Response to the reviewer #1

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

1. Comments to revised manuscript:

The paper has improved, but some context is still lacking in the introduction; a more comprehensive overview of the literature would help the reader to better understand the novelty of the findings.
In terms of content, the analysis of the ocean segment of the proposed mechanism is interesting and clearly described, but the mean temperature section is still difficult to follow. Why focus on a region with a clear variability bias between MPI and ERA5? How robust is the ensemble mean spectrum compared to a pooled spectrum? Isn't the logic a bit circular that a 5-10-year period with many warm extremes is a positive sub-decadal phase?

We thank the reviewer for his/her feedback and would like to address a few points here in general before the specific responses follow afterwards: We have carefully extended the introduction section, now mentioning the coupled ocean-atmosphere cycle and added some appropriate citations. Further, we have revised mean temperature section, reworded the corresponding sections and changed/added some figures, we hope that these changes contribute to the understanding.

Specific Comments:
L33-34: I'm not sure how this follows from the paper... Please see note on the attached document.

Thanks for bringing this up. This comment is answered in detail below in section 2 "Comment of previous reviewer #2 referring to L19" and "Comment of previous reviewer #2 referring to L29-31".

L47-48: Maybe soften "responsible for" to "associated with" or "precede". Additionally, ...which processes ... are…

We agree with the reviewer and changed "Additionally, we identify which processes in the North Atlantic Ocean is responsible for the increase of the occurrence of extremely warm summers." to "Additionally, we identify which processes in the North Atlantic Ocean precede the increase of the occurrence of extremely warm summers."

L59: we

Changed.

L59: What variables are you using?

We analyzed anomalies of the ocean heat content, the ocean heat transport, the barotropic stream function, and the ocean-atmosphere heat flux, as well as anomalies of the vertical temperature and the temperature at surface. We added these variables to to corresponding sentence and changed the sentence from "Here, we are using monthly data averaged to seasonal summer means over June, July, and August (JJA) from 1950 to 2022." to "Here, we are using monthly data averaged to seasonal summer means over June, July, and August (JJA) from 1950 to 2022 and analyzed anomalies of the ocean heat content, the ocean heat transport, the barotropic stream function, and the ocean-atmosphere heat flux, as well as anomalies of the vertical temperature and the temperature at surface."

L61: Incomplete sentence

We have to apologize, we could not find any incomplete sentence in the mentioned lines. However, we added a missing comma: "For time-lagged analyses, up to three years prior to 1950 are analyzed".

Section 2.2 Please revise this section into paragraphs. Rarely should one sentence be set off as its own paragraph in scientific writing.

Thank you for this comment. We have revised this section in a way that individual sentences no longer form a separate paragraph.

There is a break in line numbers for some reason, so the following are based to the 60:
L60+4: linearly

We changed "linear" to "linearly".

L60+8: temperature as in 2-m temperature (tas)?

We have to apologize for this imprecise wording. We changed "…we consider those JJA mean temperature anomalies…" to "…we consider those JJA mean 2m air temperature anomalies…"

L60+11-12: What do you mean by "if and where...can represent the sub-decadal timescales"? Timescales of what? By definition, isn't ERA5 the target? Why would it not be able to represent the sub-decadal timescales?

Our goal here is to investigate how MPI-GE can represent subdecadal temperature variability compared to ERA5. This investigation is intended to test the assumption that real-world processes are correctly represented in climate models (such as the MPI-GE here). For a better understanding, we changed the sentence from "We use a cross-spectral analysis, based on a multi-taper method to analyze if and where the MPI-GE and ERA5 can represent the sub-decadal time scales." to "We use a cross-spectral analysis, based on a multi-taper method to analyze how the MPI-GE can represent the sub-decadal temperature variability compared to ERA5."

L60+23-25: Can you show the following as regional average timeseries, even just for a single ensemble member?
- unfiltered JJA mean anomalies, with extremely warm summers marked
- *bandpass filtered JJA mean anomalies*

Yes, we can show such a figure. The figure illustrates that not necessarily all warm summers occur in a positive bandpass-filtered phase and show that summers which occur in a positive phase are clustered together with other extreme years.

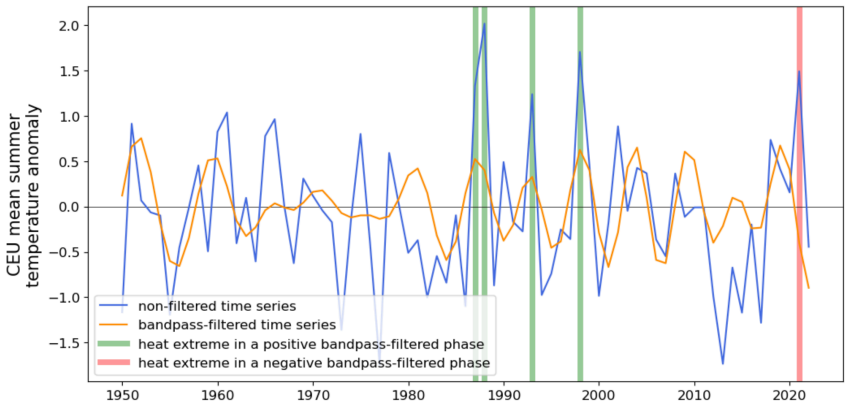

Figure A. Extremely warm European summers on sub-decadal time scales (for illustration purposes only ensembles member 50 is shown). Detrended time series of Central European surface air temperature anomalies, non-filtered (blue) and 5-10 year bandpass-filtered (orange). Heat extremes (above 90th percentile) that occur in a positive bandpass-filtered phase (therefore fulfill our selection criterium) are marked in green, while heat extremes that occur in a negative bandpass-filtered phase are marked in red.

Line numbering resumes:
L65-66: I don't understand the phrase: "the proportion ... has on the…"

To clarify this, we changed the sentence from "The scaled anomaly simply illustrates the proportion that a sub-decadal mean change has on the occurrence of an extremely warm summers compared to the overall occurrence of extremely warm summers."
to "The scaled anomaly simply illustrates the impact of sub-decadal processes on the occurrence of an extremely warm summers compared to the overall occurrence of extremely warm summers."

L73-74: But they disagree over your study region? The dominant mode in ERA5 is below 5 years, while MPI-GE shows a dominant 5-10 year mode, right? This is why I recommended decimating your ERA5 grid to the MPI-GE grid; it's not "fair" to compare a low-resolution model's "spatially average" climate to a high-resolution reanalysis and make statements about regional behavior.

We unterstand the doubts raised by the reviewer and regridded the EAR5 data therefore to the coarser MPI-GE grid. As pointed out by the reviewer, there are differences in the key region of this study. We now have added some further explantations in the manuscript that reflects the reasoning of the regional differences between the model and ERA5. However, we think that the model is able to capture the principle large-scale distribution of the sub-decadal variations compared to ERA5, thereby making it suitable to investigate the large-scale drivers. In the subsequent analysis, we concentrate on the model world to establish the mechanism. We added:"The cross-spectral analysis reveals, that MPI-GE is able to capture the large-scale distribution of the dominant sub-decadal variations compared to ERA5. This points towards the ability of the model to simulate the underlying large-scale mechanism in principle. However, still there are regional differences in the distribution, as can be found for example over South-East Europe. The reasons are yet unclear, and can be related to a regional displacement of the principal modes of large-scale variability compared to ERA5, or the limitation of the model to reflect the impact of the regional sea on extremes in this region (Beobide et al., 2023)."

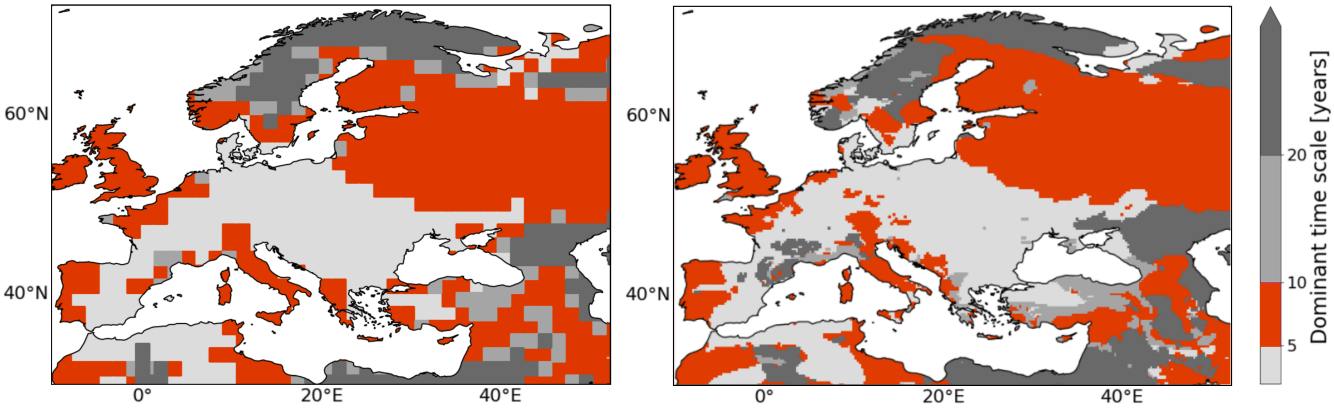

Figure B. Dominant time frequencies and their relation to extremely warm European summers. (a),(b) Cross-spectral analysis, performed using the multi-taper method, showing the dominant time scales of European surface air temperature variability in ERA5 (a) revised figure and (b) figure 1a of the main manuscript.

L78-79: Doesn't the presence of heat extremes in a particular 5-10 year period then determine that temperature is in a positive phase?

Thanks for this comment, we have to admit that this sentence is not very precise. Although extremes occurring within the 5-10 year period is not a necessary criteria, these positive phases occur over periods clustered warmer than normal, eventually extreme, summers. This means it is neither a sufficient nor a required criteria, but there may be a correlation between the two. A non-filtered extreme needs not necessarily to occur in a positive bandpass-filtered phase (see Fig. A). To clarify this we changed "Analyzing the ratio between all heat extremes and those occurring in a positive bandpass filtered phase, Central Europe stands out as the area with the highest percentages." to "Analyzing the ratio between all non-filtered heat extremes and those occurring in a positive bandpass filtered phase, Central Europe stands out as the area with the highest percentages.

Section 3.2: Please revise the one-sentence paragraphs

We have also revised this section to avoid individual sentences forming a separate paragraph.

L117-118: Is this shown in the figure?

The value is calculated as given in section 2.2, offering a possible way to interpret the values given in Fig. 3a. However, since this value is already given in the figure caption, we decided to remove this sentence from the main text.

L157: Average in time or across members or both?

We want to thank the reviewer for bringing this up. We are actually using the average in time and across ensemble members during extremely warm European summers. We changed the sentence from "In addition, the average position of the jet stream is shifted northward…" to "In addition, the average position of the jet stream, in time and across ensemble members, is shifted northward…".

L170: in the MPI-GE.

We agree with the reviewer and changed "The North Atlantic Ocean heat accumulation impacts the occurrence of extremely warm summers over Central Europe on sub-decadal timescales." to "The North Atlantic Ocean heat accumulation impacts the occurrence of extremely warm summers over Central Europe on sub-decadal timescales in MPI-GE."

L177-186: This seems like it should be in the introduction rather than only in the discussion.

We thank the reviewer and extended the introduction section. "… Further, the variability in the North Atlantic region has been shown to include a fully coupled atmosphere-ocean cycle with a period of about 7-10 years shown for various atmosphere- and ocean-related quantities, such as sea surface temperature and Gulf Stream indices (Czaja et al., 2001; McCarthy et al., 2018), ocean heat content and overturning stream functions (Martin et al., 2019), prominent winter sea-level pressure patterns (Czaja et al., 2001), and the North Atlantic Oscillation (DaCosta et al., 2002). This cycle is associated with an active role of the atmospheric heat and momentum forcing, together with a delayed effect of the redistribution of North Atlantic water masses (Czaja et al., 2001; Eden and Greatbach, 2003; Reintges et al., 2016; Martin et al., 2019). In fact, this process have a significant impact on European summer temperatures as demonstrated by Müller et al., 2020). However, the assessment of drivers for extreme temperatures on such long-term timescales is currently limited (Simpson et al., 2018; Wu et al., 2019), and their relevance for extreme summers remains uncertain."

L213-214: Please revise.

We agree and changed the sentence from "Lastly, this is a single model study which allows us to delve deeper into specific processes and model intricacies, which can contribute to model improvement and process understanding" to "Lastly, replicating this analysis for different climate models would be of great importance to sample potential model uncertainty in these results and help us gain further understanding of this mechanism".

L217-221: These sentences are repeats of the previous paragraph.

We agree and removed the doubled sentences.

L233-234: Is this a reemergence signal? If so, doesn't that connect it to European summer climate

The reviewer is right, as also stated in the text, the described mechanism can be seen to be attached to a fully coupled atmosphere-ocean cycle evolving in a 7-10 year period. Such oscillating behavior shows indeed a reemergence signal. We added this detail to the text: "Analyzing longer lags, in this case lag -7 to 0, prior to extremely warm European summers shows that the described mechanism can be seen as attached to a reemerging fully coupled atmosphere-ocean cycle evolving in a 7-10 year period."

2. Comments to answers to previous reviewer #2:

Comment of previous reviewer #2 referring to L19. Please be sure to read the paper carefully and cite it appropriately. "the main drivers of extreme heat are soil moisture deficits and moisture-temperature feedbacks" - This is not what the study found over Europe; heat in your study region is primarily adiabatic compression and subsidence.

Original comment of former reviewer #2 and four answer: *L19: Consider highlighting the recent work of Röthlisberger and Papritz (2023).*

We thank the reviewer for this comment and cited Röthlisberger and Papritz (2023).

We thank the reviewer for this clarification, which we address together with the following comment below.

Comment of previous reviewer #2 referring to L29-31. Again I'm not sure if this is a true statement based on the reference. They assess heat on shorter timescales and mention uncertainty in how the drivers will evolve given future circulation (inherently uncertain).

Original comment of former reviewer #2: *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 toned down and reworded 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 their relevance 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

We thank the reviewer for these two comments and cite Röthlisberger and Papritz (2023) now in a more appropriate way. For line 19 (now line 20) we decided to move the citation of Röthlisberger and Papritz (2023): "On time scales of days to several weeks, the main drivers of extreme heat are soil moisture deficits and moisture-temperature feedbacks (Seneviratne et al., 2006; Fischer and Schär, 2008; Vogel et al., 2017; Suarez-Gutierrez et al., 2020a), diabatic heating, adiabatic compression and advection (Röthlisberger and Papritz, 2023), and large-scale atmospheric patterns such as atmospheric blocking and the North Atlantic Oscillation (Meehl and Tebaldi, 2004; Horton et al., 2015; Li et al., 2020; Suarez-Gutierrez et al., 2020a). However, these short-term drivers of extreme temperatures could be influenced and conditioned by mechanisms on longer time scales". For line 32-34 we decided to remove the citation.

Comment of previous reviewer #2 referring to Section 2.1. What variables are you using?

Original comment of former reviewer #2: *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 added 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."

We analyzed anomalies of the ocean heat content, the ocean heat transport, the barotropic stream function, and the ocean-atmosphere heat flux, as well as anomalies of the vertical temperature and the temperature at surface. We added these variables to to corresponding sentence and changed the sentence from "Here, we are using monthly data averaged to seasonal summer means over June, July, and August (JJA) from 1950 to 2022." to "Here, we are using monthly data averaged to seasonal summer means over June, July, and August (JJA) from 1950 to 2022 and analyzed anomalies of the ocean heat content, the ocean heat transport, the barotropic stream function, and the ocean-atmosphere heat flux, as well as anomalies of the vertical temperature and the temperature at surface."

Comment of previous reviewer #2 referring to L65. What is being shown here? There is no y-axis label.

Original comment of former reviewer #2: *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 clarified in the revised manuscript which detrending method we used.

We have to apologize for this confusion. The y-axis of the figure is showing the temperature anomalies [°C].

Comment of previous reviewer #2 referring to Figure 1. Could you comment on this in the manuscript?

Original comment of former reviewer #2: *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

We unterstand the doubts raised by the reviewer and regridded the EAR5 data therefore to the coarser MPI-GE grid. As pointed out by the reviewer, there are differences in the key region of this study. We now have added some further explantations in the manuscript that reflects the reasoning of the regional differences between the model and ERA5. However, we think that the model is able to capture the principle large-scale distribution of the sub-decadal variations compared to ERA5, thereby making it suitable to investigate the large-scale drivers. In the subsequent analysis, we concentrate on the model world to establish the mechanism. We added:"The cross-spectral analysis reveals, that MPI-GE is able to capture the large-scale distribution of the dominant sub-decadal variations compared to ERA5. This points towards the ability of the model to simulate the underlying large-scale mechanism in principle. However, still there are regional differences in the distribution, as can be found for example over South-East Europe. The reasons are yet unclear, and can be related to a regional displacement of the principal modes of large-scale variability compared to ERA5, or the limitation of the model to reflect the impact of the regional sea on extremes in this region (Beobide et al., 2023)."

Comment of previous reviewer #2 referring to Figure 1. In this case, it would be interesting to see what the spectrum would be if you pooled all the ensemble members and computed the spectrum.

Original comment of former reviewer #2: *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 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 added a sentence discussing the peaks on other time scales. Further, we agree that it would be helpful to see the spectra of the individual ensemble member, however the single spectra of our 100 ensemble members are basically all over the place and thus provide no additional knowledge and would require a larger y-axis range. We decided to leave the figure as it is in order to make it not too messy and to focus with the chosen y-axis range on the ensemble mean spectrum.

We thank the reviewer for this comment and added the requested pooled ensemble spectrum below. However, we do not plan to add this spectrum to the manuscript, since the pooled ensemble features discontinuities at the points where individual members are concatenated, which leads to artificially induced errors in the resulting spectrum. Therefore, we plan to continue with the mean of all ensemble member spectra.

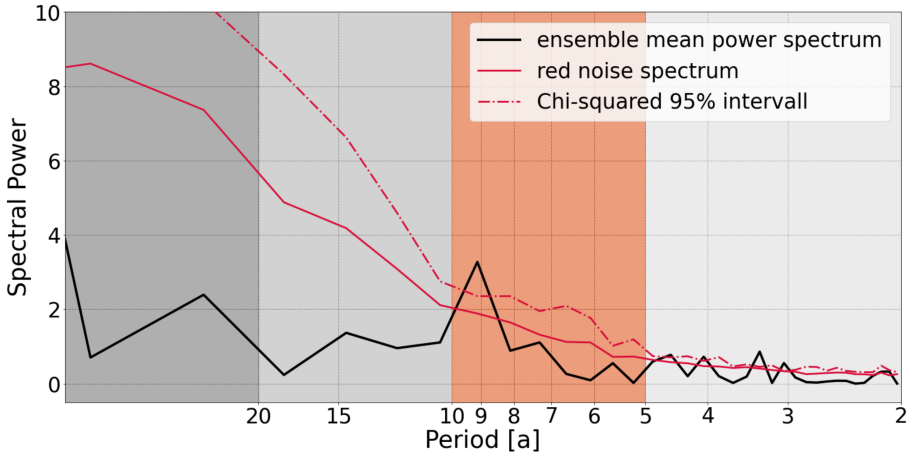

Figure C. Power spectrum of Central European surface air temperature (black line) in MPI-GE (pooled ensemble spectrum). The significance is shown via a red-noise spectrum (solid red line) and the chi-squared 95% interval (dashed red line). Period 1950-2022. (For comparison, see Fig. 1d main manuscript.)