Reply to Review of "An extreme cold Central European winter such as 1963 is unlikely but still possible despite climate change" (Reviewer #2)

The manuscript asks the question of how would the cold 1963 European winter look like if similar atmospheric conditions were to develop in the recent, warmer, climate. Furthermore, it asks whether events as cold as winter 1963 might still be possible today. For this purpose, a number of different methodologies recently develop in the context of extreme event attribution are adopted and put to this test case. They conclude that climate change has warmed an event like winter 1963 by about 1.5 degrees and an event as cold as winter 1963 could still happen today, though it this is unlikely.

I have very much appreciated the multi-method approach which is adopted in the manuscript. The overall finding that the different methods provide very consistent answers is an important result, since it helps confirming the validity of the proposed approaches in this, and other, applications. I think this is an important result, which might benefit being stressed a bit more. The overall conclusion that such cold winter conditions would be warmer today is certainly not unexpected, but a quantification is still useful. The manuscript is well written, and the figures, though packed with perhaps too much information, are nice and clear. Overall, I have very few remarks for this manuscript, which I fully recommend for publication.

We thank the reviewer for the positive evaluation of our study, and we provide a more detailed response to the issues raised below. We think that stressing a bit more, as suggested, the comparison and comparable results of the different method is a good idea for the revised manuscript.

General (minor) comments:

- There is some repetition of the description of the approaches between the "method" section, and the result section. This would be fine if the methods section were at the end of the paper, but given the structure of this journal, I think repetitions could be reduced.

Thank you for point. We will reduce these repetitions for a revised manuscript.

- The authors could consider summarising the results from the different methods in a table.

Indeed, a table would make sense. We aim to summarize the results in a Table for the revised manuscript.

Specific minor comments

Line 31: It is not clear there is a long cold tail in the temperature distribution from Fig 1a, though I agree the tail is clearly evident in Fig 3. Please clarify.

This is true. The large outlier in Fig. 1a may be an indication, but is not clearly evident. We refer in a revised manuscript to literature that discusses that winter temperature distribution in Central Europe have a long negative tail.

Line 95: remove parenthesis in the citation of Shepherd 2016

Line 106: "with the same spacing" is not clear. Please clarify.

Line 115: please specify the estimated percentile

Line 117-125: this text should be removed from here, since it discusses results and not a method.

OK, we will adjust/clarify (all 4 points above).

Line 126: it's not clear to me why 2.5 is called "unconditional". Wouldn't 2.4 be unconditional too? The exact methods differ, but it seems to be me that the main difference between 2.4 and 2.5 is that the first is based on climate model data, and the latter on reanalysis data.

We agree that the terminology is somewhat unspecific. We called 2.5 "Unconditional" because the approach in 2.5 follows exactly the World Weather Attribution statistical approach (Philip et al. 2020) (on observations or reanalysis data), which is called "Unconditional" in the extreme event attribution literature. But we agree that even 2.5 is in fact conditional on the occurrence of a tail event, which is conceptually similar to 2.4 in models (although for this specific application it is not identical, because 2.4 uses the CESM2 large ensemble, i.e. tail events across the ensemble, while 2.5 is based on all winters because of the sparse observations).

Line 131: please add mean after DJF.

OK.

Line 137: do you mean a linear function of GSAT? If not, what function?

Yes, clarified.

Line 155: "utilising importance sampling techniques" would require some more discussion on how this is implemented. It received considerable less space than the boosting methodology. For example: is the reshuffling performed daily, or over blocks of a given length to better preserve autocorrelation?

This remark is justified. It received less space, because it was used mainly in comparison with the boosting, and has been implemented and described in a recent publication focusing on this method (Cadiou et al., 2023).

Line 163: please remove continuous.

OK.

Line 175: what is the CESM2-ETH ensemble? Please clarify

The CESM2-ETH ensemble is a small large ensemble, comparable to the CESM2-LE. The boosting methodology requires bit-by-bit reproducibility (to get to the restart files before the actual event that is to be boosted), and we can achieve this required bit-by-bit reproducibility with the CESM2-ETH ensemble.

Line 175-185: this text could be reduced, since discussed in the results section.

Line 185: If the discussion of the method is repeated in the results section, such as here, then please add a reference to the section in which that method was discussed.

OK (to both points above).

Line 186: If I get it right, the winter cold extremes such as 1963 have warmed less than the DJF mean temperatures. But don't we expect cold extremes to warm faster than the mean, due to, e.g. plank feedback and polar amplification? Could you provide some discussion of this unexpected result? Can it be a consequence of the impact of the mean circulation trend on the mean temperature trend? Or other processes are at play.

This is an interesting point – indeed we expect cold extremes to increase faster than the mean caused mainly by temperature feedbacks such as lapse-rate feedback, Planck feedback and surface albedo feedback (also causing Arctic amplification) (e.g. Pithan and Mauritsen, 2014). In Central Europe, winter cold temperature anomalies are typically advected from northern regions, hence the mechanisms that cause AA likely also apply here.

Here, we analyse cold extremes on a seasonal time scale, and we indeed find this amplification in the CESM2 model (increase of about 1.4°C per 1°C GMT change, which is also larger than the mean change over Central Europe). Because we study observations, however, it is not so clear whether amplification of cold seasonal extremes can be shown or not: The sample is very short, and the (large) number of 2.5°C in the mean refers to the change in the decade 2014-2023 vs. 1951-1980, which is to some extent driven by circulation (and thus may possibly reverse in the future). We clarify in the manuscript that the large number of 2.5°C does not contradict the generally expected larger increase in the cold tail.

Line 261: Please note that method 2.5 had already answered this question. To the extent that return levels of such amplitude anomaly are still associated to a return period in the present day climate, events colder than 1963 are still possible. Please discuss.

Good point, yes, they generally seem to be possible based on the GEV analysis in 2.5. However, please note that the GEV approach bears large uncertainties because it is based on a relatively small sample of about 100 years in observations – so several (and independent) methods are needed to test this assertion. This brings us to 3.3, where we will include the comment about the GEV analysis from 2.5.

Line 239: Following on the previous point, I don't understand why the CESM2-LE approach is "conditional on the 1963 atmospheric circulation". That approach is just looking at the trend in a low percentile. It includes both changes in the magnitude and frequency of cold events. If the we assume that looking at a low percentile implies conditioning on "tail events", then the GEV approach from WWA should be equally considered a conditional analysis.

By that statement, we simply mean that the estimated warming of +1.4 to +1.6°C (independent of the approach) would yield a 1963-like winter of about -4.5° to -4.7°C below the seasonal average – which would still be very cold, but would only be possible if the 1963 circulation would return.

Line l 268-273: some redundant text, since the boosting is already discussed in the methods section.

OK.

L 300: The only information about the ability of CESM2 comes from Fig 4a, which is qualitative. Could you please quantify the mean temperature bias, and the bias in a low temperature quantile? That would be useful to add some confidence on the model results.

We will quantify these biases for a revised submission, along with the difference in the variance of seasonal temperature anomalies, which may be more relevant than the mean bias since we are working with anomalies.

Line 310: why not showing a map for the temperature anomaly associated to the empirical importance sampling? It could be showed in place of the second boosted CESM simulation which does not add much information.

We believe this is a good suggestion.

Fig 1 caption: add (blue dashed) after "atmospheric circulation" and (black dashed) after "global mean temperature".

Fig 2 caption: Please specify the method used to estimate the blu line in Fig 2a. b): replace over Germany with over Europe. The last (d) should be a (b).

Fig 3 caption: The main message seems that the different methods give consistent answer, more than there is uncertainty.

Fig 4: what is the unit of the Circulation (SLP) axis? Please specify.

Thanks for these points, we will clarify and expand the figure labels, also in response to reviewer #1.

References

Cadiou, C. and Yiou, P.: Simulating record-shattering cold winters of the 21st century in France, https://hal.science/hal-03900209, 2022.

Philip, S., Kew, S., van Oldenborgh, G.J., Otto, F., Vautard, R., van Der Wiel, K., King, A., Lott, F., Arrighi, J., Singh, R. and van Aalst, M., 2020. A protocol for probabilistic extreme event attribution analyses. Advances in Statistical Climatology, Meteorology and Oceanography, 6(2), pp.177-203.

Pithan, F. and Mauritsen, T., 2014. Arctic amplification dominated by temperature feedbacks in contemporary climate models. Nature geoscience, 7(3), pp.181-184.