Dear Alex,

Thank you for your follow up comment. We provide our responses below in blue.

Thank you Emma for these detailed comments which helped me better understand your work. It is nice to see that the results remain the same when based on summer data only. Out of curiosity, did you use the raw SLP data or did you remove an annual cycle?

We used raw SLP data so as to ensure that the analogues selected by the algorithm truly correspond to the most similar circulation patterns. Using anomalies could lead to analogues in different months displaying similar anomalies relative to the month they occur in, but potentially very different actual circulation patterns. As discussed in our previous answer, this leads to the bulk of the analogues for days in a given season falling in that same season.

You wrote "In other words, we are not taking an average of the persistence of configurations similar to those of the heatwave and assigning that average persistence to the heatwave itself. This is a very important point that we will make sure to clarify in the revised methods section of our manuscript." I agree that this is important and might cause a lot of confusion, because clearly what "persistence" means for the average atmospheric scientist is very different from your definition. In particular, persistent patterns in the "traditional sense" may not be classified as very persistent with your dynamical systems approach. This also relates to my earlier comment about heatwaves and persistence, and this is why you should be careful when writing that heatwaves in Western Europe are not associated with much atmospheric persistence (e.g., "We thus argue that atmospheric persistence is not a necessary requirement for summertime heatwaves") What you argue, if I understood well, is that the states of the circulation that tend to occur during heatwaves are not persistent from the attractor perspective. However, this does not imply that during a heatwave, the circulation does not tend to remain stuck in a small part of the attractor (and Hoffmann et al's metric would capture this case). Is that correct? If yes, the point needs to be made clearly in the manuscript.

Thank you for highlighting this point. Based also on the review comments we would, if invited to submit a revised manuscript, endeavor to clarify our interpretations, and word them carefully so as not to confuse the audience, in particular taking care to highlight when we consider different perspectives on persistence. As far as we understand, Hoffmann et. Al.'s metric is based only on the similarity between 10 consecutive maps, rather than drawing on analogues from a longer time series. The algorithm we use to compute theta (the extremal index) is computed using the time between clusters and lengths of clusters of similar maps (i.e. those maps within the 5% threshold to class a map as an analogue of our chosen day). Thus, our computation has elements which are quite similar to Hoffman et. Al.'s metric, with the key difference that we calculate our statistics based on the properties of a much larger data set. In this sense, our method could be viewed as a generalisation of Hoffmann et. Al.'s metric, with the added advantage that our metric can be applied forwards in time i.e. we can quantify for how many days we would expect today's circulation pattern to remain similar to itself. Our persistence metric does indeed provide an estimate of how many time steps a trajectory is expected to stay within an epsilon-sphere of it's starting point in phase space. Thus, it is directly a measure of how long a trajectory 'sticks' in a given phase-space region. Consequently, higher persistence indeed means that the circulation remains stuck in a small part of phase space, whilst lower persistence means that it doesn't. We do note that here we can only make comments on the entire structure of the circulation pattern in a given geographical domain, and not about specific circulation features within the domain.

For future research, it would be nice to use Z500 (with the seasonality taken out) instead of SLP. For summer heat extremes, SLP is not quite as relevant, especially in the lower half of your domain. Z500 would be a better proxy for atmospheric circulation. Heatwaves may indeed be associated with heat lows while Z500 exhibits a pronounced ridge. The SLP analogues might thus not be very physically meaningful since low SLP at other times could be associated with cyclonic activity.

Thank you for this suggestion, it is a very important point and we refer you to Comment E and our response in the discussion with Reviewer 2.

One last point (motivated by personal curiosity, no need to change the manuscript): analogues are selected based on some low percentile of L2-distance across the whole trajectory. So, the number of analogues is the same for all time steps, but the average distance between the analogues and the target value is not fixed. In particular for some rare states, it could end up being high. Or is the threshold reasonably constant with time?

You are correct in stating that some days will have analogues which have a smaller L2 distance than others. We fit the generalized pareto distribution to the negative log of distance, as detailed in the methods section, to ensure that what we call "analogues" satisfy the mathematical assumptions our estimation of theta rests on, and that our overall statistics are not affected by poor analogues. One could hypothesise that this may eventually become an issue with a changing climate, however, based on the above we do not believe that trends in SLP are strong enough to affect our analysis.