Final author response to reviewers for wcd-2023-6
https://wcd.copernicus.org/preprints/wcd-2023-6/

Title: Strengthening gradients in the tropical west Pacific connect to European summer temperatures on sub-seasonal timescales

Summary of important proposed non-textual changes:

- Recompute results with consistent temporal aggregation for WPD and T2m. Previously this was inconsistent: (3-week sst, 2-week gap, 4-week t2m), now it is consistent (4-week sst, 2-week gap, 4-week t2m). The result is minor quantitative changes to Figures 3 to 8. Our original interpretations remain unchanged.
- Introduce a new supplementary plot (Fig. R5) for clarifying the optimality of chosen sub-seasonal timescales (4-week sst, 2-week gap, 4-week t2m), and sensitivity to this choice of timescales. The plot also highlights the difference between our lagged/subseasonal approach (panel A) and the concurrent/seasonal approach of previous studies (panel B).
- Add two supplementary plots (Fig. R7 and R8) with maps showing the robustness of correlating grid cells under different temporal aggregations. This is to help the reader understand why the current WPD boxes are placed where they are placed, and to assess alternative spatial choices.
- Addition of two panels to the original Figure 8 to visualize subdivision of negative WPD phases (Fig. R1). This eliminates one previously unexplored potential reason for the apparent contradiction in explaining the western european temperature trend. (See major comment 1 of Reviewer #1).
- Addition of significance tests to the composite plots of Figure 5 and 7. The test is based on a bootstrapping procedure, and insignificant differences are masked out (see Fig. R3 and R4). (See major comment 4 of Reviewer #1)

Summary of important proposed textual changes:

- Textually stress the separation between the timescales of phenomena. Originally we framed the long-term West Pacific Warming as 'La Nina-like'. This would suggest that it has an interannual component, but instead it is a long-term strengthening of dT/dx in the basic state which can 'amplify the SST gradients that arise due to ENSO variability, specifically the La-Nina-like anomalies'. With this textual rephrasing the descriptions become truer to previous findings (see major comment 3 of Reviewer #1).
- Clarification that the t2m response classes in Western Europe are defined as terciles of a rolling 21-year t2m distribution. This means the measured connectivity is relative to the long term warming trend that is itself the consequence of forced/unforced dynamic/thermodynamic changes. This concerns the beginning of section 4, end of section 5, and discussion. (see also major comment 1 of Reviewer #1)

REVIEWER 1

This work examines the process that anomalous heating over the tropical-subtropical western North Pacific can affect European temperature variability, with the modulation from the North Atlantic SST, on subseasonal timescales in the summer season. The authors defined a West Pacific Dipole (WPD) index to characterize the combined zonal and meridional heating contrast in the western North Pacific. They found that positive WPD can trigger teleconnections towards Europe, leading to warm T2m anomalies over western Europe in 3-6 weeks. They also found that the negative phase of WPD has become more frequent since 1980, consistent with the "Western Pacific warming mode". However, the summertime temperature over western Europe has not shown any cooling trend since 1980. The authors argued that the North Atlantic SST also plays an important role in modulating the teleconnections.

The results in this manuscript potentially present valuable improvement to our understanding of subseasonal climate predictability and prediction. I have some major comments and clarification questions, that hopefully could help to improve the manuscript.

Major comments:

1. A large portion of the manuscript depends on connecting the warming SST trends in the western Pacific to the increasing occurrences of negative WPD. I am not familiar with the "Western Pacific warming mode", but my understanding is that this is for a long-term warming trend in the SST (basic-state shifts). On the other hand, the WPD phases and their teleconnections to the European T2m characterized in this manuscript are subseasonal variability.

   I guess I am confused about the connection here: one is change in the mean (shift in the distribution) and another one is change in the standard deviation (change in the shape of the distribution). For example, in L285-289, the authors stated: "though negative WPD has occurred more frequently, western Europe has not experienced a relative cooling as expected." Based on the results shown here, WPD affects the "subseasonal variability" in western European t2m. More frequent of this event is not necessarily equal to change in seasonal mean temperature, if the standard deviation has become larger (e.g., more frequent cold subseasonal events, but also more frequent and stronger warm subseasonal events).

   In short, it would be helpful if the authors clarify the different timescales across the phenomena they have mentioned in the manuscript. Also, in L290-296, the authors examined that the positive WPD phase has become more likely to result in positive t2m anomalies in western Europe. It is also helpful to show the negative WPD phase: has negative WPD phase become less likely to result in negative t2m anomalies in western Europe?

   In this comment you point to the apparent contradiction between the long-term trend towards more frequent negative WPD phases (which would result in cooler than normal western European t2m) and the obvious strong increase in summertime western European t2m. You seem to ask us three questions:
   1 Can I interpret the long-term Pacific trend as a shift in the basic state, and is this decoupled from the teleconnection which seems to be a sub-seasonal form of variability?
   We argue in the paper that yes, the long term trend is a shift in the basic state of the Pacific, but also that it is highly coupled to the sub-seasonal scale. Therefore the answer to the second part of your question is negative. The WPD index is simply a temperature difference between central and western Pacific on a 4-week timescale. Since the long term trend shifts the mean/basic state to stronger temperature differences, this leads to more frequent extreme (in this case negative) WPD states. The WPD states are based on a simple thresholding of the T-differences, which are now more often exceeded. See our response to your major comment 3 for how we clarify the wording.
   2 Is such a decoupling between long-term trend and sub-seasonal variability also applicable to the western European t2m? Meaning the sub-seasonal connection from WPD to t2m only changes variability (e.g. more short cold events), whereas the mean t2m warming would be unaffected?
   Also for the t2m response we think that such a conceptual decoupling makes little sense. The long term trend in t2m is affected by many factors, among which are forced thermodynamic changes and forced and unforced dynamical changes. Explaining long term trends would require an investigation of all those factors, and is not a goal of this paper. To make this clearer we propose to remove the last sentence of the abstract, and to rewrite/de-emphasize a few other sentences.
   Our actual goal is to show that a summertime teleconnection with sub-seasonal character seems to have emerged amid long-term changes. To this end our analysis is based on a forced decomposition. As will be explained in response to major comment 2, we implicitly detrend the response in western Europe by defining the cold and hot t2m classes within a rolling 21-

year window. In the manuscript we mention that without such a procedure any diagnostic measure of sub-seasonal connectivity in the data would be muddied by the thermodynamic contribution to the long term t2m trend. In short: the sub-seasonal connection, as we define it, cannot contradict the long term trend because it is defined relative to it. There only appears to be a contradiction because of carelessly worded sentences such as the one you quote:

"though negative WPD has occurred more frequently, western Europe has not experienced a relative cooling as expected."

We propose to clarify such phrasing as:

"as negative WPD has occurred more frequently, and significantly relates to the cold tercile of the rolling 21-year temperature distribution, it seems that without this teleconnection the warming due to other factors in this region could have been even stronger.

With this last statement we acknowledge that the teleconnection (though defined in a rolling fashion) could be one of the factors influencing the long term trend. Additionally we want to more clearly state in the beginning of section 4, the end of section 5, and in the discussion, that the boundaries of the t2m classes shift due to the combined effect of other factors.

3 Does the apparent contradiction occur because negative WPD phases (though more frequent) become less good at causing negative t2m responses.

This is an interesting suggestion. After all, in Figure 8 we present an analysis that shows that over time positive WPD phases result in more positive t2m anomalies, possibly due to the influence of the Atlantic jet. Why could the converse not also be an explanation for the heating in western Europe, namely that the negative WPD phase has become less effective in resulting in negative t2m anomalies? We propose to add this decomposition of t2m responses as extra panels in the bottom row of Figure 8:
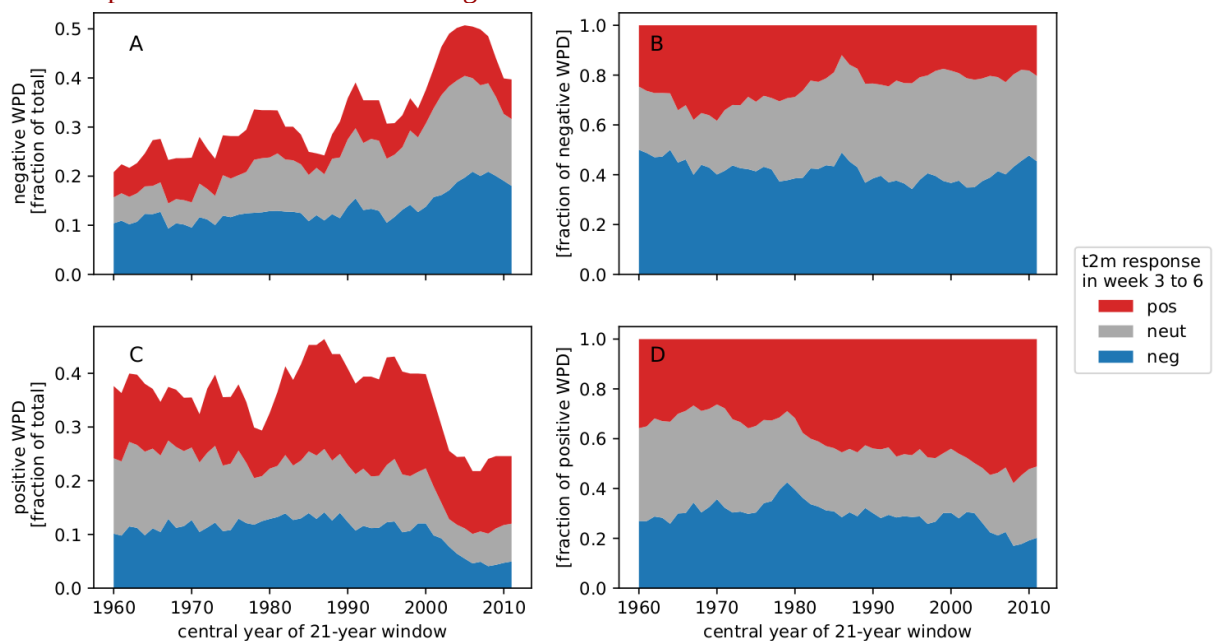

Fig. R1, New proposed panels for Figure 8, now including negative WPD phases in the top row.

In it one can see that the blue fraction in the top right panel is roughly constant over time. Negative WPD is thus as likely to be followed by a relatively cold t2m episode in the beginning of the data, as at the end of the data. Therefore the answer to your question is no. We propose to incorporate this in the part of the discussion where Figure 8 is discussed.

2. Also relevant to the timescale issue. In Fig. 4, WPD index is not detrended, but t2m over western Europe is detrended. As the main purpose is to examine how WPD affects western European t2m on subseasonal timescales, why the inconsistency here?

First, we don't think this is inconsistent, but actually in line with the respective properties of the WPD precursor and t2m response. For European t2m we look at mean temperatures in a particular region and this mean has a strong upward trend. For WPD classification we look at a dipole and thus a difference in temperature between two regions and this is far less affected

by global mean long-term warming trends.

Second, as said above, we need to implicitly detrend the t2m response because otherwise the (pos,pos) class would be artificially overpopulated in the last decades by the thermodynamic trend, and any measure of sub-seasonal connectivity would be distorted.

Third, we could treat precursor (WPD) and response (t2m) more consistently by detrending the WPD index as well. But Figure 3A already provides a hint that this will not alter results much because WPD is hardly trended. To demonstrate this we present here a version of Fig. 4 with detrended WPD (Fig. R2). Results are not greatly altered as the increase of negative WPD phases in recent decades is still present, just like the emergence of the teleconnection (red line Fig R2). There is more emphasis on positive WPD phases in the last 40 years. But the overall change remains small, and as said, we think our original approach is in line with the properties of the data. Therefore we refrain from detrending WPD.
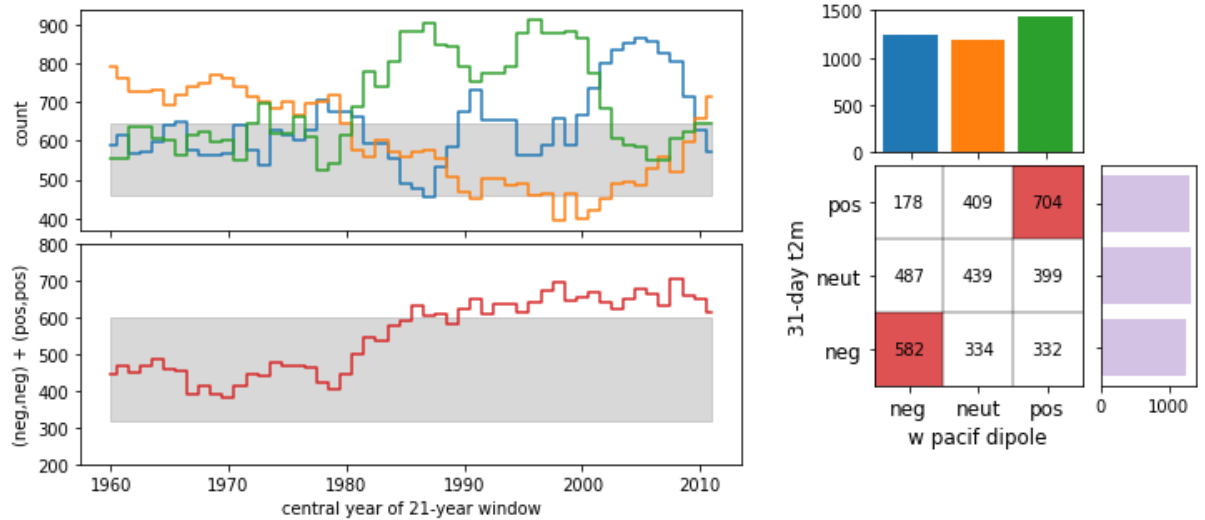

Fig. R2 Version of Figure 4 achieved when detrending WPD. Will not be adopted in the paper.

3. The "Western Pacific warming mode" frequently mentioned in this manuscript (Funk and Hoell, 2015) characterizes the warming extending from the subtropical North Pacific to the subtropical South Pacific. On the other hand, the "La Nina-like warming" has been commonly referred to the trend in the zonal SST gradient in the equatorial Pacific, especially the cooling over the equatorial eastern Pacific (e.g., Seager et al. 2022; Lee et al. 2022). It would be better to distinguish these two. I don't think the "Western Pacific warming mode", "La Nina-like warming" or the cooling over the equatorial eastern Pacific are interchangeable. (e.g., L146-149).

In terms of dominant timescales the warming mode and ENSO-variability are indeed not interchangeable (in fact, the West Pacific Warming Mode is diagnosed by Funk and Hoell (2015) in ENSO-residual sst's, implying at least some first-order independence). We agree that ENSO is a shorter term (inter-annual) variability than the decadal warming. The long term warming results in a basic state with a stronger SST gradient over the whole pacific basin, because warming is more dominant in the west pacific. This increase in overall gradient aligns with the slight negative trend in the ENSO indices, which, when viewed spatially, is most expressed in the relative cooling of the equatorial eastern Pacific. Despite the difference in timescales and different spatial emphasis the two phenomena are thus connected by a similar effect on the overall gradient: the long term warming creates a pattern that just like the ENSO trend, and especially during La-Nina episodes, leads to strong gradients over the entire pacific.

Funk and Hoell (2015) describe the long term warming as resulting in: "a stronger western Pacific SST gradient which can interact with La Niña to intensify the Walker circulation and modulate Northern Hemisphere upper circulation patterns"

To be more faithful to this view (and to not make the phenomena appear interchangeable) we

now refer to west pacific warming as 'amplifying La Nina-like gradients' instead of 'La Nina-like'.

4. Statistical significance tests should be included in the composite plots (Figs. 5 and 7). Also, whether the differences in Figs. 5 and 7 are statistically significant should be addressed.
This is a valid point, and is also raised in Major comment 3 by Reviewer #3. To include statistical significance tests we decided to adopt a block-bootstrapping procedure. The composite means in Fig 5 and 7 are based on groups delineated by the WPD and t2m tercile classes. These groups can be considered subsets of the entire population of summertime anomalies from 1980-2021. Our first step was to resample that entire population of anomalies to obtain 500 draws of potential composite means when group sizes are as in Fig 3C. The second step was to subtract those composite means to exactly re-create the situation in the middle columns of Fig. 5 and 7. The result is a sampling (null) distribution of differences that could arise by chance. Then we perform a two-sided test per gridcell, which tests whether the actual difference falls outside the empirical quantiles of the sampling distribution (q0.025 and 0.975 for alpha = 0.05, or q0.05 and 0.95 for alpha = 0.1). For grid cells where the test fails, the difference is insignificant and therefore masked. This procedure will of course be described in text and caption, and the proposed new panels for Fig. 5 and 7 are in Figs R3 and R4.
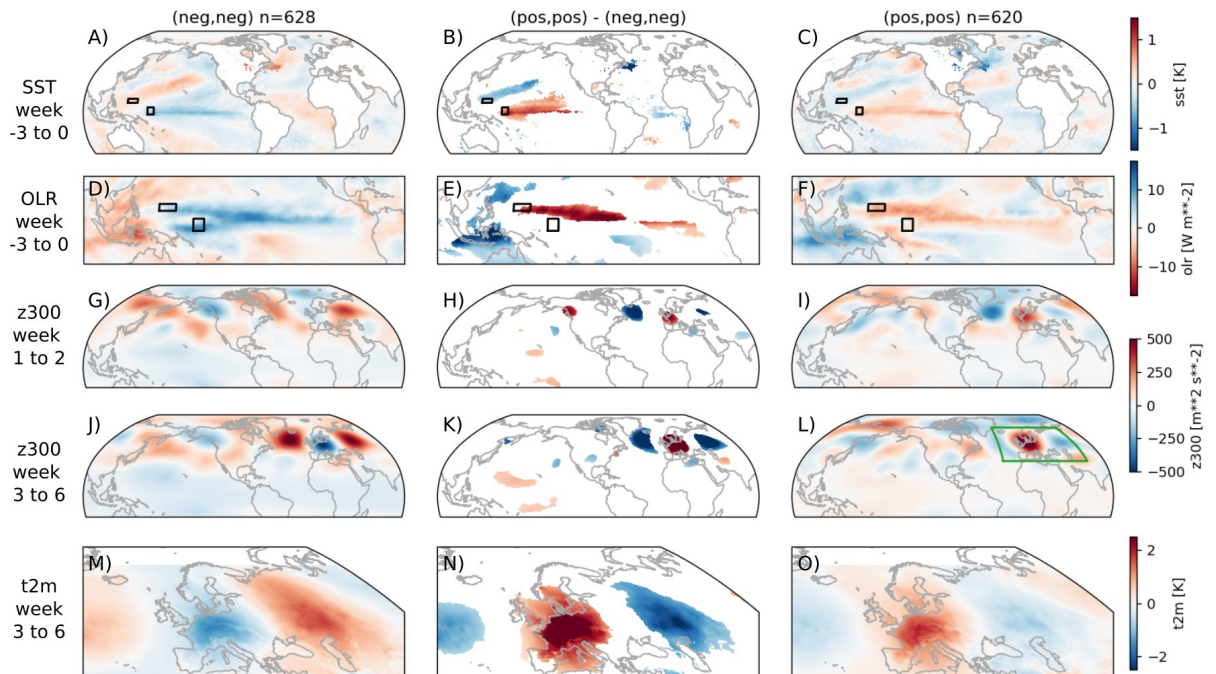

Fig. R3 proposed new panels for Fig. 5. Middle column only displays significant differences (conservative block bootstrapping, with alpha = 0.05). Also the t2m composite is extended to include eastern Europe and Russia.

We chose block-bootstrapping because of the potential temporal overlap and dependence between samples in our population. With block bootstrapping one randomly samples blocks of samples for which dependence can be expected. In this case we employ entire summer seasons as block size, which can be considered conservative because the actual groups need not contain a summer in its entirety. The result is that the null distribution might display a heightened variability, making it more difficult to pass the test. Another reason why the null distribution can be considered conservative is because it contains differences that might have arisen by more than chance alone. The samples are namely taken from the entire 40 years and can therefore contain diluted versions of the studied effect. Overall we can thus be confident that remaining (unmasked) differences are highly significant.

5. The teleconnection patterns shown in Figure 5 is quite similar to the circumglobal teleconnection pattern shown in Ding et al. (2011). As mentioned in L194, Ding et al. (2011) also showed similar heating dipole pattern. Furthermore, Ding et al. (2011) also suggested this pattern is more frequent in the developing ENSO summer (Fig. 11a in Ding et al. 2011). I

am wondering how much the patterns shown in this manuscript are similar to the CGT pattern in Ding et al. (2011). Is it possible there are actually the same thing?

They indeed seem to be the same phenomena. In both studies, the teleconnection mechanism is one of the tropics influencing the midlatitudes through Rossby waves. One further similarity is indeed the preferential occurrence in summers preceding ENSO peaks. However, we think we also shed a different light on the phenomena. We show that the pattern that shows up in concurrent seasonal analyses, as in the analysis of Ding et al. (2011), has a sub-seasonal substructure consisting of a lagged relation, as the wave pattern generally takes some time to establish. In the analysis of Ding et al. this timing aspect is masked by the fact that they take concurrent JJA averages. In this sense our approach that takes a lag of 2 weeks is closer to the physical reality of the time needed for such a connection (at least in numerical experiments (Branstator, 2014)). Quantitatively the link also seems somewhat stronger than concurrent analyses. We propose to provide evidence for this in the supplementary material. See minor comment Lines 69-71 of Reviewer #2, and major comment 2 of Reviewer #3.

6. Throughout the manuscript (especially in Section 5), the authors mentioned other studies frequently, e.g., Ding et al. (2011), Vijverberg and Coumou (2022), SEA pattern, Behera et al. 2013, O'Reilly et al. (2018) etc. The authors should provide more background information regarding these studies. Also, there are lots of consistency between your work and previous results. Which part of their results support the element you mentioned here? What are the unique parts of your results? We should not assume all the readers have read all these references or will read all of them along the way.

You are correct. In section 5 we use the composites to discuss the wave pattern, and we highlight the centres of action that match the patterns found in these other studies. We generally do not discuss differences (except for e.g. the difference with the concurrent seasonal analysis of Ding et al. (2011)). We propose to expand the text in section 5 to also cover differences.

Minor comments:

L94-97: Please include the longitudinal ranges of Nino3 and Nino4 regions. Also, how about the climatology? Or SST anomalies are relative to zonal mean (between 20N-20S), rather than relative to any climatology?

Ranges are now included. And we expand the sentence:

"Relative ENSO indices are computed like regular ENSO indices as the average of the SST anomalies within a Niño region (climate normals defined from 1981-2010), from which the average SST anomaly between 20N and 20S is subtracted. This makes the indices less distorted by global warming and more useful for describing teleconnections (van Oldenborgh et al., 2021)."

L122: "eastern" edge of the Nino4 area: "western"

replaced

L124-125: please provide the lon/lat ranges of the components 1 and 2.

Now included

L194: "Both types of heating contrasts were found to be important by Ding et al. (2011)" -> important to what?

This should be "for the large-scale circulation (z200) over Europe". We will include this in text.

L208-209 & L283: why not include t2m anomalies over Eastern Europe and Russia in Figure 5?

This was not included because we did not download that data. But now this is included in the renewed Figure 5. See Fig R3..

L250-254 & Fig. 7: it would be helpful to also include wave patterns (Z300 anomalies) in Figure 7.

We agree and propose to include panels with Z300 as in Fig. R4. In Panels G and I one sees that up to the west coast of North America the response to positive WPD is similar (in both the bottom left and

bottom right, high pressure is centered over the Alaska/Vancouver area), but that the significant difference (panel H) exists further east, with the low pressure south of greenland and the high pressure over europe. This supports our conclusion that the strength of the Atlantic jet modulates the downstream travel of the Rossby wave response.
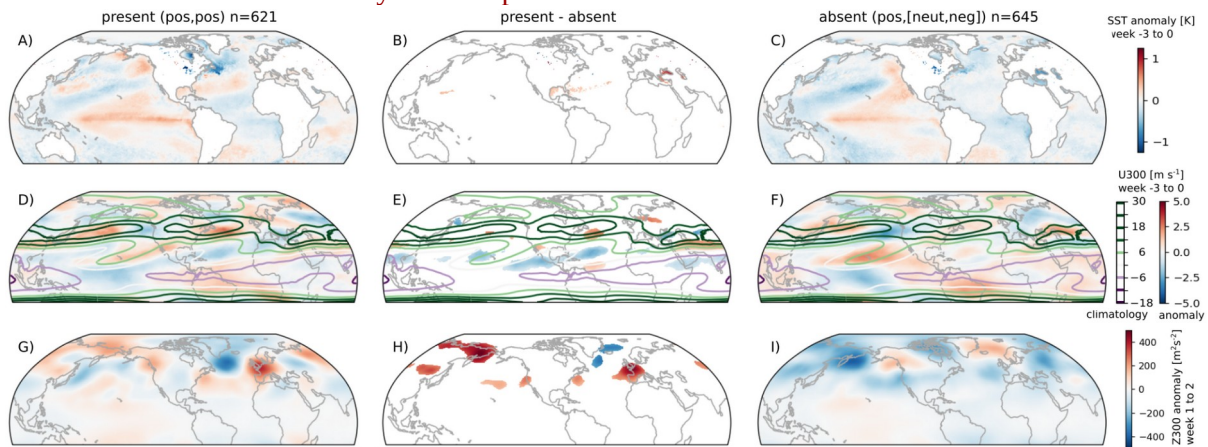

Fig. R4. Proposed new panels for Figure 7. Bottom row now displays z300 anomalies. Middle column shows only the significant differences (conservative block bootstrapping, as per major comment 4).

L281-284: Are there any evidences supporting that the more frequent warming & high pressure over Eastern Europe and Russia are statistically significant related to the increasing WPD negative phase? This evidence provided in this study is indirect, basically because the reasoning consists of a combination of Figure 4 and Figure 5. Figure 4 provides evidence that the link to western europe has become statistically significant. Figure 5 shows that the surface imprint of the teleconnection pattern is opposite for Western Europe and Eastern Europe and Russia. This is now clearer because we've added the t2m in that region to the composite plots, as per minor comment L208-209. We however consider it out of scope for the current manuscript to provide direct evidence by repeating the analysis of statistical significance behind Figure 4, but with a different, Russian response variable.

REVIEWER 2

The study by van Straaten et al investigates the link between sub-seasonal variability in tropical Pacific convection, closely related to ENSO, and summertime circulation over the Euro-Atlantic. The study shows that SST gradients associated with the dipole represent a combination of ENSO variability and West Pacific warming. These gradients in the West Pacific are followed by quasi-stationary waves that linger for multiple weeks. Situations with La Niña-like gradients are followed by high-pressure centers over Eastern Europe and Russia, three to six weeks later. The study also confirms earlier findings that the summertime connectivity between the Pacific and Europe has shifted in recent decades and partly explains the increased occurrence of high sea level pressures and summer temperatures over the European continent. The evidence the authors show is compelling and extends to sub-seasonal scales the findings reported by other authors who have analysed teleconnections from the Pacific towards Europe. I think the study is worth publishing after minor amendments to the text and figures.

Minor comments:

Line 44: There is a typo in consequence.
Fixed

Lines 69-71 Even when a two weeks lag has been used, I consider it important to show that this is the optimal time lag since it is not so obvious that the signals arrive with the same lag.
This links to the question of whether this is a sub-seasonal or seasonal/concurrent phenomena (major comment 5 of Reviewer #1).
We agree that this is indeed good to show. We propose to do so with Supplementary plot R5. We see

that strong relations are found when looking at concurrent near-seasonal averages (9-12 weeks, panel B). But equally strong or even stronger relations are also found for shorter timescales of 8 to 4 weeks (panel A), in which case the strongest link is <u>not found concurrently</u>, but exists when the variables are related with a separation of zero to one week (between the end of the WPD period and the start of the t2m period). That a certain separation is required is even clearer when we follow the correlation values horizontally along the 2-week aggregation (second row from the bottom). These values peak at lags of 2 and 3 weeks, which corresponds to the common finding in literature that Rossby waves from tropics need two weeks to fully establish in the mid-latitudes (see also Major comment 2 of Reviewer #3). Overall this leads us to conclude that the link we are studying in this manuscript could be found in concurrent seasonal aggregates, but that this signal consists of an underlying lagged sub-seasonal connection. And with the hindsight of this new figure we can conclude that the timescales, as informed by our previous study, and as used in the presented analysis (4-week aggregates, and a 2-week separation) are at least in the vicinity of the optimum.

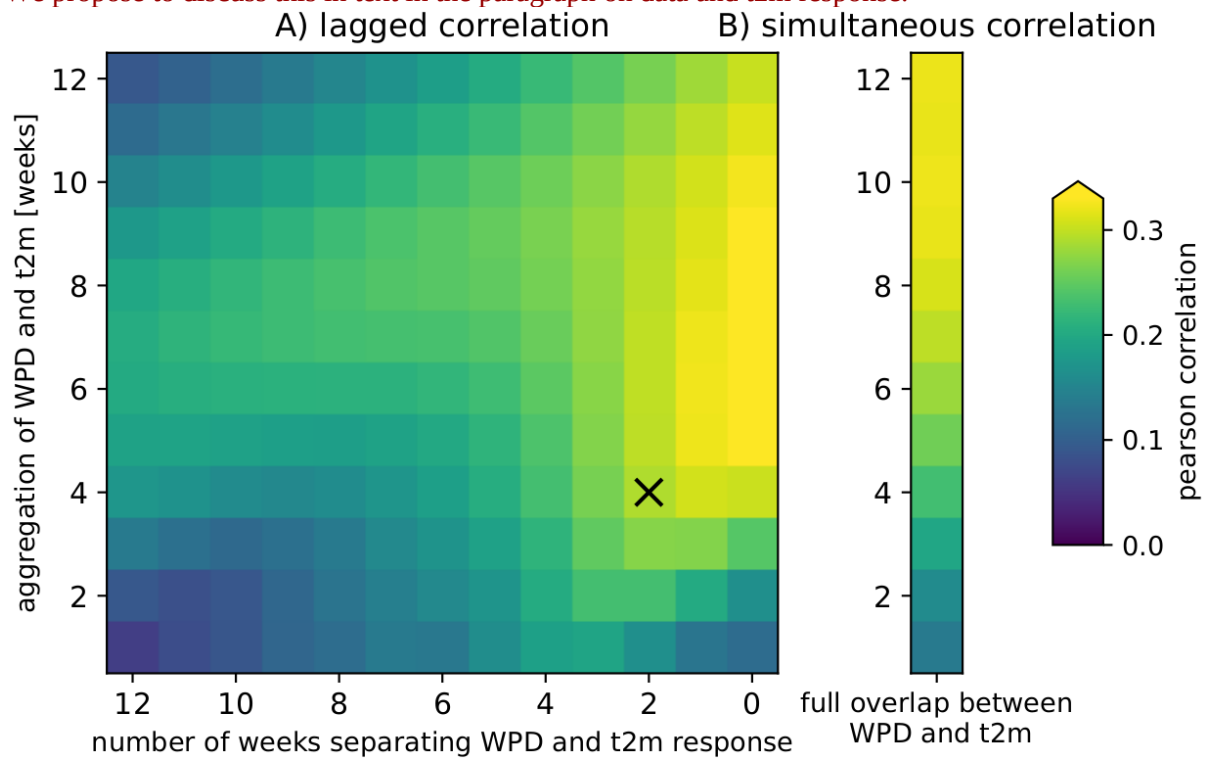We propose to discuss this in text in the paragraph on data and t2m response.

Fig R5 correlation with detrended t2m at different lags and levels of temporal aggregation. Separation is defined as the number of intervening weeks (after the end of WPD, before the start of t2m response). B) is correlation when fully overlapping.

Line 86-87: A clear reason should be stated on why the different times for averaging the different variables are used. The authors provide citations to other works, but still in my view, for clarity's sake, the authors should mention in this paper the reason for using different time lengths for averaging.

The reason for the differing timescales was the fact that we used precomputed monthly indices for relative ENSO and PDO. However, we have now recomputed all results using a 4-week ('monthly') aggregation for our manually defined SST indices (WPD and WNP), which matches the timescales of ENSO and PDO, and also that of the average t2m response in week 3 to 6.

Figure 4B. It would be better to show the result in frequency, meaning the occurrences are divided by the total amount of occurrences and not occurrences.

Good idea. Though at least for panel C we do prefer to keep counts, as this gives an idea how many samples are present. These numbers will later match the number of samples going into each composite, presented in Figure 5 and 7. The proposed new plot is Fig R6. We propose to also use the fraction of total amount in the original Fig. 8. See the new Fig R1.
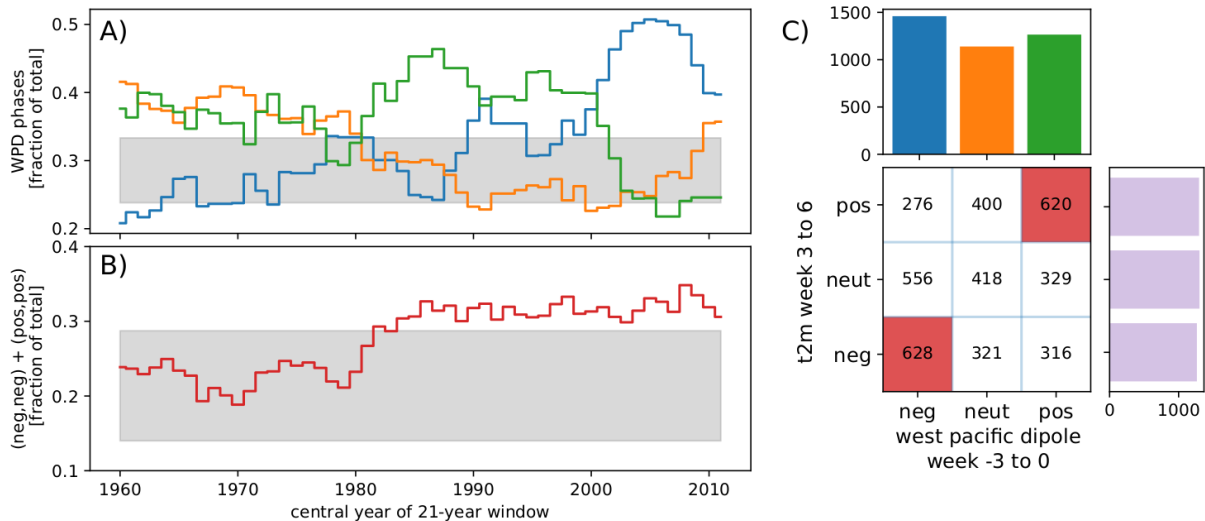
Fig R6. New version of Fig 4 with frequencies instead of absolute counts for panels A and B, and with consistent temporal aggregations (4-week wpd, 2-week gap, 4-week t2m).

REVIEWER 3

The paper by Van Straaten et al. studies the connectivity between the western Pacific and Western Europe, with a focus on the sub-seasonal predictability of summer European temperatures related to tropical Pacific sea surface temperatures (SSTs) as well as the modulation of the teleconnection by north Atlantic SSTs. The authors base their analysis on the ERA5 reanalysis and other observed sea surface temperature datasets.

I have a few comments that need to be addressed before a possible publication of the paper in Weather and Climate Dynamics.

Major Comments:

1. My first comment is that the study is based on one reanalysis only (ERA5) while other reanalyses do exist and could be used to assess the sensitivity of the results to the choice of ERA5. Given that they use the 1950-2021 period for ERA5, I am also assuming that the authors have used the preliminary ERA5 back-extension (1950-1978) and not the final ERA5 recent release since 1940. If it is not too cumbersome, I would suggest to update the analysis with the new ERA5 data (at least for the figures using years before 1979).
   The study is indeed based on the preliminary back-extension of ERA5. According to the documentation https://confluence.ecmwf.int/display/CKB/The+family+of+ERA5+datasets the preliminary back-extension could suffer from too intense tropical cyclones. As we do not study these, we think that there is no reason for concern, and that differences will be negligible. Also, the redownloading of all data is quite cumbersome on the IT infrastructure that we employ, which is why we prefer not to extend to 1940.
   We acknowledge that other reanalyses do exist and that these could be used to check the sensitivity to the data product. For aspects of this study this has already been done, such as for the west-pacific warming mode (Funk and Hoell, 2015), and the tropical central pacific cooling (Seager et al, 2022), meaning we can be confident that the increase of strong 4-week gradients as captured by negative WPD is robust. Nonetheless we acknowledge the limitation and propose to mention this in section 7.

2. In addition to spatial aggregation for the predictand, the authors use a monthly (4 weeks) aggregation for t2m and 3 weeks for SSTs. It would be interesting to see the results with a shorter temporal aggregation (10-15 days for instance, as the authors explicitly mentioned this period in their introduction as the time needed for the waves to establish) and comment on the potential differences. If these long averaging periods are required to obtain significant results (albeit with low amplitude for the correlations, 0.15–0.2), it is interesting information.

We agree that the sensitivity to temporal aggregation is interesting to test. See our response to the suggestion of Reviewer #2 (Line 86-87). In that response we propose to include Figure R5 in the supplementary materials. This figure displays that a two-week aggregation actually leads to less prominent correlation, but that at this two-week aggregation it is even clearer that one needs to employ a sub-seasonal (2-week) separation between SST and t2m response to find the clearest connection. This aligns well with the period needed to establish mid-latitude Rossby wave responses, as mentioned in the introduction of the manuscript.

The choice of timescales is now more elaborately discussed in the data and t2m section.

3. Statistical significance: no statistical significance of the composites is provided for figures 5 and 7. They need to be performed with an appropriate method (for a cautionary note about t-test, see "The Use of t values in Composite Analyses" by Brown, T.J., and Hall, B.L., Journal of climate, vol. 12. 2941-2944, 1999). It would be nice to have more details in the paper about the method used to derive statistical significance for Figure2 (without having to go to a previous publication). In relation to the false discovery rate procedure, it is not clear to me what alpha (page 6, line 119) is (in Wilks BAMS2016, alpha is the nominal level of the statistical test).

    Significance testing has been added to composite plots. As an appropriate and conservative method we chose to do block-bootstrapping. See our response to major comment 4 of Reviewer #1

    Regarding the false discovery rate procedure for the correlation plots we will add more detail to the relevant paragraph.

4. Figure 2 and the definition of the WPD index: as it is now, there seems to be a lot of subjectivity in the choice of the two boxes (comp_1 and comp_2). For instance, one could have extended the comp_1 westward and southward (including regions with significant correlations). Similarly, the comp_2 box could have been slightly extended eastward and southward. Is it possible to present a rationale for the choice of the boxes? Or at least, one would like to see a sensitivity analysis regarding the definition of the WPD index. It is a bit surprising to notice that a large spatial aggregation is performed for the predictand (t2m in Europe), contrasting with the small spatial extent of the regions used to estimate the WPD index.

    Figure 2 gives the impression that the boxes could indeed have been extended to include some more significantly correlated cells. However, whether these extra cells outside of the boxes show up, depends on the exact chosen timescales and lags. In Fig. R7 we illustrate how the correlation patterns can shift. Influential is also whether one performs the stratified cross-validated correlations on the whole reanalysis period (shown) or only on the last few decades (Fig. R8). We made a subjective choice for boxes that seemed to capture those cells that are relatively consistent across aggregation timescales, lags and subsets of the data, i.e. we chose the boxes such that they would capture 'the robust centres of the phenomenon'. In a subjective sense this means that the presented analysis is one of the more robust options among all possible placements of the boxes. Objectively it is a fact that the SSTs in these boxes were the most useful predictor in a previous study (van Straaten et al. 2023, something we already mention in the manuscript). And it is also a fact that in the current diagnostic framework the placement works, because links to Europe are found that are in agreement with previous studies. Still we want to acknowledge that alternatives could potentially also work. We propose to include Fig R7 and R8 as supplementary material to illustrate how correlation patterns can shift when different choices are made. In this way the reader can form an opinion about alternative definitions of the index. In the text of section 3 where the boxes are introduced, we propose to discuss the influence of aggregation and lag on potential placement, and to refer to the supplementary material.
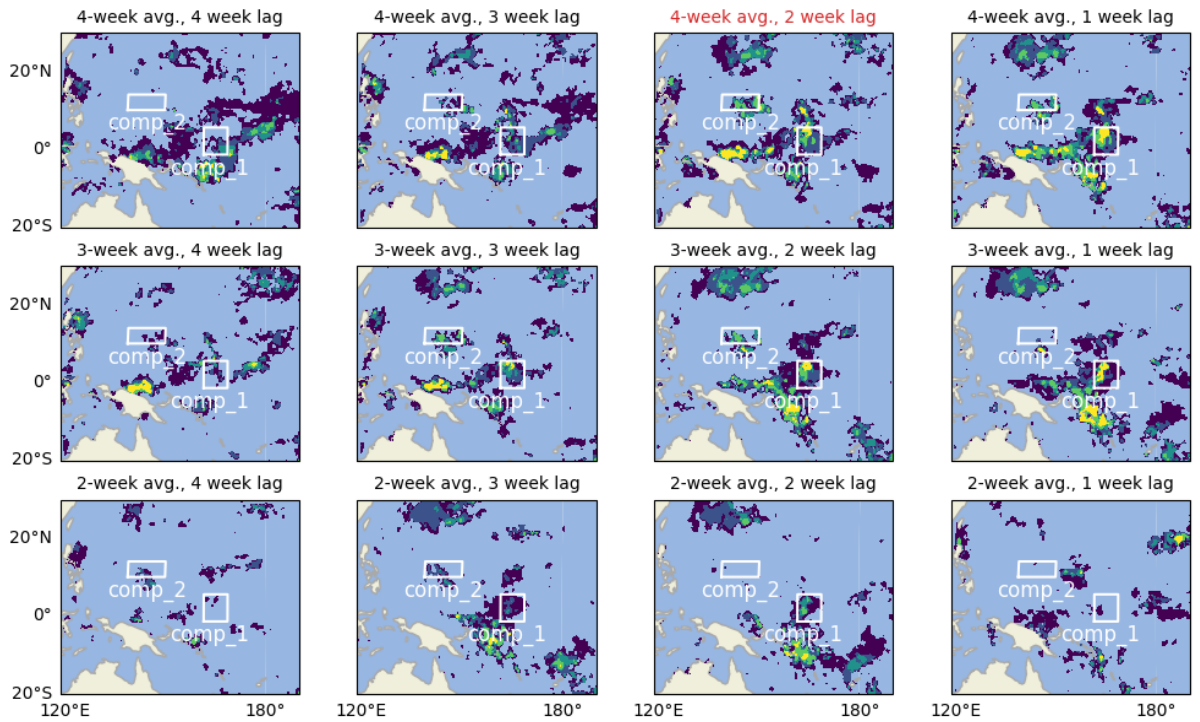
Figure R7 As Figure 2B, with partial correlation computed on reanalysis data from 1950 till 2021 (5-fold cross validation). Panels illustrate different combinations of lags and time aggregations to assess alternatives for box placement. The results of this study are based on the aggregation and lag highlighted in red.
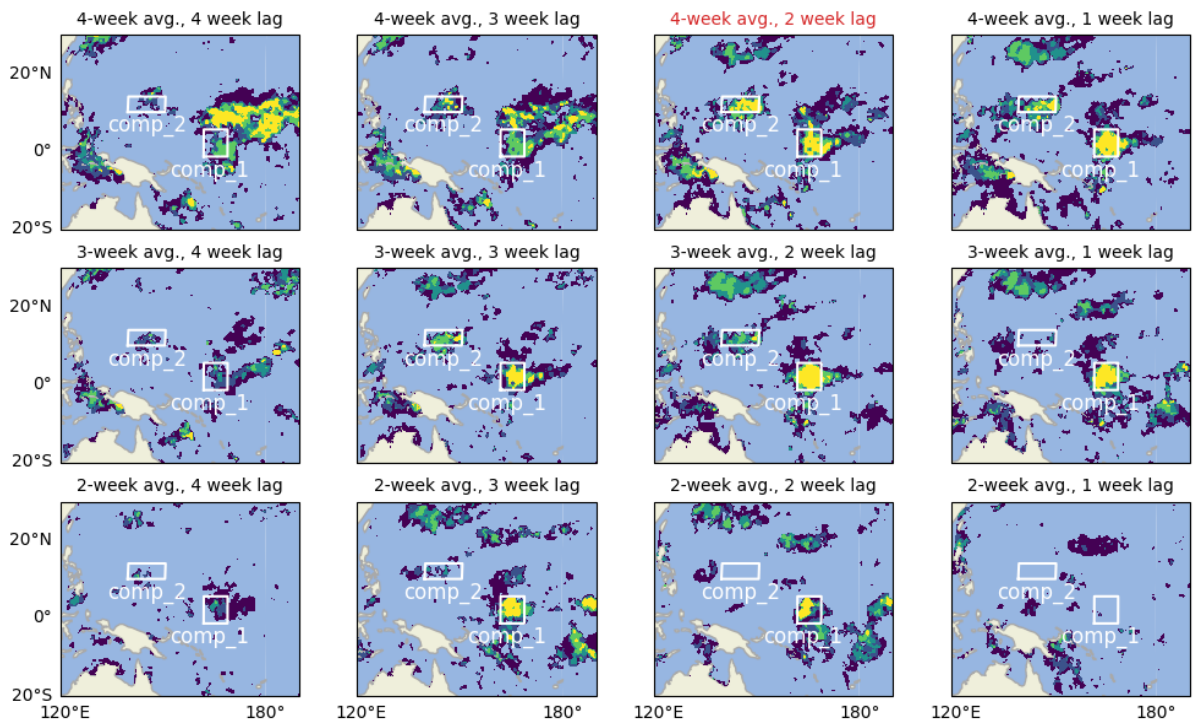

Figure R8 As Figure R6, but with partial correlation computed on data from 1980 till 2021 (5-fold cross validation). The results of this study are based on the aggregation and lag highlighted in red.

Minor comments:

- Page 2, line 34: Many studies suggest that the contribution of ocean SSTs (including therefore La Niña) to the Russian heatwave is very weak (see Dole et al., 2011; Hauser et al. 2016; Wehrli et al. 2019).
This is indeed an interesting contrast with the cited study of Schneidereit et al. (2012). Perhaps one way to understand the conundrum is that these three are modeling studies, which, as we also highlight in the introduction, can have problems simulating this type of teleconnection. The study of Schneidereit on the other hand uses reanalysis. Also, the small contribution in the case of Wehrli et al. (2019) might be a consequence of the experimental protocol, which derives SST's influence by comparing to a small collection of reference states that are neighboring in time and perhaps not so different. But nonetheless, we acknowledge that the issue is not settled. We propose to cite the mentioned studies, and state that the influence of the La Nina on the russian heatwave is debated.
- Page 2, line 51: this is a reference to a paper under review, I am not sure about the WCD policy about that.
The paper is now accepted.
- Page 4, line 100: using this PDO definition where global mean SSTs are removed might not be the best option This assumes that any anthropogenic influence on PDO behavior is fully removed by subtracting the time evolving changes in global mean SSTs from local SST changes in the PDO region. This is unlikely as it is known that there are different forced warming rates between the PDO region SSTs and the global mean SSTs. This means that an anthropogenic signal is likely to be aliased when using this PDO index definition (see Bonfils and Santer, 2011).
Thank you for this insight. This definition was employed by the institute from which we obtained the data: https://oceanview.pfeg.noaa.gov/erddap/tabledap/UW-JISAO_PDO.html But we notice now that the updating of this data is also discontinued. So instead we propose to replace it with a PDO index that does not suffer from this issue https://oceanview.pfeg.noaa.gov/erddap/tabledap/cciea_OC_PDO.html and update the data description accordingly. Replacing the index hardly changes the presented correlation coefficients in Fig. 3, and PDO remains uncorrelated to the response.
- Page 4: there is no reference to figure 1A, perhaps add it line 109.
Good suggestion
- Page 6, lines 126–129: I think that this statement (that the WPD captures the combined heating contrast …) need a bit more details, or maybe even an additional figure. It is written as if it is an obvious fact, but it is not.
True, we propose to amend the text.
- Page 6, figure 3 caption: why do you use 4 weeks instead of 3 for PDO and Niño indices?
This is now harmonized, all variables are at the 4 weeks/monthly timescale. See also comment Line 86-87 of Reviewer #2.
- Page 6, lines 131–134: I do not understand the logic here. The WPD index is based on two boxes that are located in the Western Pacific, and one of the boxes is outside the equatorial region. It is not clear to see how this can be related to the equatorial SST gradient across the whole Pacific basin. Please explain.
True, the index as defined is not purely based on equatorial gradients, but it is a combination of equatorial and meridional. As mentioned in the manuscript both types were found to be important by Ding et al. (2011). We propose to amend this text to make it clearer.
- Page 7, lines 155–159: why not using the "detrending" approach outlined for Figure 2 instead of using a 21-yr rolling window?
The procedure with the rolling 21-yr window is more involved than detrending with linear regression (the approach taken for Figure 2), but we think it is particularly suited for the data, because of the non-linearity in the long-term t2m warming in the last few decades. The reason we detrend with regression in Figure 2 is that we wanted to simultaneously correct for inflation due to autocorrelation. which is possible when both factors (autocorrelation, and long-term trend) are jointly estimated and removed within a single regression framework. That the approach of Figure 2 is different does not lead to inconsistencies, as after all, the grid-based correlation is only used to inform the placement of the boxes, after which WPD is

defined on SST anomalies that are only de-seasonalized. See also our reply to major comment 2 of Reviewer #1.

- Page 7, lines 163–165: please specify what are exactly the equatorial (SST?) gradients you are talking about.
  In this case this is the increase in gradient as seen in the long-term warming of the west pacific (WNP) and the cooling of the central pacific (Nino3). This could have been clearer indeed, and is also a gradient that is not fully equatorial. As above, we propose to amend this text.
- Figure 7: the contours are very difficult to see, can you enlarge the plots and make thicker contours?
  Contours are now made thicker. See Fig R4.

References:

Bonfils C, Santer BD (2011) Investigating the possibility of a human component in various Pacific Decadal Oscillation indices. Clim Dyn 37:1457–1468. https://doi.org/10.1007/s00382-010-0920-1

Dole, R., Hoerling, M., Perlwitz, J., Eischeid, J., Pegion, P., Zhang, T., et al. (2011).Was there a basis for anticipating the 2010 Russian heat wave? Geophysical Research Letters, 38, L06702. https://doi.org/10.1029/2010GL046582

Hauser, M., Orth, R., & Seneviratne, S. I. (2016). Role of soil moisture versus recent climate change for the 2010 heat wave in western Russia. Geophysical Research Letters, 43, 2819–2826. https://doi.org/10.1002/2016GL068036

Wehrli, K., Guillod, B. P., Hauser, M., Leclair, M., & Seneviratne, S. I. (2019). Identifying key driving processes of major recent heat waves. Journal of Geophysical Research: Atmospheres, 124, 11,746–11,765. https://doi.org/10.1029/2019JD030635