EGUsphere, referee comment RC1
https://doi.org/10.5194/egusphere-2022-1478-RC1, 2023

Comment on egusphere-2022-1478

Anonymous Referee #1

Referee comment on "Persistent warm and cold spells in the Northern Hemisphere extratropics: regionalisation, synoptic-scale dynamics and temperature budget" by Alexandre Tuel and Olivia Martius, EGUsphere, https://doi.org/10.5194/egusphere-2022-1478-RC1, 2023

The paper is generally well-written and easy to read. It deals with an interesting topic, namely finding a coherent definition and analysis approach for extratropical temperature extremes in different regions and studying their persistence. However, I have several major concerns on the methodology, value of the analysis, interpretation of the results and contextualization relative to the literature. I find the overall framing of the paper unconvincing, and while the title appears to promise a focus on persistence, the latter is never objectively defined nor quantified. I detail all these points in my comments below.

Major Comments

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. 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.

Hirschi, M., Seneviratne, S. I., Alexandrov, V., Boberg, F., Boroneant, C., Christensen, O. B., ... & Stepanek, P. (2011). Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. Nature Geoscience, 4(1), 17-21.

Lowe, R., Ballester, J., Creswick, J., Robine, J. M., Herrmann, F. R., & Rodó, X. (2015). Evaluating the performance of a climate-driven mortality model during heat waves and cold spells in Europe. International journal of environmental research and public health, 12(2), 1279-1294.

Perkins-Kirkpatrick, S. E., & Lewis, S. C. (2020). Increasing trends in regional heatwaves. Nature communications, 11(1), 3357.

Xu, Z., FitzGerald, G., Guo, Y., Jalaludin, B., & Tong, S. (2016). Impact of heatwave on mortality under different heatwave definitions: a systematic review and meta-analysis. Environment international, 89, 193-203.

2. I have a number of methodological concerns. 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. 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? 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? 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?

3. 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. 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.

4. 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. 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.

5. 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.

Other Comments

1. ll14-15 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.

2. 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.

Galfi, V. M., & Lucarini, V. (2021). Fingerprinting heatwaves and cold spells and assessing their response to climate change using large deviation theory. *Physical Review Letters*, *127*(5), 058701.

Vogel, M. M., Zscheischler, J., Fischer, E. M., & Seneviratne, S. I. (2020). Development of future heatwaves for different hazard thresholds. Journal of Geophysical Research: Atmospheres, 125(9), e2019JD032070.

Grotjahn, R., Black, R., Leung, R., Wehner, M. F., Barlow, M., Bosilovich, M., ... & Lim, Y. K. (2016). North American extreme temperature events and related large scale meteorological patterns: a review of statistical methods, dynamics, modeling, and trends. Climate Dynamics, 46, 1151-1184.

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

4. 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?

5. 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.

6. 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?

7. A few typos (although I appreciated how well-written the paper generally was):

l. 211 near-surfacediabatic

ll. 272, 328 Incorrect figure reference?

l. 328 (difference in intensity/spatial extent?)

ll. 347-348 (Figures 11) and S5)

8. 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.

9. 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).

Lehmann, J., & Coumou, D. (2015). The influence of mid-latitude storm tracks on hot, cold, dry and wet extremes. Scientific reports, 5(1), 1-9.

Takaya, K., & Nakamura, H. (2005). Geographical dependence of upper-level blocking formation associated with intraseasonal amplification of the Siberian high. Journal of the atmospheric sciences, 62(12), 4441-4449.

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

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