Weather Clim. Dynam. Discuss., author comment AC2
https://doi.org/10.5194/wcd-2021-1-AC2, 2021

Reply on RC2

Carl Thomas et al.

Author comment on "An unsupervised learning approach to identifying blocking events: the case of European summer" by Carl Thomas et al., Weather Clim. Dynam. Discuss., https://doi.org/10.5194/wcd-2021-1-AC2, 2021

In this paper, the authors introduce a new blocking index based on self organizing maps. They compare this index to other commonly used blocking indices and to a ground truth time series they build based on expert judgement. They show that this new index generally performs better than other indices. I find this paper very well written and structured, with a very thorough sensitivity testing of their index to several key parameters. I have a few comments and questions that I list below. I recommend to accept this paper after some revisions.

We thank the referee for the supportive comments and helpful critique of the manuscript.

General comments:

You say your GTD is objective, but you make some heuristic choices too (namely the region and time period), so it's not completely objective. It's also a little bit unclear to me how you label your blocking periods from the maps shown in Figures 1 and 2. Do you manually look at each group of maps and identify blocking highs?

We actually do not consider the definition of the region to be an entirely arbitrary choice. The region we are studying is primarily motivated by our interest in the European sector. We have added an explanation for our choice of domain in L 145-148:

"JJA Europe was chosen because of our interest in the role of atmospheric dynamics in the development of mid-latitude land heat waves. Europe is a region which has seen many recent significant heat extremes (Christidis et al., 2014), and the role of changes in atmospheric dynamics in those heat extremes has been an activate area of research (Cattiaux et al., 2013; Horton et al., 2015; Saffioti et al., 2017; Huguenin et al., 2020)."

The time period we have chosen is the longest time period currently available in ERA5. We have also shown in Figure 11 that the algorithm has reached an optimal performance when only 20 years are used, and after this point there is weak sensitivity to the algorithm's performance.

To label our GTD we manually look at each five day period such as those shown in Figures 1 and 2 and conclude by eye using all the variables shown if a blocking high persists across this period. This has been further clarified in L 149-150:

"The GTD has been derived by studying every successive five day period from 28 May to 4 September 1979-2019, and manually identifying whether or not a blocking high persisted across any such five-day period."

Then from these sets of classifications, the GTD has been reconstructed, where any day that has been identified as part of a blocking pattern in at least one of these five day periods has been labelled as blocked. This has been clarified in L 160-163:

"Once the total set of all 4001 consecutive five day periods across JJA 1979-2019 has been classified, persistent blocking events are reconstructed to form a time series where each day is labelled as blocked or not. If a day belongs to any one of the consecutive blocked five day periods, it is individually labelled as blocked (1), and if a given day does not belong to any of the blocked five day periods it is labelled as not blocked (0). This creates a classification of blocking patterns for each day where each blocking event has a minimum length of five days."

Could you make it clearer in the text what's the added value of the SOM-BI compared to the GTD? As it is presented, the GTD gives supposedly better results than the SOM-BI so why do you need to build the SOM-BI? This may be easier to understand once the methodology to construct the GTD is clearer.

The SOM-BI has been constructed from the GTD so that it can be used in climate models and future projections where no GTD exists. Since labelling the GTD is quite an onerous task, if an algorithm can be trained on the GTD, the labelling process can subsequently be automated, e.g. to apply the SOM-BI to subsequently collected new meteorological data or to extensive climate model data (as demonstrated here for the case of UKESM). The SOM-BI could further provide useful regional information when applied in such automated ways, because it classifies blocking patterns with different node groups which in turn have inherently different characteristics that reflect the stationarity and location of blocking patterns. We have also shown in Figure 11 that the SOM-BI can perform well when only 20 years are labelled, so future applications would not need the longer records (41 years and 101 years in reanalysis and model) to be labelled as done here. The longer labelling exercises demonstrate the sensitivity of SOM-BI to the length of training data, and the robustness of SOM-BI over multi-decadal periods.

Furthermore, the SOM-BI can provide unique dynamical information, e.g. concerning spatial information about the location of the blocking anomaly. New sections have been added to the manuscript to describe how the SOM-BI can provide this information (section 2.7), and an application of the SOM-BI for ERA5 has been included in section 3.7.

A discussion of the utility and further application of the SOM-BI is included in the discussion section (L 586-593):

"The use of SOMs as a blocking index provides opportunities for regional study that are not directly available in the other BIs. Through an additional post-processing step involving K-means clustering on blocked node groups (sections 2.7 and 3.7), we have shown that the SOM-BI can identify specific types of blocking events and provide detailed information about the changing nature of blocked events over a European subdomain (Fig. 12). The case of k = 4 has been shown in Fig. 12, but larger values of k can also be chosen to identify more distinct types of blocking pattern. Whilst the SOM-BI does not directly produce a gridded climatology of blocking patterns, we have shown that the SOM-BI can be integrated with the other BIs to develop a climatology that only considers only those days detected as blocking by the SOM-BI. This results in a SOM-BI climatology with a higher precision than the BI climatology.

We intend to apply this method to future trends across CMIP5 and CMIP6 models to better understand the patterns of blocking in models, diagnose model skill at reproducing the historic patterns of European circulation regimes and compare projections of future changes in blocking patterns. The identification of distinct blocking patterns from node groups enables a detailed study of blocking characteristics over European subdomains as shown in Fig. 12. Further quantities such as the Rossby wave breaking properties or the nature of blocking onset and decay can also be studied. This could be done by studying particular dynamical quantities on the blocked days identified by the SOM-BI, and extended by contrasting the dynamical quantities across different categories of blocking pattern identified by the SOM-BI node groups."

The sinuosity index seems a bit disconnected from the other indices since it's not a direct measure of blockings. I would advise to remove this part of the analysis from the article, especially since it's a hemispheric measure and your study applies to regional blocking indices.

The discussion of sinuosity has been removed and replaced with a discussion of K-means clustering, also in response to Reviewer #1.

I am not a specialist of SOMs but I am more used to k-means algorithm and weather regimes. You never cite this part of the literature but I think it would be important to do so since that's something familiar for a large part of the atmospheric dynamics community.

Michelangeli, Paul-Antoine, Robert Vautard, and Bernard Legras. "Weather regimes: Recurrence and quasi stationarity." Journal of Atmospheric Sciences 52.8 (1995): 1237-1256.

Vautard, Robert. "Multiple weather regimes over the North Atlantic: Analysis of precursors and successors." Monthly weather review 118.10 (1990): 2056-2081.

It would be good to discuss the differences between your approach and what is done in the context of weather regimes.

We have removed the sinuosity discussion and sinuosity in the figures and replaced it with K-means clustering for the case of k=4 has replaced sinuosity in Figures 5, 6, A1, and A2. In L 209-212 some of the literature that uses K-means clustering to study mid-latitude variability has been cited and briefly discussed, including the references that you suggest:

"K-means clustering analysis (Diday and Simon, 1980) has also been extensively used to study the Euro-Atlantic midlatitude variability and identify weather regimes (Vautard, 1990; Michelangeli et al., 1995; Cassou, 2008; Ullmann et al., 2014; Strommen et al., 2019; Fabiano et al.,2021)."

The comparison in the case studies between SOMs and K-means analysis has also replaced the sinuosity discussion in L 350-357:

"Figures 6d and 7d show a K-means clustering analysis using Z500 anomaly fields for the case of 4 centroids. As described in section 2.4, the case of K-means with 4 centroids produces a similar set of weather regimes to SOMs with 4 nodes. Consequently, the K-means analysis exhibits a similar behaviour to the SOMs discussed above but distinguishing between fewer weather regimes. One weather regime indicating

Scandinavian blocking consistently represents the 2003 European heat wave across Fig. 6d, but the Westward shift of the high pressure centre from Scandinavia on 31 July to the UK on 8 August 2003 is not described by 4 centroids. For the 2019 heat wave in Fig. 7d, all four weather regimes are represented, and the blocked period is primarily associated with a mixed weather regime. This shows that the 2019 case is also not described well by K-means clustering."

The key difference between Self-Organizing Maps (SOMs) and K-Means clustering is that SOMs use a neighbourhood function, which for each stage of the update enables neighbouring nodes to shift towards the best matching node (Figure 3). As a result, in each step not only the best matching node (equivalent to the closest centroid in k-means by e.g. Euclidean distance) is updated, but also – to a degree - neighbouring nodes on the map. Consequently, with increasing node number, SOMs represent a smooth topology across the nodes from left to right and top to bottom on the final map, a realistic continuum of weather patterns. However, for small centroid/node numbers, the difference between K-means and SOMs is small, because the SOMs will still differentiate the range of possible weather patterns, but within a few nodes, making a smooth transition across the map of nodes impossible. Consequently, the results for SOMs and K-means approach each other for small numbers of nodes (n)/centroids (k), as shown in the revised manuscript in Figure A4 for the case of k and n equalling 4 . This figure is also shown below. Therefore if K-means for k=4 was used as a blocking index in the same way as done for SOMs in our manuscript, it would give similar results to those shown in Figure 7 for n=4.

However, this difference between K-means and SOMs is large for high node numbers. We have shown from Figure 7 that the optimum node number is 20 nodes for the SOM. With k=20, the K-means algorithm will lead to a set of distinct weather regimes, whereas the SOM will create a continuum of weather patterns. The SOM-BI associates node groups with blocking patterns which is consistent with the behaviour of the SOM as it represents a smooth topology.

A comparison of the SOM nodes for and the K-means nodes has been attached to show the similarity for low node numbers. In L 245-252 these differences have been highlighted:

"This property of SOMs is also the distinguishing feature between SOMs and K-means clustering. In the case of K-means clustering, each node is updated independently and no neighbourhood function is applied. K-means tries to maximize differences between the centroids such that it does not learn a topology. This difference between K-means and SOMs is minor for low node numbers, since the sharp differences in spatial patterns are imposed on the SOMs and the neighbourhood function has a limited effect. For larger node numbers, the SOM topology becomes smoother and the K-means centroids remain distinct rather than representing a continuum of states, whereas a continuum is a more realistic reflection of the actual atmosphere (Skific and Francis, 2012). A comparison between SOMs and K-means analysis for 4 and 20 node/cluster numbers is shown in Fig. A4."

Specific comments:

The abstract is a bit misleading. The SOM-BI doesn't work well with the 2019 case and this should be acknowledged.

The abstract has been modified in L12-15 to acknowledge the identified limitation of the SOM-BI:

"We present case studies of the 2003 and 2019 European heat waves and highlight that well-defined groups of SOM nodes can be an effective tool to diagnose such weather events, although the domain-based approach can still lead to errors in the identification of certain events in a fashion similar to the other BIs."

L58: do you mean objective or subjective?

The term "objective" has often been used to contrast automated procedures for blocking identification and subjective human analysis e.g. Tibaldi and Molteni (1990). It has been removed here to avoid misunderstandings.

L.254 I think there's a mistake with R=0.03. Shouldn't it increase compared with 0.19?

We list the skill of each node group individually here rather than the skill of the set of node groups being summed together (which is probably how the reviewer interpreted it). Consequently, P can be lower for a node group while R does not necessarily need to increase accordingly – it really depends on the characteristics of the specific node group. The wording has been clarified accordingly in L289-292:

"For the 3x3 SOM learned from ERA5 data, the node group with the highest precision is [1], with P = 0.91 and R = 0.15, followed by [2] with P = 0.89 and R = 0.19 and [1, 2, 6] with P = 0.87 and R = 0.03. If only one node group is included in the set (e.g. [1] or [1, 2, 6]), there is a high P and low R, but as more node groups are added to the set of node groups ([1], [2]; then [1], [2], [1, 2, 6]), P decreases but R increases."

L.263 remove « be »

Thank you. This word has been removed.

L.309 rephrase « are not more generally »

Thank you. This has been rephrased to "are not generally".

Figure 7 and 8 look very flat to me. Since those are the curves you use to determine your optimal number of nodes I would discuss this in the paper and potentially introduce some sort of confidence interval around your optimal k.

Whilst the dependence of the F1 score appears quite flat, we show in Figures 7 and 8 that we also consider the balance of precision and recall and not just the F1 score. The procedure we use for identifying the optimal number of nodes is discussed in L 439-444:

"Common features are observed for each variable for a very small or large number of SOM nodes. For small k the SOM-BI identifies more days as blocked, such that R » P. This indicates that the SOM is under-fitting the data for European circulation patterns across the domain and so the algorithm lacks a precise delineation of blocking events. In other words, it could be beneficial to increase k to be able to represent a larger number of dynamical states and thus to detect and describe blocking events more precisely. For large k, R « P, showing that the SOM-BI is trending towards overfitting the training data. We deduce that the optimal k occurs when the difference between P and R is small and the F1 score is close to its maximum value."

We consider the minimum difference between precision and recall and the peak of F1 score, which together lead to a close approximate value for the number of nodes. We think that this procedure quite clearly outlines a way for choosing a consistent number of nodes in any given case, e.g. of SOM-BI was applied to a different domain. We would thus rather refrain from defining an ambivalent confidence interval that could confuse this message for the general reader.

L.419 and 421 I think you're talking about Figure 10 here, not Figure 9.

Indeed – thank you. This Figure label has been corrected accordingly.

Please also note the supplement to this comment:
https://wcd.copernicus.org/preprints/wcd-2021-1/wcd-2021-1-AC2-supplement.pdf