Responses to reviewers for GMDD Paper "Representation of climate extreme indices in the coupled atmosphere-land surface model ACCESS1.3b"

Review #1 from V. Brovkin

1. Statistical significance of biases and resolution of comparison: We agree that it is useful to know which biases are significant and which are not. We performed a modified t-test, as described in Zwiers and von Storch (1995), to indicate which biases are statistically significant and updated plots 1c, 2c, 3c. We tested the biases of T_{MAX}, T_{MIN} and total precipitation for statistical significance since these are the underlying data for the extreme indices. We note that the observations and model results are compared on the same resolution (the lower of the two). In most cases this is the resolution of the ACCESS run (1.25˚x1.875˚), only GPCP and HadGHCND have a lower resolution than the model and we regridded the model output correspondingly (2.5˚x2.5˚ for GPCP and 2.75˚x3.75˚ for HadGHCND).
2. PDF's in Figs. 9, 10: We agree that the text lacks detail in how we calculated these PDF's and we added more information in the manuscript. We added a section "2.3 Statistical Methods" which also describes the significance testing mentioned above. The PDF's are based on T_{MAX}(time,lat,lon), T_{MIN}(time,lat,lon) containing monthly means for the time period 1951-2011 for the corresponding season and region. We use R's kernel density function, using the default Gaussian smoothing kernel and a bandwidth estimated via Normal Reference Distribution (using a well supported rule-of-thumb which defaults to 1.06 times the minimum of the standard deviation and the interquartile range divided by 1.34 times the sample size to the negative one-fifth power unless the quartiles coincide when a positive result will be guaranteed). For the calculation of the skill score we used a bin size of 0.5˚C as in Perkins et al. (2007).
3. Abbreviations: We agree that the manuscript contains a lot of abbreviations and followed the reviewer's suggestion to improve the readability of the text. We indicate variables in italic, expand the subregions and provide a table with the datasets used.
4. T_{MAX}/T_{MIN} : Thanks for this comment, we removed all occurrences of this confusing notation and changed it to T_{MAX},T_{MIN}.
5. NCL: We included the full name of NCL to make this clearer.
6. Rx5day: we included the definition of Rx5day in the Table 1.
7. "all regions bar North America": we rephrased the text to "all regions except North America"
8. Figs. 4,5,6,7,8: we optimized the plot areas similar to the other figures and followed the reviewer's suggestion to remove the unnecessary subtitles.
9. Figs.9,10: We removed T_{MAX}/T_{MIN} from the x-axis

Review #2 from J.Mao

1. Section 2.1.1: We shortened this section and point readers to the relevant references where more information can be found.

2. ET evaluation dataset: We agree that it is useful to use a second dataset here. We included the LandFluxEVAL dataset for evaluation of Evapotranspiration/Latent Heat Flux to reduce the observational uncertainty in this variable. The result is similar, CABLE overestimates evapotranspiration almost globally. Since LandFluxEVAL provides a merged dataset using a range of data we included the Correlation Figure 13b (now 13c) using LandFluxEVAL instead of GLEAM.
3. Clouds: Franklin et al. (2013) provided a detailed evaluation of ACCESS1.3 using the satellite simulator package Cloud Feedback Model Intercomparison Project Observation Simulator Package (COSP). The differences between ACCESS1.3 and ACCESS1.3b are very small and relate to minor changes in the land surface model. The output we obtained from ACCESS1.3b cannot be directly compared to satellite data because observation time, view from above and retrieval filtering have to be taken into account (Stubenrauch et al. 2013). This has been done in Franklin et al. (2013) by using COSP. Hence, we think it is better to rely on the results from Franklin et al. (2013) instead of undertaking a potentially misleading quick comparison. However, we have taken the comments by the reviewer into account in the discussion and have added a reference.
4. We agree that it is useful to learn more about how the errors in land surface schemes and atmospheric variables propagate through calculations to affect extreme climates. We did a CABLE offline simulation with the same set-up to look into these issues and to compare online and offline results in terms of extremes. However, we think this analysis exceeds the scope of this paper and we decided to present these results separately.

References

Franklin, C.N. et al., 2013. Evaluation of clouds in ACCESS using the satellite simulator package COSP: Global, seasonal, and regional cloud properties. *Journal of Geophysical Research: Atmospheres*, 118(2), pp.732–748.

Perkins, S.E. et al., 2007. Evaluation of the AR4 Climate Models' Simulated Daily Maximum Temperature, Minimum Temperature, and Precipitation over Australia Using Probability Density Functions. *Journal of Climate*, 20(17), pp.4356–4376.

Stubenrauch, C.J. et al., 2013. Assessment of Global Cloud Datasets from Satellites: Project and Database Initiated by the GEWEX Radiation Panel. *Bulletin of the American Meteorological Society*, 94(7), pp.1031–1049.

Zwiers, F.W. & von Storch, H., 1995. Taking Serial Correlation into Account in Tests of the Mean. *Journal of Climate*, 8, pp.336–351.