This paper describes and evaluates the development and optimization of a global fire model. A key improvement compared to current approaches is a consistent approach to model multi-day fires. The paper also contributes to the understanding of fire drivers. Therefore this work can be considered important and is suitable for publication. The methods are sound and the paper is well written. The second part of the discussion give to much detail about future plans of the authors and reads more like a research proposal from then on. I suggest to remove the parts that simply describe future plans of the authors, while general possibilities of model applications may be mentioned. One rather weak point is that the authors suggest the model to be used for future projections. While the model is evaluated in space and for the interannual variability, the performance of the model on longer time scales especially with respect to the human influence is unclear. The parameters for human ignition and suppression are probably strongly constrained by the current spatial patterns, but may have a strong influence on simulations for the next hundred years.

We thank the reviewer for his feedbacks on the manuscript. We addressed concerns about the discussion beeing to focused on research plans (see below). We maintain that the model is useful to explore fire regimes under future environmental conditions. The human influence is a tricky aspect to model, especially the use of fire in agricultural and land management activities. In HESFIRE, it is dependent on countries' GDP (similar to other fire models), which clearly cannot capture the wide range of factors influencing fire use, as mentionned in the paper (Sect. 2.2.1.2). As the reviewer notes, the GDP influence is probably strongly constrained by current spatial patterns, and may not hold in the future. This "permanence" issue is common when modeling human activities which depend on rather unpredicable factors such as technological development and traditional practices. This is the case of deforestation fires in the tropics and preventive fires in sub-saharan Africa. There's not much ground to believe these practices will closely follow future GDP trends. GDP is certainly part of the equation, however (e.g. technological development for alternatives to fire use, fire suppression capabilities). Most importantly, fire projections using HESFIRE will be relevant despite this issue because they can include the resulting uncertainties, for example driving the model in a scenario with unchanged GDP (current practices assumed to continue in the future).

Title: why earth system? the model only interacts until the level of a vegetation model, no atmospheric or biogeophysical influences are discussed

We understand the reviewers concern but suggest we keep this nomenclature:
HESFIRE represents a number of interactions from human activities, ecosystems and the atmosphere, and although they are mostly one-way interactions, we feel it justifies the Human-Earth System (HES) label of the model. A fire impacts module with the implementation of HESFIRE in DGVM/ESM models is underway, and the name was chosen in anticipation of that too.
As for its use in the title ("an explicit fire model for projections in the coupled Human-Earth System"), it conveys the fact that the model can be used to explore fire regimes under contemporary drivers, including natural drivers (e.g. climate), anthropogenic drivers, and their interactions (e.g. climate change), thus the coupled Human-Earth System.

p. 10788, l. 5: what means normalized from 30-80%? are they normalized between 0 and 1 and below (above) the given thresholds the values are 0 (1)

Yes, this is how they are normalized. It is illustrated in figure 2. We added the normalization equation.

p. 10791, l. 25: As far as I am aware this is also a development of the optimization metric, other studies used least squares approaches. You might add a line to highlight this modification of the optimization metric and why you chose to define the metric by using classes not the actual values

We added a discussion of the optimization metric:
"The optimization metric was defined to minimize classification error across 7 classes of annual burned fraction (interval boundaries: 0, 1, 5, 10, 20, 35, 50+% of the grid-cell), and to maximize the correlation with observed interannual variability. Within each class, grid-cells are attributed continuous values based on linear interpolation: a grid-cell with 3% burned fraction is given the value of 2.5, being in the middle of the 2nd interval boundaries. This classification approach aims at capturing important changes that would have little weight on the optimization if using direct burned fraction value. The difference between 3% and 4% in fire-sensitive tropical forests is probably more relevant to capture than between 33 to 34% in fire-adapted grasslands of northern Australia."

p. 10793, l.1,2: is it reasonable to change the parameters to +50% and -50%? Another approach could be to increase the parameter value according to its variability, e.g. +/- its standard deviation/uncertainty derived by the optimization procedure

A number of studies have used fixed percentage changes, one parameter-change at a time, as we've done (Potter et al., 2001; White et al., 2000; Zaehle and Friend, 2010). There are a number of possible approaches, however (Saltelli et al., 2000), including the use of parameter probability density functions and runs with more than one parameter change to cover sensitivity to interactions between parameters (Quillet et al., 2013). We did run the sensitivity analysis with the standard deviation approach, as suggested, and the results are largely similar. The main change between both methods is in Africa, were some areas are now flagged as most sensitive to the anthropogenic instead of fragmention parameter (Figure 1). This is due to the standard deviation of the fragmentation parameters among the 20 optimization runs being relatively low (black horizontal line in figure 4 of the paper). We feel that this is not accurate, as the low standard deviation is actually due to the model being very sensitive to this parameter, thus finding similar values across optimization runs. Accordingly, we keep the first method in the paper, provide references to similar studies, and mention other approaches.

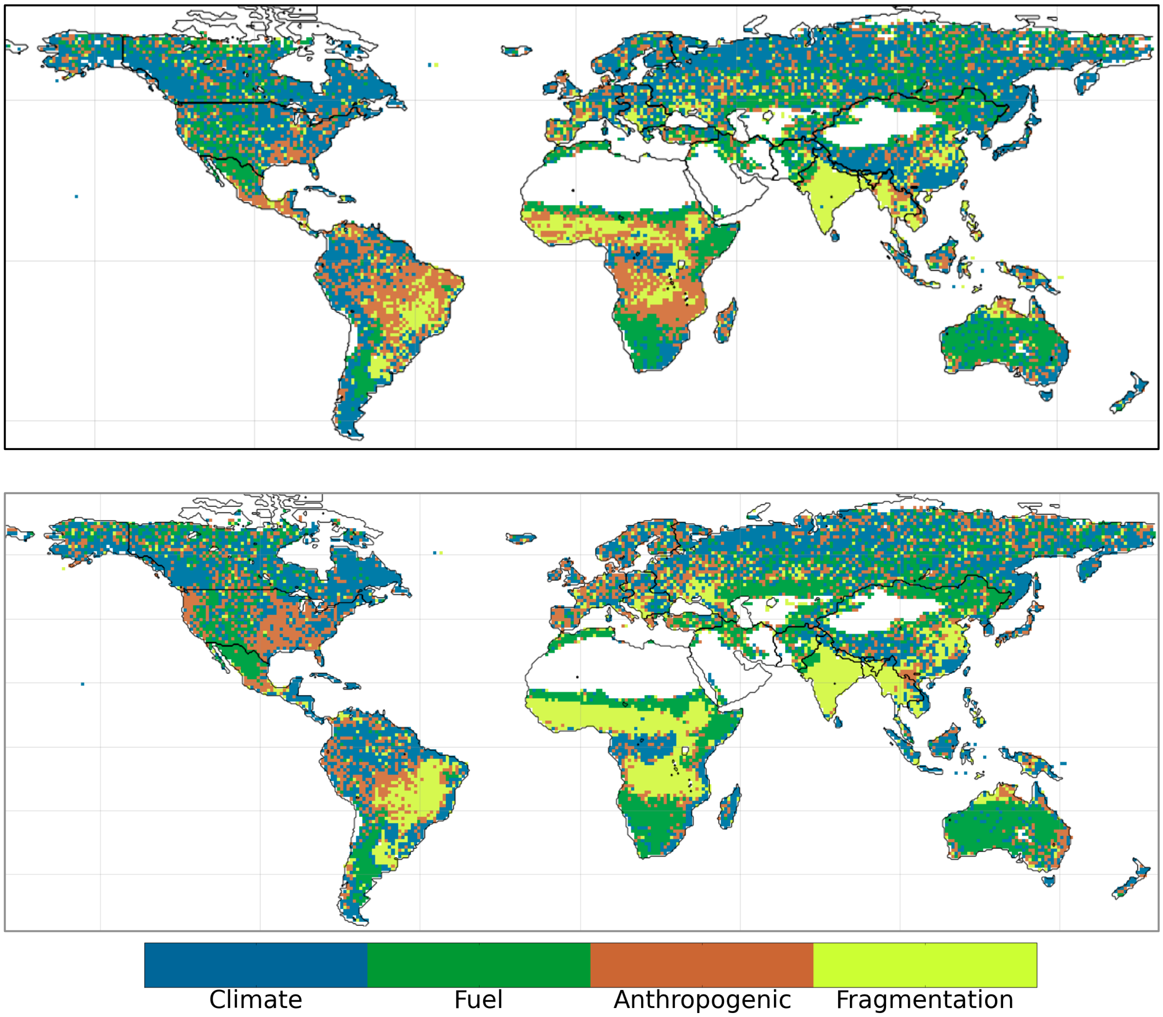

Figure 1. Comparison of the sensitivity map with parameter variation based on +/- std deviation (top) and on +/-50% as applied in the paper (bottom).

p.10792, l. 7-10: please be more precise on the criteria of the gridpoints used in the optimization, a matrix table could be useful here to show the different combinations of criteria represented in the choice of datapoints used in the optimization. Which biomes, which land use densities... were represented ?

As indicated in the manuscript, the gridpoints were selected manually, that is without any statistical method to go through the space of climate/anthropogenic/vegetation conditions. We implemented such a statistical approach but it came out as quite complex to go through the space of all variables, and causing additional issues more than anything. For example, some regions have biased input data (e.g. boreal), which erroneously influenced the optimization early on and were thus largely excluded in the final optimization grid-cell subset. Also, some environmental conditions did not influence the optimization and just a few of the corresponding grid-cells were selected. That includes desert grid-cells for example, which are below the precipitation proxy lower threshold, thus will not have any fires, whatever the optimized parameters. We thus manually selected grid-cells as optimization subsets (this was done roughly, without care for the exact location). Figure 2 shows a map of the subset used for the main optimization (without any grid-cell in South America). Note the sparse grid-cell density in boreal regions, in the Sahara, in the Himalayas, etc. This figure was added to the supplementary material.

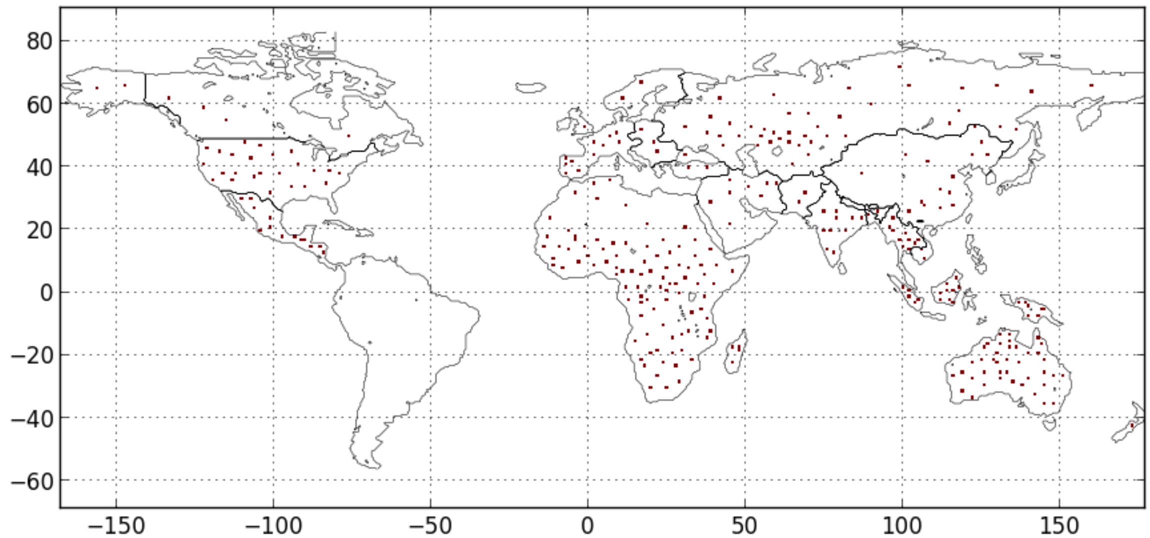

Figure 2. Grid-cell subset used for the main optimization run. Note that no grid-cell was selected over South America, and the selection was sparse over boreal regions (to avoid bias in model parameters due to biased input climate) and over arid regions were fire do not occur (e.g. deserts)

p10795, l. 11: ignition-saturated means to me that more ignitions don't lead to an increase in fire activity. I think here, it just means that more anthropogenic activity (land use) does not result in more ignitions. Moreover, do you really think that ignitions and suppression can be seperated well in your approach?

We indeed intended the meaning that more anthropogenic activities do not result in more ignitions beyond the landuse threshold. We now changed it in the manuscript: "Regarding anthropogenic sources, the optimization procedure suggests that the number of human ignitions saturates at a relatively low landuse fraction, with any additional land use beyond 2–3% of the grid-cell area having no contribution (Fig. 5a)."

Regarding the separation of human influence on ignition and suppression, we agree that it is not necessarily achieved well in the model. GDP and landuse influence both ignitions and suppression, and for GDP, the relationship to fires is negative in both cases. The parameterization can thus easily swap influences between these 2 pathways of GDP fire-driving. This is one of the reasons why we force the ignition-GDP and suppression-GDP parameters to have the same value.

In the presentation of the fire suppression equation (Eq. 12), we added: "Note that GDP_{exp} is the same parameter as in Eq. 3 for human ignitions. GDP has a negative relationship on fires through both ignitions and suppression, leading to an underconstrained optimization if maintaining 2 separate parameters."

p10796, l. 13,14: probably due to the simple representation of fuel.

Indeed, the smoother-than-observed fire incidence patterns in southern-hemisphere Africa are probably due to vegetation classes and the fuel proxy. We now refer to the specific discussion section on this issue.

p10798, l. 17: do integrated assessment models also provide GDP? figure 1. is cut off

Integrated assessment models do provide GDP, at various spatial scale depending on the model (e.g. GCAM divides the world in 14 regions, and we would thus have to apply the same GDP changes to all countries within each region). See Van Vuuren et al. (2011) for the global GDP trajectories from the 4 Representative Concentration Pathways (RCPs) of IPCC AR5 (Figure 3).

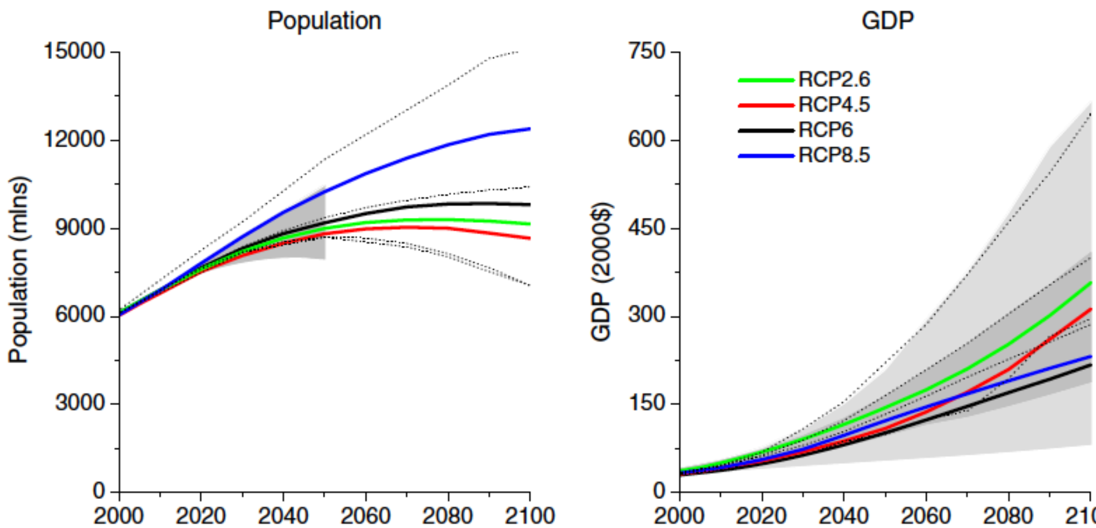

Figure 3. Population and GDP projections of the four scenarios underlying the RCPs. From Van Vuuren et al. (2011).

Figure 1 in the paper (model diagram) was voluntarily cut-off, as a way to show that the model goes on through bi-daily timesteps (Day 1, Night 1 , Day 2, etc), repeating the same computations.

p. 10801 l. 15 ff: In my opinion the description of your future plans should not be described here. The discussion should deal with the results presented here.
p.10801, l. 26/7: same
p. 10802, l. 6 ff: same
p. 10803, l. 12, whole paragraph: This whole paragraph sounds like a research pro-posal, I don't see the benefit of this discussion with respect to your results

We agree with the reviewer that the paper was too heavy on future plans. We also think that identifying major issues in the model, proposing strategies to address them, and discussing potential applications is relevant to a model-description paper. We have substantially revised the discussion section, trying to find an adequate middle ground. As part of that, we re-worded sentences which were focused on our future plans to suggest potential research areas to improve global fire models in general, and Sect. 4.2 has been reduced substantially. The whole section about the potential of regional versions of the model has been removed.

p. 10803 l. 2: I am surprised that here the interest in collaborations is expressed, I think that this is inappropriate here

The mention of potential collaborations was removed from the paper.

Figure 7: the IAV correlation is significant? The correlation is based on the annual Values ?

The correlation is based on the annual burned fraction. We now provide an indication of the significance of the correlation. Note that classic significance tests on goodness-of-fit were not applicable because of the small sample size (14 data points for each region) and non-normal distribution of the data (both in GFED and in the model). We now report the Spearman correlation (ranked correlation), which is not subject to a normal distribution assumption, and indicate its significance for $p<0.05$.

Potter, C. S., Wang, S., Nikolov, N. T., McGuire, A. D., Liu, J., King, A. W., Kimball, J. S., Grant, R. F., Frolking, S. E., Clein, J. S., Chen, J. M. and Amthor, J. S.: Comparison of boreal ecosystem model sensitivity to variability in climate and forest site parameters, J. Geophys. Res., 106(D24), 33671–33687, doi:10.1029/2000JD000224, 2001.

Quillet, A., Garneau, M. and Frolking, S.: Sobol' sensitivity analysis of the Holocene Peat Model: What drives carbon accumulation in peatlands?, J. Geophys. Res. Biogeosci., 118(1), 203–214, doi:10.1029/2012JG002092, 2013.

Saltelli, A., Tarantola, S. and Campolongo, F.: Sensitivity Analysis as an Ingredient of Modeling, Statistical Science, 15(4), 377–395, 2000.

Van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V. and Lamarque, J.-F.: The representative concentration pathways: an overview, Climatic Change, 109(1-2), 5–31, 2011.

White, M. A., Thornton, P. E., Running, S. W. and Nemani, R. R.: Parameterization and Sensitivity Analysis of the BIOME–BGC Terrestrial Ecosystem Model: Net Primary Production Controls, Earth Interact., 4(3), 1–85, doi:10.1175/1087-3562(2000)004<0003:PASAOT>2.0.CO;2, 2000.

Zaehle, S. and Friend, A. D.: Carbon and nitrogen cycle dynamics in the O-CN land surface model: 1. Model description, site-scale evaluation, and sensitivity to parameter estimates, Global Biogeochem. Cycles, 24(1), GB1005, doi:10.1029/2009GB003521, 2010.