

Response to reviewer comments

Comment 1.1 *I find the premises and motivation for the paper unconvincing. The authors argue that one of the most interesting aspects of temperature extremes is their persistence, as this is closely linked to their impacts and thus needs to be studied more (ll. 13-27). They further identify two key gaps in the literature: (i) the focus has been on short-lived temperature extremes but it is necessary to look at temperature extremes over longer periods (ll. 28 and following); and (ii) in the past, work has been conducted mostly on arbitrary regions (ll. 42 and following). I understand these to be the main limitations this study seeks to address. Concerning point (i), what I find to be missing is any evidence that looking at 3-week temperature extremes as opposed to e.g. 5-day temperature extremes actually provides a better picture of the impacts (or even of the meteorological drivers) of the temperature extremes. The 5-day or similar threshold generally captures past high-impact events and allows to identify coherent sets of meteorological drivers, and I assume that these are amongst the reasons why it has enjoyed such popularity in the literature. Can the authors actually identify a set of high-impact events that is overlooked by a 5-day minimum persistence criterion but captured by their own definition? Similarly, can they make a case for the fact that using a 5-day minimum persistence criterion confounds the meteorological drivers of longer-lasting events? After all, the main conclusions on the dynamical drivers of the temperature extremes that the authors find in many cases seem to support previous findings in the literature. A further point that I detail further in one of my other comments is that a lower 5-day threshold on heatwaves does not prevent including much longer-lived events in the analysis. Additionally, in many cases, impacts are related to duration of temperature extremes in a non-intuitive fashion (e.g. Xu et al., 2016), which again seems to go against the argument of the authors for looking at 3-week periods.*

Answer: Thank you for this important input. In response to our earlier comment, we modified the introduction at length to better motivate our approach and improve the literature discussion. Please see our reworked introduction. In particular, we added a paragraph to justify our choice of time window: *"However, previous works have generally used short time windows to define warm or cold spells, on the order of 3-5 days (e.g., Schaller et al., 2018; Plavcová and Kyselý, 2019; Jeong et al., 2021; Jiménez-Esteve and Domeisen, 2022). Heatwaves, for instance, are frequently defined as sequences of at least 3 or 5 extreme warm days, with the longest events usually lasting about a week (e.g., Perkins and Alexander, 2013; Plavcová and Kyselý, 2019). While such windows may also identify multi-week spells, they focus on the part of the spells with the most extreme temperatures. In addition, while persistent spells frequently include short periods of extreme temperature anomalies, the opposite is not true: many short-lived, extreme temperature anomalies are not part of multi-week warm or cold spells. Finally, choosing a longer window is useful to identify regions where spells tend to co-occur, as it allows to capture the trajectory of the synoptic-scale systems responsible for the temperature anomalies across their lifetime."*

Comment 1.2 *Concerning point (ii), there are several papers that have proposed regional partitions of temperature extremes based on somewhat objective meteorological criteria, some of which are cited by the authors later in the paper (e.g. Stefanon et al., 2012 and others in Sect. 6.2). Moreover, there are several studies that have chosen specific regions motivated by non-meteorological but perfectly sensible criteria, such as taking an impacts perspective (e.g.*

Lowe et al., 2015), maximizing data availability (e.g. Hirschi et al., 2011) or favouring ease of comparison with previous work (notably SREX regions, e.g. Perkins-Kirkpatrick and Lewis, 2020). The authors do mention that arbitrary regions may make sense from an impact perspective; I would argue that defining the regions from an impact perspective makes them distinctly non-arbitrary. I do see a value in defining regions on a hemispheric rather than continental level, as the authors do here, but I find the statement on l. 45 to be a gross misrepresentation of the literature.

Answer: "Arbitrary" was a poor choice of word here and we agree that previous region selection can make sense in a perspective different from ours. Here, we are interested in a purely meteorologically-driven regionalisation to highlight (i) the spatial co-occurrence of persistent spells and (ii) the influence of different drivers. We modified the introduction to highlight this argument: *"Still, much remains unknown about the geography of persistent warm and cold spells, their spatial dependence, and how the role of various identified drivers varies in space. Such information is essential for risk assessment and to improve forecasts (Perkins, 2015; van Straaten et al., 2022). Persistent spells indeed typically occur over large spatial scales, from hundreds to thousands of square kilometers, with potentially complex spatial footprints (Lyon et al., 2019; Vogel et al., 2020). Yet, most previous work on heat waves, for instance, has focused at the grid-point scale or over regions based on impacts, administrative boundaries, data availability or past observed events (e.g., Hirschi et al., 2011; Bieli et al., 2015; Lowe et al., 2015; Plavcová and Kyselý, 2019; Zschenderlein et al., 2019; Hartig et al., 2022). Case studies provide useful insights into the dynamics of persistent warm and cold spells, but their results cannot necessarily be generalised. There is consequently value in a more systematic regionalisation of persistent warm and cold spells, as has been attempted in a few studies at the regional scale (e.g. Carril et al., 2008; Stefanon et al., 2012; Sousa et al., 2018)."*

Comment 1.3 In Sect. 2.2.2 the authors define their analysis window. I understand the logic of wanting to study the onset phase until the peak of the warm/cold spells, but defining the first day as the one where the regional temperature anomaly changes sign may introduce an unrealistically long build-up phase for some events. For example, I am not sure that I buy the argument of a wintertime warm spell taking on average more than 2.5 weeks to build up, as suggested by table 1.

I also think this makes the reasoning on ll. 91-92 somewhat circular. The authors define a very generous window to analyse the build-up of temperature anomalies, and then argue for the importance of analyzing temperature deviations over a long time period based on their own definition of the build-up period.

I would be more convinced if some analysis could show the typical evolution of regional temperatures from the first day of the analysis window and make the case that this reflects a consistent, continuous build-up as opposed to e.g. a period of oscillations around weak anomalies of the same sign and then a rapid increase of absolute anomalies closer to the peak date. Presumably, both from a physical and a more impacts-based perspective, the interest lies in the build-up and persistence of somewhat large temperature anomalies, not in small anomalies, even if persistently of the same sign.

Answer: Following our earlier reply to this comment, we modified the manuscript by expanding the motivation for our choice of analysis window and deleting the comment on ll. 91-92. Please see the following addition to the methods section:

”For each event, we define an analysis window (during which to calculate circulation and temperature budget anomalies) as the period between the last day before the event on which the region-mean daily temperature anomaly was negative (for warm spells) or positive (for cold spells) and the day during the event on which the daily-mean temperature anomaly peaks (maximum for warm spells, minimum for cold spells; Figure 1). This definition follows Röthlisberger and Papritz (2023a) and Röthlisberger and Papritz (2023b). The reason why we do this, instead of using the original three-week periods that define the spells, is to capture the onset of the persistent warm or cold spells at a daily resolution, and avoid the decay part of the spells when synoptic drivers are likely disappearing. Note that with this definition the analysis window is different for each spell. By looking back at the last day when temperature anomalies changed signs, we can capture the onset and build-up period of the spells, not just their peak. Admittedly, this choice may not always be the most relevant for individual events. However, on average, it allows to capture the period during which temperatures consistently depart from the climatology. One downside is that this definition may favour unrealistically long build-up times and weaken the significance of the detected synoptic anomalies. Another limitation of this choice is that it excludes the days following the peak temperature anomaly, which can still be relevant, for instance in terms of impacts. Still, we show in Figure A2 that spells do tend to develop over multi-week time scales, slowly and persistently deviating from their climatology, which supports our choice of analysis window.”

Comment 1.4 *A separate issue I have with the methodology is that the authors weigh differently each event, and could be giving a disproportionately large weight to events with a long build-up time, regardless of whether these are particularly extreme events or not. What is the range of the analysis windows for individual events and is there a correlation between event severity and duration of the analysis windows within the single regions?*

Answer: We added the following lines at the beginning of section 4: ”There is however a large variability in analysis window length from one event to the other (coefficient of variation approx. 0.5 for most regions). Note that because we calculate synoptic anomaly maps for each single event separately (section 2.2.3), longer events do not get more weight than shorter ones in the results.”.

Comment 1.5 *Further, I am confused by the description of the data preprocessing. On l. 65 the authors mention normalization of daily temperature using the mean and standard deviation. I assume that the former part entails subtracting the mean. On ll. 137 and following they mention subtracting the long-term average. Is the procedure such that the authors first normalize the data, including subtracting a mean value, then average it regionally and then subtract again a mean value?*

Answer: Thank your for pointing this out, we clarified this point in the revision. We added the following paragraph to the data section: ”For the statistical modeling and regionalisation of warm and cold spells, we further normalise daily temperature data on each day and at each grid point by subtracting the mean and dividing by the standard deviation, both being estimated on a 30-day, 7-year moving window (as in Pfeleiderer et al. (2019)). This allows us to remove the seasonality and long-term trends so as to focus exclusively on sub-seasonal variability. We only use this normalisation in the quantile regression (section 2.2.1). For anomaly maps during warm and cold spells (section 2.2.3), we use the non-normalised temperature data.”

Comment 1.6 Finally, as far as I can tell the way statistical significance is computed is never explained in detail. For example, in Fig. 5 and following how is the 90% confidence level computed, and do the authors account for multiple testing?

Answer: Thank you for this comment. We added the following details to section 2.2.3: ” We then calculate anomaly maps of selected fields (section 2.1) averaged over the analysis windows. For each season, region and event, we average the different fields during the corresponding time window, and subtract the long-term average calculated over the same calendar days across the whole ERA5 record. Anomaly maps for each season and region are then averaged across events for warm and cold spells separately. We assess their statistical significance of anomalies by calculating at each grid point their rank among a sample of 1000 anomalies obtained from randomly generated sets of events with the same distribution of duration and timing of occurrence (calendar days) as observed events. From this rank an empirical p -value is obtained, which we correct for the false discovery rate (Wilks, 2016).”

Comment 1.7 While some of the regions defined in the paper are relatively intuitive, others look puzzling to say the least. For example, is it really the case that cold spells in the middle-east are part of a coherent region with cold spells in the south-western Sahel (region 2 (or 3, I can't really tell the colours apart) in Fig. 2a)? Similarly, Fig. 2d seems to suggest that heatwaves in Eastern Europe/Western Russia actually belong to three (or even four) separate clusters – something that I do not recall ever having seen in the literature. I am in general not against introducing new definitions to the literature; indeed, it is part of scientific progress. However, when new definitions are presented – in this case of temperature extreme regions – which appear at odds with the ”conventional” ones from the previous literature, some more robust justification and contextualization would appear necessary. I find Sect. 6.2 somewhat dismissive, by providing references to three previous regional definitions of temperature anomalies but not going any further in this direction.

Answer: The point of our regionalisation approach is to let the data define the regions themselves. Our results are meant to highlight the spatial connectivity of warm/cold spells in a systematic way (specific events may behave differently) to best analyse the corresponding physical drivers. In practice, we find that neighbouring regions generally experience warm/cold spells at different times. There is certainly some overlap, as we discussed in our previous reply, but in most cases, events in neighbouring regions overlap by less than 20%.

We also expanded section 6.2 by adding some recent references: ” It is not straightforward to compare our regionalisation results to previous findings, because (i) we take a hemispheric perspective, (ii) we look at long (3-week) cold and warm spells and (iii) we limit ourselves to regions where the quantile regression model skill is high. Most previous studies are indeed limited to specific areas (Europe in particular) and focus on shorter-term events, chiefly heatwaves. Still, over Europe during summer, our results are in general agreement with those of Carril et al. (2008) who, from EOFs of monthly temperature anomalies, found a clear tripole structure between Northwestern Europe, the Euro-Mediterranean and Eastern Europe/Western Russia similar to regions 10, 15 and 21 on Figure 3-d. Stefanon et al. (2012) and Pyrina and Domeisen (2023) likewise both found 6 spatial clusters across Europe based on the simultaneity of hot summer days, including a North Sea cluster comparable to region 21 on Figure 3-d, a Western European cluster (region 15), a Scandinavian cluster (region 7), an Eastern European cluster (region 10)

and a Russian cluster (region 11). Some of the typical European heatwave patterns of Felsche et al. (2023) are also similar to our Figure 3-d results. Regarding cold spells, Xie et al. (2017) detected three main patterns of winter cold waves in North America (East, Northwest and Centre), two of which bear some resemblance to regions 11 and 13 on Figure 3-a. Region 11 is also similar to the North American region of cold outbreaks triggered by sudden stratospheric warmings identified by Kretschmer et al. (2018).”

Comment 1.8 *A separate comment on the regionalization results is the conflation of land and sea regions. I assume that the composites shown in Figs. 5 and following are centred on the centroid of each region. The composites then show an odd combination of averages over land and sea regions. Variables such as cyclone frequency may be largely governed by specific regions spanning the storm tracks, and much less relevant for others. I suspect that more coherent results may be obtained if, especially for surface and near-surface variables, the authors tried to composite separately land and sea points.*

Answer: You are correct and these are limitations which we need to make explicit in the manuscript. Consequently, we added the following paragraph: ” *While the goal of our paper is to highlight the similarities in terms of synoptic-scale anomalies during persistent warm/cold spells across the northern hemisphere, there are some limitations relative to the averaging that must be mentioned. First, whereas summer regions are generally confined to land (Figure 3-b,d), in winter this is not the case (Figure 3-a,c). As a consequence, averaged anomaly maps shown in Figures 5, 6, 7 and 9 mix together land and ocean grid points. As the particulars of individual regions is not the main focus of the present study, we choose not to treat land and ocean separately, but one should remember that it may affect the results in the aforementioned figures. Since cold spells are advection-driven, land/ocean contrasts probably play a limited role, but for warm spells, surface heat fluxes from the ocean surface may be significant. Second, averaging over regions of different shapes and sizes, and located at different latitudes or different points of the storm tracks, also impacts the results. The analysis of individual regions will be the subject of future work.*”.

Comment 1.9 *Sect. 5 is an odd section, which seems to be midway between a results and a discussion section. Much of the section is speculative and imprecise, and generally seems to make strong statements without a clear support from the analysis presented in the study. An example is the subsection on Recurrent Rossby Wave Packets. The authors dedicate a subsection to this without ever explaining clearly whether they define in some objective way RRWPs. It is unclear to me how the reference they make to Fig. 12 would support their claims. The subsection is largely based on speculation and analogies with previous results in the literature, rather than an objective analysis. Similarly, to support the statement made on ll. 349-350, one would need to at least show precipitation anomalies in Fig. A6.*

Answer: We agree that the content of section 5 should rather be placed in a results section, and therefore we moved it to the end of section 4 (“Synoptic conditions during persistent warm and cold spells”). We also rephrased some of the paragraphs to better explain the analyses that we did (especially the RRWP analysis and the precipitation deficit figure, which we had forgotten to include).

For the precipitation analysis, we added the corresponding panel on Figure A8, as well as the following sentence: ” *Regions with particularly high precipitation deficits (Western North*

America, Western Europe, Kazakhstan and Eastern Siberia) also tend to exhibit the largest sensible heat flux anomalies (Figure A8).”

For RRWPs we included the following: ”The persistent circulation patterns during warm and cold spells can also be part of recurrent Rossby wave packets (RRWPs). In these synoptic-scale wave packets, individual troughs and ridges amplify repeatedly at the same longitudes, leading to persistent circulation anomalies frequently associated with extreme surface weather (Röthlisberger et al., 2019; Ali et al., 2022). To highlight the role of RRWPs, for each region, we look for statistically significant positive R-metric anomalies during event analysis windows in a longitude interval covering the region and up to 30°W of the region’s westernmost points (to identify potential upstream RRWPs). Results show that persistent spells in a large number of regions are associated with significant RRWP activity (Figure 12),”

Note that we do not claim to precisely quantify the role of all the physical drivers we identified. In this part we list and discuss the major drivers highlighted by our analysis (and which also concur with the literature). A complete and quantitative analysis would go beyond the scope of our paper.

Comment 1.10 Finally, in the author’s intention this section should discuss persistence of temperature extremes, but no definition or quantification of persistence is ever presented. The only place where the authors do define a timescale is in choosing the averaging period for the temperature anomalies and computing the ”analysis window length”. However, neither the persistence of the warm/cold spells (which may or may not be similar to the analysis window length if one is concerned here with the persistence of somewhat large anomalies) nor the persistence of the atmospheric circulation features associated with these is ever explicitly addressed. I struggle to see how one may make statements on the circulation features explaining the persistence of a given surface feature when persistence is never defined, nor is the duration of either feature quantified.

Answer: Thank you for this comment. As we discussed in a [recent preprint](#), persistence can be defined in many ways, and we clearly need to explicitly detail which one we choose. As you wrote, we consider a fixed timescale to define persistent warm and cold spells. The 3-week time scale is meant to capture persistence over sub-seasonal time scales. We added the following paragraph to the beginning of the methods section (2.2) in the revision:

”Our proposed regionalisation method brings together grid points where persistent warm or cold spells share a similar dependence on the large-scale circulation. There are many ways to characterise a time series, such as temperature, as ”persistent” (Tuel and Martius, 2023). What we consider as ”persistent” spells here are periods of time that stretch over sub-seasonal time scales (week to month) and during which temperature anomalies tend to remain either much above or much below zero. This approach corresponds to the ”episodic” persistence in the typology introduced by Tuel and Martius (2023). To identify such periods, we consider a fixed sub-seasonal time scale of 3 weeks, and define sub-seasonal warm and cold spells as 3-week periods during which the average temperature is above its 3-week 95th percentile or below its 5th percentile.”

In addition, in response to your [Comment 1.3](#), we added supplementary figure A2 to illustrate the development of the warm/cold spells over sub-seasonal periods.

Comment 1.11 I find the contextualisation of the findings lacking. While in Sect. 6.2 the authors attempt to make a link to previous work which has looked at coherent regions of extreme

surface temperature occurrences, what is completely missing is a discussion section making a link to the wealth of literature looking at drivers of different regional cold or warm spells. Admittedly, there is some of this in Sect. 5. However, discussing in more detail the literature related to the mechanisms is a key point needed to contextualise the current analysis relative to the literature, and I would have expected the bulk of the discussion section to be dedicated to this.

Answer: The discussion in section 4.5 (previously section 5) already provides many references to the literature looking at the drivers of warm and cold spells, and we cannot expand it too much for the sake of length. However, some aspects indeed deserve to be further detailed, notably the link of the various mechanisms to the persistence of extreme temperature spells. Consequently, we expanded section 4.5, in particular:

- Section 4.5.1 (on blocking), by adding the following: *”Atmospheric blocks are by definition persistent features (Schwierz et al., 2004), but their typical lifetime is usually on the order of 5-10 days, and longer blocks are extremely rare (e.g., Nabizadeh et al., 2019). The persistence of individual blocks is known to be fueled by upstream latent heating (Steinfeld and Pfahl, 2019), but recurrent blocks are likely also key to the persistence of cold or warm spells over sub-seasonal timescales. For instance, the month-long heatwave over the Baltic in June-July 2022 (Tuel et al., 2022) and the 2010 Russian heatwave (Drouard and Woollings, 2018) were both related to several successive blocking systems.”*
- Section 4.5.2 (on RRWPs), see our answer to your [Comment 1.9](#).
- Section 4.5.1 (on land-atmosphere feedbacks), which we reformulated as follows: *”It has long been argued that land-atmosphere feedbacks are essential for the persistence of summer warm spells (e.g., Lorenz et al., 2010; Mueller and Seneviratne, 2012; Perkins, 2015; Bartusek et al., 2021; Martius et al., 2021). Dry soils indeed increase sensible heat fluxes to the detriment of latent heat fluxes, leading to lower tropospheric warming and drying which further dries the soils and stabilises the atmospheric column (Miralles et al., 2019). This is a well-known mechanism whose role has been highlighted in several recent persistent heatwaves (e.g., García-Herrera et al., 2010; Dirmeyer et al., 2021). All our regions experience precipitation deficits during JJA warm spells, most of which are statistically significant (Figures 11 and A8-a). Over continents, these precipitation deficits translate into increased sensible heat fluxes upwards from the surface (Figure A8-b), which enhance the positive temperature anomalies and partly balance the radiative cooling seen in the negative diabatic tendencies (Figure 11-f). Regions with particularly high precipitation deficits (Western North America, Western Europe, Kazakhstan and Eastern Siberia) also tend to exhibit the largest sensible heat flux anomalies (Figure A8). Land-atmosphere feedbacks thus seem important for the persistence of summer warm spells, especially in Europe and Scandinavia, where they have already been shown to play a role (e.g., Dirmeyer et al., 2021). It remains however unclear to what extent dry soils can trigger persistent heatwaves on their own, or whether a dynamical forcing is required (e.g., Domeisen et al., 2023). In our case, we do not find significant precipitation deficits before the start of the persistent warm spells (not shown). A pre-conditioning by dry soils therefore does not appear to be required for persistent warm spells, though they may still play a role for the most extreme warm spells (Hirschi et al., 2011; Stefanon et al., 2012).”*

Comment 1.12 ll. 14-15 Perhaps the authors could explain what spatially and temporally compounding events would be in the context of temperature extremes. Presumably concurrent, geographically remote heatwaves or cold spells or successions of heat waves or cold spells at the same location within a given season? Moreover, compound events are never mentioned or discussed in the analysis presented in the paper, so it seems largely irrelevant to discuss them in the introduction.

Answer: This sentence was indeed irrelevant and we removed it.

Comment 1.13 ll. 36-37 I find this statement misleading. First of all, there are indeed papers that have treated temperature extremes on multi-week to monthly timescales (I can think of Galfi and Lucarini (2021) off the top of my head, but a careful literature search would likely bring up others). Second, while many papers do impose a 5-day or similar threshold for heatwave duration, most heatwaves end up being much more persistent than that. It is thus misleading to state that these papers "focused on short-lived events, on the order of a few days only". Indeed, using a few days in duration as a lower threshold can easily lead to identifying multi-week heatwaves, see e.g. Vogel et al. (2020) or Fig. 10 in Grotjahn et al. (2016). This also links to my previous major comment on the motivation for looking at 3-week temperature deviations.

Answer: Thank you for this comment; we modified the introduction to better reflect the literature by adding the following:

"A handful of studies have looked specifically at prolonged periods of heat and cold, especially over sub-seasonal timescales. Galfi and Lucarini (2021) modeled the marginal probabilities of persistent warm and cold spells of various lengths to describe their climatology across several regions with large deviation theory. Carril et al. (2008) analysed the circulation patterns associated with extreme warm summer months in Europe and the Mediterranean, and Li et al. (2017) did the same for persistent cold spells of at least 10 days in North America. More recently, van Straaten et al. (2022) attempted to discover sources of predictability for monthly warm anomalies in Central Europe."

and:

"[...] previous works have generally used short time windows to define warm or cold spells, on the order of 3-5 days (e.g., Schaller et al., 2018; Plavcová and Kyselý, 2019; Jeong et al., 2021; Jiménez-Esteve and Domeisen, 2022). Heatwaves, for instance, are frequently defined as sequences of at least 3 or 5 extreme warm days, with the longest events usually lasting about a week (e.g., Perkins and Alexander, 2013; Plavcová and Kyselý, 2019). While such windows may also identify multi-week spells, they focus on the part of the spells with the most extreme temperatures. In addition, while persistent spells frequently include short periods of extreme temperature anomalies, the opposite is not necessarily true, i.e. some short-lived, extreme temperature anomalies are not part of multi-week warm or cold spells (Figure A1). Finally, choosing a longer window is useful to identify regions where spells tend to co-occur, as it allows to capture the trajectory of the synoptic-scale systems responsible for the temperature anomalies across their lifetime."

Comment 1.14 ll. 63 Since the authors themselves later in the paper mention the challenges of working with 42 years of data, an obvious question is why they have chosen not to take advantage of the ERA5 back-extension, which has been available in preliminary form for over two years and in its final form since mid-2022. The major issues with the preliminary version

	Cold	Warm
DJF	10-16 (12)	8-23 (12)
JJA	5-13 (9)	8-24 (12)

Table 1: Analysis window length (in days) for 2-week warm and cold spells: inter-region range and median.

	Cold	Warm
DJF	12-20 (16.5)	11-25 (18)
JJA	8-17 (13)	12-27 (17)

Table 2: Analysis window length (in days) for 4-week warm and cold spells: inter-region range and median.

were tropical cyclones, which is not something that would have affected the analysis presented here.

Answer: It is true that the ERA5 back extension represents a wealth of new data that upcoming studies should make the most of. However, when the final version came out, we were already very advanced in our analysis, and we decided to stick to the 42-year data.

Comment 1.15 ll. 108 In Fig. 1 many of the patches with $DR > 0.4$ are fragmented. How does the PAM behave when given data with scattered "holes" in it? Does this affect the robustness of the final clustering?

Answer: The clustering algorithm does not take any geographical information as input, so whether the DR map is fragmented or not does not matter. It is true that some regions end up including some fragments (i.e., a few lone points scattered away from the region's main group of points) but these represent at most a few percent of a region's points and do not impact the results. Such fragments are unavoidable in our regionalisation since by choosing a rather small number of regions forces the algorithm to add isolated points to one of the regions. We explain at ll. 117-121 how we remove these isolated points to better visualise the regions.

Comment 1.16 ll. 190-191 It would be interesting for the readers to see these numbers for 2 and 4 week averaging times, e.g as an Appendix table.

Answer: We added the numbers for 2 and 4 week events as appendix tables.

Comment 1.17 Sect. 4.1 How do the authors determine this grouping? Is it through some objective criterion or through a subjective analysis of the individual regions?

Answer: We obtained this grouping by subjective analysis based on the anomaly maps. We made it clear in the revision: "*Comparing circulation anomaly maps across regions, we visually identify three main groups of regions with similar circulation anomalies during persistent DJF cold spells...*"

Comment 1.18 A few typos (although I appreciated how well-written the paper generally was): l. 211 near-surface diabatic

ll. 272, 328 *Incorrect figure reference?*
l. 328 *(difference in intensity/spatial extent?)*
ll. 347-348 *(Figures 11) and S5)*

Answer: Thanks, we corrected these typos in the revision.

Comment 1.19 ll. 296-297 *I am not sure I agree with this statement, and indeed the following two sentences seem to counter it. I would suggest adding a "summertime" to the first sentence to avoid confusion.*

Answer: We followed your suggestion, thanks.

Comment 1.20 l. 310 and following. *The current formulation seems to suggest that there is plenty of work on the role of upstream blocking but that the authors are the first to highlight the role of downstream blocking/processes for the occurrence of upstream temperature extremes. While upstream blocking has certainly received more attention, there are studies which also consider the role of downstream features (e.g. Takaya and Nakamura, 2005; or Lehman and Coumou, 2015, who discuss the role of downstream storm track anomalies in the context of eastern North American extremes).*

Answer: Thank you for this comment. We reformulated the sentence to avoid misleading readers on the previous literature ("*First, atmospheric blocking plays an important role in the persistence of both warm and cold spells in summer and winter. Our hemispheric perspective highlights the various influences of blocking that have been discussed in a host of previous studies...*") and added the Takaya and Nakamura citation.

Comment 1.21 Fig. 5 and following: *using green dots on a red background is not ideal for many readers (including this reviewer).*

Answer: Thank you for pointing it out. We changed to yellow dots in all relevant figures.

Comment 1.22 Fig. 12 *If the authors only look upstream of the regions (30° W of the westernmost point as stated in the figure caption) how can they diagnose downstream blocking as again stated in the caption?*

Answer: We only look 30°W upstream for the R-metric anomalies, not for the blocking (the caption states that "Hatching indicates significant positive blocking frequencies up- or downstream of the region". We added "*not restricted to 30° W*" to the caption to make this more explicit.

Comment 1.23 *On a separate point, it may be worth explaining in the methods what the R-metric is and how it is being used here.*

Answer: We added the following paragraph to the data section:

"To assess the potential role of recurrent synoptic-scale Rossby wave patterns, we use the R-metric developed by Röthlisberger et al. (2019). The R-metric is calculated from conventional Hovmöller diagrams of 6-hourly 35–65°N averaged 250 hPa meridional wind. We first apply a 14-day running mean to the series, and remove contributions outside the synoptic wavenumber range $k = 4-15$. The R-metric is then defined as the absolute value of the time- and wavenumber-filtered signal. Large R-metric values therefore tend to indicate the presence of several successive synoptic-scale wave packets, in other words recurrent Rossby waves."

Response to reviewer comments

Comment 1.1 *The manuscript is well-written and the figures are of high quality. However, I have significant concerns about the methods, the implication of the results, and the comparison with the existing literature. In my opinion, the study does not address the research question stated in the title.*

Answer: Thank you for your comment. We agree that our initial title was misleading: we do not analyse the persistence of warm and cold spells per se, but the characteristics of persistent warm and cold spells (which is different). We modified the title accordingly. We hope our replies below address your concerns as to the methodology and comparison with existing literature.

Comment 1.2 *In the Introduction, the authors identify two gaps in the literature on the processes responsible for warm and cold spells: (i) persistence and (ii) spatial dependence. With respect to the first gap (persistence), what can we learn meteorologically from longer lasting events that we don't already know from the existing literature? There is already a lot known on driving factors for heat and cold waves. I am not arguing that it is not interesting to look at the persistence, but the authors insufficiently motivate why it could be interesting to look at persistent events (apart from impacts). In addition, a minimum duration of three days does not exclude long-lasting events, e.g. in Zschenderlein et al. (2019), a heat wave with a duration of more than 40 days was included in the analysis. It is therefore not true that previous work generally focused on short-lived events only, as the authors claim.*

Answer: Thank you for your comments. We reworked our introduction to further motivate our work and discuss the choice of the 3-week time scale. Specifically, we expanded the second paragraph as follows:

"There already is substantial literature on extratropical warm and cold extremes and their driving factors (e.g., Domeisen et al., 2023). Among these, atmospheric blocking has long been recognized as a key driver of temperature extremes in the mid- to high-latitudes, for both summer warm spells and winter cold spells (Buehler et al., 2011; Pfahl and Wernli, 2012; Schaller et al., 2018; Kautz et al., 2022). Topography (Jiménez-Esteve and Domeisen, 2022) and land-atmosphere interactions (Bieli et al., 2015; Miralles et al., 2019; Wehrli et al., 2019) also play important roles, as does polar vortex variability for cold air outbreaks in mid- and high-latitudes (Kolstad et al., 2010; Biernat et al., 2021; Huang et al., 2021). However, previous works have generally used short time windows to define warm or cold spells, on the order of 3-5 days (e.g., Schaller et al., 2018; Plavcová and Kyselý, 2019; Jeong et al., 2021; Jiménez-Esteve and Domeisen, 2022). Heatwaves, for instance, are frequently defined as sequences of at least 3 or 5 extreme warm days, with the longest events usually lasting about a week (e.g., Perkins and Alexander, 2013; Plavcová and Kyselý, 2019). While such windows may also identify multi-week spells, they focus on the part of the spells with the most extreme temperatures. In addition, while persistent spells frequently include short periods of extreme temperature anomalies, the opposite is not necessarily true, i.e. some short-lived, extreme temperature anomalies are not part of multi-week warm or cold spells (Figure A1). Finally, choosing a longer window is useful to identify regions where spells tend to co-occur, as it allows to capture the trajectory of the synoptic-scale systems responsible for the temperature anomalies across their lifetime."

Comment 1.3 *With respect to the second gap (spatial dependence), the authors argue that the recent literature looked at arbitrary regions and mixed up areas where warm or cold spells are shaped by different physical processes. They are by far not arbitrary in the recent literature, either regions are motivated by impacts (which is perfectly fine and definitely not arbitrary) or by different climates (e.g. humid vs. dry climate). There are also studies looking at the different physical processes in a region, e.g. Sousa et al. (2018, Fig. 5). It is therefore not overlooked that in one region the processes can differ regionally.*

Answer: We agree that "arbitrary" was a very wrong choice of word. Here, we are interested in a purely meteorologically-driven regionalisation to highlight (i) the spatial co-occurrence of persistent spells and (ii) the influence of different drivers. We modified the introduction to highlight this argument: *"Still, much remains unknown about the geography of persistent warm and cold spells, their spatial dependence, and how the role of various identified drivers varies in space. Such information is essential for risk assessment and to improve forecasts (Perkins, 2015; van Straaten et al., 2022). Persistent spells indeed typically occur over large spatial scales, from hundreds to thousands of square kilometers, with potentially complex spatial footprints (Lyon et al., 2019; Vogel et al., 2020). Yet, most previous work on heat waves, for instance, has focused at the grid-point scale or over regions based on impacts, administrative boundaries, data availability or past observed events (e.g., Hirschi et al., 2011; Bieli et al., 2015; Lowe et al., 2015; Plavcová and Kyselý, 2019; Zschenderlein et al., 2019; Hartig et al., 2022). Case studies provide useful insights into the dynamics of persistent warm and cold spells, but their results cannot necessarily be generalised. There is consequently value in a more systematic regionalisation of persistent warm and cold spells, as has been attempted in a few studies at the regional scale (e.g. Carril et al., 2008; Stefanon et al., 2012; Sousa et al., 2018)."*

Comment 1.4 *The study promises a lot to provide new insights into the persistence and spatial dependence of warm and cold spells. However, both aspects are not really quantified in the study. I detail my comments below:*

1. *Persistence: Sections 4 and 5 are believed to address this research question. However, these sections rather analyse synoptic conditions during warm and cold spells (which is nicely done) and identify three atmospheric drivers (upstream blocking, downstream blocking, subpolar troughs). Many studies already investigated the position of the block relative to the high/low temperature extreme at the surface (e.g. Pfahl and Wernli 2012, Pfahl 2014, Bieli et al., 2015, Santos et al., 2015, Sousa et al. 2018, Zschenderlein et al., 2019). The authors should therefore make the new findings more clear. While I find the synoptic analysis very nice, I don't understand what this has to do with persistence. And this is the main aspect of the paper (at least it is stated in the title). For example, how important is the position of the block relative to the warm/cold spell with respect to the persistence in comparison to a heat/cold spell of about 5 days? I think the authors try to discuss the persistence aspect in Section 5, but not really successful. They are rather speculating about the role of blocks, RRWPs, land-atmosphere feedback, and subpolar troughs (with results from the existing literature and not their own) and not quantifying the importance of these processes with respect to persistence.*

Answer: Thank you for your comments. The choice of title was clearly misleading, and we modified it accordingly to *"Persistent warm and cold spells in the Northern Hemisphere extratropics: regionalisation, synoptic-scale dynamics, and temperature budget"*, since we do not explicitly analyse the mechanisms causing spells to persist or not. Such a detailed analysis would

indeed require investigating warm and cold spells across different timescales to understand what determines their length, and is beyond the scope of the present manuscript (but will be the subject of future work). We hope that our analysis and discussion are now more in line with the title. In particular, an important outcome of our study is the regionalisation, as well as the comparison between regions it allows to make in terms of driving mechanisms. We consequently expanded the corresponding discussion (please see our answer to your [Comment 1.5](#)).

Comment 1.5 2. *Spatial dependence: I think the authors want to address this research question with the regionalisation approach (Fig. 3). But in the end, they summarise all regions into three categories (and they don't describe how they do it, see also my minor comments). What is the added value of the regionalisation? What can we learn from it? What do the different colors in Fig. 3 imply meteorologically? In my view, Figure 3 is insufficiently discussed.*

Answer: The regionalisation approach is indeed meant to highlight the spatial connectivity of persistent warm and cold spells. This is something we should have stressed more and which we therefore developed in the revised manuscript. Discussing what the regionalisation implies meteorologically is difficult to do in detail as it would require analysing each region separately, which cannot be done in a single paper, but we expanded the discussion of Figure 3 in section 5.2 by adding the following:

"While our goal is not to discuss individual regions in detail, some general remarks can be made about the regionalisation results shown on Figure 3. First, it is important to note that, as with any regionalisation approach, our results are dependent on the number of regions we select. The distance metric we use implies that regions should be understood as areas where warm/cold spells are related to similar large-scale circulation patterns, and therefore where spells tend to occur at approximately the same time. The subsequent synoptic analysis showed that our choice of regions seems consistent with the corresponding physical drivers. Too few regions would likely have blurred the significant signals in e.g. blocking location. In practice, we also find that neighbouring regions generally experience warm/cold spells at different times. While there is some overlap, in most cases, events in neighbouring regions overlap (in time) by less than 20%. Still, our proposed regions are not necessarily the most relevant from an impacts perspective, notably because (i) the temperature threshold we consider are relative, not absolute, and (ii) the extremeness of the temperature anomalies is softened by averaging over large scales.

Second, the difference in the spatial footprint and mean zonal extent of regions between DJF and JJA, while certainly impacted by the choice of region number, is consistent with planetary waves having longer wavelengths in the cold season. This leads to more extent temperature anomalies at the surface, e.g., the well-known North American dipole (Singh et al., 2016) which may correspond to regions 11 in Figure 3-a and 15 in Figure 3-c. Another important feature of the winter season is that temperature gradients are not only meridional, but also zonal. This is much less the case in summer when meridional gradients dominate. Zonal circulations can therefore lead to strong cold advection in winter. Our temperature budget analysis shows that cold spells are primarily advection-driven, in both winter and summer. The cold air in summer essentially comes from the high latitudes, and a strongly meridional circulation is required to bring it equatorwards, consistent with rather narrow regions in Figure 3-b. By contrast, in winter, land/ocean temperature contrasts are steep, and zonal circulations can bring about strong cold (or warm) advection, consistent with the large zonal extent of some regions like e.g., region 13 on Figure 3-a. Locally, some particularly zonally elongated regions like regions 6 and 23 over Europe in Figure 3-a, are also likely related to strong zonal circulations (in this case the

North Atlantic Oscillation) that can lead to persistent surface temperature patterns. Further differences in the regionalisation between winter and summer may be due to the seasonality of blocking (Steinfeld and Pfahl, 2019) or RRWPs (Röthlisberger et al., 2019) (in terms of location, extent and intensity).”

Comment 1.6 *I am not convinced by the method for the identification of persistent warm and cold spells. While I understand that you incorporate the days prior to the spells to investigate the onset, I do not understand why your analysis window ends with the peak temperature anomaly. What if this anomaly is reached already very early during the warm/cold spell and high/low temperatures still persist after that peak?*

Answer: There are certainly limitations to our choice of analysis window, as we detailed in our initial reply. Our choice however allows to capture, on average, the onset and build-up phase of the spells. Individual events will behave differently from the average behaviour, but we are precisely interested in the average. We modified/expanded the corresponding paragraph of the methods section to further justify our choice and state its limitations:

”For each event, we define an analysis window (during which to calculate circulation and temperature budget anomalies) as the period between the last day before the event on which the region-mean daily temperature anomaly was negative (for warm spells) or positive (for cold spells) and the day during the event on which the daily-mean temperature anomaly peaks (maximum for warm spells, minimum for cold spells; Figure 1). This definition follows Röthlisberger and Papritz (2023a) and Röthlisberger and Papritz (2023b). The reason why we do this, instead of using the original three-week periods that define the spells, is to capture the onset of the persistent warm or cold spells at a daily resolution, and avoid the decay part of the spells when synoptic drivers are likely disappearing. Note that with this definition the analysis window is different for each spell. By looking back at the last day when temperature anomalies changed signs, we can capture the onset and build-up period of the spells, not just their peak. Admittedly, this choice may not always be the most relevant for individual events. However, on average, it allows to capture the period during which temperatures consistently depart from the climatology. One downside is that this definition may favour unrealistically long build-up times and weaken the significance of the detected synoptic anomalies. Another limitation of this choice is that it excludes the days following the peak temperature anomaly, which can still be relevant, for instance in terms of impacts. Still, we show in Figure A2 that spells do tend to develop over multi-week time scales, slowly and persistently deviating from their climatology, which supports our choice of analysis window.”

Comment 1.7 *And for the quantile regression I would exclude grid points in the tropics beforehand, because a low Z500 variability in this area is no surprise.*

Answer: The quantile regression is applied to each grid point separately, so there is no influence of e.g. the tropics onto the mid-latitudes. As to the regionalisation, we remove grid points with low model skill (deviance ratio below 0.4) beforehand, which includes the vast majority of the tropics.

Comment 1.8 *The results of this study are not set into context with the existing literature. An attempt is made in Sections 5 and 6.2, but it is not clear what new knowledge is introduced by the study.*

Answer: Please see our answer to your [Comment 1.4](#).

Comment 1.9 L5-10 (Abstract): *All listed processes (blocks important precursors of warm and cold spells, location of the blocks, recurrent Rossby waves, land-atmosphere feedbacks) are already described in many papers (see list in major comment 2) or in a new review paper by Domeisen et al. (2022, "Atmospheric processes" section; the paper was probably not yet published at the time the authors submitted to WCD).*

Answer: Our findings are indeed consistent with the literature. Please see our reply to your [Comment 1.4](#). We included the recent reference of Domeisen et al. to the manuscript.

Comment 1.10 L65: *Does this mean that you first subtract the mean and divide by the standard deviation?*

Answer: Yes, please see our next answer.

Comment 1.11 L66-68: *Why do you use different methods for the trend removal of T2M and Z500?*

Answer: We reformulated this paragraph to make our processing clearer. T2M is treated differently for the quantile regression analysis only. We added the following: *"For the statistical modeling and regionalisation of warm and cold spells, we further normalise daily temperature data on each day and at each grid point by subtracting the mean and dividing by the standard deviation, both being estimated on a 30-day, 7-year moving window (as in Pfeleiderer et al. (2019)). This allows us to remove the seasonality and long-term trends so as to focus exclusively on sub-seasonal variability. We only use this normalisation in the quantile regression (section 2.2.1). For anomaly maps during warm and cold spells (section 2.2.3), we use the non-normalised temperature data.."*

Comment 1.12 L70: *Is the persistence criterion of 4 days valid for an individual grid point or for the whole cyclone mask?*

Answer: It applies to the cyclone as a system (this is a Eulerian persistence criterion).

Comment 1.13 L70-80: *Should be moved to the methods section.*

Answer: These lines are indeed a mix between data and methods, but we would prefer to focus the methods section on the core methods for this paper (modeling and analysis) rather than on the data pre-processing.

Comment 1.14 L84: *Is the percentile based on daily or 3-weekly values? And please do not use different words for the same method (normalise, rescaled)*

Answer: Thank you for this comment. The percentile is indeed calculated on 3-week averages. We slightly expanded the sentence to make this clear: *"Warm and cold spells are defined as 3-week periods during which the average temperature is above its 3-week 95th percentile or below its 5th percentile. Temperature is first normalised (section 2.1) to remove seasonal and long-term trends."*

Comment 1.15 *L89: Why non-overlapping 3-week intervals?*

Answer: We want the intervals to be non-overlapping to avoid as much dependency between model input data points as possible.

Comment 1.16 *L105: How sensitive are your results with respect to the DR threshold?*

Answer: We chose this specific threshold (0.4) to capture about as much of the extratropics in all seasons as possible. A lower threshold would lead to more tropical grid points retained in the regionalisation (extratropical regions remaining essentially the same in the regionalisation algorithm). A higher threshold removes points from the extratropics and quickly tends to decrease the "optimal" number of regions.

Comment 1.17 *L112: What is a silhouette coefficient?*

Answer: We added the following to the manuscript: "*The silhouette coefficient is a common metric used to identify an "optimal" cluster number. It essentially compares the mean intra- and inter-cluster distances, with higher values indicating a better clustering (high inter-cluster distance and low intra-cluster distance)*".

Comment 1.18 *L129: Why not using the back-extension of ERA5?*

Answer: It is true that the ERA5 back extension represents a wealth of new data that upcoming studies should make the most of. However, when the final version came out, we were already very advanced in our analysis, and we decided to stick to the 42-year data.

Comment 1.19 *L135: Is the decay phase of the event not interesting?*

Answer: It certainly is, but not in the context of our paper which focuses on the mechanisms responsible for the onset and maintenance of the warm and cold spells. See our reply to your [Comment 1.6](#).

Comment 1.20 *L143: median point – do you mean centroid?*

Answer: No, here we take the actual (geographical) median point, and not the cluster centroid (which in PAM are called medoids, and minimise the distances to all other cluster points). We specified "geographical" in the text to make this clear.

Comment 1.21 *Figure 1: I don't understand the values on the y-axis. Is it normalised with the 95th percentile? I also do not understand why the grey shaded area goes from approximately day 11 to 31.*

Answer: Thanks for this comment; we reformulated the caption to make all this clearer: "*From the region-averaged normalised (i.e., unitless; see section 2.1) daily temperature series, a warm spell (gray shading) is identified as an extreme 3-week average anomaly (here, from days 11-31). We then identify the corresponding analysis window as the period between the last day before the event on which the temperature anomaly was negative (day 6) and the day during the event*".

on which the temperature anomaly peaked (day 28). Cold spells are identified in a symmetrical way. Note that the reference day value for this example was artificially set at 0.”

Comment 1.22 *Section 4.1: How have you identified the three main groups?*

Answer: We identified the main groups visually, after removing the tropical regions. We made this explicit in the revision.

Comment 1.23 *Figure 5 (and others): Mark the centroid of the regions. Add wind directions or Z500 mean values.*

Answer: By construction, the centroid of the regions is at the origin in these figures. The mean Z500 and wind fields are not as meaningful as the anomaly fields here since we average across regions located at different latitudes/longitudes.

Comment 1.24 *General remark to the composites: I personally think that compositing can be quite dangerous here, because you are averaging over areas with different sizes and shapes. Can you comment on that?*

Answer: Following our earlier reply, we added a paragraph stating the limitations relative to the averaging over regions of different sizes/shapes/latitudes and over land/ocean grid points: ” While the goal of our paper is to highlight the similarities in terms of synoptic-scale anomalies during persistent warm/cold spells across the northern hemisphere, there are some limitations relative to the averaging that must be mentioned. First, whereas summer regions are generally confined to land (Figure 3-b,d), in winter this is not the case (Figure 3-a,c). As a consequence, averaged anomaly maps shown in Figures 5, 6, 7 and 9 mix together land and ocean grid points. As the particulars of individual regions is not the main focus of the present study, we choose not to treat land and ocean separately, but one should remember that it may affect the results in the aforementioned figures. Since cold spells are advection-driven, land/ocean contrasts probably play a limited role, but for warm spells, surface heat fluxes from the ocean surface may be significant. Second, averaging over regions of different shapes and sizes, and located at different latitudes or different points of the storm tracks, also impacts the results. The analysis of individual regions will be the subject of future work.”.

Comment 1.25 *L225: Be careful because you are mixing land and ocean grid points (see red colours in Fig. 4).*

Answer: Please see our answer to your [Comment 1.24](#).

Comment 1.26 *Figure 6: The scale in Fig. 6c differs from Fig. 5c!*

Answer: The scale varies depending on the figure, otherwise the panels would not be easily readable.

Comment 1.27 *L250: These results are in line with Zschenderlein and Wernli (2022).*

Answer: Which study are you referring to? The only one we could find has been withdrawn.

Comment 1.28 *Section 4.4 and Figure 12: I am not able to follow because you haven't described the R-metric sufficiently.*

Answer: We added the following paragraph to describe the R-metric:

"To assess the potential role of recurrent synoptic-scale Rossby wave patterns, we use the R-metric developed by Röthlisberger et al. (2019). The R-metric is calculated from conventional Hovmöller diagrams of 6-hourly 35–65 N averaged 250 hPa meridional wind. We first apply a 14-day running mean to the series, and remove contributions outside the synoptic wavenumber range $k = 4-15$. The R-metric is then defined as the absolute value of the time- and wavenumber-filtered signal. Large R-metric values therefore tend to indicate the presence of several successive synoptic-scale wave packets, in other words recurrent Rossby waves."

Comment 1.29 *L290: remove the second "is"*

Answer: Corrected, thanks.

Comment 1.30 *L310: This agrees with a host ... → ok, but where is the new knowledge regarding the persistence?*

Answer: Please see our answer to your [Comment 1.8](#).

Comment 1.31 *L312-313: This is probably also true for events on the shorter time scales.*

Answer: You are right; we added the Takaya and Nakamura (2005) reference which discusses the case of East Asian cold waves.

Comment 1.32 *L318: remove half sentence in brackets at the end of the line.*

Answer: Corrected, thanks.

Comment 1.33 *L348-350: This is not clear to me. Increasing surface heat fluxes would imply diabatic heating, but we see diabatic cooling in Fig. 11f. Radiative cooling is typically dominant in the mid- to upper-troposphere.*

Answer: In Figure 11f, the diabatic heating term is calculated as the residual of the energy balance equation. Therefore it is not restricted to radiative cooling. What likely happens however is that the radiative cooling term outbalances surface sensible heat fluxes, since it has to balance not only sensible heat fluxes, but also adiabatic warming and horizontal advection. Note that our analysis also covers day and night, and that sensible heat fluxes peak during the day but are weaker at night, while radiative cooling can remain large across the clock.

Comment 1.34 *L351-352: What can we learn from this sentence?*

Answer: We mentioned this because a number of previous studies have argued that extreme warm spells were often associated with antecedent precipitation deficits. However, in our case we probably do not look at event that are extreme enough to see the connection.

Comment 1.35 *Section 5.4: You discuss how a subpolar trough develops. But I don't see how you discuss subpolar troughs and their connection to persistence of warm and cold spells.*

Answer: You are right that we do not discuss the development of subpolar troughs, but rather point out that persistent wintertime cold spells at high latitudes are likely associated with such systems (the analysis of synoptic-scale anomalies points in that direction). They connect to persistent cold spells because of the associated circulation anomalies, but also because such systems have a general tendency to persistence, as we discuss in the manuscript (cf. also Woollings et al. (2023); <https://doi.org/10.5194/wcd-4-61-2023>).

Comment 1.36 *Section 6.1: A matter of taste, but the discussion of Fig. 3 appears quite late in the study. I would prefer to discuss the results earlier.*

Answer: We expanded the discussion of Figure 3 in the revision, please see our answer to your [Comment 1.5](#).

Comment 1.37 *L383: Acronym S2S not introduced*

Answer: Corrected, thanks (S2S = "subseasonal-to-seasonal")