Most of the author response and/or manuscript changes adequately address the major comments from the last review. I particularly like the quantile regression analysis to determine the maximum tree cover for a given MAP (see comments below). I am also satisfied with the responses and changes to the paper addressing the choice of fire data and the analysis of both fire models. However, I remain concerned about the response to the "choice of JSBACH driving data", although after clarification on the aims of the analysis, I am no longer sure if the this is an issue with using ESM model output itself or the way this is used to infer areas for improvement of JSBACH. Below I respond to each original point in turn, before including a small number of additional minor corrections.

Choice of JSBACH driving data

The authors have clarified the three aims:

1. Develop a simple multivariate technique to explore the difference between modelled and observed vegetation, fire and climate.
2. Use these differences to evaluate the simulation of, and coupling between, tree cover and burnt area in JSBACH
3. To do this within the MPI-ESM framework, achieved by driving JSBACH offline but with MPI-ESM model output.

There is nothing wrong with the aims, and I like that the authors attempt to at keep the methodology relatively simple. Driving the JSBACH as configured for use in MPI-ESM, offline with ESM output also makes sense, and in their response and suggested changes to the manuscript, the authors have justified the choice of driving data. The authors also discuss the weaknesses associated with this method in the revision of section 4.1, which is also a welcome addition to the paper. However, to critique solely the land surface component in an ESM setup such as this, as if it were independent of other potential climate model biases seems to contradicts the 3rd aim above and introduces the methodological inconsistency which I don't feel have been adequately addressed. To phrase in terms of the multivariate approach, the authors have diagnosed the vegetation cover and fire axis, but not the climate axis. The authors state in their response that "*regarding the conclusions we draw from our comparison we don't see a strong point that they would be strongly affected*." Here are just some examples from the (revised) paper where climate biases could potentially affect either the results, discussion and/or conclusion:

1. *Surprisingly the observations show a higher Spearman correlation between tree cover and precipitation than the models (Table 1). The lower correlation of the modelled relationship most likely originates from the lower precipitation regions (<500 mm year-1 where the maximum tree cover is very low in the observations and both models strongly overestimate the maximum tree cover (Figure 4).*

   The correlation between MAP and other climate variables that influence tree cover could also break down in the MPI-ESM driving JSBACH. As already noted, length of dry seasons are likely to be shorter in seasonal climates. Most GCMS models (although I don't know if MPI-ESM is amongst them) also suffer from biases in downward SW (Li et al. 2013) which could influence tree cover, particularly at the higher tree cover range, where figure 5 also indicates mismatches beetween model and observation in some continents, particularly Asia.

2. *JSBACH overestimates tree cover for low precipitation on all tropical continents.*

   The drizzle problem already discussed seems like an obvious candidate to affect vegetation cover at low precipitations, either through decreasing the length of dry periods or due to associated changes in cloud cover changing evaporative demand and hence available moisture. Despite not ruling out additional climate problems, the authors use this simulated mismatch at low tree covers to justify planned changes to tree to vegetation dynamics:

   only if a 5 year average of NPP turns negative, drought effects on the dynamic vegetation take effect. Other models require a minimum of 100 mm year-1 precipitation for sapling establishment (Sitch et al., 2003). The too high excessive tree cover could be partly improved by improving the non-vegetated fraction which decreases too fast with increasing precipitation

   and

   Tree-grass competition for water could for example be improved in the model by introducing the a sapling stage of trees, which are competitively inferior to grasses (D'Onofrio et al., 2015). Including this mechanism could improve the balance between tree and grass cover, but it could also reduce the establishment rate of trees and therefore the tree cover in the dry regions with excessive tree cover. Including a PFT-specific rooting depth of vegetation would be an important extension of the model to improve the competition for water between grasses saplings and adult trees.

   These three fundamental changes to the dynamics of JSBACH are suggested without establishing that the problem is with JSBACH itself. While it is often necessary reparameterize components of ESMs to compensate for biases in other model components, this should always been done in the knowledge that it is to compensate for other these other biases, and the suggested changes to JSBACH above go beyond a standard re-parameterization.

3. *For Australia underestimation of burned area for both fire models is strong (Figure 4). In a previous evaluation where the model was forced with observed climate and vegetation cover was prescribed (in contrast to the dynamic vegetation cover and climate modelled by the MPI-ESM) JSBACH-SPITFIRE showed better results for Australia (Hantson et al., 2015). An improved response of vegetation cover dynamics to precipitation will therefore likely improve the patterns of burned area.*

The better simulation of fire in Hantson et al. 2015 could also be due to better representation of rainfall timing and distribution, temperatures or any number of climate factors from being driven by observed climate. Also, better representation of vegetation cover would hopefully have been achieved in Hantson et al. 2015 with observed rather than simulated climate. Again, parameterization of either JSBACH or SPITFIRE to account for additional climate biases may be necessary in an Earth System model, but here the author imply the the problem is with JSBACH itself.

4. *This indicates that not an improvement of the fire model but improved modelling of drought effects on the vegetation dynamics will improve the response of vegetation to climate in dry regions.*

   Again, another likely explanation is MPI-ESM rainfall distribution or the impact of other climate factors on available moisture etc.

5. *Intercontinental variation in the relationship between precipitation and maximum tree cover is much smaller for the models compared to the observations. Known variations in vegetation are not sufficiently understood to be represented in models. However our finding that models do show differences in the fire-vegetation-climate relationships between continents shows that further exploration why models show differences can be helpful to better understand causes for intercontinental differences.*

   If this is meant purely for land surface modelling, then there is little in the results of this paper to justify this statement. That there is a modelled difference in fire-vegetation-climate relationships between continents would be more valid if the authors made it clear that this statement is about the ESM setup as a whole.

6. *Overall the multivariate model evaluation highlights the potential for more targeted model improvements with respect to the interactions between climate vegetation and fire, which are crucial for our understanding of future vegetation projections.*

   Again, this is fine as a statement about the ESM setup as a whole, but not focusing solely on the land surface component.

   Of course, there are more suggest model improvements in the manuscript where inherent climate biases from MPI-ESM have (to my mind at least) no obvious impact. However, even in these cases, the authors should be careful at presenting potential new model processes without first checking for the influence in other climate biases. The apparently stronger correlation between fire and tree cover compared to observations, for example, is used to suggest inclusion of resprouting and adaptive bark thickness or fuel feedbacks that might influence fire intensity and hence tree mortality. Again, there are no end of climate biases that could affect intensity which would not be picked up by a straight MAP-tree cover-bunt area comparison. And again, these changes go far beyond standard reparameterization of a land surface model in an ESM. To be fair to the authors, they have included the statement "*exact parameterization and needs to be tested with stepwise model development and factorial simulations*" which does help mitigate some concern with model changes such as this.

The authors suggest in their response that only way to address this contradiction is the do detailed assessment of the atmospheric component of the model, or perform complex experiments or analysis using additional model driving data. This is almost certainly not the case, and it would be a shame if further revisions did make the analysis more complicated. However, some additional, simple analysis might resolve the issue. Here are some examples based on the author responses:

- *We use the standard JSBACH setup, which is the combination of JSBACH with MPI-ESM meteorology. As the fire is sensitive to a number of variables, evaluation of the model in a different setup wouldn't help to guide model development for a model that is almost only run in the coupled setup.*

  A run with observed climate obviously wouldn't be used as a basis for further model development if your aiming to improve JSBACH when driven with ESM meteorology. But it would help the authors determine if the deficiencies already identified are due to simulated climate biases or due to the vegetation component, and would place their discussion on much firmer ground. This is part of the justification for offline land surface model runs required for MIPs associated with CMIP6, e.g (van den Hurk et al. 2016; Lawrence et al. 2016).

- *our motivation here is to evaluate the land surface model, a detailed evaluation of climate biases in the ECHAM model is therefore out of scope*

  There is no need to do a detailed evaluation of climate biases in ECHAM (which would indeed be out of the scope of this paper). However, the authors should ascertain if problems in land surface simulation are caused by either problems in the simulation of the land surface or problems with the information it receives from the atmospheric component - a rather basic first order assessment of any land surface model within an earth system framework. I was able to give a few pointer to potential climate biases from my limited knowledge of climate intermodel comparison literature in my last review. The authors should be able to identify other MPI-specific climate biases that they could at least discuss if not to test. As stated in the last review, there are two of instances where climate model deficiencies are discussed (i.e, when explaining discrepancies in simulated spatial patterns and when discussing calculation of lightning ignitions). At the very least, these types of discussions should be included when critiquing the rest of JSBACH.

- *Understanding potential influences of certain climate biases (such as extremes) on the simulation would require specific factorial experiments*

  and

  While certainly more parameters influence tree cover distribution an increasing number of variables included to explain patterns would require a totally different appraoch

  Not necessarily. A first step could be to simply show if other climate information (no. dry days, downward SW etc) are causing some of the relationships you see using the exact same approach used for MAP. It may well be that this shows that using MAP alone does do a sufficient enough job as a proxy for climate space, which will then support the rest of the papers discussion. If not, then any additional climate variable that explains some discrepancy could be included in the same way that grass and tree cover are interchanged at various stages in the manuscript.

- *Mean annual precipitation explains a large part of the tree cover variability and therefore is a useful proxy for climate*

  Obviously not enough in JSBACH when driven by MPI climate data - the range of TC at a given MAP is one of the features JSBACH as driven by MPI data does not replicate, and there is no other result to help indicate how much of this discrepancy is due to simulation of vegetation cover or other climate biases.

While I do not expect the authors to address all these points in the manuscript (that would be a very long paper!), I hope that I have demonstrated they are certainly not without options. Picking up on one or two of these point, or anything else which can either show MAP really is enough by itself to account for all other climate biases or that can truly attribute problems with model performance to either JSBACH or MPI-ESM climate, will be sufficient.

The authors do make a good point in their response regarding the performance of SPITFIRE along the MAP gradient. If an aspect of the model is ineed performing well in this analysis, then given that the authors aim is to have an improve JSBACH for use within an ESM, it makes sense no other investigation is indicated - with the normal caveat that there could be a degradation in performance after development of another aspect of the model (which in this case would include MPI-ESM as a whole). The authors also make it clear that fireMIP runs of JSBACH would not be an appropriate direct source of analysis.

As the authors are only able to use MPI-ESM model output till 2005 to drive JSBACH, they have to make a rather awkward choice about comparison time periods, as identified in the last review. An ideal solution to this would be to run MPI-ESM beyond 2005, something that could be happening as part of CMIP6 simulations? However, I realise that this is probably not possible, and the MPI-ESM may well be configured differently for CMIP6 simulations. I would like to hear to authors thoughts changing the comparison periods though. The authors state that "*Using only the overlapping period (2001-2005) would decrease the robustness of the mean fire regime and climate characterization".* This is certainly true for fire regime. However, tree cover is normally more stable, and as trees take a few years to establish, the cover found during 2001-2005 would of also be a consequence of burnt area and climate before this period. Perhaps a better choice is to split comparison periods based on variable rather then on model/observation. i.e, when performing analysis, take modelled and observed burnt area and climate from 1996-2005 (climate data, MPI and GFED overlap) and tree cover from 2001-2005. While this is still a rather pragmatic solution to the mismatch in modelled and observed time periods, it might make more sense then the pragmatic solution outlined in the manuscript?

2.2 Choice of fire dataset

The explanation in the revised manuscript demonstrates that GFED without small fire is the most appropriate choice for this study.

2.3 Quantification of similarity in multivariate relationships

The quantile and local quantile regression, while simple, help strength some of the main discussion points about intercontinental differences in multivariate relationships. To be honest, I thought the authors might include a description of this in future work, so it was a pleasant surprise to see that analysis was actually included in the paper.

Re: multivariate metric

Thank you for investigating this. I agree that any further attempt to develop a multivariate metric is too much for one paper, and given an appropriate metric could not be found by either the authors or myself, the inclusion of the extra spearman rank coefficient between fire and tree cover is certainly enough for this paper.

2.4 Use of two models

It appears I overestimated the complexity of the standard model, and the additional information is more than enough. I also accept that additional comparisons between JSBACH-SPITFIRE and JSBACH-standard could become too speculative. However, I note the authors have started to make more use of JSBACH-standard through revisions in response to reviewer 1 - when describing changes in burnt area and tree cover in dry regions in section 4.1 for example.

2.5 Specific comments

I am happy with all changes except for a couple of small details:

- 39 and 45) Kelley and Harrison 2014 should probably be changed to Kelley et al. 2014:

  Kelley, D. I., Sandy P. Harrison, and I. C. Prentice. "Improved simulation of fire–vegetation interactions in the Land surface Processes and eXchanges dynamic global vegetation model (LPX-Mv1)." Geoscientific Model Development 7.5 (2014):

2411-2433.

Kelley and Harrison 2014 looks at future changes in fire, whereas Kelley et al. 2014 is the paper describing model development and benchmarking.

- 42) Was the “minor shifts between woody PFTs in a few cells” quantified? Quantifying equilibrium during spin-up should really be a requirement for any modelling study, and if authors did quantify equilibrium in any way, then it would help the cause to state how this was assessed. However, I do understand that quantifying equilibrium is unfortunately not standard practice, and finding a way to do so is well outside the scope of this paper. So if it was not quantified, then leave this sentence as is.

- 43) Thank you for including citation this. And thank you for including the reference to the r-package used for quantile regression. Statistical software development is a tough job, and I’m sure the developers will appreciate the extra citations for their work.

Hurk, Bart van den, Hyungjun Kim, Gerhard Krinner, Sonia I. Seneviratne, Chris Derksen, Taikan Oki, Hervé Douville, et al. 2016. “LS3MIP (v1.0) Contribution to CMIP6: The Land Surface, Snow and Soil Moisture Model Intercomparison Project – Aims, Setup and Expected Outcome.” *Geoscientific Model Development* 9 (8): 2809–32.

Lawrence, David M., George C. Hurtt, Almut Arneth, Victor Brovkin, Kate V. Calvin, Andrew D. Jones, Chris D. Jones, et al. 2016. “The Land Use Model Intercomparison Project (LUMIP) Contribution to CMIP6: Rationale and Experimental Design.” *Geoscientific Model Development* 9 (9): 2973–98.

Li, J-L F., D. E. Waliser, G. Stephens, Seungwon Lee, T. L’Ecuyer, Seiji Kato, Norman Loeb, and Hsi-Yen Ma. 2013. “Characterizing and Understanding Radiation Budget Biases in CMIP3/CMIP5 GCMs, Contemporary GCM, and Reanalysis.” *Journal of Geophysical Research, D: Atmospheres* 118 (15): 8166–84.