The authors have included new figure to determine if there is a bias in rainfall distribution simulated by ECHAM. This has helped rebalance the analysis across variables in their multivariate analysis. I do have some specific issues with the new figure, which I also feel could be used more in discussing weaknesses in the model performance. I also pick up on some of the author responses that could have some bearing for revision in the rest of the paper, and provide some examples of quick analysis to assess or rule out two remaining climate biases described in the last review.

Figure B1

1. The figure uses CRU-NCEP precip observations, whereas TMPA is used for observed precip elsewhere in the paper. The same climate observations should be used for both, especially as the new figure is used to diagnose differences in climate relationships between the two different climate axes in figures 2, and 4-7. Apologies if I implied in an earlier review that it would be okay to use different precip data for different parts of the analysis - I used fireMIP as an example of an offline JSBACH-SPITFIRE simulation, and thought that it would self evident that if used, observations should still be consistent across different parts of the analysis.

   While the authors may argue that choice of dataset might make little difference to the relationship, the disagreement in precipitation between observed datasets in notorious (Beck et al. 2017; Weedon et al. 2014). A quick plot of MAP vs no. dry days I conducted with CRU TS3.2 (Harris et al. 2013) (data I chose for no other reason that I already had it downloaded, not because I'm recommending it for use in the m/s) compared to CRU-NCEP used in figure B1 shows what I mean:
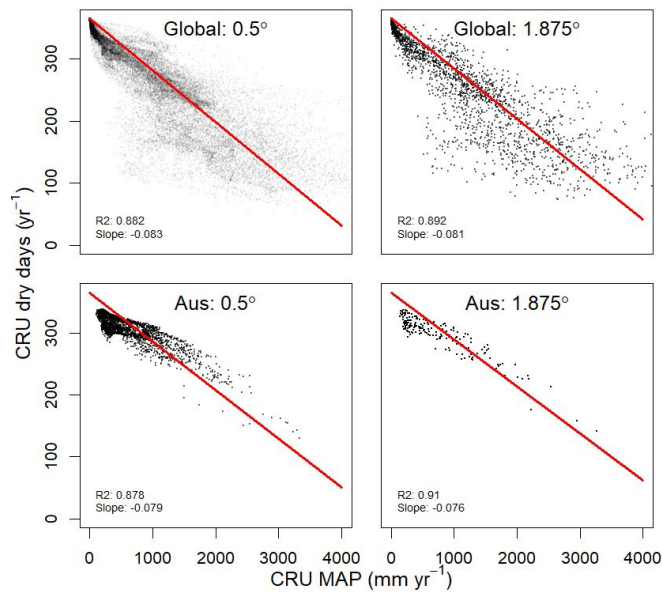
Figure 1: CRUTS3.2 MAP vs no. dry days for tropics and Australia, based on coordinates provided in section 2.

The relationship between MAP and no. drys days for CRU by itself is clearly different, and would actually agree more with ECHAM annual precip. TMPA may also show a significantly difference relationship as well.

2. The authors already produces an excellent style of figure to diagnose burnt area vs precip in figure 3 which could have been used here with the x-axis displaying precip and the y-axis cumulative days at a given precip level. This would provide more information on rainfall distribution biases. However, there is nothing particularly wrong with the simple scatter plot used, so I'll leave this as a suggestion rather than a requirement.

3. If the author's choice to stick with the scatter plots, then please add a trend line.

4. Remember that, for SPITFIRE, impact of dry days in cumulative, so cumulative dry days might also be worth considering, especially as this is another area MPI has been shown to sometimes struggle with (Sillmann et al. 2013). If the authors are able to use the variables already in figure B1 effectively though, then again this won't be

required.

5. The authors need to use this figure to help diagnose climate relationships in more detail. In their response, the authors state that "*This analysis ... shows that the number of dry days in dry regions is well comparable between model and CRUNCEP, for moister regions the number of dry days is even higher in the forcing dataset (MPI-ESM output) used here. We therefore confirm that our conclusions are unlikely affected by biases of rainfall seasonality.*" If this relationship holds once the figure is redrawn with TMPA, then the "anti-drizzle" bias in MPI is surprising. However, it is still a climate bias that will affect simulated fire and possible vegetation, and should be discussed as such in the main text. If it turns out that MPI-ESM agrees with TMPA dry days, than the text will be fine as it is.

Author responses

Author responses in italics. My response to the responses in normal font.

The main concern of the reviewer with respect to the climate biases is the seasonality of the rainfall.

I use seasonality as an example, and it was not the only or main concern.

Of courser biases always exist, here, however, it is important whether the climate biases could have such a strong effect as the reviewer claims.

This is correct. I have no idea how strong an affect the climate biases have. As the authors are presenting a new way of evaluating land surface in ESMs, they need to demonstrate that the impact of other climate biases is either negligible or can be accounted for.

Shortwave radiation does not affect the tree cover in JSBACH, we quickly tested it by applying a multivariate regression, precipitation is highly significant, radiation is not significant if only these two variables are used in a multivariate linear regression. As so far there is no discussion on shortwave radiation and how it influences the model in the paper, we did not include this in the manuscript as it would require several paragraphs to be added.

And

Radiation could have a considerable influence on the productivity of PFTs, but is very unlikely to influence tree cover in JSBACH for the tropics based on the way the model is build. We tested this also quickly with a multivariate regression TC=a1*P+a2*R for the modelled variables where the influence of radiation is not significant. It is therefore unlikely that biases in radiation would show up in tree cover. We now show that the number of dry days is not less in the ECHAM forcing. See also reply 11 and 12.

Was this test with just JSBACH, or for observed tree cover/climate as well? Obviously if SW does not have a significant effect on JSBACHs simulated tree cover but does on observed, then this would be a useful missing climate-vegetation relationship that would need to exploring. If it was tested for both model and observation, then the authors point stands.

Our proposed method clearly goes beyond the normal variable by variable comparison. Including all variables that might be important in the coupled system of fire, vegetation and climate would be optimal in a certain sense but would then suffer from the complexity of the necessary approach and difficulties in interpretation. As stated in the manuscript we use precipitation as a proxy for climate and precipiation is included as one of the axis. The same critisim, that there could be biases not in the mean but in another characteristic of precipitation, could apply to fire and vegetation cover. We simply use annual burned area as a proxy for the fire regime, but fire intensity and seasonality and extremes can be important characteristics too. For tree and grass cover we also summarized two PFTs into one variable

Although the authors have only used burnt area for the fire axis, assessment and suggested improvements have borrowed a lot from previous model assessment and literature. In response to reviewer 1s comments, they also have started exploring fire intensity (figure C1). Obviously PFT fractions are always going to be grouped into just three (tree, grass, bare) fraction types for observational comparison, but each were assessed, which gave some grounding for suggested changes in vegetation dynamics, at least from the land surface bias side. Land use experiments also help explore this impact of changing anthropogenic land cover in JSBACH - again part of the vegetation axis. There is also extensive discussion of changes in plant physiological traits and vegetation dynamics and vegetation-fire feedbacks. And this maybe the key to the problem. i.e, the number observed datasets + number of

variables assessed + past model evaluation + literature + suggested model deficiencies and potential development that has gone into the fire and vegetation axis is extensive, but there is much less detail on the climate axis. And that any mismatches in the multivariate pattern compared to observations are almost always assumed to be because of vegetation and fire biases and not climate. This can be properly balanced by proper discussion of figure B1, and/or reference to MPI climate assessment and climate biases.

A reduction in tree cover would lead to an increase in burned area, therefore what we write is correct. Or vice versa the high burned fraction observed in Australia cannot be achieved with SPITFIRE if such a high tree cover is present.

The argument that burnt area would increase with reduced tree cover is fine. That the ESM needs to reduce simulated tree cover in Australia is also fine. The problem is the statement that "*An **improved response of vegetation cover dynamics to precipitation** will therefore likely improve the patterns of burned area*" has not been demonstrated. I suspect improved vegetation response would be useful, but I also suspect that biases in MPI climate also share some of the blame. If a change in vegetation cover dynamics is induced with is used to improve fire by compensating for any climate bias, then this is not an improved response but a pragmatic tuning and should be identified as such. Figure 1c shows too much rainfall in Northern Australia, so the authors could already use some of their original analysis to diagnose precip as one potential climate bias that would affect tree cover and burnt area.

In terms of regional climate biases not taken into account by MAP, it might be that figure B1 isn't very helpful yet. In figure 1 in this review, for example, the slope of the fit line, spread of the data, and deviation from linear fit at low precips is different for Australia compared to the spread for the whole tropics.

Also the reviewer does not give any references that climate model biases can have such a big effect. Of course any of the climate parameters used can be wrong, but the same would be true for any observational dataset used as model forcing.

Apologies for not providing references in the previous review. The authors may want look at and cite (Ahlström et al. 2017). Although exploring the carbon cycle rather than vegetation cover, they did show a significant impact of precip, temperature and SW biases on simulated vegetation in CMIP5 models. Focussing on the Amazon, (Ahlström et al. 2017) showed MAP, SW and temp climate biases explain most of the simulated GPP, above ground biomass and tree cover. (Ahlström et al. 2012) also showed similar results for disagreement in projected changes in different climate variables into the future. These are just the ones I can think of off the top of my head, there is probably many more. As GCMs have been around for a lot longer, there is of course extensive literature on climate biases that could potentially lead to problems with vegetation dynamics once enabled. (Sillmann et al. 2013) might be a good starting point.

The authors could use (Li et al. 2013) to support their view that only MAP needs to be considered for tropical vegetation distribution, as they use observational constraints to show MAP is the main driver of disagreement in vegetation productivity across models in a region of similar extent to southern America used in this study. However, it should be noted that other climate biases appear to become more important at high MAPs, where vegetation productivity is predominantly limited by available radiation (Nemani et al. 2003). I don't know enough about vegetation dynamics (in model or real world) to know if this tipping point between MAP and SW limited production occurs when tree cover is already saturated. If it does, then maybe (Li et al. 2013) would suggest that other tree cover controls don't need to be considered, at least for this region.

I'm not so sure about the impact of climate biases on fire, as this is a little outside my area of expertise. However, I get the impression that, even with wind speed limitation, SPITFIRE is sensitive to variations in windspeed, especially at lower speeds (Lasslop et al. 2014), which again, GCMs struggle to adequately simulate.

In regions where fire is absent trees always win the competition in JSBACH, it is therefore impossible that other climate factors can solve this, the only reasonable reason is the absence of drought effects on vegetation cover in the model.

Again, the authors need to back these statements up by showing in some way that other climate biases are not the issue here. As they are unable to run JSBACH with climate

observations, perhaps offline runs could be referenced in other papers. For example, JSBACH seems to simulate too much tree cover at low MAPs in the offline study by (Baudena et al. 2014). If this was an appropriate test with no fundamental developmental changes compared to the JSBACH configuration used in the m/s, then the authors could cite this study to back up their suggestion of improved vegetation dynamics at MAP. The authors should have a much better idea of published JSBACH and MPI experiments and evaluation, so might also be able to think of better examples.

This comment is unclear, the variations that are mentioned are observed and the model also shows some variations. We do not see how the ESM setup as a whole comes in here.

I was just reinforcing that fact that the climate axis should be considered as much as the vegetation and fire axis. I meant "ESM setup" as a land surface model driven by ESM output that is emulating a full ESM, obviously without the land-atmosphere feedback (I'm not sure that makes it any clearer…?).

Climate biases can clearly influence the burned area, and its spatial patterns, but I do not see a way that climate biases will turn around the impact of fire on tree cover that much in SPITFIRE, except for the fire-fuel feedback mentioned by the reviewer here. This feedback is already included in the model and different climate forcing leading to different fuel loads could maybe strengthen the feedback. However, in that case it would make sense to reparameterize the model to strengthen the feedback in the Earth system model setting

The point is more to show that the cause of low tree cover is fire feedback in the first place, and not other climate biases (though the author are right that maybe the impact of climate biases on fire-feedbacks should also be a concern...?) If the authors can show climate biases beyond MAP isn't to blame, then the suggested changes fire-feedback are fine.

Precipitation is the main driver of vegetation cover in the tropics. Removing the main driver from this analysis and exchanging it with other potential climatic drivers that are correlation with Precipitation would likely lead to correlations between vegetation and the climatic driver mainly because of the correlation between the two drivers. The relationship would then still

be caused by precipitation. We do not see a way for a useful interpretation of such
relationships without removing the effect of precipitation, which would require a more
complex approach. Exchanging tree and grass cover is different as both are mainly driven by
precipitation and fire.
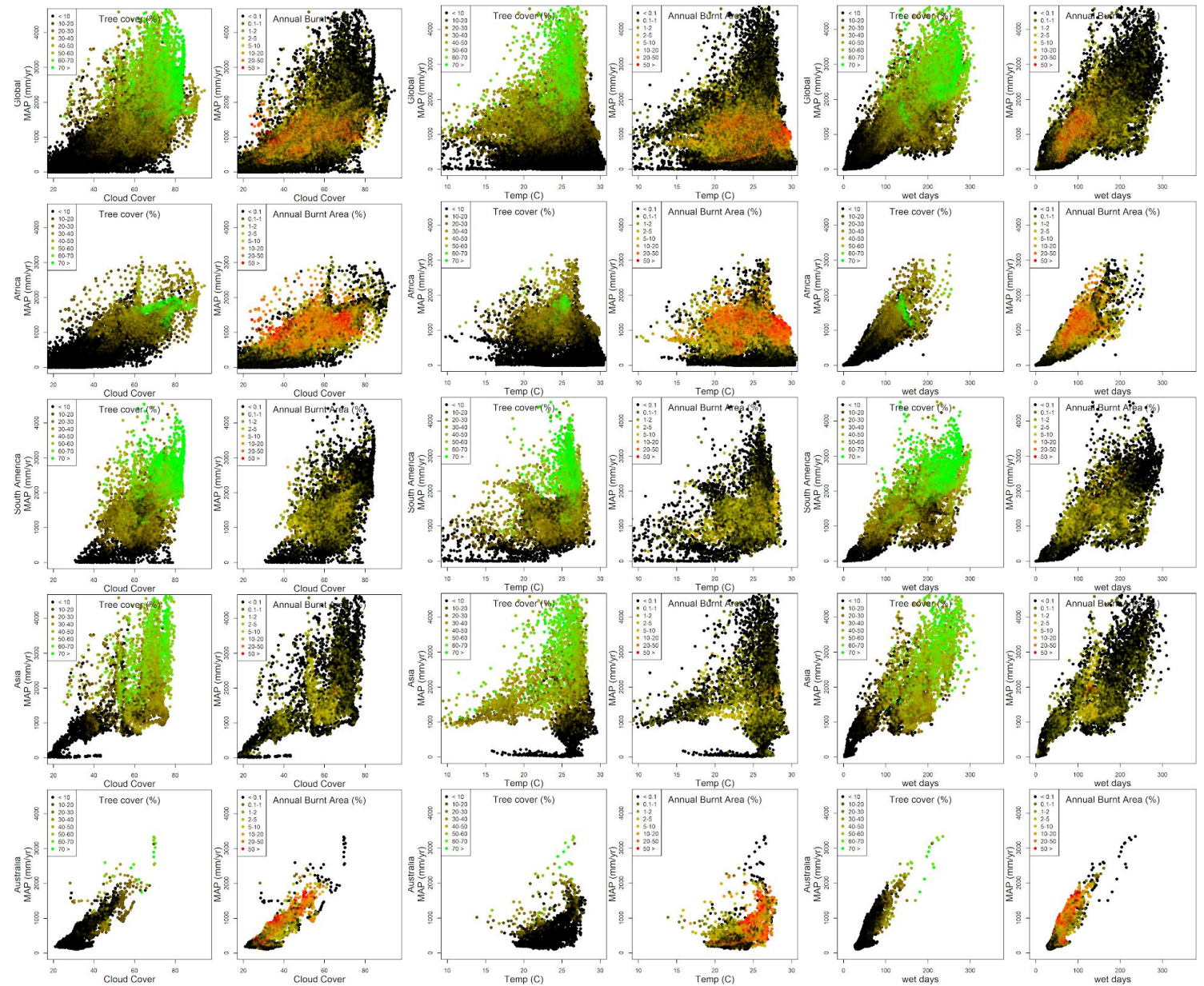
I was more thinking of some like this:

Figure 2 (Apologies for the messy style). The 2 left hand columns of the figure shows CRU TS3.2
cloud cover (roughly used a not-so-great inverse proxy for SW) vs MAP, middle shows MAT vs

MAP, and the two right show number of wetdays vs MAP. Again, I'm not recommending CRU, but just using it as a readily available example. Green column 1 and 3 shows tree cover from VCF (Dimiceli et al. 2015), and red coloured columns 2 and 4 show burnt area from GFED4s (van der Werf et al. 2017). The regions (all tropics, Africa, Southern America, Asia and Australia) are the same used in the m/s.

Even from this example, it is clear that MAP is important but not the only control on either variable. Tree cover does increase with MAP as expected, but the extent of the increase is modulated by temperature, with an ideal MAT occuring around 25 degrees C, and with a rapid drop off at warmer temperatures. The relationship can be exaggerated further in some regions. Australia in particularly has tree cover extending into very dry areas when it is cool enough. Number of wetdays also seems important for tree control in Asia and Australia. Although some of this might be explained by fire feedbacks, that only goes to show that these variables are important for the fire axis also. As I'm using different data to the authors, I won't dwell on the details in the figure above - but it is an example of using on of the technique the authors have already developed to account for more climate controls and identify which biases are appropriate to consider when. A figure like this does not need to be included in the m/s, but it could serve as a starting point to help identify important climate biases. The authors could also think about using spearman's rank or the multivariate regression they used with JBACH to rule out significant effects of short wave.

We prefer to keep the same averaging periods for all variables. If the goal was to only evaluate tree cover it would likely be a good idea. The goal here is however to evaluate the interactions. Tree cover influences the fire regime therefore having the same averaging period for these two variables seems plausible to us. Also the GFED data have more problems for the earlier years and are more reliable from 2001 on.

This is a very good point and I'm happy for averaging period to be kept as is.

References

Ahlström, Anders, Josep G. Canadell, Guy Schurgers, Minchao Wu, Joseph A. Berry, Kaiyu Guan, and Robert B. Jackson. 2017. "Hydrologic Resilience and Amazon Productivity." *Nature Communications* 8 (1): 387.

Ahlström, Anders, Guy Schurgers, and Benjamin Smith. 2017. "The Large Influence of Climate Model Bias on Terrestrial Carbon Cycle Simulations." *Environmental Research Letters: ERL [Web Site]* 12 (1): 014004.

Ahlström, A., G. Schurgers, A. Arneth, and B. Smith. 2012. "Robustness and Uncertainty in Terrestrial Ecosystem Carbon Response to CMIP5 Climate Change Projections." *Environmental Research Letters: ERL [Web Site]* 7 (4): 044008.

Baudena, M., S. C. Dekker, P. M. van Bodegom, B. Cuesta, S. I. Higgins, V. Lehsten, C. H. Reick, et al. 2014. "Forests, Savannas and Grasslands: Bridging the Knowledge Gap between Ecology and Dynamic Global Vegetation Models." *Biogeosciences Discussions* 11 (6): 9471–9510.

Beck, Hylke E., Noemi Vergopolan, Ming Pan, Vincenzo Levizzani, Albert I. J. M. van Dijk, Graham P. Weedon, Luca Brocca, Florian Pappenberger, George J. Huffman, and Eric F. Wood. 2017. "Global-Scale Evaluation of 22 Precipitation Datasets Using Gauge Observations and Hydrological Modeling." *Hydrology and Earth System Sciences* 21 (12): 6201–17.

Dimiceli, C., M. Carroll, R. Sohlberg, D. H. Kim, M. Kelly, and J. R. G. Townshend. 2015. "MOD44B MODIS/Terra Vegetation Continuous Fields Yearly L3 Global 250m SIN Grid V006." *NASA EOSDIS Land Processes DAAC*.

Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister. 2013. "Updated High-Resolution Grids of Monthly Climatic Observations - the CRU TS3.10 Dataset." *International Journal of Climatology* 34 (3): 623–42.

Lasslop, Gitta, Kirsten Thonicke, and Silvia Kloster. 2014. "SPITFIRE within the MPI Earth System Model: Model Development and Evaluation." *Journal of Advances in Modeling Earth Systems* 6 (3): 740–55.

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.

Nemani, Ramakrishna R., Charles D. Keeling, Hirofumi Hashimoto, William M. Jolly, Stephen C. Piper, Compton J. Tucker, Ranga B. Myneni, and Steven W. Running. 2003. "Climate-Driven Increases in Global Terrestrial Net Primary Production from 1982 to

1999." *Science* 300 (5625): 1560–63.

Sillmann, J., V. V. Kharin, X. Zhang, F. W. Zwiers, and D. Bronaugh. 2013. "Climate Extremes Indices in the CMIP5 Multimodel Ensemble: Part 1. Model Evaluation in the Present Climate." *Journal of Geophysical Research, D: Atmospheres* 118 (4): 1716–33.

Weedon, Graham P., Gianpaolo Balsamo, Nicolas Bellouin, Sandra Gomes, Martin J. Best, and Pedro Viterbo. 2014. "The WFDEI Meteorological Forcing Data Set: WATCH Forcing Data Methodology Applied to ERA-Interim Reanalysis Data." *Water Resources Research* 50 (9): 7505–14.

Werf, Guido R. van der, James T. Randerson, Louis Giglio, Thijs T. van Leeuwen, Yang Chen, Brendan M. Rogers, Mingquan Mu, et al. 2017. "Global Fire Emissions Estimates during 1997–2016." *Earth System Science Data* 9 (2): 697–720.