We thank both anonymous reviewers for their time and effort in reviewing our article. We respond to the comments of Reviewer 3 below.

Reviewer 3

In the following I briefly evaluate the responses separately for each of the reviewers comments. In my impression only for comment 3 it is not clear if the authors addressed the reviewers concerns appropriately.

1. I find the section 2.1.3 now clearly written even though I can't judge how reasonable the switching off of soil evaporation is. In general, I find the paper mostly clearly written.
2. The model is now well described, the motivation for its use is stated and the limitations are mentioned.
3. If I understand the comment of Reviewer 1 correctly, the authors are asked to compare their modeled precipitation to observations. As the starting date of the runs is 1st June 2014, rainfall should be compared to measurements during 1st – 5th June. The authors elaborate on why they can't use observations directly but I'm not 100% convinced. If I understand correctly, the "current" forecast is more or less a "real" forecast for 1st – 5th June 2014. Comparing this output to observations of that period should be possible and I think important to do in this study. Of course the focus of this study is not on forecast validation but on the effect of deforestation. Nevertheless, a comparison to observations would provide important context.

In our manuscript we referred to our previous paper (Crook et al. 2019) that compared this model with precipitation observations for the whole of June and July. However, we accept that a direct comparison for just the 1st 5 days of June 2014 can be made, although our simulations are not a forecast, and have now added this comparison in section 2.1 with an extra figure in the supplementary material. The supplementary figure illustrates that the modelled 5-day total rainfall does indeed show biases compared to CMORPH for certain regions, most pronounced along coastlines and topography. However, systematic regional biases that affect both our forested and deforested simulations equally will not affect the rainfall change signal linked to deforestation that we are interested in. Given this is a process study, it is the model skill in correctly capturing rainfall timing within the diurnal cycle and in representing the characteristics of convective storms (as demonstrated in Crook et al 2019, simulation V_CP4 therein) that is most important for this work. The realistic representation in timing, storm lifetimes and storm precipitation intensities provides confidence in our model results on convection responses when surface roughness and flux patterns change locally due to deforestation.

4. The authors well explain their rationale to selected the two study regions
5. Ok
6. Ok
7. Ok

Comment
I appreciate that the authors used a statistical significance test. However, there needs to be a clearer description of the method used. Importantly, did the authors account for multiple testing, i.e. was the false discovery rate controlled (Wilks, 2016)? If not, this is an issue because significance could just emerge by chance when so many tests (over the whole study domain) are done. This correction could be done, e.g. with a Benjamini-Hochberg correction (for python, see https://www.statsmodels.org/dev/generated/statsmodels.stats.multitest.multipletests.html). Also this correction could help remove some of the patchiness of some of the plots such that there is more focus on the dominant differences.

*We had not previously accounted for this. We would like to point out that our data is high resolution with N=250,478 grid points in the maps in our figures. This imposes a very strict limit on the pvalues and as a result applying this to patchy fields such as rainrate, convergence, delta θ and SW, results in not being able to find any sorted pvalues that meet the FDR test < $alpha_{FDR}$ x i/N (where $alpha_{FDR}$ is alpha*2) and therefore we cannot calculate a new alpha to use to test for significance. If we use a much-reduced region where changes have occurred (e.g. using the grid points where the individual T test pvalue<=alpha) we get all these same grid points being significant when using the FDR corrected test. For several of our variables assessed we would not expect a change over the whole region and therefore determining field significance does not seem relevant. We have, therefore, used the Benjamini-Hochberg FDR correction for maps of all variables that are not patchy (i.e. do change over a large part of the domain) and have left the plots of other variables showing the individual test significance. We have highlighted in the text where this is the case.*

Minor comments

L19 “we for the first time estimate” → I’m not a native speaker but to my ears it sounds better to say “we estimate for the first time”, or maybe remove “for the first time” completely

Corrected.

L25 “thermally induced enhanced” → “thermally enhanced”

Corrected.

L43 Unclear if you only talk about the biogeophysical changes or also the biogeochemical changes. With regard to the former it seems more precise to say “local (surface) warming”

We have added “local (near surface)”.

L79: They found **that** the enhanced (?)

Corrected.

L156: real life → reality

Corrected.

L204: Figure 2, which indicates the simulated region, shows maps of …→ Figure 2 shows maps of (...) in the target region

Corrected.

L206ff: I suggest revising this section and potentially split it in two. The first paragraph is about the statistical significance test so it could have its own section named “Statistical significance test”. The second paragraph is in principle only about the criterion for the definition of deforestation. This topic already appears in Fig 2g which is referred to in the previous section (section 2.2). Hence, the description of the criterion for deforestation could be simply added in section 2.2. In my opinion, there is no need to describe that you look at two focus regions or that you first compare albedo etc. This could be part of an introductory paragraph to the results section and not the methods section. I’m aware this is also a bit a matter of taste but I just feel it improves readability.

We agree with the reviewer that this section can be modified as suggested. Section 2.3 is now called Statistical Significance Tests and we have added information about the FDR correction procedure. The sentences regarding what results we show have been moved to the start of the Results section and the paragraph about the definition of what is counted as deforestation has been moved the end of the previous section.

Further, the first paragraph of section 2.3 suggests that for albedo, surface roughness, and initial soil moisture you also use a T-test to assess statistical significance. However, it is unclear to me how this can be done given that (if I understand the methodology correctly) for these variables you only have one field for 1950 and one for current condition (i.e. no ensemble members). If you don’t use a T-test, then there is no need to mention these variables in this section.

Albedo is calculated as the ratio of outgoing SW/ incoming SW at 13:00 UTC on each day for each ensemble member. Roughness length is output by the model daily and therefore can be determined on each day for each ensemble member. The initial soil moisture stress factor (FSMC) is calculated using the soil moisture on the first day of each ensemble member. Therefore, this did allow us to theoretically determine statistical significance. However, given that we would expect virtually the same albedo and roughness length throughout the simulation, and the same value of initial soil moisture in all ensemble members, the statistical significance is not very meaningful, and we now present these as simple differences with no mention of T tests.

L235: highFSMC → high FSMC

Corrected.

L244: Detail but I find LH as abbreviation of latent heat flux more intuitive. This would also be consistent with the abbreviation of sensible heat flux (SH)

Changed LE to LH throughout.

L250: strong,radiative → strong, radiative

Corrected.

L262ff: Sentence structure is unclear. Why does net downward long-wave radiative fluxes decrease? And what does this have to do with reduced roughness length? Is it that reduced roughness length warms the near surface (as a result of reduced heat land-atmosphere heat fluxes) which, as near surface temperature rises, leads to larger upward long-wave radiative fluxes? This will then, when downward long-wave remains constant, lead to a decrease in *net* downward long-wave radiative fluxes. If this is how we need to think about it I would appreciate a bit a clearer explanation here.

Yes, you are almost correct. The reduced roughness length reduces turbulent fluxes (land to atmosphere fluxes), warming the surface and increasing LWu which decreases LW with no change in LWd. There is also reduced cloud cover which decreases LWd so that also has some effect on decreasing LW. We have rewritten these sentences.

L268: may dominate. , → remove period

Corrected.

L271: increases long-wave emission: not sure if it is clear that this refers to long-wave emission by the surface, and not by the atmosphere. Maybe a clarification would help.

We have added "from the surface".

L360: these regions. → period missing

Corrected.

L375: whenoceanic → when oceanic

Corrected.

L377ff: I find the last part of this sentence not very clear. Maybe try: "to show that to understand rainfall changes it is crucial to analyze how deforestation affects dynamics and thermodynamics"

Corrected as suggested.

L382 and elsewhere: To improve readability, I suggest to use approximately or approx. instead of ~

We have changed all occurrences of "~" to either "approx." or "around" as suggested.

L384: "The regions of positive convergence coincide with the high rainfall patterns." Not sure if I agree. If I compare Fig. 8a and 8c I see that convergence and rainfall coincide

sometimes but not always/everywhere. Do you refer only to a certain part of the plot? If yes, it would be helpful if it was specified which part.

We agree that not all positive rainfall and convergence changes coincide. We have modified this to say "Although the convergence field is noisy, positive rainfall changes at 8-10° N tend to occur where convergence increased after deforestation."