Author response to Sergey Venevsky's review

The authors thank Sergey Venevsky *for his constructive comments. Below we provide a detailed respond in italics to each comment and suggestion.*

The paper presents evaluation of historical simulations made for the nine FireMIP models with regard to fire and vegetation properties. This is very important and necessary inter comparison study aimed to support and move further global fire modelling activities. Both methods and results are clearly presented and scientifically sounded and proven. The only methodological weakness is a bit short period of comparison of the models with the observations which could be clearly longer. I think that Discussion in this paper is the weakest part and needs more effort to make conclusions from the model intercomparison to be stronger and more clear. In particularly 1. The part about relation of areas burnt to vegetation production (lines 387 -396 and 405-411) should describe in more details why and how exactly fuel load influence burnt areas in the FIREmip nine models 2. Part on influence of areas burnt upon GPP/NPP is absent and should be added.

The global fire models participating in FireMIP use output from the vegetation models as input (e.g. different fuel loads, fAPAR, etc.) to a range of fire processes, depending on the fire model parameterization and structure. A detailed description of the fire model structure, and he interactions with the vegetation model is outside the scope of this paper and has been described in Hantson et al. (2016) and Rabin et al. (2017). However, we acknowledge the need to indicate that this interaction is present and will now modify the text to make this clearer (line 387):

"Vegetation type and stocks are input variables for the fire models, influencing fire ignition and spread in the process-based models and determining simulated burnt area in the empirical models. The occurrence of fire can, in turn, affect the vegetation type, simulated vegetation productivity (i.e. GPP, NPP) and hence the amount and seasonality of fuel build up. Our results indicate that inter-model differences in burnt area are related to differences in simulated vegetation productivity and carbon stocks. Seasonal fuel build-up and senescence is an important driver of global burnt area. Furthermore, we find that models which are better at representing the seasonality of vegetation production are also better at representing the spatial pattern in burnt area."

Comments/questions/suggestions:

Line 66 – "Willdfires and anthropogenic fires" – how you define and classify these types of fires in global models? Further on no clarifications for this important question. . .

We make the distinction here between lightning-ignited fires and fires set by humans. Some models include human ignitions as a function of population density; one model (CLM) also includes cropland fires. The specific parameterizations included in the FireMIP models are discussed in the FireMIP protocol papers (Hantson et al., 2016; Rabin et al., 2017). Since we are not documenting the models here, but only evaluating their performance we do not think that a full discussion of these parameterizations is warranted. However, we will modify this sentence to clarify this point:

"However, the representation of both lightning-ignited fires and anthropogenic fires (including cropland fires) varies greatly in global fire models."

Line 126 Lightning data 1900-1920.population density and land use 1700 where do these data come from?

The source of the data sets is given in the FireMIP protocol, which we cite here (line 133). However, we will modify this to make it clear that the information about all the drivers of the simulation are available there, as follows: "(see Rabin et al., 2017 for description of the modelling protocol and the sources of the input data for the experiments).".

Line 132 The baseline FireMIP simulation is a transient experiment starting in 1700 CE and continuing to 2013. Why simulation is only up to 2013 and comparison is only for 2002-2012? Can simulation be somehow extended to include recent years? Similarly, inter-comparison only for decade looks not so sounded, for example 1997/1998 El Ninio years are out. . . Which climate data was used?

The simulation protocol was drawn up in 2016 and at that stage there was no readily available driving data post-2013. The choice of the interval 2002-2012 for benchmarking was motivated by two considerations. First, 2002 was the first year when both MODIS sensors were operative and hence a temporally coherent high quality global burnt area dataset is available. Before 2002, and especially before 2000, the quality of the GFED4 archive is much lower. Second, two modelling groups did not run the final year of the simulation (i.e. 2013). We therefore excluded this year in order to keep the evaluation between the models as consistent as possible. CRU-NCEP climate was used as forcing data for the simulations (see Rabin et al., 2017).

We agree that there are several other evaluations of fire models that could be made and certainly testing how well they emulate the response to El Nino variability would be an interesting case study. However, this is out of scope for the current round of FireMIP.

Lines 150-159 What was the principle of selection of all these datasets? Why no global water cycle related datasets (e.g. runoff) where selected? Water status is obviously important for both fire and vegetation, I would compare at least also runoff for 2002-2012

We agree that there are many other opportunities for evaluating fire models, and it would be possible to include data sets such as runoff. However, there are a number of issues that led us not to include runoff in our evaluation: (1) runoff data in the most fire-prone regions (e.g. the Sahel) are not very reliable; (2) gauged runoff is integrated property, and thus requires models to incorporate some form of hydrological routing scheme; and (3) it is not the most relevant hydrological variable for fire -- soil moisture or litter moisture would be useful but are very heterogenous and thus global evaluation would be difficult. We agree that we could usefully modify this sentence to explain our current focus and will modify the text to read:
"Model performance was evaluated using site-based and remotely sensed global data sets of fire occurrence, fire-related emissions and vegetation properties (Figure 1; Figure S1). We include vegetation variables (e.g. GPP, NPP, biomass, LAI) because previous analyses have indicated that they are critical for simulating fire occurrence and behaviour (Forkel et al., 2019a; Teckentrup et al.,2019) and there are global data sets available. We did not consider parameters such as soil or litter moisture because, although these may have an important influence on fire behaviour, globally comprehensive data sets are not available."

Line 173 "As model benchmarking techniques become more sophisticated it would be beneficial to better evaluate the datasets the models are compared against to ensure the models are being benchmarked appropriately" Please, delete or rephrase (shorten)

We have substantially shortened and modified this paragraph based on your feedback as well as the comments by the first reviewer (see response to reviewer 1 for detailed information regarding the changes made).

Line 175-180. I think you should move formulae of NME from Supplementary back to the main text. You write in Supplementary that you applied NME for areas burnt, but it is clear from Table 3 and S1 that you apply the same metrics for other variables for benchmarking, please, correct.

Agreed. We have put the formula of NME to the main manuscript. Additionally, we have substantially rewritten the section on the benchmarking metrics to make the procedure clearer, including clarifying that NME is used for all the spatial variables.

As well Table 3 is quite difficult to read, why not to use semaphore colors (not so good- red, OK –yellow, good –green) or any other color scheme?

We thank the reviewer for his suggestion. As the objective of this table is to present the numerical scores so that readers will be able to judge differences between the models based on the actual scores, we keep these in the table as well. However, we have now added background colors to identify large-scale differences between the scores. We will modify the caption to this table as follow: "Cell are coloured blue if the benchmarking score is lower than both null models, yellow if lower than 1 null model and red when higher than both null models.".

Figure 1p. Performance in fire emissions by the models is the worst from all variables. How you can explain it? How reliable is observation data set? Please, make more explanation in Discussion part.

We discuss why the scores for emissions are worse than the scores for other aspects of fire already in the final paragraph of the Discussion. As we state there, errors in the emissions are a product of errors in burnt area, simulated biomass and combustion completeness. It is therefore natural that they are more difficult to predict accurately.

Line 219 CLM (NME: 0.63-0.80) and ORCHIDEE-SPITFIRE (0.70-0.73) are the best performing models. What makes these models to be the best in burnt areas description (for CLM as I understood it is related to cropland fires, what about ORCHIDEE-SPITFIRE ?

We focus on model structure and processes in this manuscript, and only focus on an individual model if it is the only one representing a certain process (e.g. CLM and cropland fires). In this case, both models prescribe vegetation cover, and we have shown that vegetation is one of the variables influencing performance in simulated burnt area. In fact, both models have the best scores at simulated vegetation carbon.

Line 326 to 334 ". . . the overall difference between the models (. . .JSBACH, LPJ-GUESS and ORCHIDEE..) reflect feedbacks between the fire and vegetation modules " what are these feedbacks? Where lays difference in their descriptions of these three models (in DGVMs)?

Fires in DGVMs combust biomass & grass and kill trees, hence exerting a strong impact on vegetation dynamics, which themselves drive fire occurrence and characteristics. The three models mentioned are completely different vegetation models. While fire can impact a large range of processes in LPJ-GUESS, ranging from carbon stocks, over distribution and structure. This is less the case for JSBACH and ORCHIDEE as these do not represent vegetation structure and *prescribe vegetation distribution. We provide a very detailed description of each fire model in the FireMIP protocol paper (Rabin et al., 2017) and we also provide the key reference providing the description of each vegetation model. We feel that the description of the structure of each vegetation model is outside the scope of this manuscript.*

Lines 345-349 "there is a positive relationship between simulated burnt area scores and the seasonal concentration of GPP (R2 = 0.30-0.84) and, to a lesser extent, the seasonal phase of GPP (R2 = 0.09-0.24). This supports the idea that seasonal vegetation production and senescence, which have an important influence on fuel loads, drive the interactions between vegetation and fire within each model" – I doubt this statement. It is more likely that similar dynamics of burnt areas and the seasonal concentration of GPP/ the seasonal phase of GPP are related to dependence of both areas burnt and GPP variables from soil moisture in the fire models and DGVMs. Please, either prove, or delete this paragraph.

We agree that soil moisture, vegetation productivity and fire occurrence are strongly linked to each other and it could be difficult to separate their individual influences. Some recent fire models explicitly represent fuel moisture (e.g. SPITFIRE; Thonicke et al., 2010), with the objective of diagnosing the role of fuel moisture on fire spread. However, soil moisture is only one of the multiple variables which drive fuel moisture, which results in a partial disconnection between soil moisture and fuel moisture in the models. Furthermore, the influence of differences in soil moisture dynamics between models is likely to be small in our experiments because the climate inputs controlling this (precipitation, temperature) were specified to be the same. Thus, we doubt that this feature is solely induced by soil moisture. Furthermore, in addition to this relationship between seasonal vegetation production and burnt area, we provide multiple other indicators that vegetation status impacts the performance of the fire module. This was also a conclusion from a previous study on FireMIP outputs (Forkel et al., 2019). Hence, we believe that there is enough evidence to support our statement that our results stress the importance of the interactions between vegetation and fire within each model.

References

Forkel, M., Andela, N., Harrison, S. P., Lasslop, G., van Marle, M., Chuvieco, E., Dorigo, W., Forrest, M., Hantson, S., Heil, A., Li, F., Melton, J., Sitch, S., Yue, C., and Arneth, A.: Emergent relationships with respect to burned area in global satellite observations and fire-enabled vegetation models, Biogeosciences, 16, 57-76, 10.5194/bg-16-57-2019, 2019.

Hantson, S., Arneth, A., Harrison, S. P., Kelley, D. I., Prentice, I. C., Rabin, S. S., Archibald, S., Mouillot, F., Arnold, S. R., Artaxo, P., Bachelet, D., Ciais, P., Forrest, M., Friedlingstein, P., Hickler, T., Kaplan, J. O., Kloster, S., Knorr, W., Lasslop, G., Li, F., Mangeon, S., Melton, J. R., Meyn, A., Sitch, S., Spessa, A., van der Werf, G. R., Voulgarakis, A., and Yue, C.: The status and challenge of global fire modelling, Biogeosciences, 13, 3359-3375, 10.5194/bg-13-3359-2016, 2016.

Rabin, S. S., Melton, J. R., Lasslop, G., Bachelet, D., Forrest, M., Hantson, S., Kaplan, J. O., Li, F., Mangeon, S., Ward, D. S., Yue, C., Arora, V. K., Hickler, T., Kloster, S., Knorr, W., Nieradzik, L., Spessa, A., Folberth, G. A., Sheehan, T., Voulgarakis, A., Kelley, D. I., Prentice, I. C., Sitch, S., Harrison, S., and Arneth, A.: The Fire Modeling Intercomparison

Project (FireMIP), phase 1: experimental and analytical protocols with detailed model descriptions, Geosci. Model Dev., 10, 1175-1197, 10.5194/gmd-10-1175-2017, 2017.

Thonicke, K., Spessa, A., Prentice, I. C., Harrison, S. P., Dong, L., and Carmona- Moreno, C.: The influence of vegetation, fire spread and fire behaviour on biomass burning and trace gas emissions: results from a process-based model, Biogeosciences, 7, 1991-2011, 10.5194/bg-7-1991-2010, 2010.