Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2019-42-RC2, 2019

Biogeosciences Discussions

Interactive comment on "Sensitivity of simulated historical burned area to environmental andanthropogenic controls: A comparison of seven fire models" *by* Lina Teckentrup et al.

Anonymous Referee #2

Received and published: 7 April 2019

General comments

The study is a useful compilation of the analysis of sensitivity experiments in the FireMIP output, but it is largely a technical report of the sensitivity of FireMIP model simulations of burned area since 1900. Philosophically, there is nothing really offered by the authors in terms of specific testing of improvements/changes needed with fire models beyond what has been pointed out in the literature in papers such as Van Marle et al 2017 and Andela et al 2017, and hinted at in the Hantson et al 2016 FireMIP overview paper and the Forkel et al 2019 paper.

While I appreciate the depth of the dissection of the causes for the discrepancies

among FireMIP models in this study, I find myself with no questions about FireMIP that have new or interesting answers, which is a concerning lack of momentum from the initially promising FireMIP effort. For example, did the FireMIP sensitivity experiments produce knowledge that the modeling groups could leverage for specific technical advances on, say, a future set of experiments? If anything, this paper makes me increasingly skeptical about the utility of FireMIP other than to show precisely what these authors stated in their conclusions: "Although burned area in most models compares reasonably well with satellite observations, there is a huge spread in transient simulations before the satellite era and a huge spread in the influence of the driving factors between models." Again, however, many FireMIP related papers have already pointed this out.

I recommend that the paper be published and I think that my comments fall somewhere between a minor and major revision, so I labeled it as minor revisions even though some of my comments might require some major discussion amongst the authors in terms of structuring a reply or rebuttal. The challenge that I offer to the authors is this: I do not see what we gain beyond now knowing that the sensitivity experiments are as confusingly inconclusive as the core experiments. If I were re-formulating my fire model and looking to this study, I would have little idea as to what the focus point should be other than simply acknowledging weaknesses such as the representation of human use of fire or needed better data for model parameterizations. The authors may need to make their case more clearly for this paper to stand out beyond being a technical report out.

Specific comments

Figures in the Supplement – please make larger versions of the maps in figures a1-a8. Another improvement would be to include a continuous rather than binary scale of values of the correlation coefficient in a2-a8. Painting the world with binary correlation coefficients would mask areas of potential weak and strong linear correlation. The strength of this study is the technical report-out of FireMIP sensitivity studies, so by making figures a1-a8 so hard to read, the authors are undermining the very purpose of the work. Read another way, the community may gain more with more detail in the manuscript.

Page 6 line 16-17 – authors stated they used a square root transformation to reduce the skewness of the distribution, but it is unclear why. Please expand on both the reasons and what this transformation accomplishes. Perhaps a supplemental figure?

Page 6 line 19 – major uncertainties is a subjective phrasing that requires more qualifications. Humber et al 2018 clearly discussed the nuanced and important ways that observed burned area data sets agree and disagree when using global, regional, and varying temporal scales. Looking at Figure 3 in Humber et al 2018 and Figure 1 in this paper, however, the implication is that FireMIP models have even more than "major" uncertainties in the sense that even at an annual time scale, there is more spread amongst models than amongst the observations. Furthermore, the three burned area data sets discussed in this study (GFED4, GFED4s, and FireCCI50) show that there is agreement unless the specific methodological approach is augmented with the small fires approach described in Randerson et al 2012. Is that really a major disagreement or just a difference in analysis? Please be more specific or careful in the discussion around observational uncertainties. Also, please see my comment about Figure 1 below.

Page 6 line 20-21 – please explain what is meant by 0.01 and 0.2%. I am not following what the the values refer to.

Figure 1 would benefit from being split into a two-part plot: one part could remain as is, but the other would show the present day subset of the full analysis period. This is the evaluation period, but it is buried under too many curves.

Table 3 and page 7 – are these spatial correlation coefficients that compare the grid cell to grid cell agreement on a map? Or are they temporal correlation coefficients? It does not seem that Figure 1 temporal correlation is this high, but please clarify in the

text. If this is a spatial correlation, please include the figure in the Appendix as it could be valuable to modelers in identifying regional weaknesses in the FireMIP simulated burned area.

Table A2 is missing statistics relative to GFED4s.

Page 9 – the first sentence on this page highlights a major problem in the approach with modeling. Aiming at trends without a full understanding of the drivers in the simulations is .

Table 4 – while the M-K test is likely fine, the uncertainties (standard error or confidence intervals) in the slopes need to be included to understand the results better.

Page 9 and Section 3.2.4 – I thought that FireMIP only used a repeated lightning scaled to changes in modeled convection? While there is likely something to gain in the lightning sensitivity experiment, I would like to see some clearer discussion of the important caveats in interpreting the results. For example, would it be safe to surmise that there is no sensitivity to lightning changes since 1900 only if the modeled lightning is anything close to reality? Determining a lightning climatology from an untestable climate-model based parameterization and then drawing conclusions from that testing is prone to some circular or flawed logic.

Figure 2 – please re-title these with something that is easier to quickly interpret without cross-referencing the table. For example, I suggest (a) Constant CO2 (SF2_CO2), (b) Constant Population (SF2_FPO), (c) Constant Land Cover (SF2_FLA), (d) Constant Lightning (SF2_FLI), (e) Constant climate (SF2_CLI). Also please make figure 2 much wider to avoid the visual clutter of overlaid zigzagging lines.

Figure 2 – change the y-axes ranges so they are constant. It is hard to understand the sensitivity if the plotted range is variable.

Page 11 line 9 – I agree that the statistics suggest individual trends are significant but this does not preclude the massive spread (both positive and negative) in the trends amongst models (table 4). I think this statement needs to include that caveat for an honest accounting of the FireMIP output.

Section 3.3 – the first paragraph makes no sense. What I am reading in this study is that the models barely agree on any trend, but yet the authors propose here that the models are important for understanding projected trends and supporting land management strategies. To me, a land management practice cannot be based on model trends that do not agree on trend and cannot be of much use if there is lack of agreement at country scales, let alone finer spatial scales.

Section 3.3, second paragraph – the results presented in the manuscript clearly show that models only agree in magnitude in the present day, but the quick microscope analysis of the present day trends show that observations and models do not agree in trends. Some models predict a positive slope, some negative. Unless the authors intend to propose that one FireMIP model is more physically realistic than another, then the results of the sensitivity studies are inconclusive.

Section 3.3 or 4 – it would be useful if these authors were to comment directly on fire models that did not contribute to FireMIP but that have contributed significantly to discussions of human-driven fire both in the present day and over the more distant past. This includes studies by Pfeiffer et al https://www.geosci-model-dev.net/6/643/2013/, Rabin et al https://www.geosci-model-dev.net/11/815/2018/, and Hantson et al https://journals.ametsoc.org/doi/full/10.1175/BAMS-D-15-00319.1 . All of these either echo or predict the results discussed by Andela et al 2017 and Bistinas et al 2014 related to a need to quantitatively represent the human use of fire on our planet in the modeling framework.

Conclusions – the conclusions are already evident in the Andela et al 2017 paper, so I do not see what we gain in this study. The authors conclude "further analyses are required to better disentangle" factors, but this is the same conclusion so many fire model and FireMIP papers have arrived at. Could the authors make a clearer argument

about what we gain in this manuscript?

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2019-42, 2019.