

## ***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 the reviewer for the comments and for the deep analysis of the manuscript, that helped to improve the overall quality of our work.

### **General comments**

This study investigates future changes in Atlantic-European and Pacific-North American weather regime occurrence according to CMIP projections. In a first part, the authors evaluate how well the models represent weather regimes (WRs) in their historical simulations compared to reanalysis, with a particular focus on how models from the 6th CMIP phase have improved compared to the 5th phase. In a second part, they investigate how WR frequency and persistence changes by the end of the century. In

C1

a last part, they trace these changes back to changes in the atmospheric mean state. The study provides an important basis for understanding most recent projections of future changes in surface weather from a large-scale dynamics point of view and opens interesting questions for further research. Moreover, the detailed analysis of model biases can be a good guidance for the CMIP community to further improve their models. The paper has a clear and logical structure and is comprehensibly written. The methodological procedure is thorough and transparent. Aside from one major concern, I only have a (relatively large) set of minor comments that should be addressable relatively easily by providing some further explanation or making small adjustments in the text or figures. Therefore, I suggest the paper to be published after considering this major comment and the list of minor comments.

Major comment – Section 4.2 (Drivers of future circulations): Although I appreciate the attempt to understand the origins of the projected changes in weather regimes (WRs) in more detail, I find this particular analysis in its current form not convincing enough from a causality perspective. In my opinion, this starts with calling the four indices (UTW, AA, PST, NAW) “remote drivers” and using them as statistical predictors for “predicting” WR changes based on linear regression. The reason is that there are strong dynamical links between mid-latitude storm track activity (and thus the WRs) and these indices. For instance, Ambaum and Hoskins, 2002 (<http://shorturl.at/ipDO9>) suggested a strong coupling between the NAO and the stratospheric polar vortex (which can be used as a proxy for your PST to first order) in the sense that a positive NAO can trigger a strong polar vortex, which in turn can strongly couple with the troposphere and induce persistent periods of positive NAO. Similarly, the effect of weak stratospheric polar vortex states on the troposphere (and thus on WRs) has shown to be strongly influenced by synoptic activity or WR occurrence beforehand (e.g., Kodera et al., 2016, <https://doi.org/10.1002/2015JD023359>; Domeisen et al., 2020, <https://doi.org/10.5194/wcd-2019-16>). Along the same lines, Garfinkel et al., 2015 (<https://doi.org/10.1002/2015JD023284>) showed that for instance SST anomalies (and thus the tropospheric state / storm track activity) can contribute to Arctic lower-

C2

stratospheric temperature changes (i.e., your PST). Likewise, I find it surprising to see such a strong link between the NAW and the PNA sector WRs. Although you mention that previous studies show a similar effect of North Atlantic temperatures on the troposphere in the Pacific sector – could it be that the link partly also acts the other way round, i.e. that the occurrence of certain PNA sector WRs (or a certain ENSO state) affects the SSTs in the North Atlantic and thus the NAW? At least this would be intuitive from a storm-track dynamical point of view, as the PNA sector WRs strongly influence the entrance of the North Atlantic storm track. Having these strong dynamical links between the four indices and the tropospheric dynamics in mind, I suggest that you either discuss / address this explicitly in your manuscript (by also “weakening” all the causality statements) or, optimally, that you gain some more insight in the causality by doing some kind of linear regression analysis considering time lags (similarly to Section 2.2. in Manzini et al., 2014, <https://doi.org/10.1002/2013JD021403>), if this is a possible approach in your framework. The latter analysis may help to shed some more light on this chicken-and-egg-like problem. Summarizing my comment in other words: I think it is very helpful for further research to include these four indices into your study, but I just think one should treat them more as phenomena that may be strongly and mutually coupled to the WRs themselves.

- Thank you for the comment and the many literature suggestions. We agree on the fact that in Section 4.2 (in its current form) only the link between the “drivers” and the frequency of the regimes is apparent, however the found relationship does not prove any causal link between them. In this respect, we also agree with the reviewer, that the term “driver”, without further specification, may be misleading. Therefore, we have decided to add the term “potential” to the title of Section 4.2 (“Potential drivers of future circulation changes”), to highlight this uncertainty, and use more neutral words (link, connection, relationship..) when referring in the text to the correlations found. Also, some comments on the fact that the links may be due, at least partly, to a reversed causal relationship (as in Domeisen, 2020) or to an external forcing influencing both processes in a similar way, are included to the discussion in the revised manuscript. The definition

C3

of the NAW has also been modified, and two different regions in the North Atlantic has been considered for the Atlantic warming in order to better evaluate the relative role of the tropical North-Atlantic and the subpolar gyre/warming hole region, which can be seen as a proxy for the AMOC (Liu et al. 2020, DOI: 10.1126/sciadv.aaz4876; Keil et al. 2020, <https://doi.org/10.1038/s41558-020-0819-8>). We included a new comment in the discussion, suggesting that it may also be that both the NAW and the change in the Pacific WRs frequency share a common external driver (for example a change in the ENSO forcing). A thorough analysis is needed to understand whether (or not) the found correlations imply any causal link, however such study goes beyond the scope of the present paper, and it will be carried on in a further study. Thank you for suggesting the method applied in Manzini et al. (2014). Unfortunately, a similar time-lagged correlation analysis is not suited to our case, since most processes involved in the WR transitions have typical lifetimes which are relatively short (usually synoptic). Among all potential drivers considered in Section 4.2, only the stratosphere-troposphere connection has a characteristic time scale of a few months, which is enough for the time-lagged correlations to appear. On the other hand, the dynamical link with the SSTs in the North-Atlantic or the tropical tropospheric temperature is characterized by faster time scales, not suited to that analysis.

Minor comments – L28-30: I see that several studies find a poleward shift of the upper-level jet caused by the UTW. You additionally mention an intensification of the upper-level jet due to the UTW. However, it is not obvious to me why a stronger meridional temperature gradient in the upper troposphere strengthens the jet on the same level? According to the thermal wind balance, the upper-level jet should primarily be driven by the meridional temperature gradient in the lower troposphere (as, for instance, Hasanzadeh et al., 2014, considers by looking at changes of near-surface meridional temperature gradients on jet intensity). Can you explain this from a dynamical point of view?

- We realized that we placed the wrong reference here: as you say, Has-

C4

sanzadeh et al. (2014) consider the role of the near-surface temperature gradient in affecting the mid-latitude circulation, and purposely exclude the effect of the UTW (mainly driven by increased latent heat release, not present in their dry model). We changed the reference to Barnes and Screen (2013) (<https://onlinelibrary.wiley.com/doi/epdf/10.1002/wcc.337>), which discusses the influence of the UTW on the jet stream as opposed to the Arctic Amplification (see Section “Tug-of-War: Tropics Versus Poles”). Thermal wind balance relates the vertical wind shear with the meridional temperature gradient. It is true that the wind on a constant pressure surface is not influenced by the meridional temperature gradient at the same height. At the same time this implies that the meridional temperature gradient immediately below the considered isobaric surfaces does contribute to the thermal wind. In our case we define the UTW as the temperature trend in the layer between 400 hPa and 150 hPa, which indeed spans part of the pressure heights located (along the vertical) near the jet stream core, and thus able to have an influence on the jet stream through the thermal wind balance. Moreover, the UTW is a proxy for the warming of the whole tropospheric column in the tropics and is linked with the meridional temperature gradient on a larger vertical interval. This is very clear in Fig. 1 of Shaw (2019) (<https://link.springer.com/article/10.1007/s40641-019-00145-8>), which shows that the warming in the UTW drives an increase in the meridional temperature gradient in the upper troposphere, that reinforces the jet stream. This is very clear in the Southern Hemisphere, while in the Northern Hemisphere this is moderated by the decrease of the temperature gradient at the surface due to Arctic amplification.

– L32: I would add Pithan and Mauritsen, 2014 (<https://doi.org/10.1038/ngeo2071>) to the reference of Screen and Simmonds, 2010, who discussed the mentioned “several other positive feedbacks” in more detail. – L75-76: I would add Michelangeli et al., 1995 (<http://shorturl.at/kmKW2>) here.

- Thank you for the suggestions. We added the references in the paper.

– L76: I would reword the sentence “each WR has a different impact on the climate

C5

of the downstream region” a bit, because it sounds like WRs are defined in a domain upstream, e.g., over the North Atlantic, to investigate surface weather downstream, e.g., over Europe.

- We rephrased the sentence, which now reads: “each WR has a different impact on the climate of the region considered”.

– L76: I suggest to rename the abbreviation for the Pacific-North American sector in the whole manuscript to something like PAC or PNAM, because using PNA becomes confusing later on due to the two equally named regimes PNA+ and PNA- (for instance, at the end of line 214 it is not unambiguous whether you talk about all four PNA sector WRs or only PNA+/-).

- Thank you for the suggestion, it is true that the current notation might be ambiguous. We changed from PNA to PAC when referring to the Pacific sector.

– L70-85: Could you give a very brief summary (two to three sentences) of previous studies investigating WR changes in GCM simulations / CMIP projections? I think you partly do that later in the results section, but it may be helpful to get an overview of studies here already.

- Most studies of WRs in models focused on the model performance in control/historical simulations (Dawson et al., 2012; Cattiaux et al., 2013; Weisheimer et al., 2013; Dawson and Palmer, 2015; Strommen et al., 2019; Fabiano et al., 2020a). As far as we know, changes of WRs in CMIP5 projections were only analyzed by Cattiaux et al. (2013, doi: 10.1007/s00382-013-1731-y) and by Ullmann et al. (2014, doi: 10.1002/joc.3864) for the EAT sector, which however used a substantially different approach and found different results for the regime frequency change (a stronger NAO-increase and no significant change, respectively). We added a comment on this in the introduction and a comparison of our CMIP5 results with theirs in the discussion section.

C6

– L116-118: Can you elaborate a bit more on why it is necessary to detrend the historical data with the described approach before identifying the WRs? More specifically: How robust / meaningful is the described linear trend in the Northern Hemisphere area-averaged geopotential height, considering for instance the substantial multi-decadal variability in the large-scale circulation over these 50 years? Could it be that the trend (and thus the WR identification) becomes significantly different when considering, for instance, only the last 40 instead of 50 years (which is often done when investigating the ERA-Interim period only)? How reasonable is it to detrend, for instance, the North Pacific with a linear trend that is obtained from an area average over the whole Northern Hemisphere (including the North Atlantic)? Does your WR pattern identification change if you do not detrend the historical data (which is often done in other studies to my knowledge)?

- Figure 1 shows the average geopotential height at 500 hPa (in units of meters) for all models in the historical period and ssp585 future scenario. The need for the polynomial detrending is clear from the scenario averages, which show a non-linear behaviour. When performing the scenario detrending, we judged more correct to also detrend the historical data with a linear term, since the increase in the geopotential height is already clear for the 60s to the end of the historical period. The historical increase reaches up to 20-30 m for individual models, which can certainly change some daily assignments, possibly slightly changing the frequencies and creating spurious frequency trends in the historical period. The choice to calculate the trends on the whole Northern Hemisphere was initially motivated by avoiding influences on the trend due to decadal basin-wide fluctuations, such as the AMV, which would affect less the hemispheric trend. Nevertheless, we observed that the difference in the historical and future trends when considering the whole hemispheric or the sectorial averages is very small (see Figure 2).

– L122-123: If I understand correctly, you ultimately apply the EOF analysis to unfiltered daily Z500 anomalies, right (apart from the running mean climatology you subtract)?

C7

What is your idea behind using the daily anomalies like this, without applying any low-pass filter beforehand? Would the latter change your result – did you test this?

- Yes, we apply the EOF analysis to the unfiltered daily anomalies, as it has been done in many other works in literature (e.g. Cassou, 2008; Dawson, 2012; Cattiaux, 2013; Strommen, 2019). Other works did apply a 10-day Lanczos filter (Straus, 2007; Dawson, 2015; Madonna, 2017), but the regime patterns obtained there are not significantly affected by such operation. We also checked this on ERA data and found no significant shift in the regime centroids when applying the filter. However, applying a low-pass may have an impact on the daily assignments and we judge the no-filtering approach to be more conservative in this sense.

– L123-127: Is it correct that you calculate the future Z500 anomalies (used for detecting the WRs) by subtracting the Z500 climatology (or the seasonal cycle, as you call it) based on the historical period and not based on the future period? Is this what you mean with the last sentence in this paragraph? I think this is crucial but may not be fully clear from the text.

- Yes, this is correct. Thank you for the comment, we rephrased the paragraph to make this clearer: “The seasonal cycle is computed averaging the data day-by-day at each grid point and applying a 20-day running mean to remove higher frequency fluctuations. However, the seasonal cycle computed in the historical simulations generally differs from the seasonal cycle found in scenarios. Since these differences are part of the change in the midlatitude circulation, it is important to take them into account. Therefore, for each model and scenario, the mean seasonal cycle is computed in the reference period of the corresponding historical simulation (1964-2014 for CMIP6, 1964-2005 for CMIP5).”

– L129-131: Here it would be worthwhile citing some other studies using similar regional domains for the EOF analysis.

- We added Dawson et al. (2012) and Weisheimer et al. (2014) as references for the

C8

EAT and PAC cases.

– L131-136: I am not sure if I fully understand the procedure: Do you apply an EOF analysis both on the observed anomalies (to get the 4 observed PCs), and also separately on the modeled anomalies (to get the 4 pseudo-PCs) to calculate the “computed regimes”? In the current form, it sounds like you apply the EOF analysis only on the observed anomalies, and you then project the modeled anomalies into this (observed) phase space, as a basis for both the “computed” and “projected” regimes.

- The EOF analysis is in fact applied only to the observed anomalies, and the space spanned by the 4 reference EOFs is then used as a reference phase space for all model simulations. So, for the model simulations, the daily anomaly field is directly projected on the reference EOFs, to obtain the series of pseudo-PCs. This procedure was first adopted by Fabiano et al. (2020a) and additional comments and sensitivity tests can be found there. The advantage of considering a single reference space is to allow a direct comparison of the cluster centroids from different simulations in a consistent way. We changed the paragraph to make this clearer.

– L146-148: Out of curiosity, did you calculate the “computed regimes” also for the future simulations? If yes, are they different, and do they tell us something about changes in modes of variability in the large-scale circulation?

- Unfortunately, we have not yet calculated the computed regimes for the future simulations. We agree that this would be a really interesting point to consider, but we expect the effect to be of second-order with respect to the change in frequency, and probably more difficult to assess. For this reason, we decided to focus only on the regime frequencies in this work, and leave the study of potential changes of the modes of variability to a future work.

– L160-162: Could you briefly mention in the manuscript whether a higher or lower variance ratio is generally desirable for a WR definition (independent of the comparison between observations and model)? I guess a WR definition is “better” (i.e., the WRs

C9

are more distinct) the larger the variance ratio is, because a high variance ratio implies a relatively large distance between the cluster centroids compared to the distances within a centroid, right?

- Yes, a higher variance ratio is desirable, we added an appropriate explanation to clarify this in the text. However, the comparison with the observation is key here, since models tend to have too low variance ratio on the EAT sector and too high in the PAC one. The Pacific regimes should not be so clearly defined in models, and this reflects some misrepresentation in the midlatitude circulation that needs further study.

– L195-196: Out of curiosity, do you know whether there are preferred circuits / transitions between the PNA sector WRs, considering the fact that they resemble (different states of) Rossby wave trains originating from the tropics?

- Thank you for the interest. Unfortunately we have not yet carefully studied the transition probabilities between the different PAC states. However, we computed the transition probability for the reanalysis data. At first sight, the highest transition probability from PT is towards the PNA- regime, consistent with the Rossby wave train view. For all other regimes, the transition to PT seems favoured.

– Figure 2: I really like the way you compare the WR representation / biases in Figure 2! I think it could be of interest for the CMIP community to additionally see in the Taylor diagram how individual model centers improved (or worsened) their WR representation between phase 5 and 6. Could you use a specific symbol (instead of a dot for every simulation) for the same model (or for related model simulations / centers)? Or does this make the figure too overloaded?

- Given the large number of models considered, we thought that the figure would have been simpler with just the model ensemble information. However, we are working on producing an analogous figure with specific symbols for each model, and add it to the Supplementary material. Nevertheless, the model centers represented in the two CMIP phases differ, so the check for the improvement of a particular one will not be possible

C10

for all of them.

– L212-214: I kind of see your argument of an overall improvement between CMIP5 and CMIP6 visually, but can you “proof” this with a certain measure of significance (considering the relatively small number of model simulations)? Is this the degree of overlap of the shaded blue and red ellipses (if yes, please specify in the manuscript)? I would also be careful with concluding that the two NAO WRs improve more than the others – this is not that convincing considering the large inter-model spread for instance in the NAO+. Also, in principle it can be that the same model center does not improve but rather worsen from phase 5 to phase 6, which would become visible if the symbols were changed as suggested in the previous comment (for instance, the very top-left red dot in the AR diagram probably indicates such a case). Can you elaborate a bit more on this in the manuscript and, in case you do not indicate the individual models with a symbol, say whether all (or most) models improved from phase 5 to phase 6?

- Yes, the degree of overlap gives a measure of the significance of the improvement from CMIP5 to CMIP6 ensembles, we added a comment on this in the text and relaxed the claim on the two NAO regimes improving more than the other regimes. Your suggestion of looking at the change in the metrics for individual model centers, and see how many of them improve between the two phases, would certainly be a good way to assess this more quantitatively. However, since there is not a one-to-one correspondence between the models available in the two CMIP phases, the final assessment of the two ensembles would be partial in this way too. However, we will provide some indication of how many of the “matching” models improve from CMIP5 to CMIP6 in these metrics.

– L212-214: The clearly smallest inter-model spread and generally smallest bias in the NAO- WR in the Taylor diagram may indicate a higher intrinsic predictability of the NAO- WR compared to the others (also compared to NAO+). Does this make sense, and did you think about this? And, if yes, do you have an explanation for this, or do you know whether this has been shown before?

C11

- The NAO- mean regime pattern is known to be the best reproduced in historical model simulations (see for example Strommen et al. 2019, Fabiano et al. 2020a). However, we never considered the possible connection with the regime predictability. On one side, the fact that a simulation correctly represents the observed pattern has no clear implication for its predictability. However, it seems plausible that a correct representation of the regime pattern might be due to a better representation of the processes behind the onset and persistence of the regime, which in turn could also give a better skill in predicting the regime itself. We are not aware of studies showing a better predictability for the NAO- though, but we will keep this suggestion in mind when analyzing seasonal simulations.

– On the other hand, the PNA sector WRs generally seem to be harder to capture properly, considering the large inter-model spread. Do you have an explanation for this? Could it be related to the strong dependence on the tropics, implying that models with a bad representation of the tropical-extratropical interaction perform substantially worse in terms of PNA sector WRs? I know this is beyond the main focus of this study, but I think it would be helpful to briefly discuss these aspects and speculate about possible reasons in a few sentences.

- We agree that the PAC regimes are generally more difficult to capture than for the EAT sector. This may reflect a larger natural variability in the regime structure, as suggested by the smaller values of the variance ratio of the reanalysis on the PAC sector. The link with the tropics might be key to that, since these regimes are strongly influenced by the ENSO forcing. A different representation in terms of amplitude and frequency of ENSO events in the different model simulations could perturb the regime patterns in different ways, leading to a larger spread. In this regard, there is some indication that the model performance for the EAT sector might be linked to the SST representation (see Fabiano et al, 2020a). It would be interesting in this sense to analyze the observed natural variability in these regimes, and compare it to that observed in the EAT sector. We can certainly add a brief comment on this in the paper.

C12

– Figure 3: I understand the idea of Figure 3 from a perspective of condensing information, but I do not see the scientific reason for plotting the frequency bias against the variance ratio because there is no direct link between the two. Hence, it could confuse the reader because a linear relationship may be expected by this way of plotting. If there is a link, please clarify in the text. Otherwise, I suggest showing two vertical box-whisker plots, one for the frequency bias and one for the variance ratio (with a horizontal black line for the corresponding ERA-Interim value). Furthermore, please indicate the units for the frequency bias.

- The idea of the figure was just to synthetically show two important metrics “at once”, but that might not be the best way to do this. Indeed, there is no direct link (that we know of) between the two quantities and we accept your suggestion to change this figure into a box-whisker plot in the revised manuscript.

– Figures 2 and 3: How would you relate the biases shown in your figures to the well-known blocking biases in GCMs (e.g., Davini and D’Andrea, 2020, <https://doi.org/10.1175/JCLI-D-19-0862.1>)? Do we also see this somehow in your figures?

- In the EAT sector, there is a strong link between regimes and blocking events (see Madonna et al. 2017 and Figure 10 in Fabiano et al. 2020a). In particular, most of the blocking events in the EAT sector coincide with NAO- and SBL states, with a lesser contribution of AR. Models generally are able to reproduce this correlation, but generally struggle to reach the observed intensity of the signal. Apart from the model biases in catching the regime-blocking connection, it seems plausible that a smaller frequency bias might indicate a smaller bias in the blocking representation. Also, a link between blocking bias and variance ratio has been observed in a multi-model ensemble in Fabiano et al. (2020a). As for the PAC sector, although we are not aware of a similar work linking regimes and blocking events, it is very likely that the Bering Ridge is associated with the North Pacific blocking. It might be worth commenting on this in the text, thank you for the suggestion.

C13

– L222-225: The overall different variance ratios between the PNA and EAT WRs compared to observations are indeed interesting. In the case of the PNA sector WRs, however, a higher than observed variance ratio does not necessarily mean that the models perform better than for the EAT WRs, right? It just means that the models distinguish more (too?) strongly between the different WRs. So I do not really understand this statement here, because, regarding my previous comment (about the larger spread in the Taylor diagram for the PNA sector WRs), I would rather think the models overall have more problems for the PNA sector WRs. . .

- Yes, the representation of the PAC regimes in models overestimates the variance ratio. 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.”

– L230-233: I find it somehow surprising that the model biases in reproducing the observed regimes are smaller for projected regimes than for computed regimes. Can you explain? Is it simply because applying the k-means clustering to each model simulation (i.e., the computed regimes) yields slightly different and thus “new” regimes, hence, you kind of compare apples with oranges in the Taylor diagrams in Figure 2?

- The computed regimes of each simulation are reordered as to best match the observed series. Generally there is a quite good one-to-one match, so we are still comparing apples with apples in most cases. However, the natural variability of the system

C14

is large and the k-means on a relatively short timeseries (50 years) can produce, in some cases, centroids that are shifted in phase space. Large differences were also observed between different ensemble members of the same model (Fabiano et al., 2020a), with occasionally some “rotated” regimes being produced and correspondingly very bad performance. This might be the case for the few outliers that can be seen, for example, for the PNA- pattern, with spatial correlations close to 0. This does not happen for the projected regimes since we force the cluster centroids to the observed ones and the regimes, along with their dynamical implications, are always well defined. Indeed, the pattern biases result smaller for the projected regimes (Figure S1). However, it is less obvious that the frequency and variance ratio biases should be smaller as well. Nevertheless, this is what we observe in Figure S2, hinting that the K-means might enlarge some biases by misplacing the cluster centroids, while the “real” attractors might be closer to the observed ones.

– Figures 4, 6, 7: Considering the relatively small number of models, I wonder how robust the distributions in the box-whisker plots are. Did you check whether certain distributions are skewed due to, for instance, a clustering of several model simulations from the same model center? I guess the Welch’s t-test does consider that particular problem. Nevertheless, it may be helpful for the reader to replace the box-whisker plots with violin plots additionally indicating the density within the distribution.

- The main result we want to highlight with Figures 4, 6 and 7 is the shift in the WR frequencies under increasing greenhouse forcing. The box-whisker plots are indeed enough and clear to show the multi-model ensemble mean. Besides, some information on the spread and the skewness of the distribution can be gained. We prefer not to change the plots to violin plots to avoid overloading the figure and possibly make the main result more difficult to read. Nevertheless we can show the full distributions at least for the ssp585 and historical ensembles in the supplementary material.

– Figures 4, 6, 7: I would color the historic box in black (and all the future scenarios in color, as it is), just as a suggestion.

C15

- Thanks for the suggestion. We will change the color of the historical box to help distinguish it from the scenarios.

– L257-263: The apparently non-linear response of the NAO- to the CO<sub>2</sub> forcing is very interesting! Furthermore, the temporal development in the different simulations in Figure 5 shows an interesting multi-decadal variability. You mention that this will be analyzed in further studies. Can you nevertheless speculate about some potential reasons? Could it be related to some kind of tipping points in external forcings such as the Greenland ice sheet (which could affect the Greenland high) or sea surface temperature?

- 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.

– L264: How does the temporal evolution for the PNA sector WRs in the CMIP simulations look like (analog to Figure 5)? Does it also exhibit any multi-decadal variability in specific WRs? I would suggest adding this figure to the supplement.

- The analog of Figure 5 is shown in the attached Figure 3. The temporal evolution for the PNA sector shows some multi-decadal variability, in particular for the PT regime (for which a minimum is found at the end of the historical period) and for the PNA+/-PNA-scenarios. We will add the Figure to the supplementary material. Interestingly, the RCP8.5 scenario of CMIP5 deviates from the CMIP6 projections for almost all regimes. However this is in part due to the plot construction and to differences in the historical mean frequency, since the differences with respect to the CMIP6 historical ensemble

C16



are shown here.

– L277-289: Thinking in terms of WR life cycles, the strong correlation between changes in WR frequency (Figures 4, 6) and WR persistence (Figure 7) implies that there does not seem to be changes in numbers of life cycles but rather changes in the duration of individual life cycles (which ultimately make the changes in WR frequency). Is that correct? If yes, can you discuss this with a few sentences in the manuscript? It could also be interesting / helpful to plot changes in WR frequency against changes in WR persistence. Depending on the robustness, this finding may to some degree also have implications for the (operational) predictability of WRs for instance on subseasonal-to-seasonal time scales.

- Thank you for the suggestion. We are going to produce a new figure, showing the changes in the number of regime events, which would help disentangle the persistence and frequency changes, and will add a comment on this in the revised manuscript.

– L310-311: What are the projections for future ENSO occurrence? Do we expect (significant) changes? Can you cite some of these studies here?

- A recent study of ENSO occurrence CMIP6 projections (Fredriksen et al., 2020, 10.1029/2020GL090640) shows a tendency for an increase of ENSO variability under global warming, and interestingly this change appears to be related mostly to an increase in positive El Niño events. This is in line with our hypothesis that the PT regime frequency increase might be related to a stronger tropical forcing, we will add a comment to this in the discussion.

– Figure 8: Does NML stand for the hemispheric zonal mean?

- NML stands for Northern Mid-Latitudes, which indicates the region from 30N to 90N, used to calculate the trends. However, the term might not be really adequate, since the region includes the higher latitudes as well. We clarified this in the text.

– Figure 9: Just for clarification, does the shading in this figure show the mean Z500

C17

(grid-point level) in the future simulation minus the zonal mean (at every latitude) shown in Figure 8? How does Figure 9 compare to a map that simply shows the future mean Z500 minus the historic mean Z500 (both on a grid-point level)?

- The figure shows the residual trend at each grid point in the Z500, after removing the zonal trend. Analogously, Figure 8 shows the residual zonal trend, after removing the global NML trend. The difference of future and historical Z500 is dominated by the global NML trend, which is largely positive, so it would be difficult to discern changes in the stationary eddies from that figure. We clarified this in the captions.

– L344-350: How sensitive is your analysis to the latitude / pressure boundaries used to define the four metrics?

- Apart from the North Atlantic warming (NAW), the other metrics had already been used in other works (e.g Oudar et al., 2020; Peings et al., 2018; Zappa and Shepherd, 2017). We used here the same pressure/latitude boundaries as defined in Oudar et al. (2020), very similar to those in Peings et al. (2018), but a different choice (as in Zappa and Shepherd, 2017, that consider individual pressure levels) is not expected to change the results dramatically. Unfortunately, we do not have a quantitative estimate on this. As for the NAW, we split it in two better defined subregions (tropical Atlantic and subpolar gyre), in order to evaluate their relative contributions.

– L362: Please add Figure S7 to the supplement, because it's missing.

- Thank you for pointing this out. The figure is attached here as Figure 4 and will be added to the supplementary.

– L366-376: I would add a reference to, e.g., Ambaum and Hoskins, 2002 (<http://shorturl.at/ipDO9>), who proposed a mechanism for the strong NAO-polar vortex coupling (see previous comment).

- Thank you, we will add the citation to the text.

– L388-401: The changes in WRs in a future climate must be strongly linked to changes

C18

in extratropical cyclone activity and thus the storm track. Can you briefly discuss or at least speculate here whether and how some of your results (e.g., the increase in NAO+ frequency) might relate to the expected changes in extratropical cyclone frequency, location, and intensity? I assume this question is more complex than we think, but it would be nice to at least mention these questions in the conclusions and thus make a bridge toward the cyclone research community. Because in the end, changes in WRs are also a result of changes in cyclone activity. . .

- A recent paper on the change of storm track activity in CMIP6 projections (Harvey et al., 2020, <https://doi.org/10.1029/2020JD032701>) shows an increase of activity over the North-Atlantic and central/northern Europe, with a center on the British Isles and an increased penetration of perturbations into the continent. This is in line with our finding of a strong increase in the NAO+ frequency. A corresponding decrease of perturbations at very high latitudes is also in line with a decrease in the AR regime, which tends to push the jet poleward. For the PAC sector, a strong decrease of the storm track activity is seen over the whole Northern American continent, and an increase in the central North Pacific. This agrees well with the prediction of an increased PT regime, that blocks the entrance of perturbations in the continent, and a decrease in the BR regime, that would tend to drive perturbations away from the central North Pacific.

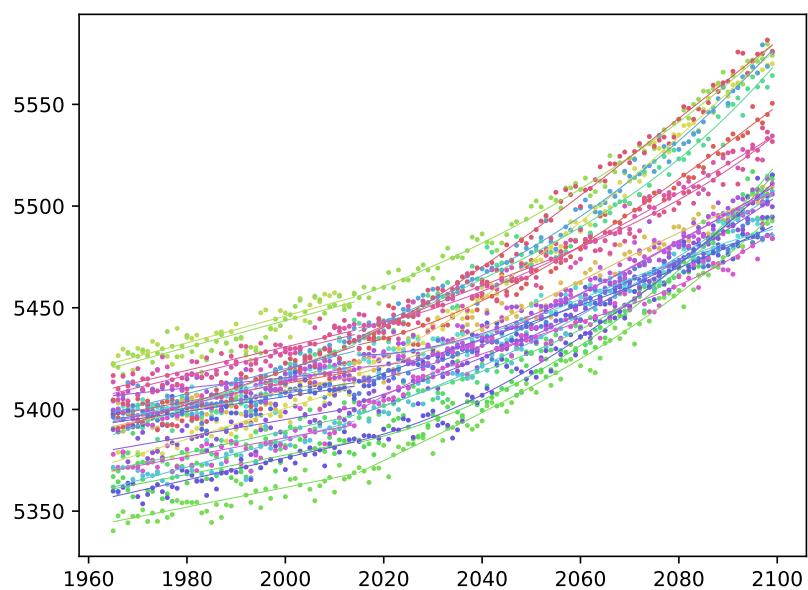
– Throughout the manuscript, you often write certain phenomena with capital letters, which I would not do. For instance, change “Weather Regimes” to “weather regimes”, “Polar Stratospheric Temperature” to “polar stratospheric temperature”, “Polar Vortex Strength” to “polar vortex strength” etc. – You misspell “Pacific Through” (instead of “Pacific Trough”) several times in the manuscript (including the Abstract) – L11: Change to “A major challenge for the climate community is to understand how a warmer climate . . .” – L13: Change to “. . . inextricably related to regional impacts . . .” – There are a few further grammatical inconsistencies throughout the manuscript, which should be detected when carefully revising the manuscript.

- Thank you for the language corrections, we implemented them in the revised version.

C19

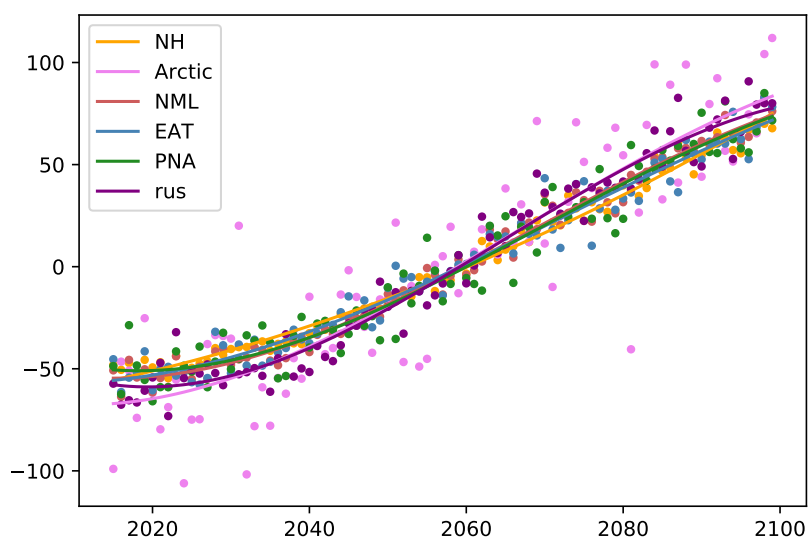
---

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



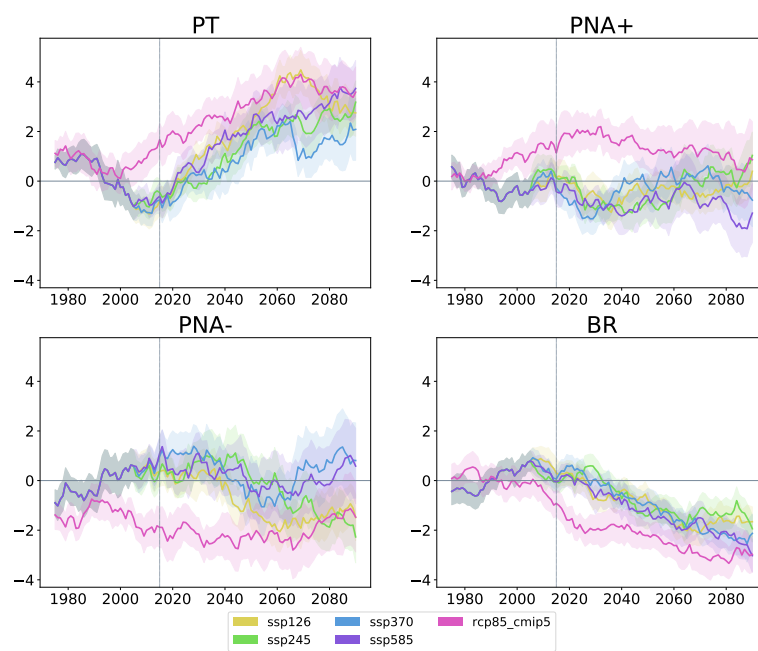
**Fig. 1.** Average geopotential field in the Northern hemisphere (30-90N) for all model simulations in the historical+ssp585 scenario (scatter) and the linear/polynomial fit for the historical/scenario (lines).

C21



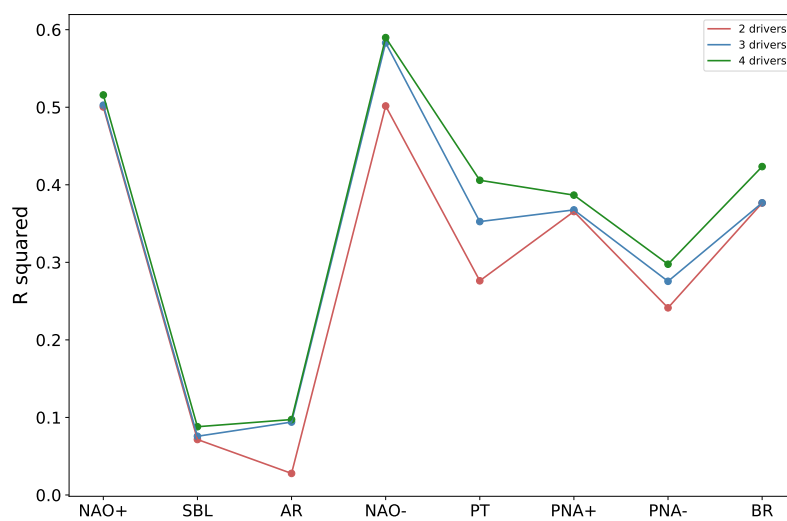
**Fig. 2.** Detrending for the ssp585 scenario simulation with EC-Earth, considering different areas for the average: NH (0-90N), Arctic (70-90N), NML (30-90N), EAT and PNA as in the paper, rus (30-90N, 40-140 E)

C22



**Fig. 3.** Analog of Figure 5, for the PAC sector.

C23



**Fig. 4.** Figure S7.

C24