Response to comments

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

Reviewer 1 (RC2)

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-Nd-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).

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).

For each 1x1 datapoint in the daily MODIS record for 2003–2021 (19 years) there are around 19*365 = 6935 data points. But there are a total of 50*40*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.

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.).

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.

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.

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.

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.

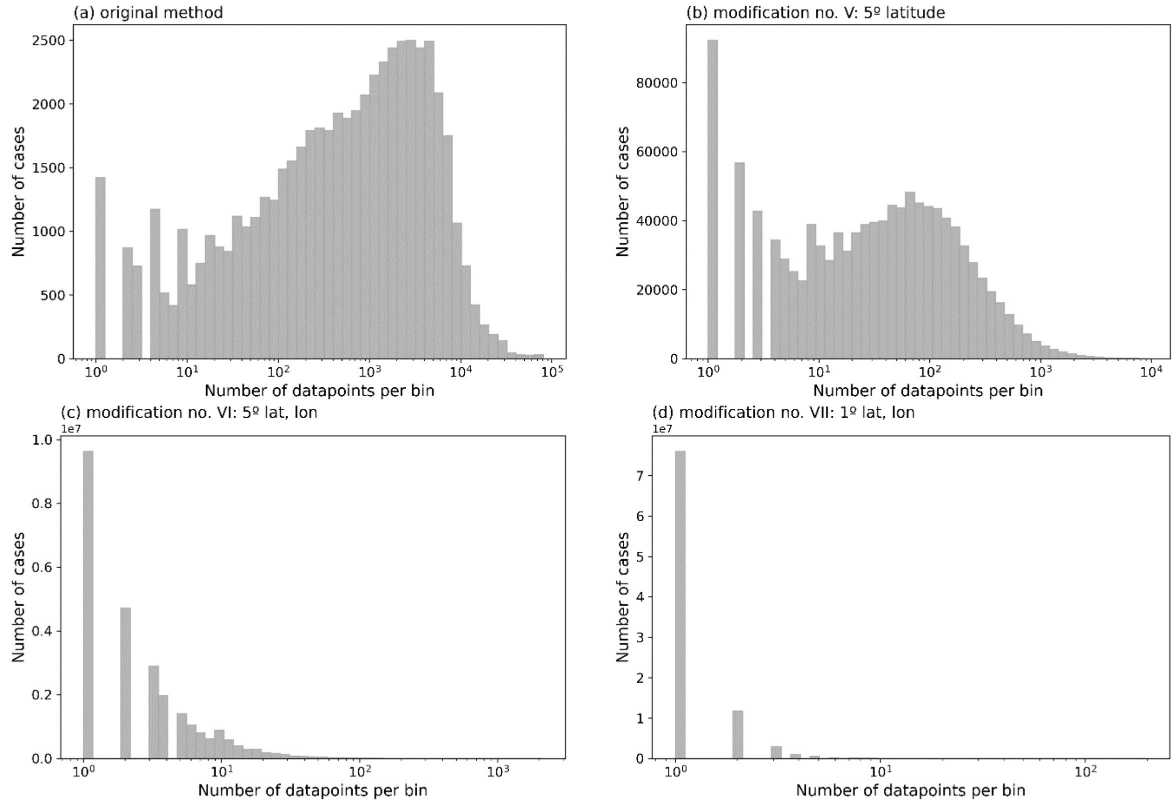

Figure R2. Distribution of the number of datapoints falling into a given CF–LWP–N_d bin for the original method (a) and modifications no. V (b), VI (c), and VII (d).

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.

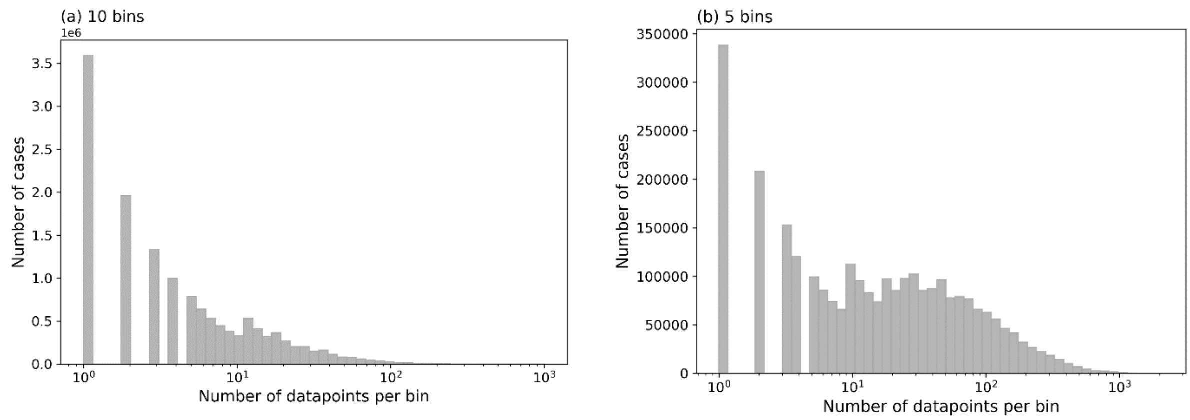

Figure R3. Distribution of the number of datapoints falling into a given CF–LWP–N_d bin for modification VII (1x1 degree grid) using 10 bins (a) and 5 bins (b) for each of the CF, LWP, and N_d variables.

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.

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:

Lines 255–263: "The most effective improvements in the estimates were achieved when αavg 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-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 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° 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

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."

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:** $\Delta\alpha < -0.02$	$\Delta\alpha > 0.02$	$\lvert\Delta\alpha\rvert > 0.02$	$\lvert\Delta\alpha\rvert \le 0.02$
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)