Dear Referee,

thank you for providing us with this constructive review.

This study reports on the multi-model LUMIP deforestation experiment. The authors show that controlled, large-scale, global deforestation may contribute global geophysical cooling of near-surface temperatures and global geochemical warming. The geo- physical effects vary latitudinally and by model, generally with warming in the tropics and cooling elsewhere, while the geochemical effects are estimated offline as warming everywhere. The geochemical effects generally are greater than the geophysical effects, leading to net warming, although a potential CO2-enhanced land sink is not included here. Land carbon losses are driven by vegetation loss. Some novel metrics for assessing and potentially estimating the effects of deforestation are also presented.
I appreciate the tremendous effort the authors have put into this study to advance our understanding of the effects of land cover change on the earth system. I have a few main comments, followed by some brief detailed comments.

Thank you for your acknowledgement of our work and your valuable feedback to our manuscript. In the following I will address your points and how we want to improve the manuscript.

1. Improving the abstract:
What are the take home messages? There are a lot of different analyses, and only one aspect is highlighted in the abstract. The abstract includes some key numbers, and at least the potential net warming statement of large scale deforestation. While this in itself is a key finding (with the associated caveat of constant CO2), there are a few other notable results to highlight. Missing are the policy/scenario implications related to time/fraction of emergence for climate vs carbon. And the potential for rough estimates of response via the sensitivity metrics (which are analogous to climate sensitivity of models). I understand that there are some limitations to the sensitivity metrics and the time of emergence estimates, but based on S13 and S18 (plus the rest of the carbon figures), it seems safe to say that the climate signals have relatively long time frames while carbon signals have relatively short time frames. However, the climate signal emergence is further complicated by observations that show large, immediate meteorological distinctions between forest and grassland. The sensitivity metrics sup- port the temperature and carbon results, and are potentially useful to the community.

Thank you for pointing this obvious finding out! We will follow your suggestions and make the abstract stronger and more significant by including these results.

2. Swap figures
In relation to comment (1) above, switching some of the regular and supplemental figures would make the paper stronger. For example, figures S22 (carbon sensitivity to deforestation) and S18 (ToE and FoE) are more relevant to the carbon points than figures 9-10, which are explanatory. Also, figure S3 is much clearer and easier to understand than figure 3, and follows the text better (you can add Tas to fig S3).

We agree that some of the figures have the potential to be shown in the main text. However, we would like to keep the number of figures as it is now and the decision for one or the other figure is difficult. We would like to keep the GPP and carbon time series figures in the main text as they show fundamental results. The decision for the surface energy balance decomposition figure has also more sides to it. We argue that the model-wise comparison in the manuscript offers an easier access to the model's performance to

simulate T_{surf}. On the other hand, a component-wise presentation provides a better inter-model comparison which might be more valuable to the reader. We therefore agree to swap Figures 3 and S3.

3. Temperature definitions

The descriptions of Tsurf-model and Tas are not complete, which makes it difficult to properly assess the temperature comparisons. While Tsurf is clearly a radiative temperature, is Tsurf-model a radiative temperature for all the models? Tsurf in some models is a canopy air temperature, at a height dictated by the displacement height and aerodynamic roughness. The 2m air temperature is often the air temperature 2m above this "Tsurf." It is important to be clear as to what and where these temperatures actually are, as shown by some of your references.

Thank you for this inquiry. We will add more specifications to the method and results section on where surface temperature is calculated in the models and what this implies.

4. MIROC

MIROC does not seem to meet the deforestation harmonization requirements, and its plots don't seem to add to the understanding of the issue. In fact, a lot of extra text is dedicated to explaining why MIROC is different from the others. It would be cleaner if it were not included.

We would like to keep MIROC in the analysis. We list the caveats of each model's execution of the experiment and highlight if this is the reason for an observed result. This in itself is a demonstration of the difficulties in carrying out harmonized land-use and land cover change-related studies with many participating models. We think that it is still interesting how strong regrowth in this model is and how small the biogeophysical responses are despite a clear disturbance. Furthermore, MIROC simulates interesting carbon dynamics and should therefore be accounted for. We argue that it would be more inclusive to leave MIROC in all analyses than to keep it in only some relevant ones (e.g. the carbon analysis).

We will add 'We nevertheless analyse results from MIROC to not only demonstrate the effect these different technical realizations of one scenario can have but to also to draw conclusions for improvements in this model.'

5) Some supplemental figures are cited out of order.
We will fix that. Thank you for your close observation.

Thank you for your further comments to improve the readability and understanding of our study. We will take them into account during the revision (see the detailed list below).

With kind regards,
Lena Boysen & co-authors

Lines 229-227: With respect to the effects of differences in initial forest cover on the implementation of deforestation (lines 229-227), the author's may be interested in this recently published paper: A.V. Di Vittorio, X. Shi, B. Bond-Lamberty, K. Calvin, A. Jones, 2020, "Initial land use/cover distribution substantially affects global carbon and local temperature pro- jections in the integrated Earth system model", Global Biogeochemical Cycles. doi: 10.1029/2019GB006383.
Reply: We have already cited this study in line 236.

Lines 326-327: his isn't clear from fig 3. fig s4 is more appropriate here.
Reply: Yes, we will add 'and Fig. S4c'.

Line 465: which figures show these regional effects?
Reply: We will add '(not shown)'.

Line 505: relate toe and implications to observation of immediate temperature differences between forest and grassland, and perceived differences
Reply: Immediate effects of deforestation are difficult to capture as variability makes it hard to pin them down. Even the time series of temperature have a 30-year moving average applied and only start after 15 years. We are therefore afraid to not be able to meet this request.

Line 536: this paragraph is out of place - it doesn't relate to the rest of the section
Reply: We will shift the paragraph to the beginning of this section.

Lines 644-646: fig S19 shows declines in GPP for CESM throughout the deforestation area, so it isn't clear how CESM has increases in GPP where the other model have decreases.
Reply: Fig S19 shows the ToE for GPP. We will add, that changes in GPP are seen in 'Fig. 9'. Relating these to Fig. S19 makes sense.

Lines 680-684: based on fig S20, it doesn't appear the MIROC can have the highest sensitivity. most of its coverage has the smallest change in c per fraction of deforestation.
Reply: Only regionally, MIROC reaches high carbon losses per deforestation fraction. Globally, MIROC is at the low end across models. We will add '…in boreal North America.'

Lines 710-713: this is unclear - you have separated your total range in two, arbitrarily, and included veg only change models in one group with other total land c change models.
Reply: We will delete 'The remaining models would yield a warming of between 0.24 to 0.87 °C.' and changed the previous sentence to 'For MIROC, IPSL and MPI (0.18 to 0.57 °C) this is the main temperature response to ΔF (with only non-significant BGP-induced effects)…'