Review of "Physical Processes Leading to Extreme day-to-day Temperatures Changes, Part II: Future Climate Change"

The work of Hamal and Pfahl is a continuation of their similar work using ERA5. They present a decomposition of the processes leading to extreme day-to-day temperature changes in a climate model (CESM), comparing a present period and a future period under global warming for extreme temperature changes in winter and summer. They provide a detailed analysis of these extreme temperature changes for several extratropical and tropical regions, looking at the synoptic conditions and especially decomposing the physical processes leading to the change of the temperature anomaly using a lagrangian backward trajectories analysis.

The paper is clear and well-written (although some parts have extensive descriptions of the atmospheric circulation that could be shortened in my opinion). I was already a reviewer in the first part of this article and I have nothing new to add to the methods, which are essentially similar. I have some technical comments below, including some statistical significance computations that should be done differently in my opinion, but apart from those I would be happy to recommend the paper after some revisions.

The only main limitation that I see is that I find the paper very descriptive without testing any physical theory. In other words, it could have been interesting to explore how the changes you see fit with some physical expectations for how those mechanisms are supposed to evolve. It is mentioned several times that Arctic amplification, and the associated change in the temperature gradient and general circulation, is expected to decrease the importance of advection and I think this kind of reasoning could be interesting to investigate further. I do not think this is a reason to reject the paper, but it could really add something on top of those descriptive mechanisms.

Major comments

1. Figure 1 and corresponding: the way the stipplings are computed does not look like a proper statistical test to me.
    a. If I understood correctly, for the present the authors flag as "significant" the grid points where 80% of members are within +/- 10% of the EAR5-derived respective quantities. I do not think this is correct: first the +/-10% for ERA5 is an ad-hoc measure of the uncertainty. Second, I do not see why 80% of the members should be a correct measure of a significant difference. I suggest to do an actual statistical test with a standard reference level of 95% significance for example. In essence you want to know whether the climate of CESM is compatible with the value for ERA5: that is what you need to test for. The climate of CESM is defined by all the members being put together: what you need to test is whether the sigma_DTDT, sigma_T and r_1,T of this climate are compatible with the same values for ERA5 (which also has an uncertainty). You could for example employ a bootstrapping approach on the ERA5 data (the values for the model are likely very well estimated given the amount of members you have) and check whether the distribution of values you obtain are compatible with the one from the model at the 95% level.
    b. Same for the future: why don't you simply test whether there is a significant difference in sigma_DTDT, sigma_T and r_1,T between the two climates by putting all members together in each period ?
2. For all your significance maps: you need to take into account correlations in statistical testing and employ a false discovery rate, see Wilks (2016).
3. Several times the authors argue that the model is doing a reasonable job in reproducing the statistics of ERA5. I am not sure this is so much the case, as exemplified by Figure 1 for example where the stipplings do not really cover most of

the regions (modulo my main comment 1). I think you should emphasize the differences more, including quantifying them when possible. One point for example is that the model seems to have a diabatic contribution larger than ERA5, which is something also found recently by Röthlisberger et al. (2025) in a different context: it seems to me that the model may be right for the wrong reasons.

Minor comments

1. Please precise which version of CESM you are using.
2. Figure 1: because of the strong meridional differences, you could plot the changes in the second column in percentage rather than absolute values.
3. Figure 3: the sigma_DTDT should be delta_T ?
4. Figure 4: the stipplings are barely visible.
5. Figure 5: I would suggest to scale the temperature and pressure differences by a global/regional warming level to see what is changing beyond the expected local warming.
6. Figure 6 and similar: maybe you could add the boxplots for the future on panels a and b also to compare the spread in each period (I do not expect the spread to be small, thus that the changes you observe are probably much smaller in intensity compared to the spread between events in each period).
7. Figure 7: why did you decide to change the position of the box for looking at extreme DTDT changes compared to Figure 4 ?

References

Wilks, D. (2016). “The stippling shows statistically significant grid points”: How research results are routinely overstated and overinterpreted, and what to do about it. *Bulletin of the American Meteorological Society*, *97*(12), 2263-2273.

Röthlisberger, M., Sprenger, M., Beyerle, U., Fischer, E. M., & Wernli, H. (2025). Advective, adiabatic and diabatic contributions to heat extremes simulated with the Community Earth System Model version 2. *EGUsphere*, *2025*, 1-32.