

Apologies for the amount of time it has taken to do this review. I did warn the editors that it would take a while, but I'm not sure any of us realised quite how long.

The authors have made detailed and convincing responses to my initial review and have made extensive revisions to the m/s. I won't go over their responses unless it related to changes still required to the m/s, but I would like to point out that they really were very good. In the paper itself, the authors do a much better job of explaining the motivation of the work, the model setup and evaluation methods and how their results link to future development priorities. The introduction could still be slightly clearer on what is actually being turned on and evaluated in the model (covered in minor comments), I am still not completely satisfied with the author response and changes with regard to scale dependency of the land surface, and I think there is a slight mismatch between some of the figures and text which might need a bit of clearing up. But these should only require slight changes to the manuscript, so I have marked this up as minor revisions.

There are also a few small corrections and specific comments.

Goodluck with future model development and (potential) CMIP7 involvement!

### **Small but not quite minor comments**

#### ***Land surface resolution dependence***

From the author's response:

*There is no inter-gridcell communication in LPJ-GUESS (ie it can be considered as a 'site model' that simulates an arbitrary list of sites) so spatial resolution does not directly affect the processes.*

You do not need inter-gridcell communication for resolution effects to be important. Many of the processes simulated by LPJ-GUESS are extremely non-linear, so simply aggregating inputs over larger scales could affect the quality of performance of LPJ-GUESS, even if using a fantastic coarse resolution model or perfect observational driving data. Infact the authors make this very point when considering likely changes in scores expected at different resolution due to homogenization of LPJ-GUESS output and observations on page 20, lines 14-16.

An example I think some of the authors might know about is fire - on fine scales you would need a model that includes fine scale processes (rate of spread etc), whereas on a coarse resolution, rate of spread becomes less important and a control based model (such as BLAZE suggested towards the end) is much easier to parameterise for broadscale controls (Burton, 2019). Soils are perhaps another example, and I think somewhere buried in soil is information on the impact of averaging soils on model hydrology.

However, in the revised m/s, there are not only a couple of instances where this is a potential problem, but these should either be changed or removed. Specifically:

Page 3 line 31: The processes could also be resolution dependent.

Page 6, line 26-29: two sentences starting with "As LPJ-GUESS..." The logic here is probably wrong. But the extra resolution in the revised m/s, and climate plots in the appendix, probably provide the tools you need to attribute between climate biases and climate aggregating. And I \*think\* it will read okay if you add "... (i.e. climate biases **or climate aggregating**)" at the end.

Page 13, line 15/16: “*indicating that this discrepancy is caused by biases in the EMAC climate at low resolution.*” If you can demonstrate this by examination of the climate plots in the appendix, then this is fine. If not, add the caveat about climate averaging again. I \*think\* you manage this a couple of sentences later so it might just involve some rearranging of these few sentences.

That's all I can spot. Other instances, Page 13, line 19/20 , for example, are fine as you have backed it up with previous assessment of climate simulation at different resolutions.

### ***Biome and tree cover bias attribution***

Specifically page 12, line 24-26.

Temperate tree cover in the EMAC simulations actually look more extensive than observations in Figure 4, though still less than the CRUNCEP simulations (Fig 4). This first sentence should be rewritten to make this clear. GPP does look slightly less than obs in some regions (Fig 3), but the difference between CRUNCEP and EMAC GPP is a lot more clear cut, so maybe use that again.

### **Minor/specific comments**

Page 1 line 7: The sentence starting “The LPJ-GUESS...” should provide a bit more detail about the processes that have been enabled and evaluated in this study. Something at the level of detail as the lines of the next sentence, which lists things not enabled.

Page 1, line 17: replace the average NME score with all three scores. Averaging across benchmarks is a bit of a controversial issue (see (Blyth et al., 2010; Kelley et al., 2013; Randerson et al., 2009)), so is best avoided seeing. There's only 3 numbers you'd need to quote so that should be fine in the abstract.

Page 2, line 10: remove “simulations”

Page 2, line 20-21: mentioning fire and phosphorous here makes it sound like your going to include them in the model. Both are future developments though (I know GlobFIRMs in the model, but you later point out that this needs replacement). So I'd leave them out of the introduction and just provide a note about how important they are in the discussion (either future work or conclusions).

Page 3, line 1: It seems like there's an “in” missing. I.e., “has already been used **in** both a global ESM”. Or maybe it's “has already been used both **in** a global ESM”. (My grammar isn't great).

Page 3, line 15: replace “both” with “all” (there are more than two ESM components and each probably claims it's own community ;))

Page 3, line 15: The sentence beginning “When development is complete....” you should mention before this sentence that you are just focussed on one-way coupling.

Page 3, line 17: “Then the full model will become a powerful tool”, slightly more cautionary language would be better. Maybe replace “will” with “should” or “We aim for the full model to become....”

Page 5 line 20: add something like “In this study...” at the start of the paragraph, just to make it clear that you are inputting on a daily timestep for this study, rather than the monthly timestep mentioned in the following sentence

Page 5 line 22: Sentence starting “In these circumstances...” Does the disaggregator involve some stochastic implementation as well? If so, say so, and as you have an extra stochastic process, maybe move the two sentences starting “All stochastic processes ...” on line 12 to a new paragraph just before 2.1.3. If not, don’t worry.

Page 6 line 6/7: Sentence starting “However in both these cases...” where does the soil moisture come from in this study? Was it simulated in LPJ-GUESS?

Page 6, line 29: Should “T63” now be replaced with “T89”?

Page 7 line 5: Should “1990-1990” be “1990-1999”?

Page 8, line 25: Figure 1 in the author responses is a good way of showing that the parts of the model being accessed here is in equilibrium -potentially quite important to demonstrate given the short spin up time. The authors may want to consider including it as a supplement or appendix figures?

Page 8 line 16: should there be an “a” before “function”?

Page 14, line 4: refer to figure 3 and 4. Also, I’m not sure I see the overestimation of either GPP or biomass. Maybe there is some work needed on the colour bar of figures 3 and 4?

Page 14 line 22/23: “low competitiveness of grass PFTs vs tree PFTs” could also be due to other processes affecting competition, and not just fire frequency? E.g soil moisture, simplistic soil depth, drought response, PFT heretical setup, establishment rates etc. And I don’t think you’ve offered evidence to suggest that it is caused fire frequency or why fire frequency should be singled out?

Page 14 line 26-28: I’m not sure I follow this. It sounds like NME scores in this study are comparable to scores in Kelley et al. 2013, which the text implies included LPJ-GUESS? A pre-GUESS version of LPJ was used in Kelley et al. 2013, so if I read this correctly, then the sentence needs adapting. If I didn’t read this correctly, the sentence could do with some clarification.

Page 14, lines 30 - Page 19 line 3: It feels like this couple of sentences should be moved up to just after the first sentence of this section.

Page 15, Figure 3: The colour scale on the top 5 figures could be altered (i.e not linear) to make spatial patterns of GPP in EMAC maps clearer.

Page 19, line 4: The sentence starting “In summary” feels better placed at the end of the section after discussion of metric scores accounting for land use, or maybe even removed as this is covered in the discussion.

Page 19, line 21: You’ll have to explain a bit better how the sentence before suggests that the disturbance rate in particular needs re-evaluating.

Page 20, line 31: I think there’s an “and” missing in “... fluxes), **and** to form...”

Page 21 line 14: I think it is worth briefly explaining why BLAZE was selected, rather than another model such as SPITFIRE or a re-parameterised GLOBFIRM.

Page 22 line 26: replace “will” with “should”

Page 23 line 6: “A future publication will present...” should be replaced with something like “Future development should focus on...”, so as not to pre-empt what journals might publish.

Page 23 line 8 onwards: That’s quite a nice way to finish.

Page 27, Figure B1: Swap the colours around so blue means more precip.

Page 29, Figure B3 caption. In not sure I understand, particularly the sentence starting “Note that these plots compare shows the radiation available..”. Are all plots adjusted by 0.17? And what is meant by “adjusted”

## References

Blyth, E., Gash, J., Lloyd, A., Pryor, M., Weedon, G. P. and Shuttleworth, J.: Evaluating the JULES Land Surface Model Energy Fluxes Using FLUXNET Data, *J. Hydrometeorol.*, 11(2), 509–519, 2010.

Burton, C.: Impacts of fire, climate and land-use change on terrestrial ecosystems, University of Exeter, 1 April. [online] Available from: <https://ore.exeter.ac.uk/repository/handle/10871/36801> (Accessed 20 December 2019), 2019.

Kelley, D. I., Prentice, I. C., Harrison, S. P., Wang, H., Simard, M., Fisher, J. B., Willis, K. O. and Others: A comprehensive benchmarking system for evaluating global vegetation models, [online] Available from: <https://www.researchonline.mq.edu.au/vital/access/services/Download/mq:26863/DS01>, 2013.

Randerson, J. T., Hoffman, F. M., Thornton, P. E., Mahowald, N. M., Lindsay, K., Lee, Y.-H., Nevison, C. D., Doney, S. C., Bonan, G., Stöckli, R. and Others: Systematic assessment of terrestrial biogeochemistry in coupled climate--carbon models, *Glob. Chang. Biol.*, 15(10), 2462–2484, 2009.