Dear Reviewer 3,

Thank you for your helpful comments. We provide our responses below in blue.

This paper examines the links between heat extremes and a metric of atmospheric persistence from dynamical systems theory. Using this metric, it is argued that there is no evidence of a link between heat extremes and anomalously persistent circulation patterns. Some assessment of the role of different terms in driving temperature anomalies is also provided and it is argued that there is an important role for zonal temperature advection in driving wintertime heat extremes, while in the summertime temperature advection does not play an important role. I can see that it's useful to quantify atmospheric persistence by this dynamical systems measure, although I have some questions about the methodology. My primary concern is about the interpretation and the conclusions when using this measure. I think there are two potential interpretations. One is that persistent circulation patterns are not closely linked to heat extremes, which is the interpretation the authors have provided. The other is that this dynamical systems metric is actually not a very good measure of persistent circulation patterns. I'm not totally convinced by the authors argument that this second interpretation can be ruled out, so my more major suggestion is that either a clearer demonstration of this being a preferred metric for atmospheric circulation persistence should be provided, or the discussion should be changed to be a bit more balanced on whether this metric is accurately measuring persistence of atmospheric circulation anomalies.

We thank the Reviewer for their valuable time and helpful comments, and detail below the edits we would make, if we are invited to provide a revised manuscript.

General comments:

As mentioned above, my primary concern is whether the interpretation that "atmospheric persistence is not a necessary requirement for summertime heatwaves" (l9) is actually correct, or whether an alternative interpretation is that this dynamical systems theory metric of atmospheric persistence does not adequately capture persistent atmospheric circulation anomalies. I feel like the demonstration in Figure 6 supports that this metric might not be very good at capturing persistent atmospheric circulation anomalies. It is clear that Figure 6b and d demonstrate a persistent block, as indeed the authors state. But it seems like it could be argued that this dynamical systems metric is, therefore, just not a very good metric for persistent atmospheric flow anomalies. I'd question whether it can really be used to argue that "highly persistent configurations are not a necessary criterion for European summertime heatwaves to occur" (l242) when clearly there is a case in Figure 6 that does have a persistant atmospheric flow configuration but is considered not persistent by this metric. I'm left a bit confused about what argument is exactly being made. If I read the paper in a cursory manner, I might think that the conclusion is that you don't need persistent atmospheric circulation patterns to produce heatwaves, but if readers think more about it and pay attention to Figure 6, it seems the conclusion should actually be that this particular dynamical systems metric of persistence doesn't really capture persistence in atmospheric flow patterns of relevance to heatwaves. I suggest the authors either need to make a clearer case for why this metric is preferable to those based on persistent flow regimes or alter the wording in places to make clear that actually this dynamical systems metric of persistence doesn't do that good a job of picking out persistant blocking highs or other flow regimes that are relevant for heatwaves. I think either conclusions is worthy of publication, but I'm confused about which one is being drawn. Another example of a confusing conclusion is lines 247-249 where it's stated that "our results appear to contrast the conventional view of heatwaves being associated with very persistent blocked configurations" when above, in reference to Fig 6, it's stated that a blocking algorithm would detect a block in the low persistence case for several days (l238). It seems, then, that this metric is not a very good metric of blocking, so how can it be used to contrast the conventional view of very persistent blocked configurations being connected to heatwaves?

Thank you for this detailed comment. We would like to argue that the choice of a ‘good’ metric contains a certain amount of perspective. We believe that this is a very important discussion to be had, and possibly the crux of bringing the mathematical and dynamics perspectives of persistence together so as to be reconciled and mutually understood. We argue that a ‘good’ metric requires a sound, non-subjective definition, and would significantly revise our manuscript so as to clarify this. As part of this discussion, we would argue that the reviewer’s evaluation that Fig. 6 “does have a persistent atmospheric configuration” is a subjective statement, while our chosen distance metric allows to make a quantitative statement. In the study we very carefully state that in Fig. 6b and d a blocking algorithm would detect a persistent block, not that there **is** a persistent block. Our whole argument is that the block in those figures should **not** be considered as persistent, at least from an SLP persistence perspective. We are not arguing that blocking algorithms are wrong, but rather that they do not necessarily match an objective and quantitative definition of continental-scale atmospheric persistence. It seems to us that the two conceptualisations of persistence are useful for different research questions, which we would clarify in a revised discussion section.

(2) One aspect of the methodology that I wondered about was is it easier to have higher persistence when there are less anomalies overall in the spatial field. I'm imagining that if you have a really large amplitude anomaly but that moves slightly, the spatial Euclidian distance between days that have a relatively small movement may end up being a lot larger than the Euclidian distance between days where there's much less going on in the spatial field. If so, then maybe this isn't a particularly good measure of the persistence of relevance to heat extremes. Persistence of nothing much going on over the region wouldn't be very meaningful, whereas having a persistant blocking high that stays around for a long time in the region, even if it moves slightly, would likely be more impactful for heat extremes. Or perhaps there's something in the methodology that prevents this from happening. I recommendt this be discussed or assessed.

This method is indeed dependent on the distance metric chosen. However, note that a given day will not necessarily be more persistent simply because it has closer analogues – as in your example of a day with a large atmospheric feature versus a day with very small anomalies over the whole domain. The estimate of theta is based on the 5% of closest analogues of each day, and the “closeness” of these analogues to the day itself does not determine the persistence of that day. Rather, it is the distribution of the analogues in time that is important. We will make sure to clarify this very important point in the revised manuscript.

Minor comments by line number:

l94: I don't think "345-45W" is correct. Maybe 15W-45E?

Thank you, this will be corrected.

Figure 1 caption: I think this could be a bit clearer if it stated "warm spells during winter (top) and heatwaves during summer (bottom)". (It took me a while to notice the winter and summer in the titles).

Thank you, this will be corrected.

l160: In the discussion of the role of advection starting here and referring to Figure 4, "the role of warm air advection toward the region of interest during warm spells" is not entirely obvious to me in a causal sense. Couldn't there also potentially be a role for the temperature anomaly itself being produced by some other cause actually leading to temperature advection anomalies. I feel like the reds next to blues in this figure may be indicative of that i.e., you get some warm anomaly set up and if the zonal flow is westerly then you end up with cold advection to the west of the warm anomaly and warm advection to the east. Even if that's not the case, the dominant role of warm advection isn't totally clear to me from the figure since it's very noisy. Is the intention that readers should be paying attention to the larger spatial scales, as oposed to the small scale noise? If so, maybe some filtering to retain only larger spatial scales could be performed?

Based also on other review comments, we are planning to significantly revise this figure and use an alternate metric, in order to clarify the arguments around warm temperature advection. Specifically, we plan to use the advection of potential temperature by 10m winds, and have included a revised figure below, see Figure 1. Ultimately this does not change our qualitative conclusion that winter time warm spells appear to be associated with warm temperature advection in all regions except Russia, whilst there is a comparatively weak signal during summertime heatwaves. We treat Russian warm spells with caution due to the small, noisy potential temperature advection signal, and because the significance stippling also appears noisier. We will also discuss in the revised text the possibility raised by the Reviewer that the temperature anomaly itself may be produced by some other cause actually leading to temperature advection anomalies.

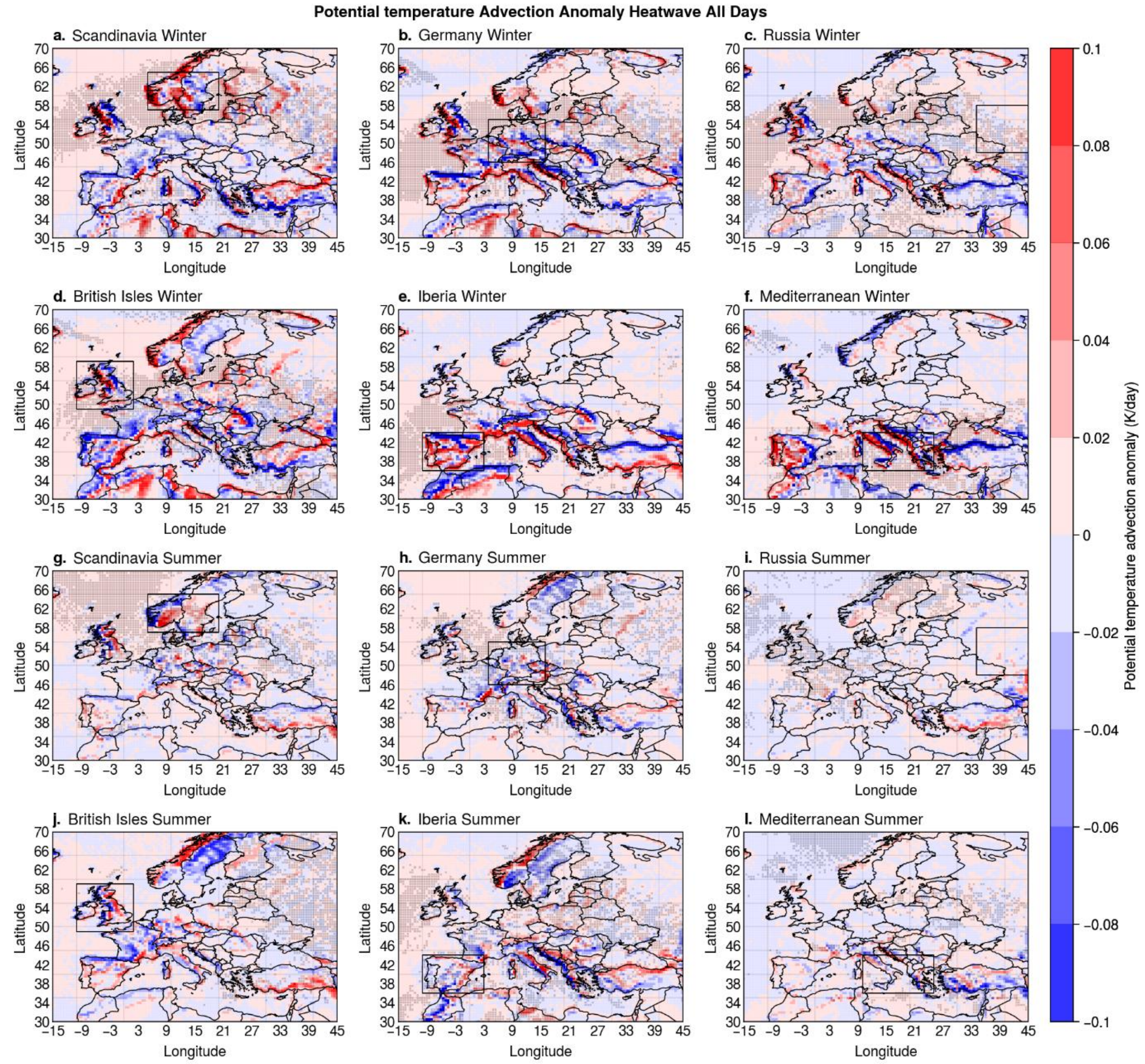

Figure 1: Potential temperature advection (K/day) anomaly during warm spell/ heatwave days in (a,g) Scandinavia, (b, h) Germany, (c, i) Russia, (d, j) British Isles, (e, k) Iberia, (f, l) Mediterranean, during winter (a–f) and summer (g–l). Statistical significance is assessed as described in Section 2 of the manuscript and shown with grey stippling.

l165-166: You mention the potential role of advection over topography for engendering large temperature anomalies here. But couldn't this also potentially be an artefact of using SLP? SLP will involve some extrapolation below the surface making an assumption about the lapse rate, I think. So SLP will be an approximation over topography and will be affected by the temperature at the surface, so is it possible that this could be playing a role here?

In light of both this comment and others, as mentioned above, we are planning on revising to use the potential temperature advection by 10m winds as a metric instead, which should not be subject to this same possible issue.