Revision of wcd-2021-6 “Reconstructing winter climate anomalies in the Euro-Atlantic sector using circulation patterns” by Madonna et al.

We would like to thank the reviewers for their constructive comments and suggestions.
In response to the main issues raised by the reviewer, we plan to:

- provide a sensitivity analysis for the choice of the thresholds (e.g. d)
- clarify how the CE has been calculated, discuss the assumptions and their validity for our calculation, and improve the presentation and application of the scaling factor.
- better highlight the importance of the study and its implications

Reviewers’ comments are in black, our replies are in blue.

Reviewer 2:

The paper presents an analysis on the ability of different circulation patterns frameworks to describe the seasonal wintertime precipitation and temperature anomalies over Europe, using ERAInterim reanalysis data. Three different frameworks are used: the two NAO phases, blocking over three different regions and five Atlantic jet stream clusters.

The paper is clear and well written in all parts, and covers a current hole in the literature regarding circulation patterns, presenting original results which can be very useful for the community. I therefore recommend the article for publication in Weather and Climate Dynamics, once the following minor comments are considered.

General comments/suggestions

- I was very pleased to see the topic of this work, which I think was missing in the literature, and really enjoyed its reading. As a minor comment, I don't really understand the authors' choice not to consider - amongst the sets of circulation patterns - the "classical" four Euro-Atlantic weather regimes framework, which has been (and still is, to my knowledge) the most used in literature. I think it would be in the authors' interest to show the main results (e.g. Fig. 6 and 7/9) also for the 4 regimes framework, which would be of interest for many works on regimes when discussing the related impacts. I respect the authors' choice and I'm not asking to repeat the analysis for the k=4 case, but maybe some comments in the text and conclusions referring to the correspondence between the k=5 and k=4 regimes would help the reader "translating" the results to that framework.

We agree that the four Euro-Atlantic winter regImes widely used in the literature. We compared the different k in a previous study (Madonna et al 2017), using the jet latitude index (3 regimes), the classical weather regimes and the jet cluster using k=3,4,5. Even if using different variables (zonal wind vs. geopotential height) and domains, in Madonna et al 2017 we showed the 4 jet clusters correspond well to the 4 weather regimes. Using 5 jet clusters, the analysis revealed that the zonal regime is a combination of central (48%) and tilted (37%) jet (their Figure 10). We are happy to include and comment and “translate” our results to this perspective.

- As discussed by the authors, one of the problems of the reconstruction is the underestimation of the seasonal anomalies, due to the limits of the mean composites. Maybe a simple estimate of how much does the variability impact the reconstruction in the different regions would be given by the ratio between the standard deviation and the mean of the composite patterns. Just a suggestion.

We thank the reviewer for the suggestion. We can compute the ratio std(x)/mean(x) for the composite pattern. We expect to have large values of standard deviation within each pattern, as noted in lines 249-25, and this can be a reason for the underestimation of the amplitude of the reconstruction.

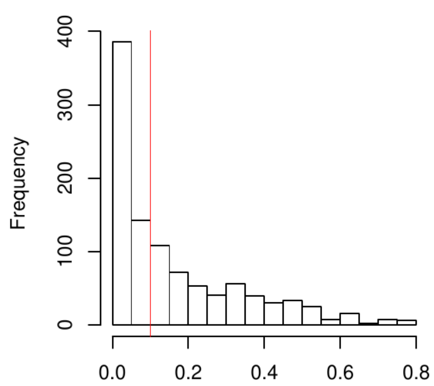
Specific comments

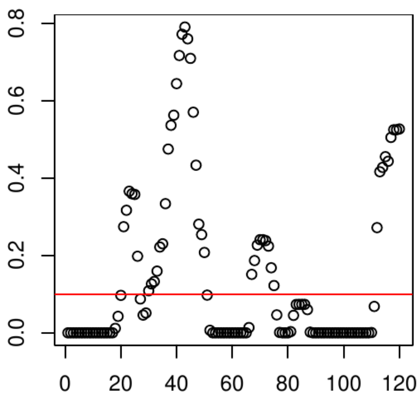
- L35. "..the North Atlantic jet stream can assume five different configurations..": the number of clusters to be considered is a matter of debate in literature and probably will not lead to a conclusive "best number", since all choices retain some level of arbitrariness. Please acknowledge here that this is the authors' choice, while other choices for the number of regimes are possible (e.g. as in Madonna, 2017).

We agree with the reviewer and will add a comment to acknowledge that.

- L70. Please briefly comment on the choice of 10%. Have you tried with other thresholds? Are the results sensitive to this choice?

The choice of 10% (i.e. 0.1) is to a certain extent arbitrary and arose as we did not want to select days where only a few gridpoints were blocked. The Figure on the left shows a histogram with the fraction of gridpoints (x-axis) in the "Greenland box" for all days with blocking, while the Figure on the right shows the fraction of blocked gridpoints for a subset of 120 days. The 10% (i.e. 0.1) threshold is marked by the red line. From the time series (right), we see that most of the points with values below 0.1 are followed by points with a larger threshold, which means that the block is building and the 0.1 threshold captures it only at a later stage. Translated to the occurence of blocked days per season, we expect that the absolute frequency depends on the threshold, but not the number of blocking events.





- L83. How is *d* calculated? Please show the formula here or add a reference.

d is the normalized inverse Euclidean distance from the centroid. The Euclidean Distance (E) for a point x from the centroid in a n-dimensional space is defined as $E = \sqrt{\sum_{i=1}^{n} (centroid_i - x_i)^2}$.

d is then the inverse of the normalised distance, d=1/Enorm, where the normalisation is Enorm = E /max(E).

- L86. Is this method and the threshold of 0.5 used elsewhere? If so, please add a reference here. Instead please briefly discuss the choice and its possible impacts on the results.

The choice of d=0.5 is somehow arbitrary but was introduced to be sure to capture only days that belong to the given centroid. We will provide some sensitivity tests to discuss the possible impact of this choice.

- L105. How close to a zero anomaly are the undefined days? I imagine the mean is very close to zero, but this may also come from cancellation of opposite anomalies. I'd suggest to show the standard deviation of the anomalies for the unclassified days (in the supplementary), to assess whether their exclusion might impact the skill of the reconstruction. The underestimation of the amplitude of the seasonal anomalies might also be linked to the filtering, since the denominator in equation 2 is always 90, but the number of assigned days is usually much less.

The anomalies of the undefined days are close to climatology. We will add them together with the standard deviation for the composites of all categories in the manuscript or supplement. The standard deviations in some regions are as large as the mean values. This can be understood from a synoptic perspective: for example, if two storms follow a similar path, the exact location where the most intense precipitation falls can be quite different.

From preliminary sensitivity tests, we don't believe our final results will change much with the number of defined/undefined days. However, we will do a more thorough sensitivity analysis to better understand the underestimation of the scaling factor in our revision.

- L129. CE>0.25. Is this threshold used elsewhere? If so, please add a reference.

A reconstruction that had no bias in the variance (a =1) and CE = 0.25 would have r=0.63, or explain 39% of the total variance; given 35 years of data (degrees of freedom), this correlation would be significant at p=0.0001. For a = 0.5, r=0.50 (p=0.0008); for a=0.25, r=0.63 (p=0.0001). Alternately, a reconstruction with r^2 = 0.5 and a=0.75 would have a CE = 0.57. As mentioned before, we will provide some more explanation for how to interpret the CE in the methods, along with some references.

- L135. Is the "/" a typo?

Thanks.

- Section 2.2.2. I found it quite hard the reading of the section before reading Section 3.3 and looking at Figure 8. To make this easier, I suggest to explicitly refer to Section 3.3 and Figure 8 in the text. Also, a possibility would be to move Section 2.2.2 to the beginning of Section 3.3, since the scaling factors are not used till then.

We agree that the introduction of the scaling factor in the method, but their application only in 3.3. make the reading and understanding of this part hard. We think that moving its introduction at the beginning of section 3.3, as suggested by the reviewer, would improve the clarity of the manuscript.

- L138. "small". with respect to the mean?

with respect to standard normal distribution

- L163. "are placed to allow easy comparison": has some automatic matching of the wind patterns been performed (e.g. maximizing the pattern correlation)?

In Madonna et al. 2017 the wind classification has been explicitly compared with blocking and NAO. There has been a temporal match (i.e. by looking at the number of overlapping days in each definition). A spatial match has been performed only between the different jet clusters in Madonna et al 2017 (i.e. for k=3,4 and 5, their Table 1).

- L174. The S-jet cluster is made up by less days, so that would probably enlarge the anomalies.

Let’s assume that the S-jet is a subsample of the NAO-. If the S-jet days are equally distributed within the NAO- days, the anomalies for S-jet should not be enlarged just because we use less days. But the anomalies would be enlarged if the subsample of S-jet belongs to the upper tail of the NAO- distribution, or reduced if they are located in the lower tail. Therefore, we do not expect the anomalies to be stronger just because fewer days are used.

- L178-179, L197. In the comparison with NAO+, the usual 4 regimes would look more natural to me. Probably a sum of the tilted and central jet states would show similar anomalies, as it is said in the text, and also correlate better with the NAO+ timeseries. It might be worth adding a comment on this in the text.

We will add a comment in the text and also add a discussion to Madonna et al. 2017, where the 4 classical weather regimes and the jet clusters (k=4 and 5) have been compared. In fact, the NAO+ regime (i.e. the zonal regime) is a mix of central (48%) and tilted (37%) jet (their Figure 10).

- L200. The unblocked days are different from the undefined/neutral days, so I won't put them in the same category. I expect (at least part of) them to correlate with positive NAO and have similar anomalies. I suggest to add a comment on this in the text.

We agree that the unblocked day category includes days that belong to the “zonal flow/NAO+ regime” and days that are undefined. This choice was intentional, as the goal was to assess how much blocking episodes influence the seasonal signal. We will add a comment on that in the manuscript.

- L201. See comment at line 105. I think it would be interesting to show the mean (even if close to zero) and standard deviation for these days in the additional material.

We will add these Figures in the supplement.

- Section 3.2. Have the reconstructions been performed for the wind also? Table 2 and L260 seem to suggest so. Why are the results not shown? It might be worth commenting on this briefly in the text.

Yes, we did that for wind as well. We presented only precipitation and temperature on the one hand because as seasonal means they have a stronger surface impact than the wind, and on the other hand because the winds/circulation is somehow included in the definition of the categories. We are happy to add the CE for wind in the supplement and add a brief comment in the text.

- L216. "that" -> than?

Thank you

- L224. Also over France the blocking method has substantially better skill.

The sentence has been modified to cover Spain and France.

- L227-8. "...regions with poor correlation skill mostly have low temperature variability..". Is this referred to the NAO phases or to all methods? I don't think this is true in general. For example Spain and southern Italy show low temperature variability but good skill for blocking/jet regimes.

The reviewer is right, and we will reformulate this part.

- L240. A large skill is also apparent over North Africa/the Mediterranean, I'd add a comment on this in the text. Also, if the figures are already available, it might be worth showing the equivalent of Figure S3 for the correlation also.

We will add the correlation for a larger domain also in the supplement and a comment regarding North Africa in the text.

- L292-297. I'd comment on the case of France as well, which is well represented only by the blocking framework.

We are happy to include a comment on that.

- L302-304. Also, this might be related to the filtering, see comment at line 105.

We will do some additional investigation on the scaling factor, as those suggested by the reviewer.

- L311. The "classical" 4 Euro-Atlantic weather regimes are usually calculated using larger domains, extending up to 40 degrees east. Do you think this could increase the skill in central/eastern Europe?

If much of the variability over easter Europe is captured by the first 5 EOFs, that could be the case. We know that the circulation over the North Atlantic plays an important role for the European region and that there is a good correspondence between the jets (k=4) and the weather regimes (from Madonna et al 2017). Using a metric that covers a larger domain might capture more of the variance over eastern Europe and increase the reconstruction skill. However, we have not tested that yet, but could be a nice additional analysis.