

The authors provided thoughtful responses to the reviewers comments to improve the clarity of the manuscript and address concerns about the separation of local and non-local effects. I particularly appreciate the comparison of the local effects between the moving window regression and chessboard methods using the MPI-ESM (Fig A6). However, I still have concerns over the overestimation of the local cooling effects in the manuscript and the implications of such overestimation for the findings.

More specifically, in the limitations section of the Discussion, the authors mention that they did not find evidence of systematic underestimation of local effects (without scaling), but that scaling to 100% could lead to overestimation of local effects compared to other methods. When performing the chessboard calculation of local effects, I'm assuming a 100% deforestation was performed on each deforested grid cell when deforesting 1 out of 4 or 2 out of 4 grid cells (Fig. A6). Given that the `deforest_glob` experiment only does partial deforestation in most grid cells, the `deforest_glob` scaled to 100% appears to be the best comparison with the chessboard experiment. I am questioning whether the comparison without scaling is relevant (in this case, for a given deforested grid cell, you would be comparing local effects resulting from partial deforestation from the `deforest_glob` experiment with local effects resulting from complete deforestation in the chessboard experiment)? If local effects were comparable across methods, shouldn't we see stronger local effects in the chessboard experiment in this case? When comparing the scaled `deforest_glob` experiment and the chessboard experiment, there is evidence of overestimation of the local effects as pointed out by the authors. The authors acknowledge factors that could contribute to the overestimation of local effects due to scaling, but isn't it possible that the choice of methodology itself (moving window regression vs chessboard) could lead to an overestimation of local effects? I don't think we can say that it is only scaling that leads to the overestimation of local effects.

Furthermore, the implications of such overestimation of local effects (regardless of where it comes from) for the findings are not discussed. How would that influence the emergent constraints discussed previously? Could this result in any changes in the relationships between the variables considered (local temperature, total temperature, albedo, latent heat flux)? For example, if the local effects are overestimated, the consistent overestimation of the local cooling response of the models in the boreal and temperate regions compared to observations may not be as pronounced as suggested in this study. Would that also explain part of the local albedo sensitivity which can't be explained by the higher percentage of snow cover in ESMs? Similarly, the authors also mention that the total cooling resulting from deforestation is overestimated in ESMs due to overestimation of the local cooling by ESMs (compared to observations), but that claim may also change if local effects are overestimated in this study. For example, if local effects are overestimated, it is

possible that non-local effects are underestimated (by a similar amount since a proportion of the cooling is just moved from local to non-local) which would then impact the relationship between local and total effects. Based on the evidence provided, particularly the stronger snow coverage in ESMs compared to observations, local (and total) cooling in northern latitudes do appear to be overestimated in ESMs compared to observations, but it may not be as pronounced as the results in this study are showing. Adding a discussion of the implications of an overestimation of local effects as part of the limitations would be important.

I also have a few additional specific comments, mainly to enhance clarity.

Line 15: Consider specifying the linear relationships between what variables in that sentence (i.e. resulting from the linear relationships between X and Y within the model ensemble).

Lines 89-106: Those are two important paragraphs setting the stage for the study. The stated goals of the study do not mention looking at the emerging relationship between local and total effect despite this being an important part of the manuscript. The stated goals focus mainly on the relationship between local temperature and biogeophysical properties within observations and models despite mentioning that knowing the emerging relationship between local and total effects within models would be useful. There seems to be a disconnect there.

Line 178: extra comma

Line 230-232: That sentence is not clear, maybe it should be rephrased. It also does not flow well with the end of the last paragraph; it is missing a transition. It is not clear how the local surface temperature sensitivities are related to the relationship between local and total temperature in this sentence.

Line 330: The link between this paragraph and the previous one where we talk about the relationship between local and total effect is not clear. Consider providing a better transition and elaborating on how Fig 9 ties to Fig 8.

Line 359: Would that still be the case if local effects are overestimated in this study, as per the main comment above?

Line 487: overestimation of local and total effects, not local and non-local effects. The results did not discuss overestimation of non-local effects.

Fig 5: FLUXNET-based surface temperature sensitivity should be green to match the green line of FLUXNET in Fig 4. Keep orange for MODIS.

Fig 6: Same comment as Fig 5. Also, why is Modis showing a negative local temperature change for all levels of snow cover, whereas it is positive in Fig 5 for the boreal region?

Fig 7: Same comment as Fig 5.

Fig 8: To be consistent with the colors chosen in other figures, MODIS sensitivity uncertainty range should be orange and FLUXNET sensitivity uncertainty range should be green.

Fig A4: Same comment as Fig 5.

Fig A6: pattern, typo in the caption. See also main comment regarding comparison between deforest_glob (not scaled) and chessboard experiments.