

Weather Clim. Dynam. Discuss., referee comment RC1
<https://doi.org/10.5194/wcd-2021-1-RC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on wcd-2021-1

Anonymous Referee #1

Referee 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-RC1>, 2021

Review of "An unsupervised learning approach to identifying blocking events: the case of European summer" by C. Thomas et al.

The manuscript presents a novel atmospheric blocking detection method based on Self Organizing Maps (SOM), but most importantly compares it against a novel ground truth for European summer blocking which has been defined in both model and observations. Especially this latter point is very interesting since it provides a quite "revolutionary" and unique dataset to work with. The manuscript is interesting and presents a novel approach to the widely known but long-lasting issue of objective blocking detection. The SOM method seems to provide better results than the currently used detection method, although some unclear passages in the presentation makes unclear to me to what extent the improvements presented are actually useful in climate model assessment. The presentation is overall clear, even in some instances some sentences are hard to follow and might need some rephrasing.

I listed below my major and minor comments which I think that the authors should address or reply to before the manuscript is suitable for publication on WCD:

Major issues

- As also commented by the author, the absence of a gridded dataset as an output is a quite significant limitation, since it implies that the SOM blocking index is valid only a certain box – and the definition of the threshold such box is arbitrary. Therefore, I wonder how can be defined - using the SOM-BI - the blocking impact over a certain European subdomain? Or how can we extract information on the Rossby wave breaking properties associated with a specific blocked event? These are only a few examples of the usages of the traditional blocking indices, so I am wondering to what extent this interesting and innovative approach can be used to investigate climate models or atmospheric blocking impact. It would be nice if the authors can comment more about this.
- In the same way, it is unclear to me how the comparison with the traditional blocking indices has been carried out: did the authors consider the whole European sector as blocked if only one grid point is blocked? The description of this method is presented in an overly simplified way at L330. This approach may explain the incredibly high frequency of blocking since for AGP (92%), which in reality is about 10 times less over Central Europe, as also shown in Figure A3 by the authors. This passage should be clarified.
- It must be kept in mind that indices that the authors are using for comparison have been developed to study mainly winter blocking, and some of them are known for not being very suited for summer study. Did the authors perform the same analysis on winter blocking? One clear example is the AGP here used, which – as the authors noticed – produce a lot of noise at low latitudes. There are versions of this index which have been tuned to run also in this season. I would encourage the authors to make use of one of the improved versions – as the one presented by Woollings et al (2018) - which gets rid of the spurious blocking events at low latitude.
- Although I must admit I am not an expert of Self-Organizing Maps, I wonder to what extent a SOM approach is different from a k-means clustering with a predefined number of patterns. In this sense, I also wonder how a canonical k-means clustering with $k=4$, a widely adopted methodology to study the Euro-Atlantic mid-latitude variability which is based on Z500 anomalies, will rank among the detection methods here presented. Indeed, k-means aims at detecting Scandinavian Blocking specifically so I would imagine that this approach might have a high skill, comparable to the SOM (although also k-means has been developed for winter circulation). It would be very interesting to see a comparison between SOM and the Scandinavian blocking regime at least for reanalysis.
- I understand the intent of the authors is to provide a comprehensive approach on the topic of the mid-latitude variability, but the discussion on the sinuosity seems out of context, especially considering that this is a hemispheric diagnostic and the study is strongly regional. I would suggest the authors remove it from the introduction and from the figures, and perhaps replace it with the k-means clustering.
- As long as I understand the authors conclude that the best skill is obtained making use of the Z500 field. However, this seems to be somehow implied by the fact that Z500 is the field which the authors have used to define the ground truth. If the blocked days ground truth have been defined on SLP or on PV, would the SOM maps always show the Z500 as the best choice possible? I am not asking to redefine the ground truth using a different variable – this should be a monster work – but it would be nice if the authors could comment on this and support in a stronger way their conclusions.

- In the abstract the authors claims that the algorithm has no arbitrary threshold and this is an advantage compared to the other objective blocking method (L10): although this is true in the strict definition of threshold, there are several arbitrary "decisions" that have been undertaken by the authors, as the domain definition, the number of modes used in the SOM, or the dataset on which to train the model (that although is shown to be weakly sensitive, it introduces a methodological difference). I don't feel as any of these arbitrary decisions as feral issues, but I would encourage the authors to tone down their statement on the objectiveness of the algorithm since it is not radically different from the "standard" blocking indices they are comparing to.
- The authors spent a lot of time describing the different sensitivities of the SOM-BI to the SOM parameters. This is certainly a good thing, but they do not focus on the dynamical meaning of the SOM-BI blocking index. It would be very interesting to see a composite of the geopotential height pattern (and/or on other dynamical fields) of the blocking events which compares the SOM-BI with other blocking detection methods, as well as the evolution on the onset and on the decay phase. I understand that the authors aim at having a robust index from the methodological point of view, but it is very important to see if the blocking identified shows physical characteristics which are reasonable.

Minor issues

L49: I found that a mention to weather regimes should be added in the introduction.

L49: Sinuosity also includes planetary waves oscillation which are not blocking, and it is a hemispheric diagnostic while blocking is regional. I would suggest removing the discussion on sinuosity here.

L50: This sentence is a bit tangled up, please clarify.

L53: typo "a SOM"

L54: please remove "and better instrumentation and observations", it is out of context

here.

L55: why are the authors referring to "surface" here?

L58: what does the author mean here with "historical"? Please specify.

L59: What the authors mean with sinuosity here? Are you referring to a specific measure of the "waviness" of the mid-latitude flow?

L64: typo, remove "of"

L73: although is clearer later in the text, the concept of "blocking index skill" is quite new since to my knowledge a definition of ground truth for blocking events is quite uncommon. I guess that the authors need an introduction to the concept they refer to when they talk about skill.

L84: this is more of a naïve question from my side: why do the authors refer to "hyperparameters" instead of simply "parameters"? It looks to me that the number of SOMs is a parameter of the blocking detection method, or I am missing something here?

L98: the discussion of the anomalies should be left for the following part: here you are referring to which field you are using, so please remove the word "anomalies".

L108: So what is the cutoff frequency of the Fourier low pass filter?

L111: The beginning of this sentence is clumsy: do the authors want to mean that a detrending is applied only over the region studied? It sounds a bit redundant information.

L121: I would remove "following IPCC AR5 definitions (Stocker et al., 2013).The northern latitude is extended to 76 N when using data on a 2x2 grid" and just saying that the region ends at 76N

L122-124: this is one of the aspects of the authors' work I struggle to understand: as long

as I see, the method works on pentads which are defined arbitrarily. What I do not understand is if there is or not a "running-mean" approach. As long as it is presented this method looks like a 5-day discretization of the dataset, for which e.g. day 1-5 are blocked, days 6-10 no, and so on. Blocking duration is thus a multiple of five? I downloaded the dataset from the Zenodo archive and it does not seem to be the case, so the authors will probably need to clarify this. I am not sure if I have misunderstood something here, but of course if a discrete approach has been chosen, this would be clearly a caveat because blocking is a continuous process, so that in this way you may lose some block events or consider some other which is only a partial event. As a consequence, the ground truth may be wrong and all the conclusions you are drawing may be re-discussed.

L380: why not use half of the dataset as training and the other half for evaluation? This is a more common approach I would say.

Figure 5, 6: colors for the different blocking indices are missing in the caption.