Reply to Anonymous Referee #2 comment on Sippel et al., 2016, Earth System Dynamics Discussions, doi:10.5194/esd-2016-48

The manuscript by Sippel et al. addresses the reduction of ensemble temperature projections by using best estimates of soil moisture-temperature coupling diagnostics under current climate conditions. Although the technique itself has been applied in several other studies, the current application is novel and the results are highly relevant for our understanding of projections of temperature extremes. The manuscript is generally well-written and results are presented in a concise manner. The work seems technically sound, and I could not detect any major flaws in the reasoning and/or analysis, although some minor points were identified that will need to be addressed. Therefore, I believe the manuscript can be accepted for publication after minor revisions.

We appreciate the positive evaluation of our study and research.

I have the following remarks/observations:

- VAC is based on the 30/70th percentile, whereas the authors consider the 90th percentile of TXx. Please motivate if and why this is justified and consistent (coupling might be different for highest percentiles).

Yes, indeed - The reviewer is correct in that coupling might be very different far in the tail of e.g. the temperature distribution (e.g. for the highest percentiles temperature extremes vs. warm, but not extremely warm conditions). This is an important caveat of our study (since we are unable to address very rare events because observations-based datasets are generally short in time).

In the end, both choices are somewhat subjective: The choice for the 30/70th percentile for determining the coupling metric has been discussed (only) briefly in the manuscript: Here, the point is that the threshold choice is basically a trade-off between having enough data while still looking at warm conditions (for both VACb and VACc). An additional analysis using the 10th/90th percentile for computing VACb and VACc yields very similar results (Figure S5), therefore increasing the confidence in our results independent of the specific threshold choice, but unavoidably throws away more data.

For TXx, we look at both ensemble mean TXx and the 90th percentile TXx across the ensemble (cf. Fig. 4d for TXx ensemble mean and Fig. 4f for TXx 90th percentile). While the metric "ensemble mean TXx" is quite natural, the consideration to choose the "90th

percentile TXx" arose mainly from considering the "upper end" of projected TXx values (similar metrics based on ensemble spread are also taken as the uncertainty bounds for heat extremes, see e.g. Seneviratne et al., 2016, *Nature*). Again, changes in ensemble mean TXx and 90th percentile TXx are consistent - i.e. the changes induced by the constraint have the same sign, but are more pronounced for the 90th percentile of TXx. Therefore, we believe that these choices are well-justifiable, and we make these considerations more clear in the revised manuscript.

However, the inherent subjectivity of these choices also means that there is scope for additional research that would look at coupling characteristics under very strong heatwaves e.g. in a small number of models with a large number of ensemble members to test the within-model variability in land-atmosphere coupling and its relation to extreme events.

- While the manuscript has a balanced number of display items, I found the link between the information displayed and that discussed in the text weak. Many sub-panels are never mentioned or discussed, and too much is left for the reader to interpret. Please make sure all relevant information in the figures is referred to, as well as all figures and sub-panels themselves. In particular a more in depth- discussion of the results in Figures 5 and 6 is needed.

Thanks for these suggestions. We have restructured the discussion section and put more emphasis on the discussion of each single display item (please see also similar comments made by Reviewer #1). In particular Fig. 5 and Fig. 6 are discussed in significantly more detail. Also, in the revised manuscript we refer to the individual sub-panels of the figures to make the connection between the discussion and the relevant figure sub-panels clear.

- The selection of references doesn't always to justice to work that other groups have been doing in this area or on this specific topic. In the introduction on weighing models in large ensembles (Page 3, lines 8–12), some examples are provide but interestingly the ones most relevant to the current work are not cited (i.e. Fischer et al., 2012 and Stegehuis et al., 2013). In this way, the suggestion is made that this study is the first to apply model selection on temperature extremes. Please include references to these works.

Thanks for pointing this out. Our intention was by no means to claim that there has been no application of other model constraints on soil-moisture temperature coupling (as the Reviewer correctly points out, e.g. H in the Stegehuis et al., 2013, paper; Interannual temperature variability in the Fischer et al., 2012, paper). We have discussed and cited both papers mentioned in the discussion section of our manuscript, but it is true that we should have referred to them also in the motivation. In the revised version we have fixed this.

Also, model selection/weighing has been applied to other aspects/fields such as snow albedo feedback (Hall and Qu, 2006) and hydrological drought projection (Van Huijgevoort el al., 2014).

Thanks for these references. They are indeed highly relevant to the study and we refer to them both in the motivation section in the revised manuscript.

When discussing the vegetation-atmosphere coupling index (VAC), the authors refer to previous work from the group (e.g. Seneviratne et al., 2006; Lorenz et al., 2012) from which VAC was developed, but not to other alternative indices that are based on a similar philosophy (for instance the metric developed by Miralles et al., 2012, although this paper is cited in a different context).

Thanks for this suggestion and the reference to the Miralles et al, 2012 paper. We have also fixed this.

References

Hall, A. & X. Qu (2006) Using the current seasonal cycle to constrain snow albedo feedback in future climate change. Geophys. Res. Lett., 33, L030502, doi:10.1029/2005GL025127.

Van Huijgevoort, M.H.G., et al. (2014) Identification of changes in hydrological drought characteristics from a multi-GCM driven ensemble constrained by observed discharge. J. Hydrol., 512, 421–434.

References

Fischer EM, Rajczak J, Schär C. Changes in European summer temperature variability revisited. Geophysical Research Letters. 2012;39(19).

93 Seneviratne SI, Donat MG, Pitman AJ, Knutti R, Wilby RL. Allowable CO2 emissions based on
94 regional and impact-related climate targets. Nature. 2016;529:477-483.

95 Stegehuis AI, Teuling AJ, Ciais P, Vautard R, Jung M. Future European temperature change
96 uncertainties reduced by using land heat flux observations. Geophysical Research Letters.
97 2013;40(10):2242-5.