## Response to the reviewer #2:

We highly appreciate and are very thankful for the time and effort that was invested in reviewing our manuscript. The detailed and constructive feedback will help us to improve the manuscript. In the following, we provide an answer to each comment brought up by the editor and reviewers. The original comments are in italic red while our responses are in black.

*General Remarks*

*The study covers a lot of ground, connecting 5-10 year European summer temperature variability to ocean heat content variability, and posits a plausible and interesting relationship between the two in the MPI-GE. There are sections where the writing is clear and sections that could use some work (e.g. with formality, dangling comparisons). My main concern with the study is the lack of context given along the way from a lead-lag relationship between climatic fields to the causality implied in the title "Extremely Warm European Summers driven by Sub-Decadal North Atlantic Heat Inertia" and throughout. The results are based on a single climate model, run with relatively low resolution, in a region that has some notoriously "obstinate" SST biases likely tied to unresolved oceanic processes (e.g. Athanasiadis et al. 2022). In addition, the North Atlantic sector has been identified as a region where the ocean-atmosphere coupling is weaker in models than in observations, creating issues for NAO variability, blocking, and decadal prediction (e.g. Simpson et al. 2018; albeit focused on winter, but that is the season the coupling is stronger to begin with). For the conclusions made in this paper to stand as they are written, there will need to be a convincing argument made for each link in the causal chain that the MPI-GE lack significant biases with respect to observed fields and that X indeed induces Y. For example, the "driver" is concluded to be a build-up of heat in the North Atlantic Current, but that accumulation ultimately can be traced back to atmospheric variability, right? If you were to impose the ocean heat anomaly in the model, would the atmosphere respond as you describe? I would recommend toning down the conclusions to reflect what is really being explored, the relationship between North Atlantic OHC variability and periods of exceptionally warm summers in the MPI-GE.*

Thank you for your detailed and constructive review of the study. We appreciate your insightful comments and concerns regarding the context, limitations, and uncertainties of our findings. We acknowledge that our study relies on a single climate model with relatively coarse resolution, which can introduce biases and affect the representation of oceanic processes. We agree that a comprehensive assessment of causality requires further investigation and a convincing argument for each link in the causal chain, including the influence of atmospheric variability on the build-up of heat in the North Atlantic current. We will carefully consider these suggestions and will provide a more nuanced interpretation of our results, focusing on the relationship between North Atlantic ocean heat content variability and exceptionally warm summer. As an example we have toned down the conclusion and the title of the manuscript, which is now "Extremely Warm European Summers preceded by Sub-Decadal North Atlantic Heat Accumulation".

*Specific Comments*

*L3: Re: "... remain unexplored": Some work appears to be done in the realm, including by the co-authors (e.g. Müller et al. 2020).*

We agree with the reviewer that this sentence is misleading. Indeed we cite several other studies later in the introduction section. Therefore we will change "…and the mechanisms controlling such sub-decadal variations remain unexplored." to "…and determine a mechanisms controlling extremely warm summers on sub-decadal time scales."

*L4-5: Please revise this sentence for clarity.*

For more clarity we will split this sentence into two parts: "We show that extremely warm summers over Europe, occurring in sub-decadal periods, are related by a strengthening of the North Atlantic ocean subtropical gyre, an increase of meridional heat transport, and an accumulation of ocean heat content over the North Atlantic several years prior to the extreme event episode. The ocean warming affects the ocean-atmosphere heat fluxes, leading to a weakening and northward displacement of the jet stream and increased probability of occurrence of atmospheric blockings over Scandinavia."

*L19: Consider highlighting the recent work of Röthlisberger and Papritz (2023).*

We thank the reviewer for this comment and plan to cite Röthlisberger and Papritz (2023).

*L29-31: I'm not convinced this is true. I've included a few potential references, but I feel a deeper dive into the literature is warranted.*

We agree with the reviewer that the statement of this sentence is perhaps a bit too overstated and therefore will tone down and reword this sentence: "However, the assessment of drivers for extreme temperatures on long-term timescales is currently limited (Simpson et al., 2018, Wu et al., 2019), and the relevance of specific year to multi-year timescales for extreme summers remains uncertain (Röthlisberger et al., 2023)."


*Simpson, I. R., Deser, C., McKinnon, K. A., & Barnes, E. A. (2018). Modeled and Observed Multidecadal Variability in the North Atlantic Jet Stream and Its Connection to Sea Surface Temperatures. Journal of Climate, 31(20), 8313–8338. https://www.jstor.org/stable/26508075*

*Röthlisberger, M., Papritz, L. Quantifying the physical processes leading to atmospheric hot extremes at a global scale. Nat. Geosci. 16, 210–216 (2023). https://doi.org/10.1038/s41561-023-01126-1*

*Wu, B., Zhou, T., Li, C. et al. Improved decadal prediction of Northern-Hemisphere summer land temperature. Clim Dyn 53, 1357–1369 (2019). https://doi.org/10.1007/s00382-019-04658-8*


*L36-37: Best in what way?*

"One of the best" means one of the most adequate representations of observed historical temperatures among the climate model large ensembles available at the time of the study. We will specify this: "MPI-GE offers one of the most adequate representations of observed historical temperatures among single-model initial condition large climate models available at the time of the study (SMILES; Suarez-Gutierrez et al., 2021).

*L39: Re: "Including some of the most extreme European summers": What do you mean by this? Extreme compared to what?*

"Most extreme" means one of of the most exceptional European summer temperatures ever recorded. We will add this for clarification.

*Section 2.1 Model Description: Maybe in this section, you could also note your study domain and the fields you will use for each part on the analysis*

Thanks for this suggestion. We agree with the reviewer and will add the temporal resolution, as well as the study domain to the method section: "Our research focuses on seasonal summer means (JJA) over Central Europe, defined as an area of 15°-35°E/ 45°-65°N as well as the whole North Atlantic Ocean area."

*L55: Parenthetical should be its own sentence.*

We thank the reviewer for this suggestion and will have the parenthetical as its own sentence.

*Section 2.2: This section is structured in a very atypical way. Please revise and avoid the use of sub-headings.*

We agree with the reviewer and plan to rewrite this section as a continuous text without sub-headings.

*L63: How precisely do you determine the significance of spectral peaks?*

To determine the significance of spectral peaks in the presence of red noise spectrum with a 95% confidence interval, we compare the amplitude of the peak to a threshold value derived from the red noise spectrum. If the amplitude of the peak exceeds the threshold, it is considered statistically significant at the 95% confidence level.

*L65: Is this a linear detrending? Is that appropriate for "all of [y]our data"?*

In our case, we have chosen a linear detrending to allow comparisons to ERA5. Both linear detrending and removing external forcings by subtracting the ensemble mean yield similar results in this case, as shown in Fig. B. Especially, for the heat extremes as peaks of the time series there is no difference in the timing of their occurrence. In contrast to other results (e.g. Borchert et al., 2021), a linear detrending in the MPI Grand Ensemble seems appropriate and does not distort the results. This difference may be due to the model type of model used here, an un-initialized fully coupled Earth-System-Model. We plan to to clarify which detrending method we used.

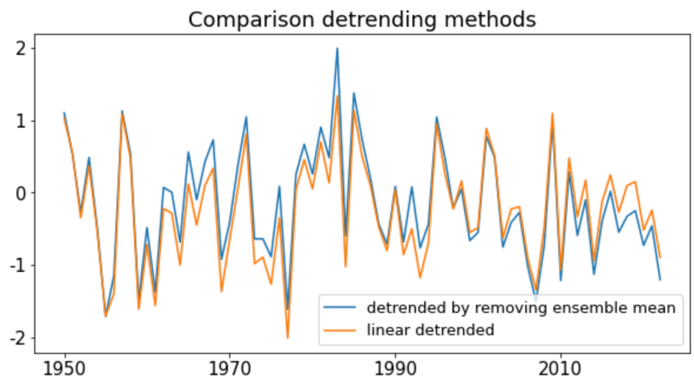


**Figure A:** Comparison between detrending methods in MPI-GE. Exemplary the Central European mean summer temperature time series for both detrending methods, linear detrended and detrended by subtracting the ensemble mean are displayed. Here, for simplicity, only one ensemble member and one variable are displayed, however the authors statement is also true in a broader context.

*L69: Randomly composed arrays of..?*

Here „randomly composed arrays of the corresponding variable“ is missing. We will add this.

*L74/76: I'm a bit lost. What is meant by: "total summer mean variability during extremely warm summers"? Is this interannual variability? Decadal? Assessed only during warm periods? How are "for times when heat extremes occur" defined?*

The "total summer mean variability during extremely warm summers" is given by the standard deviation of unfiltered summer (JJA) mean anomalies (calculated with respect to their long-term averages) for years when extremely warm European summers. This means it is the inter-annual summer variability assessed for warm periods. We agree that our phrasing appears to be unnecessary complicated and therefore we will rewrite this section, such as "…we scale the band-pass filtered summer mean anomalies by the standard deviation of unfiltered summer (JJA) mean anomalies during extremely warm European summers, calculated with respect to their long-term averages.“

"For times when heat extremes occur" refers to years that match with our definition for extremely warm summers, to make this clear, we will add a definition of extremely warm European summers to the method section.

*L79-80: "Also, timescales between 10 and 20 years are dominant in only a few more grid points compared to timescales above 20 years." What do you mean by this?*

We refer in this sentence to the number of significant grid points found for the different time intervals in the cross-spectral analysis and want to express that for the 10-20 year interval only few grid points are found, but for the interval above 20 years even fewer are found and here reanalysis and GE are not consistent. We apologize for the confusion and will revise this sentence: "On time scales between 10 and 20 years, only a few grid points are dominant. Even fewer dominant grid points are found on time scales greater than 20 years.“

*Figure 1:*

- *I thought the ERA5 grid was decimated to match the MPI-GE grid?*

  We have to apologize for this confusion. We decided to maintain the figure as it is without regridded data and will remove this erroneous statement from the manuscript.

- *There seems to be disagreement on the dominant timescale of SAT variability in your study region between ERA5 and the MPI-GE. Could you comment on that?*

  Indeed the dominant time scales in the reanalysis and the model disagree on the broader region of sub-decadal dominance. However, assuming that certain real-world processes may be simulated by climate models correctly albeit for the wrong regions, we find the agreement between the the model and the reanalysis very striking. Although the model simulates the dominance of sub-decadal timescales for temperature in a wider and slight more eastward region, it still captures its effect. Therefore, the model can still be useful to understand this mechanism and its drivers, accounting for the biases in the region of influence. We have expanded our discussion section to elaborate on this issue. Our mechanism still has great relevance for the real world, even if in a somewhat deviated/shifted region. The results from Müller et al. (2020) as well as first results of our current ongoing research confirm the validity of our statements to the real world.

*Müller, W. A., Borchert, L., & Ghosh, R. (2020). Observed Subdecadal Variations of European Summer Temperatures. doi: 10.1029/2019gl086043*

- *There are 6 extremely warm summers per 5–10-year period in each ensemble member? How can (almost) every summer be extreme?*

  We must apologize for this confusion. The number of extremely warm European summers per ensemble member refers to the entire time period (73 years) not to 5-10 year intervals. This means that not approximately every summer is extreme, but only every 10th summer. We can see that on average over all 100 members more extreme summers occur over Central Europe than in other regions. We agree with the reviewer that this figure is perhaps not completely clear, this confusion is in our opinion mainly due to the unclear description of the unit. Therefore, we plan to revise the figure, for instance by indicating the number of extremely warm European summers per ensemble member per year and include a definition of our calculation in the figure.

- *It may make the figure too messy, but it would be nice to see the power spectra of each individual member, maybe in a supplement? And isn't the dominant variability cycle at around 15 years?*

  We agree with the reviewer that it would be helpful to see the spectra of the individual ensemble member and will adjust the figure accordingly. Further, we agree that the peak at about 15 years is the most dominant one, however, with this figure we want to show that the sub-decadal time scales also have significant peaks over Central Europe and thus we have a good reason to analyze them further. We plan to make this clearer in the text.

*Section 3.3: What's missing here is validation that the low-resolution ocean model can capture these processes.*

In fact a validation analysis for MPI-GE is pending. However, there are indications that the processes are valid in forced-ocean experiments (such as those in Müller et al. (2020), using the same ocean model in their figure S1) and other coupled climate models (e.g. Martin et al. (2019). These results underline the importance of ocean heat content accumulation for summer mean climate, and their relation to a damped sub-decadal oscillation behavior in the coupled North Atlantic. However, a detailed analysis is beyond the scope of this study but of importance for further research.

*L116-117: What initiates this? How is it related to the AMO?*

We thank this reviewer for this question. For this study we analyzed which mechanisms in the North Atlantic Ocean drives extremely warm European summers up to three years prior their occurrence. Indeed the influence of external drivers leading to this anomalies within the North Atlantic Ocean is pretty interesting, but to include a thorough analysis accounting for more modes of variability we find beyond the scope of the paper.

*L149-150: How do you know this is "bottom-up" driven and not "top-down" driven?*

Since the mechanism we analyzed evolves over several years in the North Atlantic ocean, which includes the accumulation of heat, the natural conclusion is that the ocean is warming the atmosphere via the ocean-atmosphere heat flux rather than the atmosphere is cooling the ocean.

*Figure 4: Are these the ensemble means?*

Yes, the jet-stream position is calculated by the ensemble mean once for members showing and extremely warm European summer and once for members showing no extremely warm European summer

*Figure 5: Including the OHC branch of the mechanism in the schematic would be helpful.*

We agree with the reviewer and plan to add the OHC branch to the corresponding figure.

*Citations to consider*

*Athanasiadis, P. J., and Coauthors, 2022: Mitigating Climate Biases in the Midlatitude North Atlantic by Increasing Model Resolution: SST Gradients and Their Relation to Blocking and the Jet. J. Climate, 35, 6985–7006, https://doi.org/10.1175/JCLI-D-21-0515.1.*

*Simpson, I. R., Deser, C., McKinnon, K. A., & Barnes, E. A. (2018). Modeled and Observed Multidecadal Variability in the North Atlantic Jet Stream and Its Connection to Sea Surface Temperatures. Journal of Climate, 31(20), 8313–8338. https://www.jstor.org/stable/26508075*

*Röthlisberger, M., Papritz, L. Quantifying the physical processes leading to atmospheric hot extremes at a global scale. Nat. Geosci. 16, 210–216 (2023). https://doi.org/10.1038/s41561-023-01126-1*

*Hall, R.J., Jones, J.M., Hanna, E. et al. Drivers and potential predictability of summer time North Atlantic polar front jet variability. Clim Dyn* ***48****, 3869–3887 (2017). https://doi.org/10.1007/s00382-016-3307-0*

*Osborne, J. M., M. Collins, J. A. Screen, S. I. Thomson, and N. Dunstone, 2020: The North Atlantic as a Driver of Summer Atmospheric Circulation. J. Climate, 33, 7335–7351, https://doi.org/10.1175/JCLI-D-19-0423.1.*

*Wu, B., Zhou, T., Li, C. et al. Improved decadal prediction of Northern-Hemisphere summer land temperature. Clim Dyn 53, 1357–1369 (2019). https://doi.org/10.1007/s00382-019-04658-8*

We want to thank the reviewer for providing these citations. Indeed we find them very helpful and plan to implement them in the revised version of this manuscript.