Reply to the reviewer

I appreciate the effort of the authors in answering my comments / criticism and for implementing many of them in the new version of the manuscript. It think that the quality of the paper improved, and especially, the results are not just reported, but also explained and interpreted in the paper in a more critical way. I understand that it is difficult to perform attribution studies due to the limited amount of observations and different model biases, however this does not exclude a critical interpretation and a qualitatively high presentation of the results in the manuscript. I do not question that the message of the paper is very relevant and that action from authorities is urgently needed in order to prepare for the increasing frequency of heat events.

After reading the response of the authors and the new version of the manuscript, I still have a few comments / suggestions:

1) I think that the sentence "*it could either occur by chance or nonlinear effects have made such heatwave possible*" (L. 280-281) is an oversimplification leading to misunderstandings, thus it needs some clarification should be rephrased.
We often model extreme events as random variables, but the climate is essentially a chaotic deterministic system. It behaves at certain scales as if it would be random (this is why the approximations using random variables often works), nonetheless extreme events do not happen by chance. What the authors actually mean, I suppose, is that we observe "by chance" a very low-probability event in a relatively short time series. This is explained at the beginning of Sec. 3, but this and similar sentences still appear in the manuscript.
I am also not convinced that the only alternative to an event "by chance" are "nonlinear effects" or, as mentioned in the abstract, "new nonlinearities". It is possible that global warming makes some well-known processes more (or less) possible thus changing the distribution of extreme events. "New nonlinearities" sounds for me too specific considering that the authors did not study this issue directly.

We would like to thank the reviewer for pointing to the possible misinterpretation of these terms, as what the reviewer explains is exactly what we meant. We partly already explained it in the first paragraph of Sect **Observational analysis: return time and trend.** We now also change:

"With this approach we still assume this was an event happening by chance." into "With this approach we still assume this was an event happening by chance, that is, the behaviour is in line with that of a chaotic deterministic system in a warming climate and by chance we observe a low-probability event in this short time series."

In Sect **Probability of a chance event** we changed

"it could either occur by chance or nonlinear effects could have made such a heatwave possible" into

"it could either occur by chance (a low-probability event) or, for instance, nonlinear effects that have not been observed at this location before could have made such a heatwave possible"

And in the abstract:

new nonlinearities -> nonlinear interactions and feedbacks.

2) L. 439-440 and L. 210-211"*Also further research is needed into the limitations of standard GEV analysis on annual maxima with short records and seemingly non-stationary behavior*."
The GEV approach is formulated for independent, identically distributed, random variables. It can be still applied to correlated data, in case the correlations are weak enough and the block maxima are uncorrelated. However, it is not surprising that it does not work if these conditions are strongly violated due to the non-stationarity induced by global warming. This issue is often addressed by assuming a time dependence of GEV parameters, however there is no theoretical support for these kind of dependencies. It is well known as well that the method is quite data- hungry because it considers only the maximum of each block. Furthermore, it is an asymptotic method, thus it is valid only in case the block size is large enough, and the convergence to the asymptotic distribution can be extremely slow. I do not question the advantages of the method, but the above mentioned application issues are well known in case the data does not satisfy the necessary conditions. Thus, care is needed when applying this method to observational data sets and non-stationary model simulations. But, again, these problems are well-know, thus I do not see the usefulness of the mentioned future studies.
These issues are thoroughly explained in:
Coles, "Introduction to Statistical Modeling of Extreme Values", *Springer*, New York, NY, USA, 2001.

For convergence issues and the problem of limited data size see, for example:

Vannitsem, "Statistical properties of the temperature maxima in an intermediate order Quasi-Geostrophic model," *Tellus, Series A: Dynamic Meteorology and Oceanography*, vol. 59, no.

1, pp. 80–95, 2007.

Felici, Lucarini, Speranza, and Vitolo, "Extreme value statistics of the total energy in an intermediate-complexity model of the midlatitude atmospheric jet. Part I: Stationary case", *Journal of the Atmospheric Sciences*, vol. 64, no. 7, pp. 2137–2158, 2007.

Gálfi, Bódai and Lucarini, Convergence of extreme value statistics in a two-layer quasi-geostrophic atmospheric model, *Complexity*, 5340858, 2017

We changed the sentence "*Also, further research is needed into the limitations of standard GEV analysis on annual maxima with short records and seemingly non-stationary behavior*." into "*Also, further research is needed on how to overcome the known limitations of standard GEV analysis on annual maxima with short records and very extreme values.*" as this better describes what the intention of the further research is, and acknowledges that the limitations are already known.

Note that we do not think that the challenges we encountered with the 2021 data point are caused by the general inappropriateness to fit a GEV to the TXx temperature data in this region, as the fit with GMST-dependence introduced works well to describe all data up to but excluding this event (Fig. 6). The difficulties encountered here are instead probably linked to shortness of the time series and the extremity of the event, or in addition, mechanisms previously not encountered in this region coming into play.

3) The vast number of data sets and methods used in attribution studies seem to lead to shortcomings in the presentation of these methods and the interpretation of the results. If this could be avoided, however, it would make attributions studies more accessible for a broader audience and would reduce the risk of misinterpretation.
I am aware that there is a vast body of relevant scientific literature and the authors cannot explain

every detail of their methods in the manuscript. However, I firmly believe that the relevant methods should be properly explained, even if very concisely. It is not reader friendly at all to give a list of papers and expect the reader to go through a whole body of literature just to understand one manuscript. Nonetheless, I think that also from this aspect the manuscript improved with respect to the previous version.

We are pleased that the reviewer noticed an improvement with respect to the previous version. In this version, we added further detail in the statistical methods and synthesis sections to expand on our method.

In the statistical methods section we clarified some text and also add:

Uncertainties corresponding to the statistical-model uncertainty, are obtained using a non-parametric bootstrap procedure. With this GEV distribution, first the PR and intensity change are calculated from observations, as well as the return period in the current climate. Next, the return period is used as a threshold to specify the event magnitude for the models. For this return period, the PRs and intensity changes between 2021 and the counterfactual climate are calculated from different models. This is, however, only done for models that pass our validation tests on the seasonal cycle, the spatial pattern of the climatology, and the scale and shape parameters of the GEV distribution, see Section 4. Finally, both observational and model results are synthesised into a consistent attribution statement, see Section 5.

In the synthesis section we clarified some text and also add:

"The uncertainty due to differences in model set up and physics is represented by model spread --- the average departure of each model from the mean model best estimate. This is added in quadrature to the model natural variability as white extensions to the light red bars in the synthesis figures. The uncertainty in the model average (bright red bar) consists of a weighted mean uncertainty, where the contribution from each model is inversely proportional to the uncertainty due to natural variability squared, plus the model spread term added in quadrature to the uncertainty in the weighted mean. Please see e.g. Kew et al. (2021) for more detailed information on the synthesis technique including how weighting is calculated for models."

Note that we intentionally use many different data sets and methods to increase the robustness of the results, i.e. we aim to capture something of the method-related uncertainty, represented by the intermodel spread. The synthesis figures show the results from each method as well as the synthesised combination, so it can be clearly seen if there are outliers. We hope that the general message - that warm extremes in this region are becoming more frequent and more extreme and action should be taken - is accessible to a broader audience. We are also aware that some of the challenges we encountered are being investigated in other studies.

4) *Fig. 15* is too small, it is impossible to see the arrows illustrating the wind direction.

The figure was indeed intended to cover the full width of the page, which was not true in the manuscript. We also enlarged the arrow size so that the arrows are more visible.

5) The language of the manuscript improved as well, but it still needs some revisions in terms of understandability and typos.

We have improved readability and corrected typos as well as unclear sentences.