–Legend—
Referee text is in black
Author response text is in blue
— — — —

Referee #1

The authors fail to solve the issue associated with the method: the new method yields a zero longwave forcing at noon (10:00-14:00), which is not reasonable. Unless there is no contrail at all at noon (which is untrue), the longwave forcing of contrails should not be zero. Furthermore, the surface temperature is higher at noon, so theoretically the longwave forcing of a contrail should be significant during 10:00-14:00. Therefore, the new method is not reliable.

We appreciate the reviewer's perspective here, and performed a deeper causal inference literature search leading to further auditing to investigate this concern about the methodology. This process has revealed that our data sampling for the "local hour" plot in Appendix E had a sampling issue which caused the majority of pixels for those 10:00-14:00 hour bins to have almost no flight density, violating positivity, one of the core causal inference assumptions. Fixing this data sampling issue removes the artifact where the longwave forcing near local noon was estimated to be close to 0. We have also corrected the data sampling for Figure 9 and Figure 10 (cumulative longwave forcing by advection age H) for the advection ages H < 12. The data generated in service of H=12 results (and other results outside of these three figures) did not use the afflicted codepath, so were *not* affected by the sampling issue and therefore have remained unchanged in this submission. To be explicit, our central estimated longwave $oRF_{H=12}$ of 46.9 GJ/km over the Americas has not changed.

We discovered this sampling error in the process of evaluating the "positivity assumption" using the hat-value diagnostic described by (Moodie & Schulz, 2025).The "positivity assumption" being assessed by the diagnostic dictates that subgroups in a trial/study need to have both a meaningfully positive probability of being exposed to the treatment and not exposed to treatment. That is in this case, subgroups having advected trace density ($A_H > 0$) or not ($A_H = 0$). A "positivity violation" is then referring to a subgroup that has such a low/high probability of being exposed to treatment that no data exists in the dataset to allow inference of what would have happened in the counterfactual case. When applying the Moodie & Schulz hat-value diagnostic to the aggregated $oRF_{H=12}$ estimation dataset, no violations are detected. However, when applying the diagnostic to the "local hour" subgroup analysis, it showed clear positivity violations during the local hours around noon which previously had implausibly close to zero estimated longwave forcing. When investigated, it became apparent that far more data points than intended had $A_H = 0$ for H=3 and this was biasing the OLS fitting results especially in the hours around local noon. During dataset generation for H=12 we had correctly sampled at most

1% of ABI grid pixels having $A_H = 0$, but we had neglected to filter out pixels having $A_H = 0$ for values of H less than 12.

This is now fixed; Figure 9, Figure 10, and Figure E1 have been updated with the corrected sampling, and we took this opportunity to also increase the sample size of Figure E1's data to be drawn from 183 days in panel (a). Finally, Appendix E has been updated with details of the positivity diagnostic results.