

—Legend—

Referee text is in black

Author response text is in blue

— — — —

Referee #2

General Comments

This manuscript presents a causal-inference framework for estimating contrail longwave radiative forcing using GOES-16 OLR data, advected flight tracks, and regression against confounding variables. The approach is innovative, the manuscript is supported by appropriate references, and the synthetic dataset validation adds credibility.

At the same time, the framing of the new metric “oRF12” and several methodological details are not sufficiently clear. In particular, the analysis is restricted to longwave-only forcing, which risks being misinterpreted as a net effect. The role of the 12-hour window is not well explained upfront, and some key aspects of data handling and model comparison (cloud-phase confounding, CoCiP setup) are underspecified. These issues should be addressed before publication to ensure the results are reproducible and properly contextualized.

Overall, I find the manuscript to be a promising and potentially valuable contribution, but I recommend revision to sharpen the scope, clarify limitations, and provide more methodological detail.

We thank the reviewer for their detailed review of the manuscript and helpful comments and suggestions. Please find below point by point responses.

Specific Comments

1. The manuscript refers to “long-lived contrail cirrus” (e.g., line 47) but does not provide an operational definition (e.g., $> N$ hours). Longevity is only implied by the monotonic growth of oRFH up to 12 hours (Fig. 9). Please provide a clear definition and link it to how oRFH should be interpreted.

We have clarified our usage of 'long-lived' contrail cirrus is intended to mean both linear and non-linear contrail cirrus (in the abstract as well as first usage in the Introduction section).

2. The paper emphasizes that the method does not require a contrail mask. In practice this means it is best applied to fleets or regional averages rather than individual flights. Please make this explicit and provide two or three concrete application scenarios (e.g., sectoral averages, model validation datasets).

We have added this explicitly (along with some examples) to the "Limitations and Future Work" section.

<please note, there was no comment number 3 in the Reviewer comments provided to us by the AMT review system>

4. The workflow (ADS-B + ERA5 + GOES-16 -> advection -> rasterization -> regression -> conversion to oRFH) is currently scattered across several sections. A simple flowchart would greatly improve clarity for readers and aid reproducibility.

Thank you for the suggestion, we have added a flowchart as figure 2.

5. The regression framework replaces the explicit “unaffected region” used in earlier studies. Please clarify more directly how the counterfactual is constructed statistically (i.e., which confounders are controlled, and how). It would also help to summarize the permutation test result in the main text, since this is crucial evidence against spurious correlation.

We have added more context and discussion in the Methods Section 2.1 on how Model 1 is constructing the statistical counterfactual for when there's no flight density ($A=0$), and for any counterfactual baseline flight density $A=a$, vs the observed $A=a+d$.

As the reviewer rightly notes, the permutation test is evidence against spurious correlation of our estimate. To make this more explicit, we have added a summary to the Insight from synthetic datasets section (Section 3.1), which discusses what particular coefficients are estimated to be zero in the model.

6. The paper uses GOES-16 COIN OLR as the outcome variable. Please consolidate the description into one place: which bands/channels are used, how COIN OLR is generated, and what assumptions might bias contrail-specific estimates.

We have consolidated the requested details about COIN OLR in section 2.2.1. Regarding the Referee's request for more details about how bias might affect contrail-specific estimates, we have augmented the "Limitations and Future Work" section to contain a general discussion of OLS regression input bias (including both COIN and confounder controls data).

7. The authors acknowledge that ERA5 alone cannot distinguish contrails from natural cirrus, and that GOES-16 cloud phase adds limited separation because contrails and natural cirrus both fall into the “ice” category. This is a central limitation. Please explicitly state in the manuscript that the method does not significantly improve contrail-cirrus separation, unless there is concrete evidence that it does.

We have added this clarification to the "Limitations and future work" section.

8. Relatedly, Table 1 shows large coefficients in the clear-sky category, partly attributed to cloud-phase misclassification. It would strengthen the paper if the authors quantified the sensitivity of oRF12 to plausible misclassification rates.

We agree that quantifying the sensitivity of our estimate to cloud phase misclassification strengthens the paper. To address this, we conducted a sensitivity analysis using our synthetic dataset where the ground truth forcing is known. We have added the results of this analysis as a new appendix in the revised manuscript (Appendix C: Sensitivity of $\text{oRF}_{H=12}$ to Cloud Phase Misclassification), which includes a detailed explanation of the methodology and a new figure (Figure C1). We also refer to this analysis in the discussion of Table 1. The results show an inverse relationship between the accuracy of the cirrus cloud phase classification and the error in our forcing estimate. This provides quantitative support for our hypothesis that misclassification is a primary driver of the large coefficients observed in the clear-sky category in Table 1. The analysis also shows that our methodology is robust under realistic levels of misclassification.

9. The comparison between oRF12 and CoCiP longwave iRF is interesting but not fully contextualized. Please clarify which adjustment processes are captured by oRF12 within the 12-hour window and which are excluded relative to ERF.

We are not aware of any studies to date which have quantified the timeline of (individual or collective) adjustment processes of contrails. The most similar studies from which to speculate are possibly CO₂ and sea surface temperature rapid adjustment studies such as <https://journals.ametsoc.org/view/journals/clim/22/11/2009jcli2652.1.xml> which reports in some study variants that cloud rapid adjustment response was "statistically consistent with the equilibrium value" by day 5 after the perturbation. We have updated the manuscript to include further discussion of this.

10. The study focuses exclusively on longwave forcing. This should be emphasized more prominently in the Abstract, Introduction, and figure captions to avoid misinterpretation. At minimum, please provide an order-of-magnitude estimate or citation for shortwave effects in the study domain, so readers can understand whether oRF12 represents an upper bound of the net effect or only a partial contribution.

We have added emphasis throughout about this analysis being longwave only, and added a CoCiP shortwave forcing estimate for the same spatio-temporal region (and brief discussion of whether oRF12 represents an upper bound of the net effect) to the "Contextualizing the Forcing Estimate" section.

11. The CoCiP setup used for comparison is insufficiently described. Please state explicitly the interpolation method for meteorological inputs (linear, nearest, etc.), the model time step (10, 30, or 60 minutes), and any regional or temporal subsampling. These choices are critical for reproducibility and for interpreting differences between oRF12 and CoCiP estimates.

Thank you for catching this omission, we have added these details to the "Synthetic dataset validations" section.

12. The manuscript does not provide any indication of computational cost. A simple case study (e.g., one day over CONUS) with approximate runtimes for advection, rasterization, and regression would help readers assess scalability and practical use.

Thank you for the suggestion, we have added these details as a new subsection "Computational expense" in the Methods section.

Technical Corrections

1. Define acronyms ADS-B at first use.

Done.

2. In Fig. 10 (diurnal cycle), explain the longitude-to-local-time mapping and why the x axis spans 48 h.

Done.

3. What is the oRF3 in the caption of Figure 10 and Line 329? It is better to clarify again with $H=3$.

This is the observational RF for the cumulative first three hours of advection estimated using Model (3) setting $H=3$. We have changed the notation throughout to $oRF_{H=3}$ instead of oRF_3 to make this explicit.

4. Abstract Line 4: “beyond a few hours” --> consider citing a specific range (3–6 h) with reference support.

3-6 hours has been added as the range in the Abstract, supported by adding to the Introduction a sentence citing the e-folding time from (Vázquez-Navarro et al., 2015).

5. Is it possible to provide more information in Section 3.3 for better understanding results in Fig. 9?

We have expanded Section 3.3, please don't hesitate to ask if something is still unclear.