Weather Clim. Dynam. Discuss.,
https://doi.org/10.5194/wcd-2020-37-AC2, 2020

Interactive comment on "A regime view of future atmospheric circulation changes in Northern mid-latitudes" *by* Federico Fabiano et al.

Federico Fabiano et al.

f.fabiano@isac.cnr.it

Received and published: 11 December 2020

We thank Reviewer 2 for the constructive comments to the manuscript, which stimulated a deeper understanding.

This paper deals with the changes in weather regimes in the Euro-Atlantic and Pacific-North American sectors, in both CMIP5 and CMIP6 models. A comparison of simulated and observed weather regimes is done and then the changes in frequency and persistence of the weather regimes are characterized. Potential drivers of those changes are finally discussed. This is a nice paper, well written, with interesting results. The paper is concise, despite the large amount of work necessary to do these analyses. I recommend publication of the paper after some moderate revisions.

Main comments

– The weakest part of the paper (but also, potentially, the most interesting one) is probably section 4.2. The analyses themselves are interesting but I think the interpretation of the results should be more cautious. It is not because a significant (?) correlation between changes in weather regimes frequency and what the authors call the "drivers" is found that a causal relationship exists. The weather regimes could themselves impact these drivers, and other factors, not studied by the authors may physically explain both the changes in weather regimes and in the drivers. More drivers could also have been considered: e.g. many studies have shown the potential impact of SST anomalies (especially in the Tropics, but not only) on large scale circulation, including on weather regime occurrence. I think it is difficult to discuss the potential drivers of weather regime changes without investigating the role of regional SST changes.

- Thank you for the comment. We agree with the reviewer that the term "driver" might be ambiguous because it suggests a causal relationship, which cannot be directly implied by our analysis. However, our choice has been somehow inspired by recent literature (e.g. Oudar et al., 2020; Zappa and Shepherd, 2017), where processes that can potentially affect the mid-latitude circulation under increased forcing are referred as "drivers". In response to this comment and to a similar argument by Reviewer 1, we added the term "potential" to the title of Section 4.2 and referred to the correlations found as "link/relationship/connection". We further discussed in the paper the possibility that the found relationships might work in the opposite direction too, or that both the changes might be independently driven by external forcing. To assess the role of regional SSTs changes, we investigated the correlation patterns between WR frequency trends and changes in the surface temperature (tas) across the multi-model ensemble, for the ssp585 scenario. This has been done calculating the correlation between the tas trends of all models at each point of a common grid and the regime frequency trends of all models. Both the tas and frequency trends were divided by the global temperature trend before computing the correlations. For the EAT regimes no significant

pattern appears, except for the NAO+, which shows significant correlations in some areas of the central and western tropical Pacific. The situation is more interesting for the PAC regimes, which show significant correlations in the tropical North Atlantic and in the North Atlantic subpolar gyre. Indeed, the North Atlantic warming (NAW) was considered as a potential driver in Section 4.2. However, we decided to split the NAW in two parts, in order to assess the relative importance of the tropical and subpolar North Atlantic. In addition, significant correlations for the PNA+/- regimes are also found in Northern Africa, the Indian ocean, the Northern Pacific and some smaller regions in the Southern ocean. However, the understanding of these correlations requires further analysis that goes beyond the scopes of the present paper.

– I also think that a discussion on the potential causes of the differences seen between CMIP5 and CMIP6, which are sometimes quite large, should be added to the paper. An interesting analysis would be to evaluate whether differences between the changes in weather regimes are linked to differences in the changes of the drivers between CMIP6 and CMIP5 (if the authors are right about the drivers, it should be the case).

- The differences between the changes observed in RCP85 and ssp585 might be mostly related to differences in the forcing. In fact, in terms of CO2 concentration, the CMIP5 RCP85 scenario represents an intermediate narrative between CMIP6 ssp370 and ssp585 (Meinshausen et al., 2019; Tebaldi et al. 2020, https://doi.org/10.5194/esd-2020-68). Moreover, a recent study with the EC-Earth model finds that about half of the difference in warming by the end of the century when comparing CMIP5 RCPs and their updated CMIP6 counterparts is due to difference in effective radiative forcings at 2100 of up to 1 Wm-2 (Wyser et al., 2020; doi:10.1088/1748-9326/ab81c2). Also, the models considered for the CMIP5 scenario differ from the CMIP6 ones.

– Many methodological choices are necessary in a weather regime analysis: e.g. preprocessing through EOF analysis or not, if yes number of principal components retained, precise algorithm (many variants of k-means exist), choice of the number of clusters, whether all the days are classified or not, if not what is the criteria for not

classifying some days, which variable is used (slp, zg), how to best take into account the mean increase in zg due to global warming etc. I somewhat understand why the authors don't want to discuss all these aspects, show sensitivity tests etc. The paper would be too long and less interesting. But nevertheless some aspects deserve to be better justified, at least with a sentence. Sensitivity tests could also be added in SI (as the authors decided to add SI to the paper). For example, the authors have chosen to first apply an EOF decomposition to zg and then to retain the 4 leading EOFs. They give no justification for this pre-processing step. The use of principal components for cluster analysis has been common in the past but I suspect that in many old studies, it was more a question of dimension reduction in order to run classification algorithms on computers with limited resources. The use of principal components is also sometimes intended as a kind of filter. If it the case here, the authors should explain their objective(s). It is also unusual to retain such a small number of principal components for clustering, and therefore such a low level of explained variance. I would like to see a discussion on this point and potentially the result of a sensitivity test: e.g. changes in weather regimes without EOF analysis, or with enough principal components to keep at least 90% of the variance etc.

- We totally agree with the reviewer that many methodological choices are necessary for the Weather Regimes analysis. Excluding the detrending and the idea of projected regimes, which here were necessary to take into account the huge (transient) changes in mean climate in the different scenarios, most of the choices adopted in this paper build on Fabiano et al. (2020a). In that paper, a more detailed methodological analysis and discussion are presented, regarding: a) the impact of different number of EOFs; b) changes in the selected region; c) changes in the construction of the climatology; d) the impact of the projection on the reference phase space to obtain pseudoPCs. In particular, the changes in the Weather Regimes when considering for example 10 EOFs instead of 4 turn out to be negligible for the EAT sector (Section 3.1 in Fabiano et al, 2020a). The same was found for the PAC regimes by Straus et al. (2007), for the case with no filtering applied. This is due to the fact that the regimes are large-scale

patterns, well explained by the first 4 EOFs.

– Finally, it would also be interesting to characterize the impact of internal variability on the changes in weather regime occurrence, as signal-to-noise may be small regarding atmospheric circulation changes. It would help to better evaluate the importance of future changes, of the differences between CMIP6 and CMIP5, and between scenarios etc. There are a few CMIP6 models with at least 10 members for which Zg is provided (for future scenarios). It would be very interesting to use the members of these models to characterize the impact of internal variability (for a single scenario, ssp585 for example, this is sufficient).

- Thank you for the suggestion, we agree that this would be a very interesting addition. However, the computational and additional storage resources necessary to collect and analyze the daily geopotential height field from all available single-model ensembles is non negligible, considering also that overall we have already analyzed more than 150 daily datasets. To take into account this issue, we are estimating the impact of the internal variability on our results by looking at a single model ensemble. We will provide a comment on this in the main text.

Specific comments

– L15. Please define precisely what is meant by "low-frequency" in the paper.

- We intend low-frequency as the variability on scales longer than about 5 days, which in the EAT sector is mainly related to latitudinal shifts of the jet stream. This was clarified in the revised manuscript.

– L103. I think that Table S1 with the models and members used should be in the main document.

- Thank you for the suggestion, we will move it to the main text.

– L105-108. I'm not sure it is good idea to combine the results of two reanalyses (ERA40 and ERA Interim). Moreover, both of these reanalyses have now been su-

perseded by ERA5, with major improvements in ERA5. I understand that ERA5 is not (yet) available before 1979, but I'm not sure why the authors absolutely need to start in 1964. They could use the period 1979-2015. It would lead to a smaller sample, it is true, but I also think that the hypothesis of a linear trend (section 2.2.1) may be more reasonable on a shorter than on a longer period.

- The use of ERA5 would have been preferred if the whole period had been available in time for this analysis. The need for covering a longer observed period is dictated by the fact that the internal decadal variability in quantities like the WR frequency is quite large, and we evaluated that a period of 50 years would be necessary to assess significant changes. As it is reported at line 175 of the original manuscript, the variability of the frequencies on this period is estimated to be around 1.6%, which is satisfactory for our scopes. We made an exception to this choice only for the historical simulations from CMIP5, which stops at 2005 by construction and therefore spans only 40 years instead of 50. In this regard, the combination of the ERA40 and ERAInterim datasets might be seen as a standard workaround, already adopted in several works analysing mid-latitude variability (e.g. Schiemann et al, 2017; Davini and D'Andrea 2020). In particular Dawson et al. 2012 used a combination of these reanalyses to compute Euro-Atlantic clusters, using a methodology very similar to that presented here and showed that the cluster patterns computed using the NCEP reanalysis are almost identical (see Table 1 in Dawson et al. 2012). We have also checked the ERA40-ERAInterim combined dataset vs NCEP on the period considered and found almost no-difference in cluster centroids, with a pattern correlation of about 1.

– L118. Please explain (in the paper) why you need to "detrend" the data (by the way, I'm not sure that detrend is the good word considering what is done). And please justify why you use the average on the Northern Hemisphere (30-90N), rather than the averages on the domains of classification, as it is usually done. I think it may have some non-trivial implications, for example if the mean increase in geopotential height, largely controlled by warming, is different between the Euro-Atlantic and Pacific-North

American sectors, as it may favor artificially some weather regimes in both domains.

- Thank you for the comment. This point has been now clarified in the text. Figure 1 shows the average geopotential height at 500 hPa (in units of meters) in the Northern Hemisphere (30-90N) for all models, merging the historical and scenario simulations. The need for the detrending is due to the fact that anomalies associated with the WRs are of the order of 100 meters, which are comparable to the average increase in the mean geopotential height field seen in the scenarios. Even if to a lesser extent, also the historical simulations show such an increasing trend. A method to remove the trend - in order to not influence the regime detection has been therefore developed: a linear detrending has been applied to the historical trend and a polynomial detrending has been applied to the scenarios - in order to take into account the fact that scenarios exhibits a non-linear behaviour. In order to retain decadal basin-wide fluctuations, such as the AMV (which we do not want to remove), we decided to calculate the trends on the whole Northern Hemisphere, and not for the separate domains. The difference in the future trends when considering the whole hemispheric or the sectorial averages is very small (see Figure 2). On the other hand, this choice may have a larger effect on the historical trends, which are calculated on 50 years only (compared to 85 for the scenarios). However, following the above-mentioned argument on the decadal basin-wide fluctuations, the evaluation of the hemispheric trends is more reliable than the equivalent one in sectorial regions.

– L132-134. See my main comments. Why do the authors only use 4 EOFs for around 50% of explained variance?

- We added a clarification in the text, mentioning the sensitivity tests done in Fabiano et al. (2020a): “Sensitivity tests performed in Fabiano et al. (2020a) for the EAT sector show that the changes in the regime patterns when considering for example 10 EOFs instead of 4 are negligible.”

– L136. I’m not sure that it is a very good justification, but OK. . .

- We are not interested here in assessing the “right” number of clusters to be used for the two sectors and we acknowledge at line 74 (original manuscript) that this number is still a matter of debate. We then adopt the most common choices in the literature, which are 4 clusters for both the EAT (Michelangeli et al., 1995; Cassou, 2008; Dawson et al., 2012; Madonna et al., 2017; Strommen et al., 2019; Fabiano et al., 2020) and PNA sectors (Straus et al., 2007; Weisheimer et al., 2014), in order to set up a framework to discuss future changes in the circulation.

– L142. How is the centroid exactly defined: average, median, real day closest to the average etc?

- The centroid is defined in phase space as the average of all days (PCs) assigned to a certain cluster. This has been now clarified in the revised text.

– L145. Are all days classified and why? In some studies, the transition days are not classified, for example.

- Yes, we classify all days here. Although not classifying all days may be a legitimate approach, it requires the definition of a rule for excluding some of the days (threshold on the deviation from the mean state, on the velocity in phase space, ecc.) and the definition of an arbitrary parameter to set the strength of the filtering. We judged it more conservative not to exclude any day from the clustering.

– L146. Computed regimes: how are the computed regimes associated with the observed regimes? Is the associated observed regime the one with the stronger spatial correlation? Are there ambiguities? (e.g. a computed regime that looks like something intermediary between two observed regimes)

- Thank you for the comment. The matching of the computed and observed regimes is done minimizing the average RMS between all regime couples. This usually coincides with the best matching obtained maximizing the spatial correlation. However, as you point out, it may happen that some ambiguities arise when a computed regime is very

far from the observed ones: this usually happens to two regimes at a time, which turn out to be a mixture of the two observed regimes they should have reproduced. This is quite rare, but when it happens, the corresponding metric in the Taylor plot is poor: the outliers in Figure 2 (with pattern correlation close to zero) are probably examples of this. The potential instability of the computed regimes is one of the reasons why we decided to use the projected regimes when considering the changes in the future scenarios.

– L176-177. Not clear to me.

- These estimates on the variability on the 50-yr window are the standard deviation of the mean of the observed regime frequency and persistence in individual seasons (so the standard deviation divided by the square root of n_season - 1). This applies if we assume the consequent seasons to be independent. The actual variability on 50-yr windows might be larger than this due to, for example, decadal fluctuations in the WRs. We rephrased the sentence in the text to explain this.

– L223-22. I'm not sure to understand the reasoning.

- We agree that the sentence has not been formulated well. Actually, our hypothesis here is that the tropically-induced modulation of the North Pacific regimes might be too strong in models. Therefore the regime structure turns out to be too "deep" and there is less room for larger deviations from the attractors. Molteni et al. (2020) showed that the response of the NAO index to the tropical Pacific forcing is well represented in models, while the teleconnection of the Atlantic sector with the Indian Ocean is not well caught. However, this does not really help our argument here. We changed the sentence in the revised manuscript as follows: "It's worth noting that – opposite to the EAT sector – models tend to produce larger variance ratios for the PAC regimes than it is observed. We speculate that this may be due to an excess in the tropically-induced modulation of the PAC regimes in models."

– L263. "Further analysis. . ." I understand, but it is unfortunate. It is a very interesting

(and strange) results. Discussing some hypotheses would have been nice. Note that it may be relevant regarding the discussion about the "drivers", as the "drivers" could be implicated in these variations (if they really are "drivers").

- The most promising hypothesis is related to the aerosol forcing, that could have a role in driving in-phase AMV oscillations in the model simulations. This has been hypothesized for the observed AMV (see e.g. Zhang et al. Have aerosols caused the observed Atlantic Multidecadal Variability? J. Atmos. Sci. 70, 1135–1144 (2013); Qin et al. 2020, DOI: 10.1126/sciadv.abb0425). In turn, the AMV perturbs the observed frequency of the NAO+/- regimes (a positive AMV increases the NAO- frequency). It is not clear whether a similar process might be at work for the future scenario period, but the way seems promising.

– L315. I think the mean changes in geopotential height at 500 hPa (maps) should be shown somewhere, maybe as the first figure of the paper.

- Thank you for the suggestion. The changes in the geopotential height at 500 hPa are dominated by the global positive trend, shown in Figure 3 of this response. This is the only component of the change in the geopotential height that we are not showing in the paper, but a corresponding figure (similar to Figure 3, left panel) could be added to the Supplementary material. The residual zonal and local trends of zg500 for ssp585 are shown in Figures 8 and 9. We judged these residual trends more interesting from a dynamical point of view, since they reflect the changes in the circulation that we observe.

– L320 and legend of figure 8. It is not clear what the "trend" is here. In section 2.2.1 the trend is defined as filtered area-weighted average Northern Hemisphere geopotential height, and used to detrend the local geopotential height. Therefore, if my understanding is correct, here the authors look at the zonal mean "trend of detrended" data, right? It is not clear, and bit awkward from a vocabulary point of view. Please indicate the period used to compute the trends in the legend.

- Thank you for the comment. Yes, this is correct. The quantity shown differs from the zonal trend of the original geopotential height fields only by a constant, represented by the global NH trend. The complication here is that we used a polynomial detrending for the scenario data, so that's not simply a linear term. The period is the full scenario period 2015-2100.

– L323: zonal mean trend or zonal mean trend anomaly, as said in the legend?

- We refer here to the zonal mean trend anomaly, the text was corrected.

– L351. For the predictors I suppose that you also use the trends, as for the predictands (it is not said explicitely)? What is the period on which the trends are calculated for the predictors and the predictands? 2015-2100 I guess, but it is not explicitly said I think (also in the legend of Fig 10).

- Yes, we also use the trends for the predictands, divided by the global mean surface temperature trend. All trends are calculated on the 2015-2100 period. We will add this information to the main text and to the legend.

– Section 4.2. Are there correlations between changes in EAT and PNA regime frequencies? I think knowing that might be useful for the discussion in this section.

- Thank you for the suggestion. We did not compute correlations between the two sectors frequencies, but we agree this would be interesting to explore. We will compute them and add a comment in the text.

– L352. Why 2 or 3? Please justify.

- The scope of Section 4.2 is to find the most significant set of potential drivers for the two sectors. In this regard, a lower number of predictands is desirable. We were inspired by Peings et al. (2018) and Oudar et al. (2020), that consider 2 and 3 drivers respectively, to search for the best 2 and 3 drivers models for the two sectors.

– Are (i) all the models with the ssp585 scenarios used, or, at it seems to be the

case based on Table S1 (ii) only the models with the 4 scenarios are used even for this analysis? I think (i) would be much better as it would lead to a larger sample of models, which is quite small with (ii). It is even truer since some of the models are nearly duplicates: models at different resolutions, ESM and AOGCM from the same group etc., which decreases the "effective" sample size.

- The models with available daily zg dataset for ssp585 were 22 at the time when the analysis was done. This number is reduced to 19 with the constraint on the availability of all ssps, which were used for all analysis in the paper, including Section 4.2. We understand that the number of models is somehow limited, nevertheless it can give some indication in terms of correlated quantities.

– L362. I don't see figure S7. Are these correlations significant (with the issue of effective sample size mentioned above it is difficult, or impossible, to do the test right, but it is still an interesting indication).

- Thank you for pointing this out, and sorry for forgetting to add the figure to the Supplementary. The figure is attached here as Figure 3. The significance of the correlations at 99% and 95% level is indicated by the big and small white circles in Figure 10. The issue of the effective sample size might effectively reduce the significance of these correlations, a comment on this issue is included now in the revised manuscript.

– L387. Any idea of the reason(s) that might explain the improvements in CMIP6? Please discuss.

- CMIP6 models have been shown to improve in many regards with respect with the previous CMIP generation. This has been observed also for the northern mid-latitude circulation. In particular, models have been shown to have smaller biases in the blocking frequency (Davini and D'Andrea, 2020) and in the representation of storm-tracks (Harvey et al., 2020, https://doi.org/10.1029/2020JD032701). This finding is consistent with the improvement observed for the weather regimes. A possible reason for the better performance of the CMIP6 models might lie in the refined horizontal resolutions,

which are significantly larger in CMIP6 models compared to CMIP5. In this regard, in Fabiano et al. (2020a) the models' response to increased resolution has been analysed in terms of the simulation of weather regimes and an overall improvement in the representation of regime patterns and variance ratio was found.

Interactive comment on Weather Clim. Dynam. Discuss., https://doi.org/10.5194/wcd-2020-37, 2020.

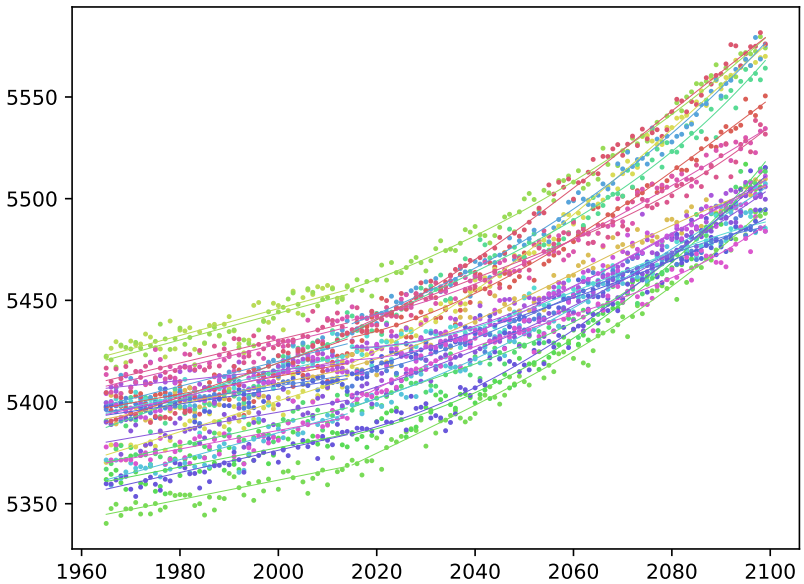

Fig. 1. Average geopotential field in the Northern hemisphere (30-90N) for all model simulations in the historical+ssp585 scenario (scatter) and the linear/polynomial fit for the historical/scenario (lines).

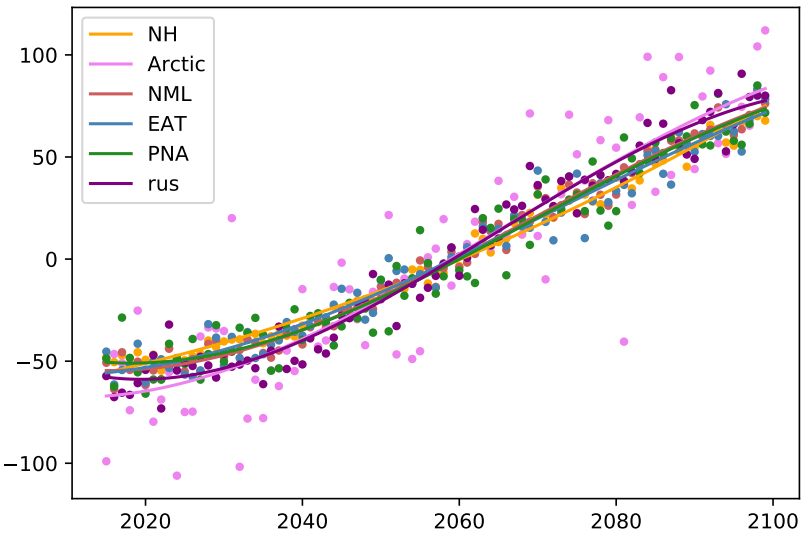

Fig. 2. Detrending for the ssp585 scenario simulation with EC-Earth, considering different areas for the average: NH (0-90N), Arctic (70-90N), NML (30-90N), EAT and PNA as in the paper, rus (30-90N, 40-140 E)

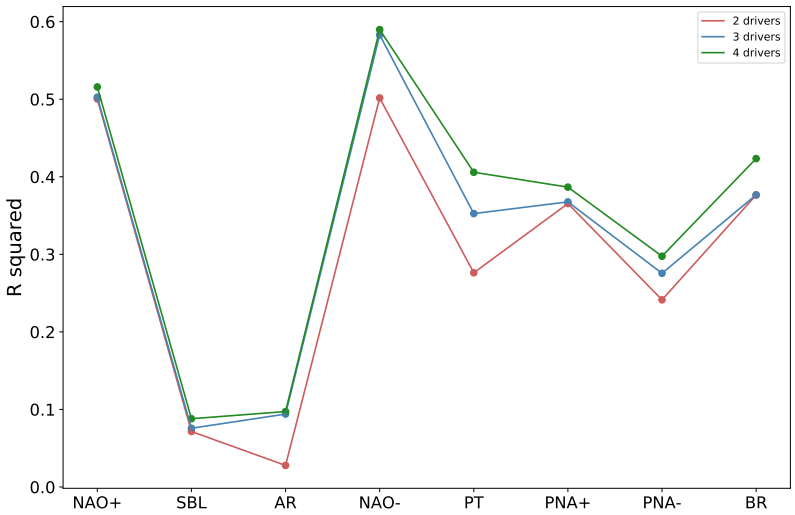

Fig. 3. Figure S7.