Reply to Reviewer #3

This is a clear and concise paper, using a range of appropriate methodologies, to examine an important and useful question. I appreciate the discussion of the potential role of natural variability in the dynamical trends, and find the results and conclusions to be clear and well communicated. I recommend publication, subject to some minor comments below.

We thank the reviewer for the positive evaluation of our study, and we provide a more detailed response to the issues raised below.

Line 26. I suspect the c in O'conner should be capitalized? (also line 37 and other references to O'Conner (1963)

Yes, thanks.

Line 31. The long cold tail isn't immediately apparent from Fig. 1, I realized after looking at Fig. 3 that you do have the distribution shown on the far right, but this isn't referenced in the caption or labelled in any way, and so is easy to miss.

We will clarify, thanks (Reviewer #1 made a similar comment).

Line 80. Why SLP for the ERA5 model and z500 for the CESM LENS trained model? Is daily SLP not available in the CESM LENS? Do you think this influences the effectiveness of explaining surface temperature extremes?

There is no particular reason except to use to different published methods, based on SLP (Sippel et al., 2019) and based on z500 (Singh et al., 2023). Both methods show qualitatively similar behaviour, including the comparison of trends, with a somewhat slightly smaller explanatory power for the CESM-trained method. This is to be expected because of the dataset shift from training to application. We clarify in the respective section.

Line 95. Citation shouldn't be in parentheses

Thanks.

Line 130. I assume you mean 90-day running means?

Yes, fixed.

Line 174. CESM2-ETH is only defined later, on line 275.

Thanks, fixed.

Line 180. I think you mean Fig. 6 here? Also, it's a little confusing to reference Fig. 5 (or 6) before Fig. 4, although I understand that it's just to give some more information on the methodology.

We agree and will remove the early mention of Fig. 6 here.

Section 3.1: You show that average winter temperatures over Germany have increased by 2.5C, and explain that some of this is due to dynamical impacts, but it would be useful to have a quantitative assessment of how much thermodynamical warming has occurred over Germany

over this time period. In section 3.2 you show that the coldest winter (1963) would be only 1.4C warmer in the present climate, but also make the argument that the coldest winters occur when there is advection from regions that are experiencing greater thermodynamical warming. This would suggest to me that we might expect the coldest extremes to warm faster than the average temperature, but it is hard to make this comparison without know the thermodynamical contribution to average Germany winter temperatures.

This is a very important point. Reviewer #2 made a similar point, and we refer to the longer discussion in this reply to R2: Yes, we expect cold extremes to increase faster than the mean. We see this in the CESM2 model analysis, but it is difficult to show this concretely in observations because of the relatively short sample size and because we study seasonal extremes (rather than the coldest day per year or so). The 2.5°C increase is a large number (and is the total multidecadal change) and heavily influenced by the dynamical trend during this multidecadal period. We will clarify and discuss this point in a revised manuscript.

Sections 2 and 3: There could be clearer separation of the results between different sections. For example, I think section 2.4 reports results that might be (more) useful in section 3. Similarly, in line 245 you state that the winter 1963 event has a return period of 371 years in 2021, but then in the next section, 3.3, ask whether such an event as the 1963 winter would be possible in today's climate – it seems like you already answered this in the previous section if it has a return period of 371 years. I understand that you look at other methodologies in Section 3.3, but I suggest some re-arrangement to help this flow a little better.

Indeed the point about the GEV analysis around line 245 is important. We will clarify and rearrange such that it will become more clear that the section 3.3 is about the worst-case sampling methods (rather than the GEV fit), but we will mention that the GEV fit already indicates that 1963 would be possible (but unlikely) today.

Line 280. I don't think the word 'bias' is quite correct here. I would recommend 'residual' or similar.

We mean bias in the sense of the statistical model being able (or not) to explain the output of the climate model. But indeed this word may lead to confusion, so "residual" may be a better word here.

Fig. 4. The x-axis labels of circulation and albedo for the top row are closer to the plots below, and so look more like titles for those plots than x-axis labels for the upper plots.

Will be fixed.

Section 4. I would consider renaming this section as Discussion and Conclusions, as the first paragraph seems more discussion than conclusion.

Section 3 is already called "Results and Discussion". To account for the comment, we plan to rearrange such that the text related to discussion will end up in the "Results and Discussion" section.

References

Singh, J., Sippel, S., and Fischer, E.: Circulation dampened heat extremes intensification over the Midwest US and amplified over Western Europe, https://doi.org/10.21203/rs.3.rs-3094989/v1, 2023.

Sippel, S., Meinshausen, N., Merrifield, A., Lehner, F., Pendergrass, A. G., Fischer, E., and Knutti, R.: Uncovering the Forced Climate Response from a Single Ensemble Member Using Statistical Learning, Journal of Climate, 32, 5677–5699, https://doi.org/10.1175/JCLI- D-18-0882.1, publisher: American Meteorological Society Section: Journal of Climate, 2019.