Review of "Past evolution and recent changes in Western Europe large-scale circulation" by Blanc et al. submitted to Weather and Climate Dynamics (WCD-2021-69)

General comments

This study investigates the evolution of geopotential height over Central Europe during the past 170 years based on different reanalyses. The authors first analyze how similar the geopotential height fields are in terms of shape among the different reanalysis products. They further investigate how intrinsic characteristics of large-scale atmospheric states such as stationarity, uniqueness, and gradients compare between different reanalyses and how they developed over the past 170 years. They ultimately stratify these flow characteristics according to the four major Central European flow patterns. The findings are used to provide potential explanations for changes in Central European surface weather over the past century found in previous studies. The manuscript has a clear structure, the underlying method is simple and relatively elegant, and the various methodological steps have been performed and described thoroughly. Overall, the findings of this study are important to further our understanding of historic (and thus potentially future) changes in Central European weather. Furthermore, they are valuable for users of any of these reanalysis products because they highlight the partly substantial differences between these "observational" datasets that are often treated as the "truth". Despite these valuable insights, the study is quite detailed, descriptive, and thus on the edge of being too lengthy. Furthermore, it partly misses to better highlight the novelty of its findings and how they can specifically be used for further research. Also, some of the methodological steps and figures require more careful description, particularly for readers who are not familiar with the used method. Therefore, I recommend publishing this manuscript after major revisions.

Major comments

Interpretability and implications of the results (1): Although the method (TWS and derived descriptors) is elegant in my view, I struggle to interpret some of the findings obtained with it and the implications it has for further research. For instance, Fig. 4 is very interesting and contains lots of important information and food for thoughts. Nevertheless, it is hard (at least to me) to "translate" the differences between reanalyses and between periods into more intuitive concepts of the large-scale circulation (i.e., geopotential height, storm track, weather regimes, etc.). For instance, what does a change in the summer celerity anomaly from –0.06 to 0.0 (in 20CR mean) over the last 170 years mean? Is this a large change in stationarity we would for instance see in weather regime statistics as well? And should we worry about these differences when using these reanalysis products for certain applications (and for which applications)? Likewise, what does the large difference in singularity and relative singularity in the early 20th century between 20CR and ERA20C tell us? Which of the two reanalyses should we "trust" more? Can you discuss this in your manuscript, if possible? Related to this, I think you should briefly discuss the limitations of the TWS in general. For instance, it does not tell us how exactly two geopotential fields differ from each other. You partly overcome this by additionally showing differences in geopotential fields (e.g., Fig. 3), but you could still discuss more clearly what added value we get from using this score and what limitations it has.

Interpretability and implications of the results (2): I assume the multi-annual/-decadal fluctuations in the different atmospheric descriptors (Figs. 2 and 4) is caused by both trends/changes in the observations that are assimilated and multi-annual/-decadal variability in the Atlantic-European large-scale state (for instance, a decade of predominantly persistent negative NAO phases might manifest in a lower celerity; cf., e.g., Woollings et al., 2015). Can you estimate or at least speculate which of the two might be more important, particularly in the late 19^{th} and early 20^{th} centuries? Have other studies investigated this question? It might be a quite relevant question for users of the different reanalysis products.

Significance testing: You perform careful and important significance tests for detecting trends/changes in the atmospheric descriptors, which is great, but you don't do the same for the differences in the geopotential anomalies shown in, e.g., Fig. 6. Could you perform such a test and highlight the significant regions? For instance, I doubt that the mentioned reinforcement of the Atlantic anticyclone in winter (L303) in Fig. 6 (top left) is really significant…

Investigated domain: I understand the reason for using the study region centered over Central and Southern Europe (rectangle in Fig. 1a), as you are primarily interested in focusing on weather over France. However, I think some of your results for the descriptors might be quite sensitive to the choice of this domain. Did you test this? For instance, I assume increasing the region over whole Europe or even including the North Atlantic might yield results that are more linked to the typical weather regimes in these regions and might thus be easier to interpret. If you do not want to discuss this in too much detail in the manuscript, could you at least extend the anomaly maps in Figs. 1a, 3, and particularly 6 to a larger domain including the main centers of action of the NAO and the further modes of variability (of course by leaving the black rectangles to see the study region)?

Method: It took me some time to fully understand the TWS and the derived descriptors (I did not know it before), which I assume other readers might experience, too. Although your schematic in Fig. 1b is very helpful, there are some methodological details you could add to the text. For instance, on L176: How exactly do you compute the TWS between two reanalyses when referring to Eq. 1? What are the days t_k and $t_{k'}$ in this context? Do you go through each day (for instance in a specific season) and compute the corresponding TWS between the field of reanalysis 1 and reanalysis 2, which gives you one TWS value for each day, which you then average over the season? This might be a trivial question to you, but it is not obvious from considering Eq. 1, which refers to a day k and another day k'. Furthermore, relating to Fig. 4: Again, I am not sure if I completely understand the details here. Let me explain how I understand it: for the celerity of a specific season between 1850 and 2010 (e.g., for 20CR), you go through each day k during each DJF season and compute the celerity with respect to the previous day k-1. This yields as many celerity values as you have days in this time period. You finally compute a 5-year running mean over this daily celerity time series, which is plotted in Fig. 4 (as anomalies relative to the celerity climatology between 2006 and 2010). Is this correct? Likewise, for the singularity and relative singularity you go through each day and compute the two terms by using a Q of 0.5% of all days available? Are the 0.5% relative to the whole period 1850 to 2010, or do you split it in sub-periods as you mention on L160? And finally, why do you compute the index anomalies against a climatology between 2006 and 2010 only? Would it not be much more robust to use the climatology over the whole investigated period?

Minor comments

Title: I would write "in Western European large-scale circulation", but I'm not fully sure.

L20: You could slightly extend this sentence in the sense that the LSC provides the general dynamical setting for various local weather phenomena, not just via the airflow towards a region but also via modifying stability and moisture availability.

L21: Replace "Over the large scale" with "The"

L23: This sounds a bit odd, because the further (second, third...) modes of variability are also active throughout the year, i.e., not just the NAO.

L25: "by intensifying westerlies"

L25-27: This is an important point, that additional modes of variability are needed to explain surface weather modulation particularly in Central Europe. You could add some further references here, for instance Pasquier et al. (2019), who investigated the link between precipitation extremes in Europe and a higher number of weather regimes than the classic 4 regimes.

L28: “associated with wet conditions at the northern flank”

L38: Rather something like “low pressure systems developing over the Atlantic and Iberian Peninsula”?

L39: “associated with strong southwesterly to southerly flows”?

L43-45: I would distinguish here between stationary cyclones that lead to persistent moisture transport towards the Alps and thus accumulated precipitation and large-scale situations associated with high instability (for instance induced by upper-level cut-off lows; e.g., Portmann et al., 2021) and thus strong forcing for thunderstorms, which are also crucial for (convective) heavy precipitation.

L49: Remove “long-lasting” as it is already stated with “persistence”.

L51-52: You could for instance add Zschenderlein et al. (2019) who investigated the processes behind European heat waves in detail and climatologically.

L55: “may have been linked to”

L88: “500 hPa geopotential height ranges from”

L89: Rather “pressure distribution”?

L93: How exactly is the ensemble for the 20CRv2c created, i.e. which kinds of “perturbations” are they based on?

L96: Could you add one or two sentences about the quality of this reanalysis, for instance during the 19th century? I assume continuous pressure / SST / sea ice observations from ships were rather rare at that time?

L103: “referred to as full-input”

L112 and later: I’m not so convinced about the term “atmospheric influences”. As they basically refer to the direction of the flow towards Southern France, could you replace it with something like “flow directions”, “flow patterns”, "weather patterns", or similar? Also, you might switch 3.1 and 3.2, because you mainly start by analyzing the atmospheric descriptors and only then use the atmospheric influences. If done so, you could also split Fig. 1 into two, with a first figure showing the schematic in current Fig. 1b, and a second figure showing the patterns in current Fig. 1a. Furthermore, I would find it very useful to see the links of these four patterns to well-known North Atlantic-European weather regimes / modes of variability. You could partly achieve this by increasing the plotted domain (see one of my major comments). Furthermore, you could write a few sentences about it or refer to other studies that did.

Fig. 1a: Could you increase the wind vectors in Fig. 1a, or make them thinner? It is hard to see their directions. Furthermore, are the names Northeast and Anticyclonic really appropriate, as you primarily consider the influence of the pattern on the flow towards Southern France? Would it not rather be something like Northern (for Northeast) and Northeast (for Anticyclonic)?

L120: “trends being poor” sounds odd – you should rephrase to “small”.

L178: “to allow for the computation”

L162: Why do these three measures only focus on shape? As far as I understand the TWS, it considers the (sum of the) pressure difference of each grid point with respect to each other grid point in a certain domain, and thus indirectly also accounts for differences in pressure magnitudes, right?

L166: I am a bit confused about the notation in Eq. 5: Does max_j stand for the maximum value of all grid points j within the rectangle domain? So MPD is the difference of the maximum pressure and minimum pressure (one grid point each) on a certain day k?

L168: I assume the weak relation of your MPD to the NAO is simply because your rectangular domain (Fig. 1a) is too small to include the centers of action of the NAO, right? So, if you would increase your domain enough a large MPD would start to coincide with a strong positive or negative NAO, right (see also one of my major comments)?

L170: What exactly do you mean with "per season"? Do you compute all four atmospheric predictors (cel, sing, rsing, mpd) on a daily basis and you then average them over the seasons, or do you compute the atmospheric predictors for the already averaged seasonal geopotential fields? I guess it must be the former, but you should specify this.

L183: At first sight, it seems counter-intuitive that the geopotential fields agree less well in summer than in winter, considering the generally larger baroclinic activity in the investigated region in winter probably associated with a larger potential for "errors" in the reanalysis. Do you have an idea why they agree less well in summer? Could it be associated with (smaller-scale) convection?

L188: Why are the shapes more similar between 1850 and 1900 than between 1900 and 1950? I would have assumed that the shapes become generally less similar the fewer observations are available (i.e., the further back you go in time), but this is not the case. Is it because you generally increase the degrees of freedom from 1900 on, with the introduction of further types of observations and the 4D-Var? Can you explain?

L194: Did you plot the same as in Fig. 3 but with the absolute fields (i.e., 4 fields for each season: 20CR late-century, 20CR early-century, ERA20C late-century, ERA20C early-century)? If yes, could you show them in your response?

L210: Does this mean that 20CR should only be used with caution for studying any climate change / trends over the past 170 years, as it misses a dynamically very relevant change of meridional pressure gradients over the Atlantic-European region (assuming that ERA5 is the best product we have, which is a reasonable assumption)?

L217-219: I find it intuitive that the higher resolution and "more details" in ERA5 generally increase the MPD compared to the other reanalyses, but I do not find it obvious for the other descriptors. Wouldn't a higher celerity for instance mean generally less persistent atmospheric states? How do you explain this? Is it because you get more smaller-scale features in ERA5 which tend to be less persistent? Likewise, does the much higher relative singularity in ERA5 imply much more rare/extreme atmospheric states in ERA5 compared to the other reanalyses? Why is this – due to the same reason as above? I think it would be helpful for the reader if you could discuss this a bit.

L274: "according to ERA5" or something similar

Figs. 5, 7, and 8: These are nice figures, but they contain a lot of information. To simplify slightly, would it make sense to replace the classic box plots (right and left of the difference) with violin plots but remove the difference plots instead? You could even make the blue and red violins transparent and overlay them, which would allow directly seeing the absolute values and differences. This would reduce the figure from currently 48 boxes to 16. Could you test this and check whether it is useful?

Fig. 5: I'm not sure if I understand correctly: You say that the percentiles are defined with respect to all days within 1950 and 2019. So, shouldn't the sum of the blue and red bars consequently give a distribution that is somewhat centered around the 50^{th} percentile on the y-axis? For many descriptors, such as the celerity in winter or summer, however, this does not seem to be the case. Why is that?

L296: "meaning new anticyclonic patterns are less explored over the present period" – this sounds a bit strange, although I know what you mean referring to the singularity. It sounds like we had completely different atmospheric patterns in the earlier period, which do not exist anymore. Can you rephrase somehow?

L300: “associated with high pressure”

L310: “reproducibility of Atlantic circulations” sounds a bit like a model that fails to reproduce these patterns, although we are dealing with reanalysis. Could you rephrase somehow?

Figs. 9 and 10: These two figures appear out of a sudden in Section 4.2.3, without introducing the reasons behind showing them (for instance, why only for summer and autumn, and the Mediterranean circulation?). Furthermore, the description in the caption of Fig. 10 is hard to understand and I had to read it several times. In general, I wonder if these two figures are really needed or whether you could simply state some of the most important changes in, for instance, the occurrence frequency of the most extremely stationary states? Although the idea behind Fig. 9 is good, for instance, I find it hard to grasp a message from looking at this figure, as the two distributions are overall rather similar. Removing these two figures would further help to reduce the content of the rather long manuscript…

L337-339: What might be the physical reason to obtain such a “seasonal shift” as you call it (considering the fact that the solar cycle, i.e. the main driver of the seasonal cycle, should not be affected by any change)? Isn’t it more the northward shift of the storm track that we might see here, which influences the Mediterranean differently during the different seasons because its overall year-round meridional oscillation also changes with the northward shift?

Conclusions: Coming back to one of my major comments, I miss some more specific implications / ideas for further research in this last paragraph. When reading the conclusions, it is not clear whether your main take-home message should be the usefulness of your method to better understand surface weather extremes in France, or whether it is more about better understanding historic large-scale changes over Europe in general with this special approach and highlighting differences in reanalysis products. To me, it is kind of both, but you should state/distinguish this more clearly. Also, you could again refer some of your results to some other specific studies or ongoing debates. For instance, you could relate your findings about celerity (i.e., stationarity) to the widely discussed (but partly disputed) increase in jet stream waviness due to Arctic amplification, potentially leading to more persistent and thus extreme weather (e.g., Kornhuber & Tamarin-Brodsky, 2021).

References

Kornhuber & Tamarin-Brodsky, 2021: Future changes in Northern Hemisphere summer weather persistence linked to projected Arctic warming. *Geophysical Research Letters*, https://doi.org/10.1029/2020GL091603

Pasquier et al., 2019: Modulation of atmospheric river occurrence and associated precipitation extremes in the North Atlantic Region by European weather regimes. *Geophysical Research Letters*, https://doi.org/10.1029/2018GL081194

Portmann et al., 2021: The three-dimensional life cycles of potential vorticity cutoffs: a global and selected regional climatologies in ERA-Interim (1979–2018), *Weather Clim. Dynam.*, https://doi.org/10.5194/wcd-2-507-2021

Woollings et al., 2015: Contrasting interannual and multidecadal NAO variability. *Clim Dyn,* https://doi.org/10.1007/s00382-014-2237-y

Zschenderlein et al., 2019: Processes determining heat waves across different European climates. *Q J R Meteorol Soc,* https://doi.org/10.1002/qj.3599