The reviewers' comments are highlighted in blue, the responses are in red.

Response to Report #1 (Referee #2)

New table "XX" added following Table A1: For the "Aggregation" column I believe the text for the FLUXNET row should read "true monthly mean".

Table A2 has been adjusted accordingly.

Response to Report #2 (Referee #1)

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.

We are thankful for the reviewer's constructive comments, and apologize that some of our responses in the first Revision (R1) were perhaps not clear.

In regards to evaluating possible uncertainties due to choice of method for separating local and non-local biogeophysical effects due to deforestation in the ESMs, we compared our results (based on a modified moving window regression) to the findings from Winckler et a. (2019) (based on chessboard experiments) in Revision 1. We agree with the reviewer's comment that for this comparison at grid cell level, scaling to 100% is the most appropriate way and have revised Fig. A6 accordingly (see revised ms). The comparison to the chessboard pattern simulation with 1 out of 4 and 2 out of 4 grid cells being deforested was motivated by the similarity in the degree and spatial extent of deforestation in these simulations and *deforest-glob* (see Fig. R1)

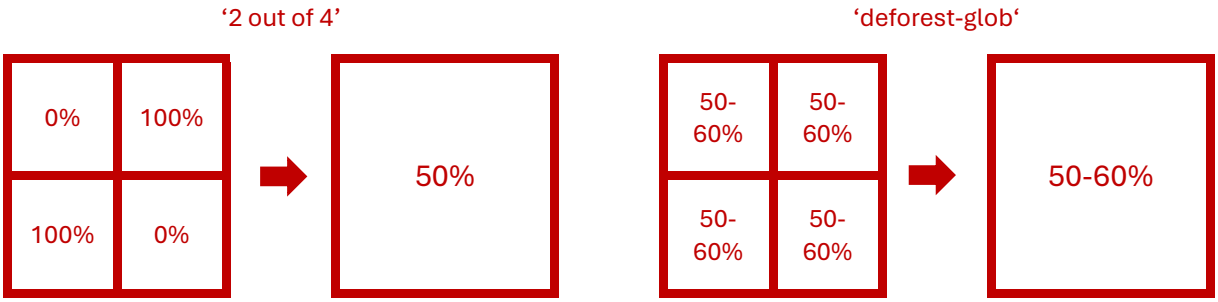

Fig. R1: Comparison of degree of deforestation in chessboard pattern simulations from Winckler et al. (2019) and *deforest-glob*.

The comparison between local temperature sensitivities due to deforestation based on the two different separation methods do show some deviations including in *deforest-glob* a stronger cooling over large portions of Northern Eurasia and a stronger warming over central portions of the Amazon and African rainforests (Fig. A6 in ms). But the *deforest-glob* simulation also shows weaker cooling over parts of boreal North America and similarly weaker warming over northeast Brazil, southeast Africa and southeastern Asia (Fig. A6 in ms). Thus, while these results underscore uncertainties when separating local and non-local effects, they provide no evidence for systematic over- or underestimation of the local effects with our applied moving window regression separation method. To reflect on these results, we made the following changes in the revised ms in the Limitations section (Section 5):

"An additional analysis comparing temperature sensitivities to deforestation based on the linear regression method applied in this study and the chessboard pattern deforestation experiments of Winckler et al. (2019a) did reveal some differences in corresponding patterns but did not provide evidence of systematic under- or overestimation of local effects (Fig. A6)."

We would like to point out that *deforest-glob* and the chessboard experiments in Winckler et al. (2019) are different in many aspects (e.g. degree of deforestation, deforestation pattern), and a comparison of the results has, therefore, limited utility. To reflect on these aspects more clearly, we included under Limitations the following: "However, comparisons of the local biogeophysical effects among different deforestation scenarios – even within the same model framework - are challenging as local effects are influenced by the degree of deforestation (e.g., partial or complete), and the initial forest cover (Li et al., 2016; Winckler et al., 2017)."

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.

We agree with the reviewer, that a discussion on the implications of overestimation of local effects has not been included adequately. In the case of an overestimation of the local effect due to applying linear scaling (as done here), the validity of the emergent constraints – i.e., the linear relationship between local and total temperature sensitivities – is not affected as the same scaling approach is applied to each of the ESMs. If scaling is applied to observations as well (as in the case of MODIS observations) than the proximity between the ESM-based local effects and observational constraints would shift. We illustrated this in Fig. R2.

If a systematic overestimation or underestimation of the local effect is caused by the method used for separating local and non-local effects in ESMs, the inter-model comparison and emergent constraint would be still valid but the proximity to observed constraints would be affected. For example, if the local ESM temperature sensitivities in the boreal regions would be systematically overestimated or underestimated, their proximity to observed local temperature sensitivities would diminish or become larger (after bias correction), respectively. This is because the ESM based local sensitivities show enhanced cooling (see Fig. R2). Yet, we find no evidence of systematic biases of the local temperature sensitivities in ESMs based on our choice of separation method (see above).

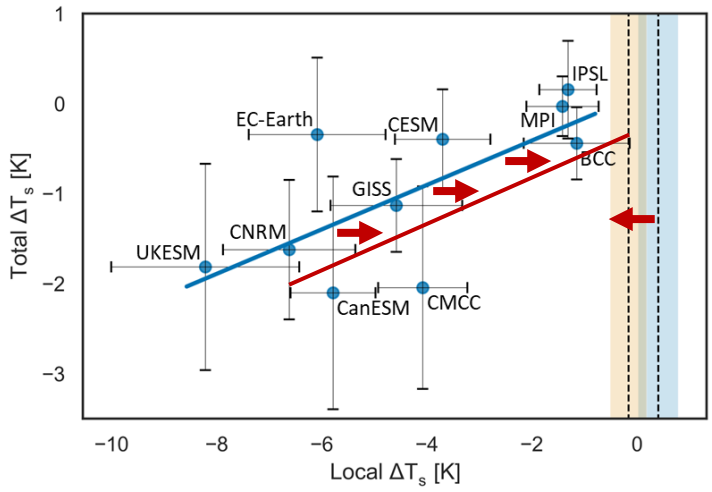

Fig. R2: Hypothetical effect of overestimation of the local temperature sensitivity in the boreal region due to scaling (see Fig. 8a in ms). The observed constraints are highlighted in orange (Bright et al., 2017) and blue (MODIS).

To reflect on those points, we have added the following paragraph under Limitations: “Importantly, potential uncertainties associated with scaling or method of separating local and non-local effects do not alter the relationship between ESM-based total and local temperature sensitivities (Fig. 8 in ms) - as all models experience the same ‘bias’ - and therefore the slope of the linear relationship central to the emergent constraint concept is not affected. However, a

possible under- or overestimation of the ESM-based local effects could lead to shifts in the proximity to observational constraints."

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).

Changed to "(i.e. resulting from the linear relationships between local albedo and surface temperature 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.

We highlight the emerging relationship between local and total effects in the end of the first paragraph: "By knowing the local effects from observations and the emerging relationship between local and total effects from model experiments, we could use these relationships to constrain the total effects of deforestation on the near-surface climate."

To put more emphasize on this aspect, we have changed lines 103-105 as follow: "The objective is to provide observationally based emergent constraints for local surface temperature, albedo and latent heat flux sensitivities to deforestation, against which both the local and total (local and non-local) ESM based sensitivities can be compared."

Line 178: extra comma

Comma removed.

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.

Lines 230-232 rephrased as follow: "By showing the linear relationship between local and total surface temperature change due to deforestation, we are able to constrain the overall response of ESMs to deforestation, for which no observations are available, as the range of plausible local surface temperature sensitivities is narrowed down by observations and total surface temperature responses are related to those by statistically and physically meaningful relationship.". This paragraph is moved up to improve the flow of the text.

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.

Added at line 333: "As the local cooling effect of deforestation is strongest in the colder months, it is important to consider whether the overestimation can be observed also during the warmer months, when the albedo effects are not that pronounced."

Added at line 339: "confirming that the overestimation of cooling (Fig. 8) persists also during summer."

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

The statement will still be true because the spread between the models (which is independent of a potential overestimation bias as discussed above) is too large, so it is not possible that all ESMs fall within the boundaries of the observational constraints.

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

Fixed in revised manuscript.

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

Updated in revised manuscript.

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?

In Fig. 5 we show the annual temperature effect, while in Fig. 6 only March and April are considered, hence the difference in the MODIS-based temperature. The figure is updated in the revised manuscript.

Fig 7: Same comment as Fig 5.

Updated in revised manuscript.

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.

Updated in revised manuscript.

Fig A4: Same comment as Fig 5.

Updated in revised manuscript.

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

Typo fixed.

Literature

Li, Y., De Noblet-Ducoudré, N., Davin, E. L., Motesharrei, S., Zeng, N., Li, S., & Kalnay, E. (2016). The role of spatial scale and background climate in the latitudinal temperature response to deforestation. *Earth System Dynamics*, *7*(1), 167-181.

Winckler, J., Reick, C. H., & Pongratz, J. (2017). Why does the locally induced temperature response to land cover change differ across scenarios?. *Geophysical Research Letters*, *44*(8), 3833-3840.