 hPa; not shown).”

95 As the analysis is focused on liquid clouds, this filtering is consistent with the intended cloud regime and does not affect the scope of the study.

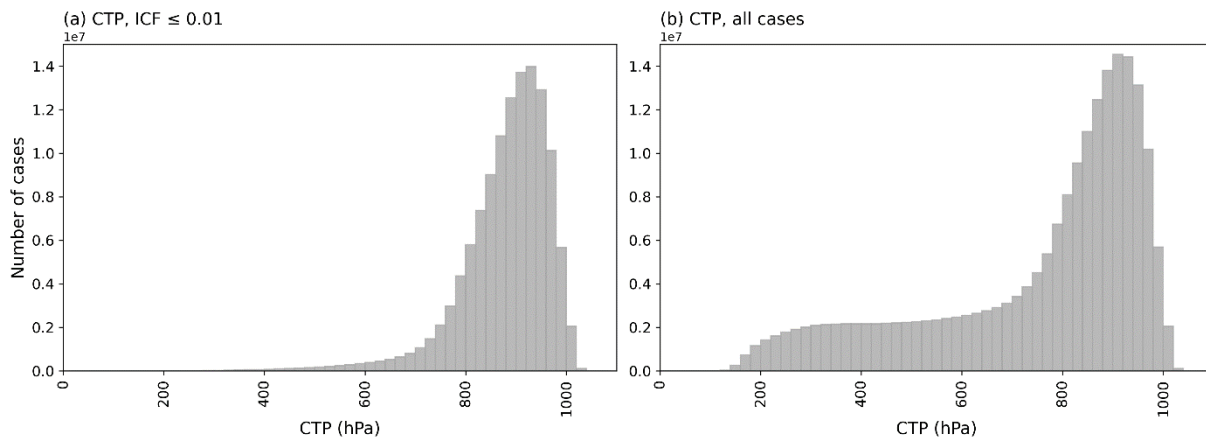


Fig. R1. Distribution of cloud top pressure (CTP) in the 2003–2021 MODIS (Terra and Aqua) dataset, with (a) and without (b) filtering by ice cloud fraction.

100

Line 96 – it would be useful to reiterate that this is 1x1 degree data. E.g., “The daily gridded 1x1 degree data...”.

Reply: The suggested information has been added.

105

Line 98 – “For each bin, the average albedo (α_{avg}) was then calculated as a multi-year mean value of all pixels across the globe that fall into the same bin of CF, LWP, and N_d .”

- It’s not clear from the methods section how you calculate the albedo of the pixel using CF, LWP and N_d . Presumably, it is as mentioned on line 62 (“Using a joint histogram/kernel approach from Gryspeerdt et al. (2019),”)? But this should be described in the methods section too. Some details on how the method works should be provided too.

110

Reply: Thank you for this comment. We have clarified the method description and removed the citation from line 72 (“approach from Gryspeerdt et al. (2019),” as this was misleading. We now explicitly describe how the kernel is constructed and how it functions to calculate albedo based solely on CF, LWP, and N_d .

115

As described in our response to an earlier comment (lines 67–86 of this response document), we have revised the kernel description in the manuscript (lines 111–118) to clarify that the kernel is a 3D joint histogram constructed from global observational data, where mean albedo is calculated for discrete bins of CF, LWP, and N_d , and that this approach is purely diagnostic and does not require additional variables or radiative transfer calculations.

120

- Otherwise one might think that you could use the bin-center values to calculate albedos for each bin using the kernel method without having to do it for every datapoint and then averaging? But I think this comes from the lack of explanation about the kernel method. Presumably the kernel method requires more information so that this is not possible? It would be good to talk about that a little.

125

Reply: This point has been addressed in our previous reply regarding the kernel construction (see above).

130

- It would also be good to say that the average albedo values for each bin are the ones that could form the “look-up table” that might enable rapid albedo calculations based on just CF, N_d and LWP, which would be a lot easier than doing radiative calculations and (presumably) easier than doing the kernel calculation. And then that this approach needs to be tested against CERES (following onto the next sentence). This would be useful to the reader because it is reiterating the aim of the paper, but at the point in the text where you have explained the approach.

135

Reply: The functionality mentioned by the reviewer is already implemented in our kernel method. Specifically, the average albedo values for each CF–LWP– N_d bin, as constructed in the kernel, effectively serve as a reference for estimating albedo from these three cloud parameters. This approach enables rapid albedo calculations based solely on CF, LWP, and N_d , which is considerably easier than performing full radiative transfer calculations.

As clarified in the revised manuscript text (lines 111–118), as well as described in our response to an earlier comment regarding the kernel method in lines 67–86 of this response document, the approach is conceptually similar to the cloud-feedback kernels of Zelinka et al. (2012). It is derived from observational relationships and does not require radiative calculations. The validation of this approach against CERES observations forms the core aim of this study.

- The word “pixel” here is a bit confusing too – “1x1 degree daily datapoints” would be clearer.

Reply: Corrected, as suggested.

Line 132: “Underestimates of $\Delta\alpha < -0.02$ are particularly frequent around 40°latitude in both hemispheres,”

- It looks to me like the frequencies are high at latitudes greater than 40 deg?

Reply: This has now been corrected.

Line 140: “as suggested by the faint diagonal lines visible in Figure 4b.”

- I can’t really see any faint diagonal lines? I can see some straight lines that look like artefacts, though.

Reply: We have removed this part of the sentence to avoid confusion. The diagonal lines we referred to (aligned with the satellite orbits) were most likely artifacts related to sun glint.

Line 160: “which in Figure 4a appear predominantly brighter than other cloud scenes with similar CF–LWP– N_d characteristics”

- Could be worded better. Fig. 4a more suggests that they “are observed by CERES to be brighter than calculated from the CF–LWP– N_d values using the kernel approach”, or similar.

Reply: Rephrased to “cases of thick stratocumulus, which in Figure 4a are observed by CERES to be brighter than calculated from the CF–LWP– N_d values using the kernel approach”.

Line 175: “This explains the significant number of strong underestimates also visible in Figure 3.”

- It might also suggest why there are underestimates at high latitudes in Fig. 4a.

Reply: This has been now included.

Line 190: “Secondly, in order to ensure that the number of bins (50) was sufficient to reflect the characteristic U-shaped distribution of cloud fraction (with very small or nearly complete cloud cover occurring most frequently, while intermediate values appear relatively rarely), an alternative estimation was also performed using a much larger number of bins – 1000 (modification no. II).”

- Presumably, this is separate to modification no. I? It would be good to make that clear here.

Reply: The sentence has been revised so that modification no. II is introduced at the beginning of the sentence.

Line 216 – “Figure 7a-b shows the histogram of $\Delta\alpha$ after applying this correction.” – it’s not clear which correction you are referring to here. From the text and table I think that this is just the correction using the mean $\Delta\alpha$ within each SZA_{\max} interval (modification IV) and not also modification III?

Reply: We were referring here to modification no. IV, which has now been specified in the text.

Line 232 (and 235) – “These results show that the reconstructed albedo of a scene of clouds based on the mean cloud field properties exhibits systematic biases”

- “based on the mean cloud field properties” is a bit confusing here since you are basing it on the actual CF, LWP and N_d cloud properties – the issue rather seems to be that using a single global mean “look-up table” with mean albedo values for each bin leads to systematic biases?

Reply: The text has been revised.

- Reading on to line 260 makes it clearer what you mean here since you have now explained that there are likely factors other than the cloud properties (CF, N_d and LWP) at play. However, this was not so clear at the start of the section where you should explain the use of the global mean

190 albedo look-up-table (as mentioned in the previous bullet point) and mention that by “mean cloud properties” you mean CF, N_d and LWP only.

Reply: Amended, the beginning of the section has been clarified.

195 **Line 268** – “to build a simplified CF–LWP– N_d – α kernel” – not sure if this is a very descriptive way of describing it. “a simplified method to calculate albedo based only on CF, LWP and N_d values” or similar would be better.

Reply: Rephrased, as suggested.

200 **Line 268** – “spatial differences in albedo-to-cloud-sensitivity.”. This is also not clear – do you mean “spatial differences in the sensitivity of albedo to cloud properties”?

Reply: Clarified.

205 **Line 270** – “It was demonstrated that the number of biases in reconstructed albedo can be as high as ~60% of cases, when aiming for the accuracy in estimates (absolute difference between expected for the given CF–LWP– N_d conditions and measured by CERES albedo) at 0.02; which corresponds to about 10% of relative difference (Fig. 3).”

- This would be better as “It was demonstrated that the percentage of datapoints in which the reconstructed albedo biases (relative to the measured CERES albedo) were $>+/-0.02$ (a relative bias of around $+/-10\%$) can be as high as ~60% (Fig. 3).”

210 Reply: Rephrased, as suggested.

Line 280 “can be achieved when the average albedo in the given CF–LWP– N_d conditions is calculated at a 1° grid resolution.”

- Would be better as “can be achieved when the average albedo for each CF, LWP and N_d bin is calculated at a 1° grid resolution.”

215 Reply: Rephrased, as suggested.

Line 284: “on a pixel level” – again, better as “at a 1 degree resolution”

Reply: Amended, as suggested.

220

Line 286 – “the mean cloud field properties” – again, it would be good to say that you mean CF, N_d and LWP here.

Reply: Amended.

225 **Typos**

Line 55; “bins the” -> “bins in the”

Line 90: “explaining” -> “explain”

230 **Line 95:** “Resulting subset of cases considered in this study is pictured at Figure 1.” -> “The resulting subset of cases considered in this study is pictured in Figure 1.”

Line 189: “may have an larger influence” -> “may have a larger influence”

Line 272: “showed” -> “shown”.

Line 279: “in attempt” -> “in an attempt”

Line 282: “with modest” -> “with a modest”

235 **Line 283:** “showed” -> “shown”

Line 283: “in CF–LWP– N_d – α ” -> “in the CF–LWP– N_d – α ”

Reply: All of the above typos have been corrected.

References

245 Duran, B. M., Wall, C. J., Lutsko, N. J., Michibata, T., Ma, P.-L., Qin, Y., Duffy, M. L., Medeiros, B.,
and Debolskiy, M.: A new method for diagnosing effective radiative forcing from aerosol–cloud
interactions in climate models, *Atmos. Chem. Phys.*, 25, 2123–2146, <https://doi.org/10.5194/acp-25-2123-2025>, 2025.

250 Zelinka, M. D., Klein, S. A., and Hartmann, D. L.: Computing and Partitioning Cloud Feedbacks
Using Cloud Property Histograms. Part II: Attribution to Changes in Cloud Amount, Altitude, and
Optical Depth, *J. Clim.*, 25, 3736–3754, <https://doi.org/https://doi.org/10.1175/JCLI-D-11-00249.1>,
2012.

Response to comments

Note: All line numbers refer to the revised manuscript (with track changes hidden).

5 Reviewer 1 (RC2)

10 After thinking about the approach used in the paper a little more I (Referee #1) have a further, potentially more serious, concern about the bias calculations for the albedo estimated from the CF- N_d -LWP binned approach. It particularly applies for the estimate from the 1x1 degree resolution binned estimates, but it would be worth looking into for the global estimates too. I'm afraid that this may affect the conclusions of the paper and require some additional analysis and re-writing (although likely only for the 1x1 degree resolution part I think).

15 For the 1x1 deg correction is there a chance that each bin in the CF-LWP- N_d space is only filled once so that the mean over time for each bin contains only one value. Then when matching the daily MODIS datapoints to a value from the look-up-table, the value selected will not be the time-average of several points, but simply the same value again? This might then lead to the very small errors observed (assuming that the kernel method is a good match to the CERES observation). So this then becomes a test of the kernel method rather than testing the utility of using a time averaged look-up-table estimate. I.e., it would not be a good test of how well the look-up-table approach would work for new datapoints given just the values of CF, N_d and LWP (without the extra information that presumably goes into the kernel calculation).

20 For each 1x1 datapoint in the daily MODIS record for 2003–2021 (19 years) there are around $19 \times 365 = 6935$ data points. But there are a total of $50 \times 40 \times 30 = 60,000$ bins in the CF- N_d -LWP space used. This might make it likely that some (all?) of the bins are only used once – although some bins are likely more populated than others.

25 Therefore, you should examine how many datapoints are being used to calculate the average albedo in each bin for the 1x1 degree binned look-up-tables and come up with a statistical measure for how many datapoints you would need in a given bin for the bias estimate to be useful (i.e., a useful measure of how good the look-up-table approach would be for estimating the albedo of datapoints that weren't used in the look-up-table (using only CF, N_d and LWP). It would be good to do this for the other estimates too (e.g., the global mean binned look-up-table, 5x5 degree, etc.).

30 Perhaps a better approach would be to build the look-up-table using only part of the MODIS/CERES record and then to calculate the biases vs CERES using the other part of the MODIS/CERES record. This would ensure that the same data is not used for the kernel calculation and the bias testing.

35 **Reply:** We would like to thank the reviewer for this detailed comment. We have examined how many times each CF-LWP- N_d bin is filled and used to construct the kernel in the main method and in modifications no. V (5° latitude bands), VI (5° latitude–longitude grid), and VII (1° latitude–longitude grid). The distribution of the number of data points per bin is shown in Figure R2.

40 For the 1x1 degree latitude–longitude modification, we confirm that the majority of CF-LWP- N_d bins are populated only once over the full 2003–2021 period. As noted by the reviewer, in this case the reconstructed albedo is effectively drawn from the same datapoint, that is later used in the comparison, meaning that the resulting bias estimates do not represent an independent test of a time-averaged kernel. We agree that this is a significant limitation of the 1x1 degree modification and that the very small errors obtained in this case should therefore be interpreted with caution.

45 This limitation is specific to the highest-resolution (1x1 degree) modification, where the combination of a large number of spatial points as well as CF-LWP- N_d bins leads to strong undersampling of the kernel. For coarser spatial aggregations (the main method and modifications no. V and VI), the bins are filled

by multiple data points much more frequently, so the kernel more often represents an average over multiple cases rather than a single one.

55

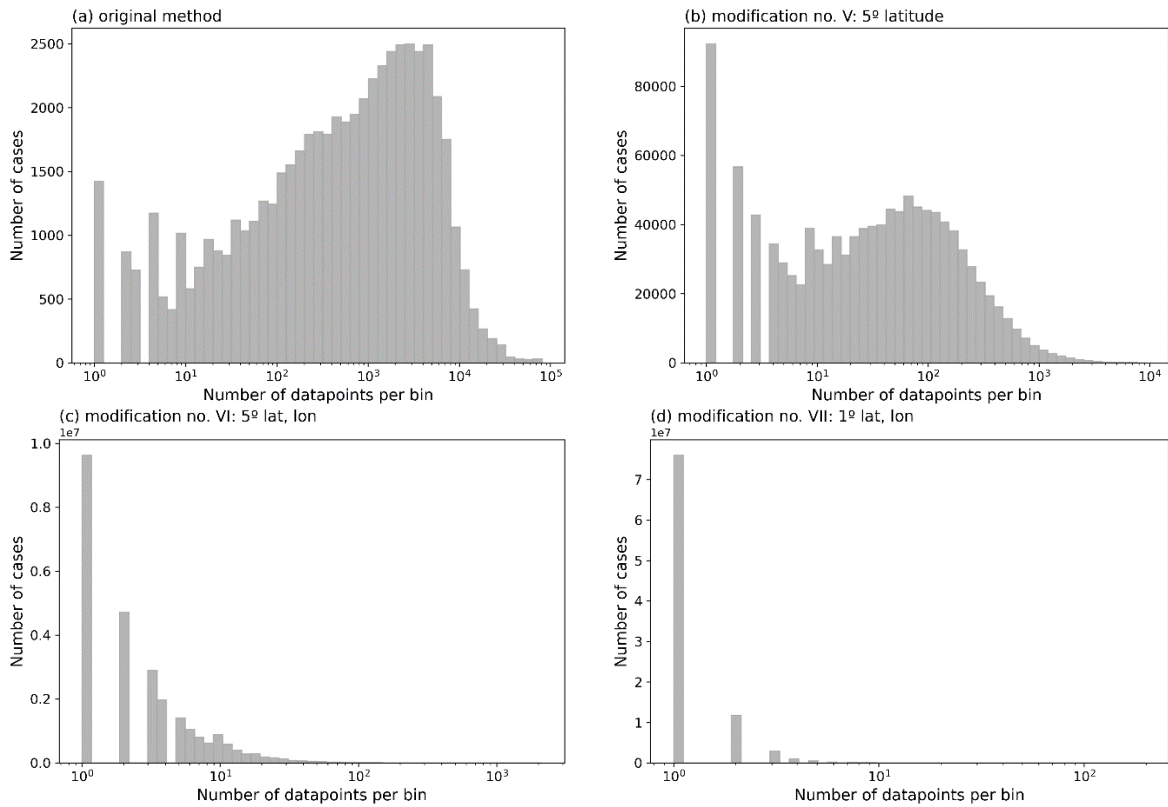


Figure R2. Distribution of the number of datapoints falling into a given CF-LWP- N_d bin for the original method (a) and modifications no. V (b), VI (c), and VII (d).

To address the sparsity issue in the 1x1 degree modification, we tested alternative bin configurations using wider CF-LWP- N_d intervals. We reduced the number of bins from 50x40x30 to 10x10x10 (Fig. R3a) and 5x5x5 (Fig. R3b) for each of the three variables. With these coarser bin structures, the CF-LWP- N_d bins become substantially better populated, reducing the problem of single-datapoint bins. This suggests that for high-resolution spatial grids, using a reduced number of wider bins may provide a more robust kernel representation, though at the cost of reduced sensitivity to fine variations in cloud properties.

60

65

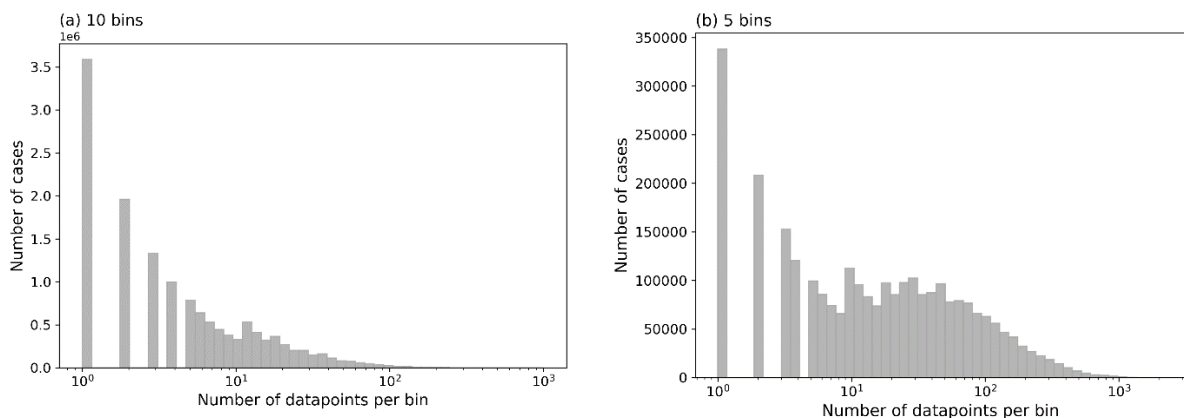


Figure R3. Distribution of the number of datapoints falling into a given CF-LWP- N_d bin for modification VII (1x1 degree grid) using 10 bins (a) and 5 bins (b) for each of the CF, LWP, and N_d variables.

70 We further tested how these coarser bin configurations (10x10x10 and 5x5x5) affect the accuracy of
 albedo reconstruction for modification VII. The results are shown in Table R2. While the wider bins are
 indeed better populated, the accuracy of the albedo estimates is only marginally better than the original
 global kernel method, particularly for the 5x5x5 bin configuration. This suggests that simply widening
 the bins does not fully resolve the trade-off between spatial resolution and kernel robustness.

75 In the revised manuscript, we have retained the original 50x40x30 bin configuration for modification
 VII (now referred to as VIIa), but have also included modifications VIIb and VIIc (using 10x10x10 and
 5x5x5 bin configurations, respectively) with their results presented in Table 1 in the manuscript. The
 manuscript has been revised to acknowledge this limitation and to clarify that the 1x1 degree results
 should be interpreted with caution. Specifically, we have added the following text:

80 Lines 255–263: “The most effective improvements in the estimates were achieved when α_{avg} was
 calculated separately for individual grid cells (modification VIIa). However, the results of this
 modification were significantly affected by bin sparsity, which in many cases have only been filled once.
 As a result, the reconstructed albedo is effectively drawn from the same datapoint, that is later used in
 the comparison, meaning that the resulting bias estimates do not represent an independent test of a time-
 85 averaged kernel. In order to reduce strong undersampling of the kernel, for modification VIIa two
 alternative bin configurations were examined. While modification VIIa retained the original 50x40x30
 bin configuration, modifications VIIb and VIIc used coarser bin structures (10x10x10 and 5x5x5,
 respectively) to improve bin population. As shown in Table 1, these coarser bin configurations reduced
 the sparsity problem but yielded only marginally better accuracy than the global kernel method,
 suggesting that simply widening bins does not fully resolve the trade-off between spatial resolution and
 90 kernel robustness.”

Lines 289–294: “As discussed in the results section, the finest spatial resolution modification (1°x1°)
 suffers from significant bin undersampling, with many bins filled only once over the study period,
 limiting its validity as an independent test of the kernel. Among the tested alternatives, the 5°x5°
 95 modification possibly represents the most reliable high-resolution approach, maintaining adequate bin
 population while capturing meaningful regional variation in the CF–LWP– N_d –albedo relationship.
 Future developments might address undersampling through alternative methodologies (e.g., machine
 learning approaches), potentially enabling robust use of finer spatial resolution data.”

and

100 Lines 340–345: “At the finest (1°x1°) spatial resolution, the reconstructed albedo biases must therefore
 be interpreted with caution, as the kernel is frequently undersampled and the apparent reduction in bias
 partly reflects self-matching of individual datapoints. Among the tested spatial resolutions, the 5°x5°
 modification provides a more reliable high-resolution estimate, maintaining adequate bin population
 while still substantially outperforming the coarser modifications, suggesting that if bin undersampling
 could be resolved, the 1°x1° results would indeed represent the true optimal performance.”

105

Table R2. Ratio of underestimated ($\Delta\alpha < -0.02$) and overestimated ($\Delta\alpha > 0.02$) cases of albedo for methodological
 modification VII using the original 50x40x30 bins (VIIa), 10x10x10 bins (VIIb), and 5x5x5 bins (VIIc).

Methodological modification	Ratio of cases (%) with:			
	$\Delta\alpha < -0.02$	$\Delta\alpha > 0.02$	$ \Delta\alpha > 0.02$	$ \Delta\alpha \leq 0.02$
Original method	27.7	31.4	59.1	40.9
VIIa: 1° lat, lon	5.6 (–22.1)	5.6 (–25.8)	11.2 (–47.9)	88.8 (+47.9)
VIIb: 10 bins	21.8 (–5.9)	23.5 (–7.9)	45.3 (–13.8)	54.7 (+13.8)
VIIc: 5 bins	25.8 (–1.9)	28.9 (–2.5)	54.7 (–4.4)	45.3 (+4.4)

Response to comments

Note: All line numbers refer to the revised manuscript (with track changes hidden).

5 Reviewer 2 (RC3)

Wojciechowska and Gryspeerdt present a simplified framework for estimating marine liquid cloud albedo using cloud fraction (CF), liquid water path (LWP), and cloud droplet number concentration (N_d). Their approach combines climatological, collocated MODIS cloud properties with CERES albedo to construct a reduced cloud albedo kernel. The authors show that this formulation yields robust albedo estimates, with more than 80 percent of samples differing from CERES by less than 0.05. The results also exhibit a clear spatial structure in regions where albedo is systematically underestimated or overestimated.

15 The manuscript is clearly written, and I find no major issues with the methodology or the interpretation of results. My comments concern a few missing details and clarifications. I also suggest a straightforward way to incorporate cloud morphology into the algorithm, which may further enhance the albedo estimates. I recommend acceptance of the manuscript after the following minor comments are addressed:

20 **Line 57:** One of the key factors that controls the relationship between albedo and CF and/or LWP and/or N_d /CER is the cloud morphology or sub-grid cloud heterogeneity (Goren et al., 2023; McCoy et al., 2023; Choudhury and Goren, 2024).

25 It is still unclear to me what the main motivation is for assessing how accurately cloud albedo can be estimated from cloud microphysical and macrophysical properties. Is the aim to evaluate these relationships because they are used to quantify changes in cloud albedo driven by changes in cloud properties, as in Zhang et al. (2021) or Wall et al. (2022)? If so, then this study seems to assess whether kernel-based decompositions of cloud albedo are justified. Could the authors clarify this point in the introduction?

30 **Reply:** We have clarified the motivation of the study in the introduction and added the suggested references. We now state specifically that the purpose of this work is to assess whether kernel-based decompositions of cloud albedo are justified, and how well mean cloud properties (CF, LWP and N_d) can be used to reconstruct the albedo of cloud scenes across the globe. This has been framed in the context of the cloud morphology and sub-grid cloud heterogeneity.

35 The revised text in the introduction (54–75) now reads:

“However, these kernels/relationships assume some linearity between cloud properties and albedo. While this can be partially addressed by using a larger number of bins in the radiative kernel (e.g. Gryspeerdt et al., 2019), this approach does not account for the sub-gridbox distribution of cloud properties. The impact of cloud heterogeneity on radiative transfer has been extensively studied, with cloud inhomogeneity shown to introduce biases in grid-averaged albedo calculations (Cahalan et al., 1994; Oreopoulos and Davies, 1998) and significant effects on the Earth’s radiation budget through mesoscale cloud variations (Rossow et al., 2002). Climate models have developed parameterizations to account for subgrid-scale cloud variability (Barker and Räisänen, 2005), though uncertainties remain in how well mean properties capture scene albedo. For example, with a non-linear relationship between LWP and cloud albedo, the mean LWP of a field of clouds does not uniquely determine the mean cloud albedo (Zhang and Feingold, 2023), as cloud morphology and sub-grid cloud heterogeneity play key roles in controlling the relationship between albedo and CF, LWP, and/or N_d is (Choudhury and Goren, 2024; Goren et al., 2023; McCoy et al., 2023).

45 Relationships between cloud properties and albedo are commonly used to quantify changes in cloud albedo associated with variations in cloud macrophysical and microphysical properties, for example in the context of aerosol forcing and cloud feedback studies (Wall et al., 2022; Zhang et al., 2021). It remains unclear to what extent scene albedo is accurately captured by a three-parameter decomposition

(CF, LWP and N_d). While climate model results suggest this is sufficient, they do not represent the details of the sub-gridbox cloud distribution needed to accurately test this assumption.

55 This study aims to assess whether kernel-based decompositions of cloud albedo are justified and how well mean cloud properties (CF, LWP and N_d) can be used to reconstruct the albedo of liquid cloud scenes across the globe. Using a joint-histogram/kernel approach, this work reconstructs albedo from average cloud properties at 100 km scales, characterising regional variations in the error in the reconstructed albedo compared to observations. Different methods for accounting for these biases were
60 assessed, providing recommendations for future observation-based calculations of aerosol forcing and cloud feedbacks.”

Line 73: Which cloud fraction is used? Is it the cloud fraction derived only from pixels with successful cloud property retrievals, as is common in many aerosol–cloud interaction studies, or the total cloud fraction including partially cloudy pixels and those with failed retrievals?
65

Reply: The cloud fraction used in this study corresponds to the MODIS “Cloud Retrieval Fraction Liquid” and is derived only from pixels with successful cloud optical property retrievals. This information has now been added to the text (lines 83–84).

70 **Line 79:** Were CERES Aqua and Terra products considered separately and paired with their respective MODIS–cloud retrievals, or were the datasets averaged across the two platforms?

Reply: The text has been revised to clarify that CERES and MODIS retrievals were paired by satellite platform and not averaged across platforms (see lines 95–96). The revised text now reads: “CERES and MODIS retrievals were paired by satellite platform (Aqua–Aqua and Terra–Terra) and analysed jointly.”
75

Section 3.1: I recommend including an additional global map showing the geographical distribution of albedo bias (potentially as an extension of Figure 4). This figure should be supported by a short description summarizing the mean bias and its spatial pattern, noting regions where the discrepancies are especially pronounced.
80

Reply: The map has been added to Figure 4, together with a short description summarizing the mean bias and its spatial pattern:

Lines 162–165: “Figure 4c presents the geographical distribution of absolute albedo bias (cases with either underestimated or overestimated albedo by more than 0.02). Generally, albedo biases are mostly evenly distributed spatially, regardless of the number of datapoints at a given location (Figure 1).”
85

Section 3.1: Does the bias decrease when the data are resampled to the monthly scale, which is the temporal resolution typically used in studies that construct such kernels?

Reply: Thank you for this suggestion. We have introduced two additional modifications of the method related to the temporal aggregation used in the kernel construction: modifications IX and X. In the original method, the kernel is constructed using CF, N_d , and LWP values from the full annual time series. In modification IX, the kernel is instead constructed from monthly-averaged CF, N_d , and LWP, while in modification X – from daily-averaged values of CF, N_d , and LWP. When evaluating the difference between estimated and observed albedo, the reconstructed albedo is compared against observations corresponding to the same month (modification IX) or the same day (modification X), respectively.
90

95 Both temporal aggregations lead to an improvement in the method (Table 1 in the manuscript and Table R1 below). The ratio of correct albedo reconstructions increased from 40.9% for no temporal aggregation (original method without modifications) to 57.7% when the kernel is constructed from the monthly-averaged CF, N_d , and LWP and 60.6% when the kernel is constructed from the daily-averaged CF, N_d , and LWP.

100 The revised text in the manuscript now reads as follows:

Lines 232–237: “In all of the above modifications, the kernel was constructed using CF, N_d , and LWP values from the full annual time series. The last two modifications were related to the temporal aggregation used in the kernel construction. In modification IX, the kernel is instead constructed from monthly-averaged CF, N_d , and LWP, while in modification X it is constructed from daily-averaged

105 values of CF, N_d , and LWP. When evaluating the difference between estimated and observed albedo, the reconstructed albedo is compared against observations corresponding to the same month (modification IX) or the same day (modification X), respectively.”

110 Lines 267–273: “When applying modifications IX and X, an improvement in the method is observed, with the ratio of correct albedo reconstructions increasing from 40.9% for no temporal aggregation (original method without modifications) to 57.7% when the kernel is constructed from the monthly-averaged CF, N_d , and LWP and to 60.6% when it is constructed from the daily-averaged CF, N_d , and LWP (Tab. 1). However, it should be noted that the daily-averaged version (modification X) may also be subject to bin undersampling, as the number of datapoints available on individual days is considerably smaller than over longer averaging periods. Therefore, the daily results should be interpreted with similar caution as the finest spatial resolution case.”

This response and Table R1 are also included in response to a similar comment from the other reviewer.

Table R1. Ratio of underestimated ($\Delta\alpha < -0.02$) and overestimated ($\Delta\alpha > 0.02$) cases of albedo for methodological modifications no. IX and X. Values in brackets indicate the difference with respect to the original method.

Methodological modification	Ratio of cases (%) with:			
	$\Delta\alpha < -0.02$	$\Delta\alpha > 0.02$	$ \Delta\alpha > 0.02$	$ \Delta\alpha \leq 0.02$
Original method	27.7	31.4	59.1	40.9
IX: monthly-averaged	20.0 (–7.7)	22.3 (–9.1)	42.3 (–16.8)	57.7 (+16.8)
X: daily-averaged	19.2 (–8.5)	20.2 (–11.2)	39.4 (–19.7)	60.6 (+19.7)

120 **Figure 6b:** How does the solar zenith angle extend beyond 65 degrees when this threshold was used in the G18 sampling strategy for estimating N_d ?

125 Reply: We would like to thank the reviewer for this comment. We have now excluded from the analysis all cases with solar zenith angles exceeding 65°. All figures and Table 1 have been updated accordingly, using only data with $SZA_{\max} < 65^\circ$. This additional filtering, however, has not impacted the study results. The cases with SZA_{\max} exceeding 65° likely resulted from gridboxes in which observations from multiple orbits (and hence a range of solar zenith angles) were averaged.

Line 213: “... more restricted to location ...” Perhaps “cloud morphology” is intended here?

Reply: The text has been revised as suggested.

130 **Line 224:** What explains the seasonal variability? Did you construct the kernel at each grid point as well as for all days of the year separately? If not, this might help reduce the bias further.

135 Reply: We have now introduced two modifications related to the temporal aggregation, as described in the lines 88–118 of this response document, and have added the relevant description to the manuscript (see lines 232–237 and 267–273).

140 I also suggest adding thin cloud fraction (F_{thin} , a proxy for cloud morphology) to the kernel. F_{thin} can be computed from the MODIS Level-3 “Cloud_Optical_Thickness_Liquid_Histogram_Counts” dataset by using $COT < 3$ or < 5 . Including F_{thin} alongside CF, LWP and N_d /CER may improve the albedo estimation (McCoy et al., 2023).

145 Reply: We thank the reviewer for this suggestion. Following the recommendation, we have now implemented a four-dimensional kernel that incorporates the fraction of optically thin clouds (F_{thin}) computed directly from the MODIS Level-3 “Cloud_Optical_Thickness_Liquid_Histogram_Counts” dataset. We computed F_{thin} using two COT thresholds: $COT < 3$ and $COT < 5$, representing clouds with optical thickness below these values. The results of this analysis are presented in Table R3.

150 Our findings show a small but consistent improvement in albedo reconstruction accuracy when including F_{thin} alongside CF, LWP, and N_d (Tab. R3). When using F_{thin} with $COT < 3$, the fraction of underestimated cases slightly decreased from 27.7% to 27.5%; the fraction of overestimated cases – from 31.4% to 30.7%. Using $COT < 5$ yields marginally better results: 27.1% of underestimated and 30.4% of overestimated cases.

While this improvement is modest, it demonstrates that cloud morphology information (as captured by the presence of optically thin clouds) does provide additional value for albedo estimation. However, we note that adding F_{thin} as a fourth dimension does increase bin sparsity in the kernel, which already is to some extent a limitation of the implemented method. Given this trade-off, we have retained the three-parameter approach in the main manuscript as our primary method, but we acknowledge that incorporating F_{thin} is a promising avenue for future refinements.

Table R3. Ratio of underestimated ($\Delta\alpha < -0.02$) and overestimated ($\Delta\alpha > 0.02$) cases of albedo for different methodological modifications: original three-parameter kernel, and four-dimensional kernels incorporating F_{thin} computed for $\text{COT} < 3$ and $\text{COT} < 5$.

Methodological modification	Ratio of cases (%) with:			
	$\Delta\alpha < -0.02$	$\Delta\alpha > 0.02$	$ \Delta\alpha > 0.02$	$ \Delta\alpha \leq 0.02$
Original method	27.7	31.4	59.1	40.9
F_{thin} for $\text{COT} < 3$	27.5 (-0.2)	30.7 (-0.7)	58.2 (-0.9)	41.8 (+0.9)
F_{thin} for $\text{COT} < 3$	27.1 (-0.6)	30.4 (-1.0)	57.5 (-1.6)	42.5 (+1.6)

Could the authors speculate on why albedo is underestimated in stratocumulus regions and overestimated in the tropics?

Reply: We have added a brief discussion addressing this point in the manuscript:

Lines 283–287: “These contrasting biases may reflect differences in cloud morphology and sub-grid heterogeneity: in stratocumulus regions, nonlinear radiative effects and unresolved horizontal variability can lead to an underestimation of scene albedo when using mean cloud properties, whereas in the tropics, highly heterogeneous convective cloud fields may cause mean properties to overrepresent optically thick elements, resulting in albedo overestimates.”

Language edits:

Line 55: “more bins in ...”?

Reply: Amended.

Line 55: “sub-pixel” or “sub-gridbox”

Reply: Changed to “sub-gridbox”.

Line 101: By “pixel,” do you mean the 100 km gridbox?

Reply: Rephrased to “1x1 degree gridbox”.

Line 123: Do you mean “left-skewed”?

Reply: Amended.

Line 270: Please rephrase. The current wording is unclear.

Reply: The sentence has been rephrased.

References:

Choudhury, G., & Goren, T. (2024). Thin clouds control the cloud radiative effect along the Sc-Cu transition. *Journal of Geophysical Research: Atmospheres*, 129, e2023JD040406. <https://doi.org/10.1029/2023JD040406>

Goren, T., Sourdeval, O., Kretschmar, J., & Quaas, J. (2023). Spatial aggregation of satellite observations leads to an overestimation of the radiative forcing due to aerosol-cloud interactions. *Geophysical Research Letters*, 50, e2023GL105282. <https://doi.org/10.1029/2023GL105282>

McCoy, I. L., McCoy, D. T., Wood, R., Zuidema, P., & Bender, F. A.-M. (2023). The role of mesoscale cloud morphology in the shortwave cloud feedback. *Geophysical Research Letters*, 50, e2022GL101042. <https://doi.org/10.1029/2022GL101042>

195 Wall, C. J., Norris, J. R., Possner, A., McCoy, D. T., McCoy, I. L., and Lutsko, N. J.: Assessing effective radiative forcing from aerosol-cloud interactions over the global ocean, *Proc. Natl. Acad. Sci. U. S. A.*, 119, 1–9, 2022.

Zhang, Y., Jin, Z., and Sikand, M.: The Top-of-Atmosphere, Surface and Atmospheric Cloud Radiative Kernels Based on ISCCP-H Datasets: Method and Evaluation, *J. Geophys. Res. Atmos.*, 126, e2021JD035053, <https://doi.org/https://doi.org/10.1029/2021JD035053>, 2021