

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.

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.

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).

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

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.

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.
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.
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.
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.
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.
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.
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.

Technical Corrections

1. Define acronyms ADS-B at first use.
2. In Fig. 10 (diurnal cycle), explain the longitude-to-local-time mapping and why the x axis spans 48 h.

3. What is the oRF3 in the caption of Figure 10 and Line 329? It is better to clarify again with $H=3$.
4. Abstract Line 4: “beyond a few hours” --> consider citing a specific range (3–6 h) with reference support.
5. Is it possible to provide more information in Section 3.3 for better understanding results in Fig. 9?